Hybrid Uncertainty and Sensitivity Analysis of the Model of a Twin-Jet Aircraft

Nicola Pedroni, Enrico Zio

Hybrid Uncertainty and Sensitivity Analysis of the Model of a Twin-Jet Aircraft

Nicola Pedroni*
Ecole Superieure d'Electricitè (Supelec), Gif-Sur-Yvette, France, 91192

Enrico Zio†
Ecole Centrale Paris (ECP), Chatenay-Malabry, France, 92295
Politecnico di Milano, Milano, Italy, 20133

The mathematical models employed in the risk assessment of complex, safety-critical engineering systems cannot capture all the characteristics of the system under analysis, due to: (i) the intrinsically random nature of several of the phenomena occurring during system operation (aleatory uncertainty); (ii) the incomplete knowledge about some of the phenomena (epistemic uncertainty). In this work, we consider the model of a twin-jet aircraft, which includes twenty-one inputs and eight outputs. The inputs are affected by mixed aleatory and epistemic uncertainties represented by probability distributions and intervals, respectively. Within this context, we address the following issues: (A) improvement of the input uncertainty models (i.e., reduction of the corresponding epistemic uncertainties) based on experimental data; (B) sensitivity analysis to rank the importance of the inputs in contributing to output uncertainties; (C) propagation of the input uncertainties to the outputs; (D) extreme case analysis to identify those system configurations that prescribe extreme values of some system performance metrics of interest (e.g., the failure probability). All the tasks are tackled and solved by means of an efficient combination of: (i) Monte Carlo Simulation (MCS) to propagate the aleatory uncertainty described by probability distributions; (ii) Genetic Algorithms (GAs) to solve the numerous optimization problems related to the propagation of epistemic uncertainty by interval analysis, and (iii) fast-running Artificial Neural Network (ANN) regression models to reduce the computational time related to the repeated model evaluations required by uncertainty and sensitivity analyses.

Nomenclature

S	= mathematical model of the system
n_{inp}	= number of model input parameters
p_i	= i-th input parameter (i = 1, 2, …, n_{inp})
$\boldsymbol{p}$	= vector of system model input parameters ($\boldsymbol{p} = \{p_i: i = 1, 2, \ldots, n_{inp}\}$)
$\boldsymbol{d}$	= vector of system design variables
n_{out}	= number of model output parameters
n_{int}	= number of model intermediate variables
$h_j(\cdot)$	= j-th intermediate mathematical model (j = 1, 2, …, n_{int})
$\boldsymbol{p}^j$	= vector of input parameters to the j-th intermediate mathematical model $h_j(\cdot)$ (j = 1, 2, …, n_{int})
$x_j = h_j(\boldsymbol{p}^j)$	= j-th model intermediate variable (j = 1, 2, …, n_{int})
$\boldsymbol{x}$	= vector of model intermediate variables
g_o	= o-th model output parameter (o-th system requirement metric), $g_o = f_o(\boldsymbol{x}, \boldsymbol{d})$ (o = 1, 2, …, n_{out} = 8)

* Assistant Professor, EDF Chair of System Science and the Energy Challenge at Supelec, Plateau du Moulon - 3 rue Joliot-Curie, 91192, Gif-Sur-Yvette, France, e-mail: nicola.pedroni@supelec.fr

† Full Professor, EDF Chair of System Science and the Energy Challenge at ECP, Grande Voie des Vignes, 92295, Chatenay-Malabry, France, e-mail: enrico.zio@ecp.fr. Also: Department of Energy, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133, Milano, Italy, e-mail: enrico.zio@polimi.it

$\boldsymbol{g} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{d})$ = vector of model output parameters (vector of system requirement metrics)
$\boldsymbol{\theta}_i$ = internal coefficients of the probability distribution of the i-th input parameter
$n_{p,i}$ = number of internal coefficients of the probability distribution of the i-th input parameter
$\theta_{i,l}$ = l-th internal coefficient of the probability distribution of the i-th input parameter ($l = 1, 2, \ldots, n_{p,i}$, $i = 1, 2, \ldots, n_{inp}$)
$q^{p_i}\left(p_i \middle| \boldsymbol{\theta}_i\right)$ = Probability Density Function (PDF) of the i-th input parameter
$F^{p_i}\left(p_i \middle| \boldsymbol{\theta}_i\right)$ = Cumulative Distribution Function (CDF) of the i-th input parameter
$\Delta_{\theta i,l}$ = interval of variation of $\theta_{i,l}$ ($l = 1, 2, \ldots, n_{p,i}$, $i = 1, 2, \ldots, n_{inp}$)
$\boldsymbol{\Delta}_{\boldsymbol{\theta}_i}$ = vector of the intervals $\Delta_{\theta i,l}$ of variation of $\boldsymbol{\theta}_i$ ($l = 1, 2, \ldots, n_{p,i}$, $i = 1, 2, \ldots, n_{inp}$)
$PB^{p_i}(p_i) = \left\{F^{p_i}\left(p_i \middle| \boldsymbol{\theta}_i\right): \boldsymbol{\theta}_i \in \boldsymbol{\Omega}_i\right\}$ = distributional p-box of the i-th input parameter
$\boldsymbol{\Omega}_i = \prod_{l=1}^{n_{p,i}} \Delta_{\theta i,l}$ = space of variation of $\boldsymbol{\theta}_i$ (case of epistemic independence)
$E[\cdot]$ = expected value
$V[\cdot]$ = variance
$P[\cdot]$ = probability
m = mean of the Beta distribution of p_1
s^2 = variance of the Beta distribution of p_1
$N(\cdot,\cdot)$ = Normal distribution
μ_i = mean of the i-th Normal distribution
σ_i^2 = variance of the i-th Normal distribution
$U[0, 1)$ = uniform distribution on [0, 1)
a_i = location parameter of the i-th Beta distribution
b_i = scale parameter of the i-th Beta distribution
$\boldsymbol{\theta}^{x_j}$ = vector of the epistemically-uncertain parameters/coefficients of the inputs to $h_j(\cdot)$
$\boldsymbol{\Omega}^{x_j}$ = space of variation of $\boldsymbol{\theta}^{x_j}$
$\boldsymbol{\theta}^{all}$ = vector of all the epistemically-uncertain parameters/coefficients contained in the entire system model S
$\boldsymbol{\Omega}^{all} = \prod_{j=1}^{n_{int}} \boldsymbol{\Omega}^{x_j}$ = space of variation of $\boldsymbol{\theta}^{all}$
$\boldsymbol{x}_1^d$ = vector of real random realizations of variable x_1
n_{d1} = size of vector $\boldsymbol{x}_1^d$
$w(\boldsymbol{p}, \boldsymbol{d}) = \max_{1 \le o \le 8} g_o$ = system worst-case requirement metric
$J_1 = E_{\boldsymbol{p}}[w(\boldsymbol{p}, \boldsymbol{d})]$ = expected value of $w(\boldsymbol{p}, \boldsymbol{d})$
$J_2 = 1 - P[w(\boldsymbol{p}, \boldsymbol{d}) < 0]$ = system failure probability
$\hat{F}^{x_1, n_d}(x_1)$ = empirical CDF of x_1 built on n_d realizations
$\overline{F}^{x_1}_{KS(\alpha, n_d)}(x_1)$ = Kolmogorov-Smirnov upper bound on $\hat{F}^{x_1, n_d}(x_1)$, with statistical confidence $100(1 - \alpha)\%$
$\underline{F}^{x_1}_{KS(\alpha, n_d)}(x_1)$ = Kolmogorov-Smirnov lower bound on $\hat{F}^{x_1, n_d}(x_1)$, with statistical confidence $100(1 - \alpha)\%$
$D(\alpha, n_d)$ = one-sample Kolmogorov-Smirnov critical statistic for intrinsic hypotheses for confidence level $100 \cdot (1 - \alpha)\%$ and sample size n_d
$\boldsymbol{\Omega}^{x_1}_{n_d}$ = space of variation of $\boldsymbol{\theta}^{x_j}$, improved by means of n_d empirical data
Q = generic output quantity of interest
$U_{\boldsymbol{p}}(Q)$ = amount of epistemic uncertainty contained in Q
$\boldsymbol{\varphi}_i$ = generic epistemically-uncertain 'factor' in the uncertainty model of p_i
$\boldsymbol{\varphi}_i^*$ = generic constant value to which $\boldsymbol{\varphi}_i$ can be fixed
$\boldsymbol{\varphi}_i^k$ = k-th value to which $\boldsymbol{\varphi}_i$ can be fixed (during computation of global sensitivity index) ($k = 1, 2, \ldots, N_e$)
$U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ = amount of epistemic uncertainty in Q when $\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*$
$\boldsymbol{\theta}_{-i}^{all} = \left[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i-1}, \ldots, \boldsymbol{\theta}_{n_{inp}}\right]$ = vector of all the epistemically-uncertain parameters/coefficients except $\boldsymbol{\theta}_i$
$\boldsymbol{\Omega}_{-i}^{all}$ = space of variation of $\boldsymbol{\theta}_{-i}^{all}$

$E_{\varphi i}[U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i)]$ = expected amount of epistemic uncertainty in Q when $\boldsymbol{\varphi}_i$ is fixed to a constant
$S_i(Q)$ = sensitivity (index) of (the epistemic uncertainty in) Q to (the epistemic uncertainty in) p_i
$A_{\boldsymbol{p}}(x_j)$ = area contained in the distributional p-box of x_j
$PB^{x_j}(x_j) = \left\{F^{x_j}\left(x_j\middle|\boldsymbol{\theta}^{x_j}\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$ = distributional p-box of x_j
$\overline{F}^{x_j}(x_j)$ = extreme upper CDF bounding the distributional p-box of x_j
$\underline{F}^{x_j}(x_j)$ = extreme lower CDF bounding the distributional p-box of x_j
N_{train} = size of the training set of an Artificial Neural Network (ANN)
$\boldsymbol{x}_t = \{x_{1,t}, x_{2,t}, \ldots, x_{j,t}, \ldots, x_{n_{in},t}\}$ = t-th input pattern of the training set of an ANN (t = 1, 2, …., N_{train})
$\boldsymbol{g}_t = \{g_{1,t}, g_{2,t}, \ldots, g_{l,t}, \ldots, g_{n_{out},t}\}$ = t-th output pattern of the training set of an ANN (t = 1, 2, …, N_{train})
$D_{train} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, \ldots, N_{train}\}$ = training set of an Artificial Neural Network
n_h = number of hidden layers in an ANN
N_{val} = size of the validation set of an ANN
$D_{val} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, \ldots, N_{val}\}$ = validation set of an ANN
N_{test} = size of the test set of an ANN
$D_{test} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, \ldots, N_{test}\}$ = test set of an ANN
R^2 = coefficient of determination
p_i* = generic constant value to which p_i can be fixed
$PB^{x_j}(x_j|p_i*) = \left\{F^{x_j}\left(x_j\middle|\boldsymbol{\theta}^{x_j}, p_i*\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$ = distributional p-box of x_j obtained keeping $p_i = p_i$*
$\underline{F}^{x_j}\left(x_j\middle|p_i^*\right)$ = extreme lower CDF bounding the distributional p-box of x_j obtained keeping $p_i = p_i$*
$\overline{F}^{x_j}\left(x_j\middle|p_i^*\right)$ = extreme upper CDF bounding the distributional p-box of x_j obtained keeping $p_i = p_i$*
$A_{\boldsymbol{p}|p_i^*}^{x_j,over}$ = area of overlap between the p-box of x_j built using the original uncertainty models of $\boldsymbol{p}$ and the one built keeping $p_i = p_i$*
$\varepsilon_{p_i^*}^{x_j}$ = (fractional) lack of overlap between the p-box of x_j built using the original uncertainty models of $\boldsymbol{p}$ and the one built keeping $p_i = p_i$*
$[\underline{J}_1, \overline{J}_1]$ = range of J_1
$[\underline{J}_2, \overline{J}_2]$ = range of J_2
$[\underline{J}_1, \overline{J}_1]^{ANN}$ = range of J_1 obtained by replacing the original system model by an ANN model
$[\underline{J}_2, \overline{J}_2]^{ANN}$ = range of J_2 obtained by replacing the original system model by an ANN model
$[\underline{J}_1(p_i^*), \overline{J}_1(p_i^*)]$ = range of J_1 obtained setting $p_i = p_i$*
$[\underline{J}_2(p_i^*), \overline{J}_2(p_i^*)]$ = range of J_2 obtained setting $p_i = p_i$*
$L_{\boldsymbol{p}|p_i^*}^{J_1,over}$ = length of overlap between the intervals $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_1(p_i^*), \overline{J}_1(p_i^*)]$
$L_{\boldsymbol{p}|p_i^*}^{J_2,over}$ = length of overlap between the intervals $[\underline{J}_2, \overline{J}_2]$ and $[\underline{J}_2(p_i^*), \overline{J}_2(p_i^*)]$
$\varepsilon_{p_i^*}^{J_1}$ = (fractional) lack of overlap between the intervals $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_1(p_i^*), \overline{J}_1(p_i^*)]$
$\varepsilon_{p_i^*}^{J_2}$ = (fractional) lack of overlap between the intervals $[\underline{J}_2, \overline{J}_2]$ and $[\underline{J}_2(p_i^*), \overline{J}_2(p_i^*)]$
N_{p*} = number of constant values p_i* selected in the sensitivity analysis
$p_{i,k}^*$ = k-th value of p_i* (k = 1, 2, …, N_{p*})
$\boldsymbol{\Omega}_{1,red}$ = reduced (improved) space of variation for the epistemically-uncertain parameters/coefficients of p_1
$\boldsymbol{\Omega}_{5,red}$ = reduced (improved) space of variation for the epistemically-uncertain parameters/coefficients of p_5
$\boldsymbol{\Omega}_{12,red}$ = reduced (improved) space of variation for the epistemically-uncertain parameters/coefficients of p_{12}
$\boldsymbol{\Omega}_{21,red}$ = reduced (improved) space of variation for the epistemically-uncertain parameters/coefficients of p_{21}
$\boldsymbol{\Omega}_{red}^{all}$ = reduced (improved) space of variation for $\boldsymbol{\theta}^{all}$
$\boldsymbol{\theta}_{1,low}^{all}$ = realization of $\boldsymbol{\theta}^{all}$ for which $J_1 = \underline{J}_1$

$\boldsymbol{\theta}^{all}_{1,up}$ = realization of $\boldsymbol{\theta}^{all}$ for which $J_1 = \overline{J}_1$
$\boldsymbol{\theta}^{all}_{2,low}$ = realization of $\boldsymbol{\theta}^{all}$ for which $J_2 = \underline{J}_2$
$\boldsymbol{\theta}^{all}_{2,up}$ = realization of $\boldsymbol{\theta}^{all}$ for which $J_2 = \overline{J}_2$
$R_i(Q)$ = sensitivity rank of parameter p_i evaluated according to indicator $Q = x_j, J_1, J_2$
$R_{acc,i}$ = $R_i(J_1) + R_i(J_2)$ = accumulated sensitivity ranking of p_i evaluated as the sum of the rankings of J_1 and J_2

I. Introduction

THE quantitative analyses of the phenomena occurring in safety-critical (e.g., civil, nuclear, aerospace and chemical) engineering systems are based on mathematical models[1-3]. In practice, not all the characteristics of the system under analysis can be captured in the model: thus, uncertainty is present in both the values of the model input parameters and hypotheses. This is due to: (i) the intrinsically random nature of several of the phenomena occurring during system operation (aleatory uncertainty); (ii) the incomplete knowledge about some of the phenomena (epistemic uncertainty). Such uncertainty propagates within the model and causes uncertainty in its outputs: the characterization and quantification of this output uncertainty is of paramount importance for making robust decisions in safety-critical applications[4-6]. Furthermore, the identification by sensitivity analysis of the model parameters and hypotheses that contribute the most to the output uncertainty plays a fundamental role in driving resource allocation for uncertainty reduction[7-10].

In this work, we consider the mathematical (black-box) model of a twin-jet aircraft described in Ref. 11, which includes twenty-one inputs and eight outputs. The inputs are uncertain and classified into three categories: (I) purely aleatory parameters modeled as random variables with fixed functional forms and known coefficients; (II) purely epistemic parameters modeled as fixed but unknown constants that lie within given intervals (contrary to Bayesian-based approaches such intervals are *not probabilistic*, i.e., they do not define a uniform probability density function); (III) mixed aleatory and epistemic parameters modeled as distributional probability boxes (p-boxes), i.e., as random variables with fixed functional form but epistemically-uncertain (non probabilistic *interval*) coefficients. Within this context, we tackle the following issues raised by the NASA Langley Multidisciplinary Uncertainty Quantification Challenge (MUQC)[11]: (A) improvement of the input uncertainty models (i.e., reduction of the corresponding epistemic uncertainties) based on experimental data; (B) sensitivity analysis to rank the importance of the inputs in contributing to output uncertainties; (C) propagation of the input uncertainties to the outputs; (D) extreme case analysis to identify the epistemic realizations that prescribe extreme values of two performance metrics of interest (i.e., the mean of the so-called worst-case requirement metric and the system failure probability).

In more detail, in task (A) the challengers provide 'real' empirical realizations of one of the model outputs; on the basis of this information the uncertainty models of five input parameters belonging to categories (II) and (III) have to be improved (i.e., the corresponding epistemic uncertainties reduced). This issue is here tackled within a constrained optimization framework. First, a free p-box (i.e., a couple of bounding upper and lower cumulative distribution functions-CDFs) for the output of interest is built by means of the empirical data provided: to this aim, a non-parametric approach based on the Kolmogorov-Smirnov (KS) confidence limits is considered[12-13]. Then, the improved (i.e., reduced) uncertainty model (in practice, range or space of variation) of the epistemically-uncertain parameters/coefficients is optimally determined as the one producing a distributional p-box for the output with the following (possibly conflicting) properties: (i) it contains the maximal 'amount' of epistemic uncertainty (here quantified by the area included between the corresponding upper and lower CDFs)[14-15]; (ii) it is contained by the non-parametric, free p-box constructed on the basis of data. Notice that the resulting reduced uncertainty model has the following characteristics: (i) it might *not* be a *connected* set; (ii) contrary to Bayesian-based approaches, it is *not* a *probabilistic* set. In this paper, only an *empirical* map of discrete *sampling points* belonging to the reduced set is generated, rather than a rigorous, mathematically defined set in the continuum of the epistemic uncertainty space.

The task of sensitivity analysis (B) is here tackled by resorting to two different conceptual settings[9]. In the first (namely, 'factor prioritization') we rank the category (II) and (III) 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. In the second (namely, 'factor fixing') we look for those parameters that can be assumed constant without affecting the output of interest. In order to address the first issue in the 'factor prioritization' setting, a novel sensitivity index is introduced in analogy with variance-based Sobol indices[9, 16-19]: in this view, the most important category (II) and (III) parameters in the ranking are those that give rise to the *highest expected reduction* in the amount of epistemic uncertainty contained in the outputs of interest when the corresponding epistemically-uncertain parameters/coefficients are considered *fixed constant* values (i.e., when the amount of their epistemic uncertainty is reduced to *zero*). Notice that the 'amount' of epistemic uncertainty is here defined in *different* ways according to the *different* requests by the challengers: in subproblem (B.1) we quantify it by the *area* included between the upper and lower CDFs of the model outputs of interest, whereas in subproblems (B.2) and (B.3) the challengers define it as the *length* of the intervals of two relevant performance metrics (i.e., the mean of the worst case requirement metric and the system failure probability, respectively). Instead,

in the 'factor fixing' setting sensitivity analysis aims at finding those parameters that *minimally* affect the outputs, i.e., that can be assumed to take on a fixed constant value without incurring in significant 'error': in this context, we quantify the 'error' as the *mismatch* (i.e., *lack* of *overlapping*) between the output quantities of interest obtained using the original uncertainty models and those produced by fixing one of the parameters to a constant value (again, depending on the subproblem the quantites of interest may be represented by the p-box *distributions* of the model output variables or by the *intervals* describing the epistemic uncertainty in the mean of the worst case requirement metric and in the system failure probability). This problem is solved within an optimization framework. In particular, for each parameter we *exhaustively* explore its *entire* range of variation to find the corresponding (constant) values that give rise to the *maximal mismatch* (i.e., maximal error) between the output quantities of interest. If such maximal error is sufficiently *small* (e.g., lower than 1% in the present paper), then there exists *no* realization of the parameter under analysis that affects appreciably the output: in other words, the parameter can be considered *not important* and can be thus *neglected* in the system model. In all the tasks related to sensitivity analysis, the original (black-box) mathematical model of the system is replaced by a fast-running, surrogate regression model based on Artificial Neural Networks (ANNs), in order to reduce the computational cost associated to the analysis[20-25]: in particular, the *computational time* is reduced by about *three orders of magnitude*.

Finally, tasks (C) and (D) are here tackled *together* by solving the (optimization) problem of identifying the values of the epistemically-uncertain coefficients of the category (II) and (III) parameters that yield the smallest and largest values (i.e., the ranges) of the two performance metrics defined above[26-29]; during the optimization search the (aleatory) uncertainty described by probability distributions is propagated by standard Monte Carlo Simulation (MCS)[30-31].

Finally, notice that all the tasks involved in the challenge require the solution of several nonlinear, constrained optimization problems, which are efficiently tackled by resorting to *heuristic* approaches (i.e., evolutionary algorithms): such methods deeply explore the search space by evaluating a large number (i.e., a population) of candidate solutions in order to find a near-optimal solution[32-33]. Notice that the population-based nature of such evolutionary algorithms allows an efficient exploration and characterization of abrupt and disconnected search spaces, which is the case of the present challenge.

The remainder of the paper is organized as follows. In Section II, the main characteristics of the mathematical system model under analysis are outlined; in Section III, the NASA Langley Multidisciplinary Uncertainty

Quantification Challenge (MUQC) Problems is addressed: the approaches adopted to tackle the problems of are described in detail and the results obtained are reported; finally, conclusions are drawn in the last Section.

II. The System

In Section II.A, we detail the mathematical model used to describe the dynamics of the Generic Transport Model (GTM), a remotely operated twin-jet aircraft developed by NASA Langley Research Center; in Section II.B, we characterize the aleatory and epistemic uncertainties affecting the input parameters to the model[11].

A. The Mathematical Model

We consider the mathematical model S that is employed to evaluate the performance of the multidisciplinary system under investigation and evaluate its suitability. Let $\boldsymbol{p} = \{p_i: i = 1, 2, \ldots, n_{inp} = 21\}$ be a vector of $n_{inp} = 21$ parameters in the system model whose value is uncertain and $\boldsymbol{d}$ a vector of design variables whose value can be set by the analyst (in the following, it is kept constant). Furthermore, let $\boldsymbol{g} = \{g_o: o = 1, 2, \ldots, n_{out} = 8\}$ be a set of $n_{out} = 8$ requirement metrics used to evaluate the system's performance. The values of $\boldsymbol{g}$ depends on both $\boldsymbol{p}$ and $\boldsymbol{d}$. The system is considered *requirement compliant* if it satisfies the set of inequality constraints $\boldsymbol{g} < \boldsymbol{0}$. For a fixed value of the design variables $\boldsymbol{d}$, the set of $\boldsymbol{p}$-points where $\boldsymbol{g} < \boldsymbol{0}$ is called the *safe domain*, while its complement set is called the *failure domain*[11].

The relationship between the inputs $\boldsymbol{p}$ and $\boldsymbol{d}$, and the output $\boldsymbol{g}$ is given by several functions, each representing a different subsystem or discipline. In particular, the function prescribing the output vector $\boldsymbol{g} = \{g_o: o = 1, 2, \ldots, n_{out} = 8\}$ of the multidisciplinary system is given by

$$g_o = f_o(\boldsymbol{x}, \boldsymbol{d}),\ o = 1, 2, \ldots, n_{out} = 8 \tag{1}$$

where $\boldsymbol{x} = \{x_j: j = 1, 2, \ldots, n_{int} = 5\}$ is a set of intermediate variables whose dependence on $\boldsymbol{p}$ is given by

$$x_1 = h_1(p_1, p_2, p_3, p_4, p_5) = h_1(\boldsymbol{p}^1) \tag{2}$$

$$x_2 = h_2(p_6, p_7, p_8, p_9, p_{10}) = h_2(\boldsymbol{p}^2) \tag{3}$$

$$x_3 = h_3(p_{11}, p_{12}, p_{13}, p_{14}, p_{15}) = h_3(\boldsymbol{p}^3) \tag{4}$$

$$x_4 = h_4(p_{16}, p_{17}, p_{18}, p_{19}, p_{20}) = h_4(\boldsymbol{p}^4) \tag{5}$$

$$x_5 = h_5(p_{21}) = h_5(\boldsymbol{p}^5) = p_{21}\,. \tag{6}$$

For the sake of compact notation, in (2)-(6) we call $\boldsymbol{p}^j$ the vector of the inputs to function $h_j(\cdot)$, j = 1, 2, …, 5: for example, $\boldsymbol{p}^3$ = {p_i: i = 11, 12, …, 15} in (4). Input parameters $\boldsymbol{p}$ = {p_i: i = 1, 2, …, n_{inp} = 21} are affected by uncertainties whose nature is characterized in the following Section II.B.

B. Aleatory and Epistemic Uncertainties in the Model Input Parameters

The uncertain parameters $\boldsymbol{p}$ = {p_i: i = 1, 2, …, n_{inp} = 21} are classified into three categories (Table 1): (I) purely aleatory parameters modeled as random variables with probability distributions of fixed functional form $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$ (resp., Cumulative Distribution Function-CDF $F^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$) and known coefficients $\boldsymbol{\theta}_i$ = {$\theta_{i,l}$: l = 1, 2, …, $n_{p,i}$, i = 1, 2, …, 21}, where $\theta_{i,l}$ is the l-th internal coefficient of the aleatory probability distribution $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$ of the i-th parameter and $n_{p,i}$ is the total number of internal coefficients of p_i: this probability model is irreducible (see parameters p_3, p_9, p_{11} and p_{19} in Table 1); (II) purely epistemically-uncertain parameters modeled as fixed but unknown constants that lie within given intervals Δ_{p_i}: these intervals are reducible as new information (e.g., data) about the corresponding parameter is gathered (see parameters p_2, p_6, p_{12} and p_{16} in Table 1); (III) mixed aleatory and epistemic parameters modeled as *distributional* probability boxes (p-boxes), i.e., as random variables with probability distributions of fixed functional form $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$ (resp., CDF $F^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$) but epistemically-uncertain coefficients $\boldsymbol{\theta}_i$. In this case, coefficients $\boldsymbol{\theta}_i$ are assumed to lie in bounded intervals $\boldsymbol{\Delta}_{\boldsymbol{\theta}_i}$ = { $\Delta_{\theta i,l}$: l = 1, 2, …, $n_{p,i}$, i = 1, 2, …, 21}, where $\Delta_{\theta i,l}$ is the range of the l-th internal coefficient of the aleatory probability distribution of the i-th parameter: again, these intervals can be reduced (see parameters p_1, p_4, p_5, p_7, p_8, p_{10}, p_{13}, p_{14}, p_{15}, p_{17}, p_{18}, p_{20} and p_{21} in Table 1). The distributional p-box for a generic parameter p_i is indicated as $PB^{p_i}(p_i) = \left\{F^{p_i}\left(p_i|\boldsymbol{\theta}_i\right): \boldsymbol{\theta}_i \in \boldsymbol{\Omega}_i\right\}$ and represents in practice a *bundle* of probability distributions with the *same shape* (e.g., exponential, beta, normal, …) but *different internal coefficients* (e.g., different values of the mean, variance, …). By way of example and only for illustration purposes, Figure 1 shows four CDFs belonging to the distributional p-box $PB^{p_1}(p_1)$ of parameter p_1 (dashed lines); also, the figure reports the *extreme upper and lower* CDFs, $\overline{F}^{p_1}(p_1) = \max_{\boldsymbol{\theta}_1 \in \boldsymbol{\Omega}_1}\left\{F^{p_1}\left(p_1|\boldsymbol{\theta}_1\right): \boldsymbol{\theta}_1 \in \boldsymbol{\Omega}_1\right\}$ and $\underline{F}^{p_1}(p_1) = \min_{\boldsymbol{\theta}_1 \in \boldsymbol{\Omega}_1}\left\{F^{p_1}\left(p_1|\boldsymbol{\theta}_1\right): \boldsymbol{\theta}_1 \in \boldsymbol{\Omega}_1\right\}$, $\forall p_1 \in \Re$, *bounding* the distributional p-box $PB^{p_1}(p_1)$ (solid lines). It is worth mentioning that when the uncertainty in a parameter is represented by a p-box, some quantities of interest, such as percentiles or exceedance probabilities, are *not* represented by *single point values*, but rather by *intervals*. For

example, with reference to Figure 1, the probability $P[p_1 > p_1^* = 0.9]$ that parameter p_1 exceeds $p_1^* = 0.9$ is given by $\left[[\overline{F}^{p_1}]^{-1}(p_1^*), [\underline{F}^{p_1}]^{-1}(p_1^*)\right] = [0.0072, 0.4318]$.

Notice that if the internal coefficients $\theta_{i,l}$, $l = 1, 2, \ldots, n_{p,i}$, of the distribution $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$ of parameter p_i are epistemically-independent (i.e., their uncertainty models are built using independent information sources, e.g., different experts, observers or data sets), then the entire (joint) space of variation $\boldsymbol{\Omega}_i$ of the coefficients vector $\boldsymbol{\theta}_i$ is given by the Cartesian product of the intervals $\Delta_{\theta i,l}$, i.e., $\boldsymbol{\Omega}_i = \prod_{l=1}^{n_{p,i}} \Delta_{\theta i,l}$.[2, 34-42] For example, referring to Table 1, the space of variation $\boldsymbol{\Omega}_1$ of the internal coefficients $\boldsymbol{\theta}_1 = [m, s^2]$ of the Beta distribution $q^{p_1}\left(p_1|\boldsymbol{\theta}_1\right) = \text{Beta}(m, s^2)$ of parameter p_1 is given by $\boldsymbol{\Omega}_1 = \Delta_{\theta 1,1} \times \Delta_{\theta 1,2} = \Delta_m \times \Delta_{s^2} = [3/5, 4/5] \text{ x } [1/50, 1/25]$. For the sake of compact notation, in the following we call $\boldsymbol{\theta}^{x_j}$ the vector of the epistemically-uncertain parameters/coefficients related to the input vector $\boldsymbol{p}^j$ to function $h_j(\cdot)$ and $\boldsymbol{\Omega}^{x_j}$ the corresponding (joint) space of variation. For example, with reference to $x_1 = h_1(\boldsymbol{p}^1) = h_1(p_1, p_2, p_3, p_4, p_5)$, we have $\boldsymbol{\theta}^{x_1} = [\boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5] = [m, s^2, p_2, \mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho]$ and $\boldsymbol{\Omega}^{x_1} = \prod_{i=1, i\neq 3}^{5} \boldsymbol{\Omega}_i = \Delta_m \text{ x } \Delta_{s^2} \text{ x } \Delta_{p_2} \text{ x } \Delta_{\mu_4} \text{ x } \Delta_{\sigma_4^2} \text{ x } \Delta_{\mu_5} \text{ x } \Delta_{\sigma_5^2} \text{ x } \Delta_\rho$. Finally, the vector of *all* the epistemically-uncertain parameters/coefficients contained in the *entire* system model S is indicated as $\boldsymbol{\theta}^{all} = [\boldsymbol{\theta}^{x_1}, \boldsymbol{\theta}^{x_2}, \boldsymbol{\theta}^{x_3}, \boldsymbol{\theta}^{x_4}, \boldsymbol{\theta}^{x_5}]$ and the corresponding (joint) space of variation as $\boldsymbol{\Omega}^{all} = \prod_{j=1}^{5} \boldsymbol{\Omega}^{x_j}$.

		Uncertainty model	
Symbol	**Category**	**Aleatory component**	**Epistemic component**
p_1	III	$q^{p_1}(p_1\|\boldsymbol{\theta}_1)$ = Beta(m, s^2), $m = E[p_1]$, $s^2 = V[p_1]$	$m \in \Delta_m$ = [3/5, 4/5], $s^2 \in \Delta_{s^2}$ = [1/50, 1/25]
p_2	II	/	$\Delta_{p_2} = [0,1]$
p_3	I	$q^{p_3}(p_3\|\boldsymbol{\theta}_3)$ = U[0, 1)	/
p_4, p_5	III	$q^{p_4}(p_4\|\boldsymbol{\theta}_4)$ = $N(\mu_4, \sigma_4^2)$, $\mu_4 = E[p_4]$, $\sigma_4^2 = V[p_4]$ $q^{p_5}(p_5\|\boldsymbol{\theta}_5)$ = $N(\mu_5, \sigma_5^2)$, $\mu_5 = E[p_5]$, $\sigma_5^2 = V[p_5]$ Correlation coefficient: ρ	$\mu_4, \mu_5 \in \Delta_{\mu_4}, \Delta_{\mu_5}$ = [–5, +5] $\sigma_4^2, \sigma_5^2 \in \Delta_{\sigma_4^2}, \Delta_{\sigma_5^2}$ = [1/400, 4] $\|\rho\| \le 1$, i.e., $\rho \in \Delta_\rho$ = [-1, 1]
p_6	II	/	$\Delta_{p_6} = [0,1]$
p_7	III	$q^{p_7}(p_7\|\boldsymbol{\theta}_7)$ = Beta(a_7, b_7)	$a_7 \in \Delta_{a_7}$ = [0.982, 3.537], $b_7 \in \Delta_{b_7}$ = [0.619, 1.080]
p_8	III	$q^{p_8}(p_8\|\boldsymbol{\theta}_8)$ = Beta(a_8, b_8)	$a_8 \in \Delta_{a_8}$ = [7.450, 14.093], $b_8 \in \Delta_{b_8}$ = [4.285, 7.864]
p_9	I	$q^{p_9}(p_9\|\boldsymbol{\theta}_9)$ = U[0, 1)	/
p_{10}	III	$q^{p_{10}}(p_{10}\|\boldsymbol{\theta}_{10})$ = Beta(a_{10}, b_{10})	$a_{10} \in \Delta_{a_{10}}$ = [1.520, 4.513], $b_{10} \in \Delta_{b_{10}}$ = [1.536, 4.750]
p_{11}	I	$q^{p_{11}}(p_{11}\|\boldsymbol{\theta}_{11})$ = U[0, 1)	/
p_{12}	II	/	$\Delta_{p_{12}} = [0,1]$
p_{13}	III	$q^{p_{13}}(p_{13}\|\boldsymbol{\theta}_{13})$ = Beta(a_{13}, b_{13})	$a_{13} \in \Delta_{a_{13}}$ = [0.412, 0.737], $b_{13} \in \Delta_{b_{13}}$ = [1.000, 2.068]
p_{14}	III	$q^{p_{14}}(p_{14}\|\boldsymbol{\theta}_{14})$ = Beta(a_{14}, b_{14})	$a_{14} \in \Delta_{b_{14}}$ = [0.931, 2.169], $b_{14} \in \Delta_{b_{14}}$ = [1.000, 2.407]
p_{15}	III	$q^{p_{15}}(p_{15}\|\boldsymbol{\theta}_{15})$ = Beta(a_{15}, b_{15})	$a_{15} \in \Delta_{b_{15}}$ = [5.435, 7.095], $b_{15} \in \Delta_{b_{15}}$ = [5.287, 6.954]
p_{16}	II	/	$\Delta_{p_{16}} = [0,1]$
p_{17}	III	$q^{p_{17}}(p_{17}\|\boldsymbol{\theta}_{17})$ = Beta(a_{10}, b_{10})	$a_{17} \in \Delta_{b_{17}}$ = [1.060, 1.662], $b_{17} \in \Delta_{b_{17}}$ = [1.000, 1.488]
p_{18}	III	$q^{p_{18}}(p_{18}\|\boldsymbol{\theta}_{18})$ = Beta(a_{10}, b_{10})	$a_{18} \in \Delta_{a_{18}}$ = [1.000, 4.266], $b_{18} \in \Delta_{b_{18}}$ = [0.553, 1.000]
p_{19}	I	$q^{p_{19}}(p_{19}\|\boldsymbol{\theta}_{19})$ = U[0, 1)	/
p_{20}	III	$q^{p_{20}}(p_{20}\|\boldsymbol{\theta}_{20})$ = Beta(a_{20}, b_{20})	$a_{20} \in \Delta_{a_{20}}$ = [7.530, 13.492], $b_{20} \in \Delta_{b_{20}}$ = [4.711, 8.148]
p_{21}	III	$q^{p_{21}}(p_{21}\|\boldsymbol{\theta}_{21})$ = Beta(a_{21}, b_{21})	$a_{21} \in \Delta_{a_{21}}$ = [0.421, 1.000], $b_{21} \in \Delta_{b_{21}}$ = [7.772, 29.621]

Table 1. Uncertain input parameters[11]. $E[\cdot]$ = expected value, $V[\cdot]$ = variance, m = mean of the Beta distribution of p_1, s^2 = variance of the Beta distribution of p_1, $N(\cdot,\cdot)$ = Normal distribution, μ_i = mean of the i-th Normal distribution, σ_i^2 = variance of the i-th Normal distribution, U[0, 1) = uniform distribution on [0, 1), a_i = location parameter of the i-th Beta distribution, b_i = scale parameter of the i-th Beta distribution

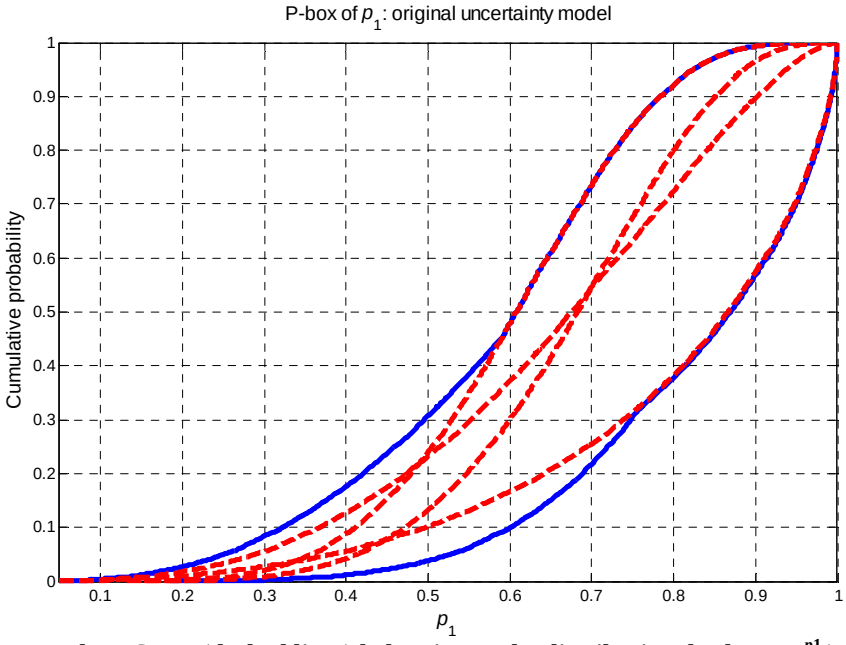

Figure 1. Four exemplary CDFs (dashed lines) belonging to the distributional p-box $PB^{p1}(p_1)$ of parameter p_1 (see Table 1); the extreme upper and lower CDFs, $\overline{F}^{p_1}(p_1)$ and $\underline{F}^{p_1}(p_1)$ bounding the corresponding distributional p-box are also shown as solid lines

III. Approaches and Solutions to the NASA Langley Multidisciplinary Uncertainty Quantification Challenge (MUQC) Problems

In the following, the approaches used to tackle the NASA Langley Multidisciplinary Uncertainty Quantification Challenge (MUQC) Problems are presented together with the corresponding results obtained: in particular, Sections III.A, III.B, III.C and III.D deals with Subproblems A (namely, Uncertainty characterization), B (namely, Sensitivity Analysis), C (namely, Uncertainty propagation) and D (namely, Extreme case analysis).

A. Subproblem (A): Uncertainty Characterization

In this subproblem, the main task of interest is as follows:[11] using a vector of observations of x_1 (2) (provided by the challengers), improve the uncertainty models of category (II) and (III) parameters p_1, p_2, p_4 and p_5, i.e., reduce the corresponding epistemic uncertainty. Notice that the observations of x_1 (2) correspond to its 'true uncertainty model', i.e., a model where p_1 is a well defined Beta random variable, p_2 is a constant and p_4 and p_5 are described by two possibly correlated Normal random variables. In Section IV.A.1, the approach adopted is illustrated in detail; in Section IV.A.2, the results of the application of the method to Subproblem (A) are reported.

1. The Proposed Approach

This subproblem is tackled by performing the following two main conceptual steps:

1) a free p-box, i.e., a couple of bounding upper and lower Cumulative Distribution Functions (CDFs) for the intermediate variable of interest $x_1 = h_1(p_1, p_2, p_3, p_4, p_5)$ (2) is built by means of a set of empirical observations of x_1: to this aim, a *non-parametric* approach based on the Kolmogorov-Smirnov confidence limits is considered;[12-13]

2) the improved (i.e., reduced) ranges of (i) the epistemically-uncertain coefficients $\boldsymbol{\theta}_i$ of the probability distributions $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$ of the category (III) input parameters p_i, i = 1, 4, 5 (Table 1) and (ii) the epistemically-uncertain category (II) input parameter p_2 (Table 1) are optimally determined as those producing a *distributional* p-box $PB^{x_1}(x_1)$ for x_1 that is *coherent* with the data available, i.e. with the non-parametric *free* p-box built at step 1. above: in particular, we look for the distributional p-box containing *all* the CDFs of x_1 that are *bounded* by the non-parametric, free p-box constructed on the basis of data.

In more detail, if a vector $\boldsymbol{x}_1^d$ of n_d observations of random variable x_1 (2) is available, an empirical CDF $\hat{F}^{x_1,n_d}(x_1)$ for x_1 can be constructed; however, the shape of this CDF is affected by significant "sampling uncertainty", which arises because of the finiteness (and typically limitedness) of the random sample employed.[12-13] We account for this uncertainty by building the Kolmogorov-Smirnov (KS) confidence limits $\overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)$ and $\underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)$ to provide upper and lower bounds, respectively, to the empirical CDF $\hat{F}^{x_1,n_d}(x_1)$ with a statistical confidence of $100\cdot(1-\alpha)$%:[43-44]

$$\begin{aligned}\overline{F}^{x_1}(x_1) &= \overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1) = \min\left(1, \hat{F}^{x_1}(x_1) + D(\alpha, n_d)\right)\\ \underline{F}^{x_1}(x_1) &= \underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1) = \max\left(0, \hat{F}^{x_1}(x_1) - D(\alpha, n_d)\right)\end{aligned} \tag{7}$$

where $D(\alpha, n_d)$ is the one-sample Kolmogorov-Smirnov critical statistic for intrinsic (two-sided) hypotheses testing for confidence level $100\cdot(1-\alpha)$% and sample size n_d. "Analogous to simple confidence intervals around a single number, these are bounds on a statistical distribution as a whole. As the number of samples becomes very large, these confidence limits would converge to the empirical distribution function (although the convergence is

rather slow)"[13].‡ It is worth recalling that the critical statistic $D(\alpha, n_d)$ is computed as $K(\alpha)/\sqrt{n_d}$, where $K(\alpha)$ is the $(1 - \alpha)$-th quantile of the Kolmogorov distribution K, i.e., the value $K(\alpha)$ such that $P[K \le K(\alpha)] = 1 - \alpha$.[43-44] Tabled values of $D(\alpha, n_d)$ can be found in Ref. 45 for α = 0.1, 0.05, 0.02 and 0.01, and n_d = 1, 2, …, 100: these values are the result of a synthesis, development and improvement of the research work by Refs. 43, 44 and 46 and are still in use today (such values are available in many softwares, e.g., MATLAB®, that is used in the present work).

Given the empirical bounds $\overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)$ and $\underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)$ (7) on $x_1 = h_1(p_1, p_2, p_3, p_4, p_5)$ (2), the improved uncertainty models (i.e., the reduced *sets* describing the epistemic uncertainty) of the corresponding category (II) and (III) input parameters p_1, p_2, p_4 and p_5 could be *rigorously* obtained by *exhaustively* searching for *all* the possible combinations of values of the epistemically-uncertain coefficients ($\boldsymbol{\theta}_i$, i = 1, 4, 5) and parameters (p_2) that produce a distributional p-box $PB^{x_1}(x_1)$ for x_1 *coherent* with the available data, i.e., with the empirical bounds $\overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)$ and $\underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)$ (7). In other words, we should look for the distributional p-box $PB^{x_1}(x_1)$ containing *all* the CDFs of x_1 that are *bounded* everywhere by the non-parametric, free p-box $[\underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1),\ \overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1)]$ (7) constructed on the basis of data.[14-15] This amounts to solving the following problem of *feasible region* identification:

$$\begin{aligned} &\text{Find all } \boldsymbol{\theta}^{x_1} \in \boldsymbol{\Omega}^{x_1}, \text{i.e., } \boldsymbol{\theta}_1 = [m, s^2] \in \boldsymbol{\Omega}_1,\ p_2 \in \Delta_{p_2}, \boldsymbol{\theta}_4 = [\mu_4, \sigma_4^2, \rho] \in \boldsymbol{\Omega}_4, \boldsymbol{\theta}_5 = [\mu_5, \sigma_5^2, \rho] \in \boldsymbol{\Omega}_5: \\ &\underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1) \le F^{x_1}(x_1 | \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5) = F^{x_1}(x_1 | \boldsymbol{\theta}^{x_1}) \le \overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1),\ \forall x_1 \in \Re \end{aligned} \tag{8}$$

where $F^{x_1}(x_1 | \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5)$ indicates the CDF of $x_1 = h_1(p_1, p_2, p_3, p_4, p_5)$ (2) obtained when the (epistemically-uncertain) internal coefficients of the probabability distributions of the corresponding category (III) input parameters p_1, p_4 and p_5 and category (II) input parameter p_2 are fixed to constant values within their ranges $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_4$, $\boldsymbol{\Omega}_5$ and Δ_{p_2}, respectively. The *subset* $\boldsymbol{\Omega}^{x_1}_{n_d}$ of $\boldsymbol{\Omega}^{x_1}$ – i.e., the *set* containing *all* the values of $\boldsymbol{\theta}^{x_1} \in \boldsymbol{\Omega}^{x_1}$ for which $\underline{F}^{x_1}_{KS(\alpha,n_d)}(x_1) \le F^{x_1}(x_1 | \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5) = F^{x_1}(x_1 | \boldsymbol{\theta}^{x_1}) \le \overline{F}^{x_1}_{KS(\alpha,n_d)}(x_1),\ \forall x_1 \in \Re$ – represents the requested feasible region, i.e., the *improved, reduced* uncertainty model for parameters $\boldsymbol{\theta}^{x_1} = [\boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5] = [m, s^2, p_2, \mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho]$. With respect

‡ It is worth mentioning that similar confidence bounds on x_1 could be obtained also by well-known *resampling* techniques (such as *bootstrap*) that are commonly used to "build confidence" in statistical estimates and to quantify the effect of sampling uncertainty (in particular, in presence of small-sized datasets). This has been verified by the authors also in the present case, but not shown here for brevity sake.

to that, it is very important to remember that $\boldsymbol{\Omega}_{n_d}^{x_1}$ is found using a *single* data set $\boldsymbol{x}_1^d$, which introduces an *epistemic dependence* between the values that $\boldsymbol{\theta}_1$, p_2, $\boldsymbol{\theta}_4$ and $\boldsymbol{\theta}_5$ may assume: thus, differently from the initial $\boldsymbol{\Omega}^{x_1}$, in general $\boldsymbol{\Omega}_{n_d}^{x_1}$ *cannot* be expressed as the Cartesian product of the separate ranges of variation of $\boldsymbol{\theta}_1$, p_2, $\boldsymbol{\theta}_4$ and $\boldsymbol{\theta}_5$. In passing, also notice that from a strictly mathematical viewpoint, solving problem (8) is equivalent to finding all the CDFs $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$ that result in a p-value p_{val} larger than or equal to α in a KS *statistical test* with the empirical CDF $\hat{F}^{x_1,n_d}\left(x_1\right)$ constructed with *one sample* $\boldsymbol{x}_1^d$ of size n_d. In particular, we test the "null hypothesis" that sample $\boldsymbol{x}_1^d$ of size n_d comes from distribution $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$: the corresponding test statistic is then the well-known Kolmogorov-Smirnov statistic $D\left(n_d\right) = \max_{x_1}\left|\hat{F}^{x_1,n_d}\left(x_1\right) - F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)\right|$ (i.e., the maximal 'vertical' distance between the two CDFs). It is worth recalling that the p-value p_{val} is used in the context of "null hypothesis testing" in order to quantify the idea of statistical significance of evidence. More rigorously, the p-value is the probability of obtaining a test statistic result D at least as extreme or as close to the one that is actually observed ($D\left(n_d\right)$), assuming that the "null hypothesis" is true (i.e., assuming that sample $\boldsymbol{x}_1^d$ actually comes from $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$): in this case, $p_{val} = P[D \geq D(n_d)]$. When the p-value p_{val} turns out to be less than a predetermined significance level α, then the "null hypothesis" is rejected: actually, such an outcome indicates that the observed result (i.e., the empirical CDF $\hat{F}^{x_1,n_d}\left(x_1\right)$ constructed with sample $\boldsymbol{x}_1^d$) would be highly unlikely if the "null hypothesis" was true (i.e., if $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$ was the real underlying distribution of x_1). Finally, it is worth admitting that in the present case *also* the "null hypothesis" distribution $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$ is obtained by plain random *sampling*, i.e., by propagating $N = 100000$ realizations of parameters $p_1, p_2, \ldots, p_5$ through the model $x_1 = h_1(p_1, p_2, p_3, p_4, p_5)$ (2): thus, a two-sample KS test should be rigorously carried out instead of a one-sample test. However, since N is very large, then "null hypothesis" CDF $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$ can be considered as the "reference" one with acceptable approximation: actually, as verified by the authors but not shown here for brevity sake, the results obtained in the two different cases are practically identical.

With respect to the approach proposed, it is worth mentioning that[12-13]: (i) KS bounds are *distribution-free* constructions, i.e., they do *not* require any *knowledge* about the *real* shape of the underlying distribution (which is the case for the intermediate variable x_1 under analysis); (ii) KS limits require the assumption that the samples $\boldsymbol{x}_1^d$

are *independent* and *identically distributed* (which is verified for variable x_1); and (iii) KS limits are *not certain* bounds, but only *statistical* ones: the associated statistical statement is that $100\cdot(1-\alpha)$% of the times such bounds are constructed from n_d random samples, they will totally enclose the *true* distribution $F^{x_1}(x_1)$ of x_1.

In this paper, we tackle problem (8) by resorting to a *population-based, heuristic* optimization technique, i.e., a Genetic Algorithm (GA)[32,33]. In the present case, the search space is represented by the entire space of variation $\boldsymbol{\Omega}^{x_1}$ of the epistemically-uncertain coefficients/parameters $\boldsymbol{\theta}^{x_1} = [\boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5]$ and the objective function to optimize (in particular, to maximize) is the p-value p_{val} obtained in a statistical KS test between the CDFs $F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right)$ and $\hat{F}^{x_1,n_d}(x_1)$:

$$\begin{aligned} &\text{Find } \boldsymbol{\theta}^{x_1} \in \boldsymbol{\Omega}^{x_1}, \text{i.e., } \boldsymbol{\theta}_1 = [m, s^2] \in \boldsymbol{\Omega}_1,\ p_2 \in \Delta_{p_2}, \boldsymbol{\theta}_4 = [\mu_4, \sigma_4^2, \rho] \in \boldsymbol{\Omega}_4, \boldsymbol{\theta}_5 = [\mu_5, \sigma_5^2, \rho] \in \boldsymbol{\Omega}_5: \\ &\bar{p}_{val} = \max_{\boldsymbol{\theta}^{x_1}} \{p_{val}\} = \max_{\boldsymbol{\theta}^{x_1}} \{P[D \geq D(n_d)]\}, \quad D(n_d) = \max_{x_1} \left| \hat{F}^{x_1,n_d}(x_1) - F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5\right) \right| \end{aligned} \tag{9}$$

In the present paper, GAs are tailored to the particular problem of identifying a feasible region: in particular, during the GA evolution towards the optimum, *all* the candidate solutions $\boldsymbol{\theta}^{x_1}$ that are found to satisfy the property in (8) are stored; at the end of the search, the *ensemble* of the *feasible* solutions found and stored during the optimization search constitute an '*empirical map*' of the *feasible region* $\boldsymbol{\Omega}_{n_d}^{x_1}$. Notice that the resulting (empirical) reduced uncertainty model has the following characteristics: (i) it might *not* be a *connected* set; (ii) contrary to Bayesian-based approaches, it is *not* a *probabilistic* set. Several considerations are in order with respect to the proposed approach. In this subproblem, GAs are *not* used with the *main* purpose of identifying a global optimum (i.e., $\bar{p}_{val}$): instead, their population-based nature and their genetic operators (relying on the criterion of survival of the fittest) are rather exploited for intelligently and thoroughly exploring the entire space of variation $\boldsymbol{\Omega}^{x_1}$ of $\boldsymbol{\theta}^{x_1} = [\boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5]$ in order to find as many feasible candidates as possible (and, thus, to make the 'empirical map' of the feasible region $\boldsymbol{\Omega}_{n_d}^{x_1}$ as complete and reliable as possible). Although this is *not* the traditional *intended* use of GAs, applications in this direction can be found in the literature, see, e.g., Ref. 47. In addition, it has to be acknowledged that the proposed approach *cannot* solve task (8) in a rigorous mathematical way. Actually, we *cannot* find *all* the combinations of $\boldsymbol{\theta}^{x_1} = [\boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5]$ for which the property in (8) holds, but rather we are *only* able to find *some* combinations by means of a GA, used in this case as an "intelligent" sampling approach. Given that we can only

identify a *finite* number of combinations (but we do not know what happens in between sampling points), we cannot prescribe mathematically the *set* $\boldsymbol{\Omega}_{n_d}^{x_1}$ in the *continuum* of space $\boldsymbol{\Omega}^{x_1}$ that has *infinitely* many elements. In facts, the property in (8) provides only the means to *test* the *membership* of a candidate $\boldsymbol{\theta}^{x_1}$ to $\boldsymbol{\Omega}_{n_d}^{x_1}$, but *not* the means to *calculate* mathematically the desired set. In order to do that, set bounding approaches, such as those presented in Refs. 26, 27, 28 and 29, should be adopted. On the other hand, it has to be also considered that this limitation does *not* absolutely impair the quality and validity of the results of the following subproblems of the challenge. Actually, all the tasks related to sensitivity analysis (III.B), uncertainty propagation (III.C) and extreme case analysis (III.D) are based on a GA optimization search within the continuum of space $\boldsymbol{\Omega}^{x_1}$. In this framework, the identification of only those solutions that belong to the (mathematically not prescribed) set $\boldsymbol{\Omega}_{n_d}^{x_1}$ is guaranteed by introducing the property in (8) as a *hard constraint* in the GA: only those candidates that satisfy such property are retained in the genetic evolution, whereas the others are discarded.

2. Application Results

Figure 2 top left shows the empirical CDF $\hat{F}^{x_1,n_{d1}}(x_1)$ (dot-dashed lines) built using a vector $\boldsymbol{x}_1^{d1}$ of $n_d = n_{d1} = 25$ real observations of x_1 (provided by the challengers) and the corresponding KS bounds $\overline{F}_{KS(\alpha,n_{d1})}^{x_1}(x_1)$ and $\underline{F}_{KS(\alpha,n_{d1})}^{x_1}(x_1)$ obtained with $\alpha = 0.01$ (resp., confidence $1 - \alpha = 0.99$) (solid lines). In addition, the figure reports the *extreme* upper and lower CDFs, $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$, bounding the distributional p-box of x_1 (i.e., $\overline{F}^{x_1}(x_1) = \max\{PB^{x_1}(x_1)\}$ and $\underline{F}^{x_1}(x_1) = \min\{PB^{x_1}(x_1)\}$, $\forall x_1 \in \Re$), before (dashed lines) and after (dotted lines) the improvement of the input parameters uncertainty model. It can be see that the area contained between the bounding upper and lower CDFs $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$ is significantly reduced; in particular, it is 0.2407 and 0.1860 before and after the update of the input uncertainty models, respectively, which means a reduction by 22.73% in the epistemic uncertainty of x_1.

In order to validate the results obtained, a *new* empirical CDF $\hat{F}^{x_1,n_{d2}}(x_1)$ is built using a *new* vector $\boldsymbol{x}_1^{d2}$ of $n_d = n_{d2} = 25$ real observations of x_1, extracted from the pool of $n_{d3} = 50$ data; then, KS statistical tests are performed between $\hat{F}^{x_1,n_{d2}}(x_1)$ and the CDFs belonging the 'updated' distributional p-box of x_1, $PB^{x_1,n_{d1}}(x_1) = \left\{F^{x_1}\left(x_1 \middle| \boldsymbol{\theta}^{x_1}\right): \boldsymbol{\theta}^{x_1} \in \boldsymbol{\Omega}_{n_{d1}}^{x_1}\right\}$. In more detail, two GA searches are carried out within the updated space $\boldsymbol{\Omega}_{n_{d3}}^{x_1}$ to calculate the maximum and minimum p-values, respectively, resulting from these KS statistical tests:[48] the corresponding values turn out to be 0.9837 (i.e., larger than the test significance level $\alpha = 0.01$) and $3 \cdot 10^{-4}$ (i.e., lower than the test

significance level α = 0.01) respectively. Taking as reference the smallest p-value within the reduced epistemic space (i.e., the value for which the null-hypothesis is the weakest), the reduced uncertainty model $\boldsymbol{\Omega}^{x_1}_{n_{d1}}$ would *not* be validated. With respect to that, for illustration purposes Figure 2 bottom shows: (i) the KS bounds $\overline{F}^{x_1}_{KS(\alpha,n_{d1})}(x_1)$ and $\underline{F}^{x_1}_{KS(\alpha,n_{d1})}(x_1)$ obtained using α = 0.01 (resp., confidence 1 – α = 0.99) and the data $\boldsymbol{x}_1^{d1}$ employed to improve the uncertainty model (solid lines); (ii) the extreme upper and lower CDFs of the p-box of x_1 after the improvement of the input parameters uncertainty model (dotted lines); (iii) the KS bounds $\overline{F}^{x_1}_{KS(\alpha,n_{d2})}(x_1)$ and $\underline{F}^{x_1}_{KS(\alpha,n_{d2})}(x_1)$ for x_1 obtained using α = 0.01 and the vector $\boldsymbol{x}_1^{d2}$ of validation data (dashed lines). It can be seen that for x_1 ranging within [0.1, 0.3], a consistent part of the 'improved' p-box of x_1 (dotted lines) 'lies outside' the KS bounds of the validation data set (dashed lines): in other words, some CDFs of x_1 are *not* bounded everywhere by the $(1 - \alpha)\cdot 100\%$ = 99% confidence limits associated to the validation dataset $\boldsymbol{x}_1^{d2}$ (correspondingly, the p-values of the related KS statistical tests will be smaller than α = 0.01). A possible explanation for this lack of model validation is as follows. It can be observed that the two sets of data provided by the challengers, $\boldsymbol{x}_1^{d1}$ and $\boldsymbol{x}_1^{d2}$, are concentrated in different ranges. The first dataset $\boldsymbol{x}_1^{d1}$ (used to improve the uncertainty model) is mostly concentrated within [0, 0.1] and [0.3, 0.4], whereas a large part of the second dataset $\boldsymbol{x}_1^{d2}$ (used to validate the model) is located in the range [0.05, 0.2]. Thus, it is not unexpected that a model calibrated by data lying mostly in [0, 0.1] $\cup$ [0.3, 0.4] fails to "describe the uncertainty" in data mostly concentrated in [0.05, 0.2] (correspondingly, as expected and highlighted above, the maximal discrepancy between the 'improved' p-box of x_1 and the KS bounds of the validation data set $\boldsymbol{x}_1^{d2}$ is observed for $x_1 \in$ [0.1, 0.3] where the calibration dataset $\boldsymbol{x}_1^{d1}$ is 'poorer' of evidence). Finally, in order to have a very rough measure of the discrepancy between the two p-boxes, we compute the percentage fraction of the area of the 'improved' p-box of x_1 that does *not* overlap with the KS bounds of the validation data set $\boldsymbol{x}_1^{d2}$: this fraction turns out to be only 7.29%.

Then, the uncertainty models of parameters p_1, p_2, p_4 and p_5 are further improved by using all the $n_{d3} = n_{d1} + n_{d2}$ = 50 data available, $\boldsymbol{x}_1^{d3}$ = [$\boldsymbol{x}_1^{d1}$, $\boldsymbol{x}_1^{d2}$]. As before, Figure 2 top right shows the CDF $\hat{F}^{x_1,n_{d3}}(x_1)$ (dot-dashed lines), the corresponding KS bounds $\overline{F}^{x_1}_{KS(\alpha,n_{d3})}(x_1)$ and $\underline{F}^{x_1}_{KS(\alpha,n_{d3})}(x_1)$ for α = 0.01 (solid lines) and the extreme upper and lower CDFs, $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$, bounding the distributional p-box of x_1 (i.e., $\overline{F}^{x_1}(x_1) = \max\{PB^{x_1}(x_1)\}$ and

$\underline{F}^{x_1}(x_1) = \min\{PB^{x_1}(x_1)\}$, $\forall x_1 \in \Re$), before (dashed lines) and after (dotted lines) the parameters update. In this case, the area included between the bounding CDFs is 0.1409, which means a reduction of 41.46% in the epistemic uncertainty of x_1 relative to the initial condition and a reduction of 24.25% relative to the results obtained using $n_d = n_{d1} = 25$ data: thus, with respect to an increase of 100% in the size of the data set (i.e., from $n_{d1} = 25$ to $n_{d3} = 50$), we obtain a relative improvement in the uncertainty model (i.e., a reduction in its epistemic uncertainty) of *only* 24.25%.

Finally, only for illustration purposes Figure 3 depicts two exemplary scatterplots representing the 'empirical maps' of the two-dimensional projections on the plans m-μ_4 (left) and m-μ_5 (right) of the improved (i.e., reduced) *joint* eight-dimensional space of variation $\boldsymbol{\Omega}^{x_1}_{n_{d3}}$ of the epistemically-uncertain coefficients/parameters $\boldsymbol{\theta}^{x_1} = [\boldsymbol{\theta}_1, p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5] = [m, s^2, p_2, \mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho]$, obtained after the update carried out by means of the data set $\boldsymbol{x}_1^{d3} = [\boldsymbol{x}_1^{d1}, \boldsymbol{x}_1^{d2}]$ (of size $n_{d3} = 50$). It is worth noting the *epistemic dependence* between the estimates of the epistemically-uncertain coefficients that is generated by the update of the corresponding uncertainty models by means of the *same* data set: differently from the initial condition where coefficients m and μ_4 were allowed to range within the corresponding intervals Δ_m and Δ_{μ_4} with no restrictions (Table 1), now it is not possible to have, e.g., low values of m and low values of μ_4 at the same time. Notice that these empirical maps have been generated by GAs and they contain approximately 500000 points. In order to avoid that the patterns observed are the result of the manner GA searches for the optimum (and not of the true dependency among variables), the following strategies have been implemented: (i) GA is *repeated* several times (say, ten times) with different random seeds and different settings of the GA operations (e.g., different crossover points and mutation rates), and an *approximate* feasible region is found and recorded for *each* repetition (as the number of repetitions of GA is increased, the approximate feasible regions approach the true feasible regions); (ii) the capability of GA of thoroughly exploring the entire search space (technically speaking, of maintaining a high "genetic diversity" in the population of candidate solutions) is guaranteed by a proper setting of its parameters, mainly based on the experience of the authors in the use of GAs[49,50]: for example, high population sizes (i.e., $N_{pop} = 200$) and high mutation rates (i.e., $p_{mut} = 0.025$) are employed; (iii) since in the present subproblem A the *main* purpose of the GA search is that of finding *many* feasible solutions instead of a *single* global optimum, the GA evolution is stopped only when a certain (large) number of generations (e.g., $N_{gen} = 500$) is achieved. Finally, the validity of these GA-based maps has been further checked by

generating about 500000 samples belonging to $\boldsymbol{\Omega}_{n_{d3}}^{x_1}$ by a *standard sampling method*: as verified by the authors, but not shown here for brevity, the resulting pattern of dependence is almost identical to the one produced by GAs.

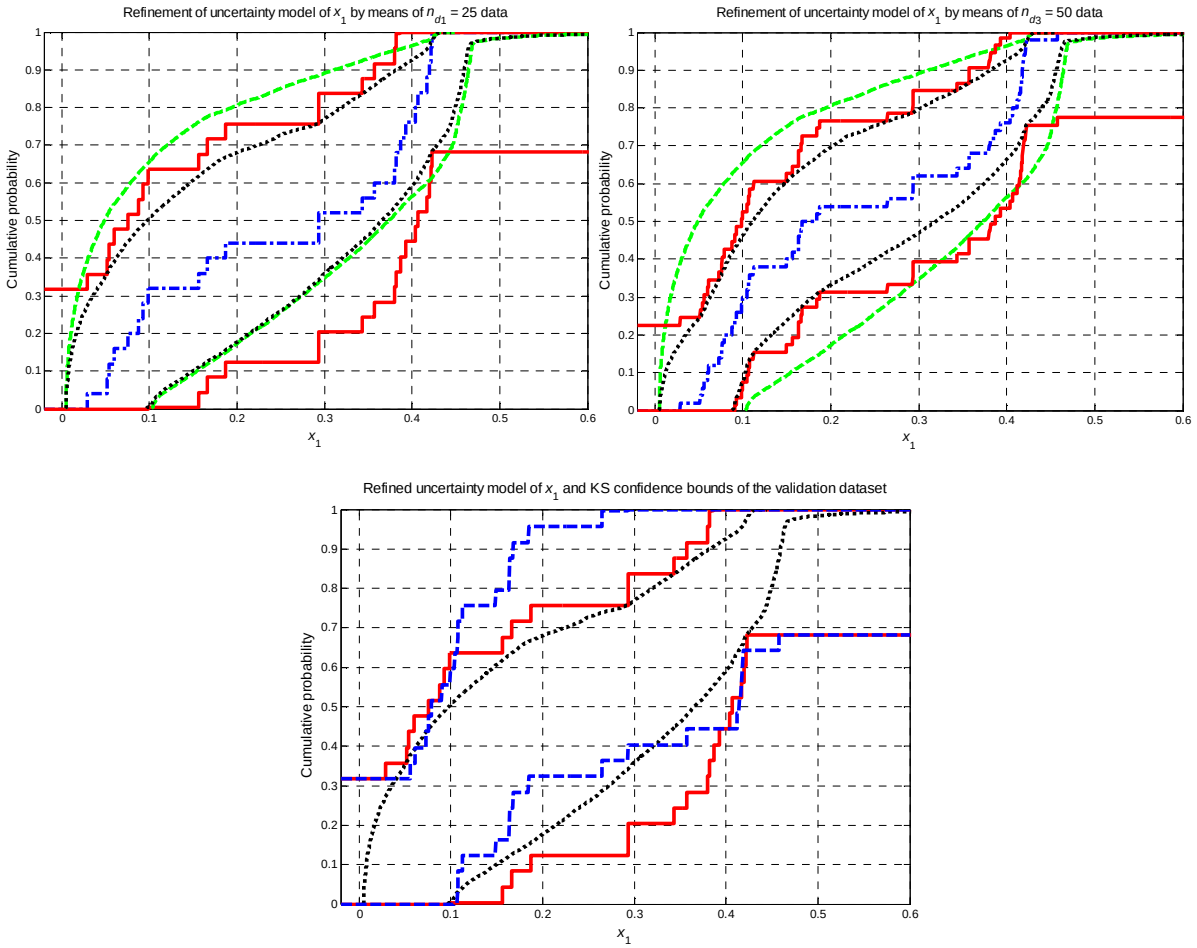

Figure 2. Empirical CDF $\hat{F}^{x_1,n_d}(x_1)$ (dot-dashed lines) built using a vector x_1^d of n_d real observations of x_1, the corresponding KS bounds $\overline{F}_{KS(\alpha,n_d)}^{x_1}(x_1)$ and $\underline{F}_{KS(\alpha,n_d)}^{x_1}(x_1)$ for α = 0.01 (resp., confidence 1 – α = 0.99) (solid lines), the extreme upper and lower CDFs, $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$, bounding the distributional p-box of x_1 (i.e., $\overline{F}^{x_1}(x_1) = max\{PB^{x_1}(x_1)\}$ and $\underline{F}^{x_1}(x_1) = min\{PB^{x_1}(x_1)\}$, $\forall x_1 \in \Re$), obtained before (dashed lines) and after (dotted lines) the improvement of the input parameters uncertainty model by means of $n_d = n_{d1}$ = 25 (top, left) and n_{d3} = 50 (top, right) data. Bottom: KS bounds $\overline{F}_{KS(\alpha,n_d)}^{x_1}(x_1)$ and $\underline{F}_{KS(\alpha,n_d)}^{x_1}(x_1)$ obtained with α = 0.01 (resp., confidence 1 – α = 0.99) for the calibration and validation datasets x_1^d (solid lines) and x_1^{d2} (dashed lines); the improved p-box of x_1 (dotted lines) is also shown

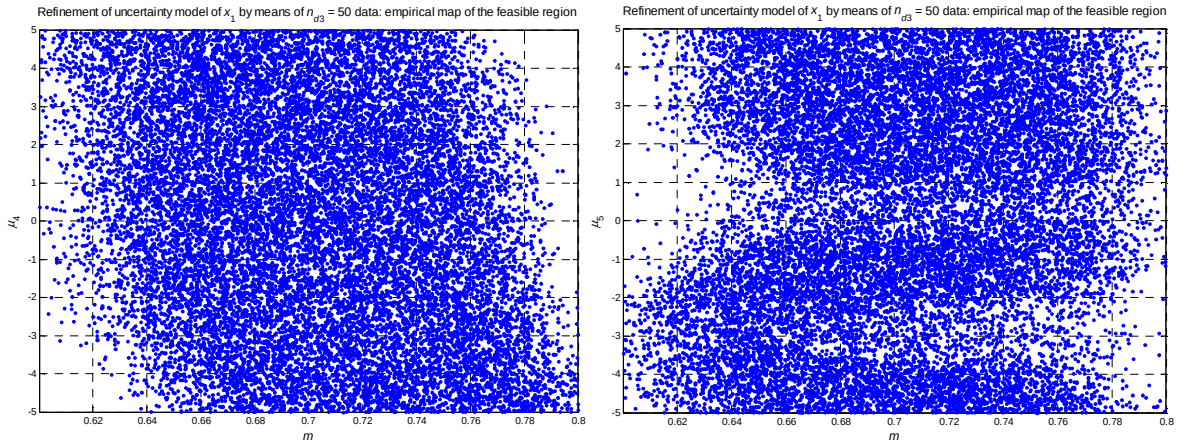

Figure 3. Scatterplots represenenting 'empirical maps' of the two-dimensional projections on the plans m-μ_4 (left) and m-μ_5 (right) of the improved (i.e., reduced) joint eight-dimensional space of variation $\Omega^{x_1}_{n_{d3}}$ of the epistemically-uncertain coefficients/parameters $\theta^{x_1} = [\theta_1, p_2, \theta_4, \theta_5] = [m, s^2, p_2, \mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho]$ obtained after the update carried out by means of the data set $x_1^{d3} = [x_1^{d1}, x_1^{d2}]$ of size $n_{d3} = 50$

B. Subproblem (B): Sensitivity Analysis

Sensitivity analysis is the general term for a systematic study of how the inputs to a model influence the results of the model. Sensitivity analyses are conducted for two fundamental reasons: (i) to focus future empirical studies so that effort might be expended to improve estimates of inputs that would lead to the most improvement in the estimates of the outputs, and (ii) to generally understand how the conclusions and inferences drawn from an assessment depend on its inputs (and on the basis of the results, possibly simplify or even remove from the model those inputs that turn out to be less influential)[14].

In this light, two different types of analysis are here performed: in the first (namely, 'factor prioritization' analysis), the objective is to identify those parameters $\boldsymbol{p}$ whose *epistemic* uncertainty contributes more to the 'amount' of *epistemic* uncertainty contained in some output quantities of interest: in other words, we try to rank the category (II) and (III) input parameters according to degree of reduction in the output epistemic uncertainty which one could hope to obtain by refining their uncertainty models (i.e., by reducing the epistemic uncertainty associated to them) (Section III.B.1). In the second (namely, 'factor fixing' analysis), determination has to be made as to whether these output quantities of interest are sufficiently *insensitive* to any given parameter such that that parameter can be assumed to take on a fixed constant value without incurring in significant errors: in other words, we aim at finding those parameters that *minimally* affect the outputs (Section III.B.2). In all the analyses, the first five parameters $\{p_i: i = 1, 2, \ldots, 5\}$ are modeled according to the results from task (A.3) (Section III.A.2), i.e., $\theta^{x_1} = [\theta_1,$

$p_2, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5] = [m, s^2, p_2, \mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho] \in \boldsymbol{\Omega}^{x_1}_{n_{d3}}$, whereas the remaining sixteen parameters $\{p_i: i = 6, 7, \ldots, 21\}$ are modeled according to Table 1 (Section II.B): the entire space of variation of all the epistemically-uncertain parameters/coefficients $\boldsymbol{\theta}^{all}$ is then $\boldsymbol{\Omega}^{all} = \boldsymbol{\Omega}^{x_1}_{n_{d3}} \text{x} \prod_{j=1}^{4} \boldsymbol{\Omega}^{x_j}$.

1. Sensitivity Analysis in a 'Factor Prioritization' Setting

In Section III.B.1.1, the general approach adopted to rank category (II) and (III) parameters according to their contribution to output epistemic uncertainty is illustrated in detail; in Section III.B.1.2, the results of the application of the method to the tasks of Subproblem (B) are reported.

1.1 The Proposed Approach

The use of sensitivity analysis to learn where focusing future empirical efforts would be most productive requires estimating the value of additional (hypothetical) empirical information. Of course, the value of information not yet observed cannot be measured, but it can perhaps be predicted. One strategy to this end is to assess how much less epistemic uncertainty the model outputs of interest would have if extra knowledge about an input were available. This might be done by comparing the epistemic uncertainty before and after '*pinching*' an input, i.e. replacing it with a value without (or with less) eistemic uncertainty. Of course, one does not generally know the correct value with certainty, so this replacement must be conjectural in nature. To pinch a parameter/coefficient means to hypothetically reduce its uncertainty for the purpose of the thought-experiment. The experiment asks what would happen if there were less epistemic uncertainty about this number. 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 final result[14].

In more detail, let $U_{\boldsymbol{p}}(Q)$ be an indicator of the 'amount' of epistemic uncertainty contained in a generic quantity Q of interest to the analysis. The subscript '$\boldsymbol{p}$' suggests that indicator $U_{\boldsymbol{p}}(Q)$ is computed over *all* the input parameters $\boldsymbol{p}$ (and over the space of variation $\boldsymbol{\Omega}^{all}$ of *all* the corresponding epistemically-uncertain internal coefficients $\boldsymbol{\theta}^{all}$). We want to assess the effect that a refinement of the uncertainty model of the generic input p_i (i.e., a reduction in its epistemic uncertainty) has on the amount of epistemic uncertainty $U_{\boldsymbol{p}}(Q)$ of Q. For the sake of notation generality, let $\boldsymbol{\varphi}_i \in \boldsymbol{\Omega}_i$ be the epistemically-uncertain 'factor' in the uncertainty model of parameter p_i:

thus, if p_i is a category (II) parameter (e.g., p_2 or p_{12}), then simply $\boldsymbol{\varphi}_i = p_i$; instead, if p_i is a category (III) parameter (e.g., p_1, p_4 or p_5), then $\boldsymbol{\varphi}_i$ represents the vector $\boldsymbol{\theta}_i$ of the epistemically-uncertain coefficients of the corresponding aleatory probability distribution $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$, i.e., $\boldsymbol{\varphi}_i = \boldsymbol{\theta}_i$. For example, for category (III) parameter p_1 we have $\boldsymbol{\varphi}_1 = \boldsymbol{\theta}_1 = [m, s^2]$. In order to address the issue above, a novel sensitivity index is introduced in analogy with variance-based Sobol indices[9, 16-19], which *generalizes* the approach presented in Ref. 14. Imagine that we fix $\boldsymbol{\varphi}_i$ at a particular value $\boldsymbol{\varphi}_i^* \in \boldsymbol{\Omega}_i$. Let $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ be the resulting amount of epistemic uncertainty in Q, taken over all parameters $\boldsymbol{p}$ and keeping the epistemically-uncertain 'element' $\boldsymbol{\varphi}_i$ fixed at $\boldsymbol{\varphi}_i^*$ (instead, all the *other* epistemically-uncertain coefficients $\boldsymbol{\theta}_{-i}^{all} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i-1}, ..., \boldsymbol{\theta}_{21}]$ are allowed to range in their corresponding space of variation $\boldsymbol{\Omega}_{-i}^{all}$). We would imagine that having frozen one potential source of epistemic uncertainty ($\boldsymbol{\varphi}_i$), the resulting indicator $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ will be lower than the corresponding total (or unconditional) one $U_{\boldsymbol{p}}(Q)$. One could therefore conceive of using $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ as a measure of the relative importance of p_i, reasoning that the smaller $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$, the greater the influence of p_i. However, notice that this approach makes the sensitivity measure *dependent* on the *position* of the point $\boldsymbol{\varphi}_i^*$ for each input factor, which is impractical. Thus, we take the *average* of the measure $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ over *all* the possible points $\boldsymbol{\varphi}_i^* \in \boldsymbol{\Omega}_i$, which removes the dependence on $\boldsymbol{\varphi}_i^*$. The resulting indicator is then written synthetically as $E_{\boldsymbol{\varphi}i}[U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i)]$ and represents the *expected* amount of epistemic uncertainty contained in output Q when the epistemically-uncertain coefficient/parameter $\boldsymbol{\varphi}_i$ 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 $E_{\boldsymbol{\varphi}i}[U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i)]$, the more important the corresponding parameter p_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 Q (as highlighted above, the consideration of "average sensitivities" is due to the fact that $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ is in general *strongly dependent* on the *position* of the point $\boldsymbol{\varphi}_i^*$: this suggests the necessity to calculate the *average* of the measure $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ over many possible points $\boldsymbol{\varphi}_i^* \in \boldsymbol{\Omega}_i$ in order to obtain robust and reliable sensitivity rankings). Finally, the sensitivity $S_i(U_{\boldsymbol{p}}(Q))$ of the epistemic uncertainty of the output Q to the epistemic uncertainty of parameter p_i can be synthesized with an expression like

$$S_i\left(U_{\boldsymbol{p}}(Q)\right) = 1 - \frac{E_{\boldsymbol{\varphi}_i}\left[U_{\boldsymbol{p}}\left(Q|\boldsymbol{\varphi}_i\right)\right]}{U_{\boldsymbol{p}}(Q)}. \tag{10}$$

Index $S_i(U_{\boldsymbol{p}}(Q))$ (10) is an estimate of the *value* of *additional* empirical information about the input p_i in terms of the *fractional reduction* in epistemic uncertainty that might be achieved in Q 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. In principle, one could also pinch multiple inputs simultaneously to study interactions: however, this aspect is not considered in the present paper. It is worth noting that $S_i(U_{\boldsymbol{p}}(Q))$ (10) has the advantage of being a *global* sensitivity index because: (i) the effect of the *entire* space of variation $\boldsymbol{\Omega}_i$ of the epistemically-uncertain parameter/coefficient $\boldsymbol{\varphi}_i$ whose epistemic uncertainty importance is evaluated, is considered; (ii) the importance of this input parameter/coefficient is evaluated with *all* other input parameters varying as well: actually, for each *fixed constant* value of $\boldsymbol{\varphi}_i$ the computation of $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i)$ is carried out by letting *all* the *other* epistemically-uncertain parameters/coefficients $\boldsymbol{\theta}_{\sim i}^{all} = \left[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i-1}, ..., \boldsymbol{\theta}_{21}\right]$ range within the corresponding space of variation $\boldsymbol{\Omega}_{\sim i}^{all}$; (iii) this sensitivity index is "*model free*" because its computation is independent from assumptions about the model form, such as linearity, additivity and so on[51]. Finally, note that index $S_i(U_{\boldsymbol{p}}(Q))$ (10) is nicely scaled between 0 and 1; however, unlike the factorizations used by variance-based sensitivity analyses, these reductions will not generally add up to 1 for all the input variables. Other approaches to sensitivity analysis in the presence of mixed aleatory and epistemic uncertainties can be found in Refs. 52-54.

In this paper, the sensitivity index $S_i(U_{\boldsymbol{p}}(Q))$ (10) related to the generic parameter p_i is straightforwardly estimated as follows:

1) letting *all* the epistemically-uncertain parameters/coefficients $\boldsymbol{\theta}^{all}$ range within the entire space of variation $\boldsymbol{\Omega}^{all}$, *propagate* the mixed aleatory and epistemic uncertainty from the inputs $\boldsymbol{p}$ to the output of interest Q and evaluate the resulting (total, unconditional) amount of epistemic uncertainty $U_{\boldsymbol{p}}(Q)$ in Q (notice that technical details about the uncertainty propagation phase are not given here in order to not interrupt the flow of the presentation concerning sensitivity analysis: the reader is referred to the following Section III.C);
2) select (deterministically or stochastically) N_e values $\boldsymbol{\varphi}_i^k$, k = 1, 2, …, N_e, of the epistemically-uncertain 'factor' $\boldsymbol{\varphi}_i$ under analysis within its space of variation $\boldsymbol{\Omega}_i$ (as already mentioned, if p_i is a category (II) parameter, then simply $\boldsymbol{\varphi}_i = p_i$; instead, if p_i is a category (III) parameter, then $\boldsymbol{\varphi}_i$ represents the vector $\boldsymbol{\theta}_i$ of the epistemically-uncertain parameters of the corresponding aleatory probability distribution $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$,

i.e., $\boldsymbol{\varphi}_i = \boldsymbol{\theta}_i$: for example, for category (III) parameter p_1 we have $\boldsymbol{\varphi}_1 = \boldsymbol{\theta}_1 = [m, s^2]$). These N_e realizations of epistemic uncertainty $\boldsymbol{\varphi}_i^k$, k = 1, 2, …, N_e, should be chosen in such a way to evenly cover the corresponding uncertainty space $\boldsymbol{\Omega}_i$: in this paper, a grid of equally spaced points is adopted to this aim;

3) fixing the value of $\boldsymbol{\varphi}_i$ to $\boldsymbol{\varphi}_i^k$, k = 1, 2, …, N_e, and letting *all* the *other* epistemically-uncertain parameters/coefficients $\boldsymbol{\theta}_{-i}^{all}$ vary within $\boldsymbol{\Omega}_{-i}^{all}$, *propagate* the mixed aleatory and epistemic uncertainty from the inputs $\boldsymbol{p}$ to the output of interest Q and evaluate the resulting (conditional) amount of epistemic uncertainty $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^k)$ in Q. Notice that in the computation of $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^k)$ for category (III) parameters, we condition the event to *multi-dimensional* realizations of the epistemic space. For example, for p_1 we fix *both* the mean m *and* the variance s^2, i.e., $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_1^k = [m^k, s^{2,k}]$; for parameter p_4 we fix the mean μ_4, the standard deviation σ_4 and the correlation coefficient ρ, i.e., $\boldsymbol{\varphi}_4 = \boldsymbol{\varphi}_4^k = [\mu_4^k, \sigma_4^k, \rho^k]$;

4) estimate the sensitivity index $S_i(U_{\boldsymbol{p}}(Q))$ (10) as $S_i\left(U_{\boldsymbol{p}}(Q)\right) \approx 1 - \frac{1}{N_e}\sum_{k=1}^{N_e} \frac{U_{\boldsymbol{p}}\left(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^k\right)}{U_{\boldsymbol{p}}(Q)}$.

As already mentioned above, the computation of $U_{\boldsymbol{p}}(Q)$ depends on the nature of Q. In subproblem (B.1) the task is to identify those input parameters $\boldsymbol{p} = \{p_i: i = 1, 2, \ldots, 21\}$ that lead to the greater refinement in the distributional *p-box* of the corresponding intermediate (output) variables $\{x_j = h_j(\boldsymbol{p}_j): j = 1, 2, \ldots, 5\}$ (2)-(6): thus, in this case the output quantity Q of interest is the intermediate variable x_j itself, j = 1, 2, …, 5. We propose to define the amount of epistemic uncertainty $U_{\boldsymbol{p}}(Q) = U_{\boldsymbol{p}}(x_j)$ in x_j as the *area* $A_{\boldsymbol{p}}(x_j)$ (11) included between the *extreme* upper and lower CDFs, $\overline{F}^{x_j}(x_j)$ and $\underline{F}^{x_j}(x_j)$, bounding the distributional p-box $PB^{x_j}(x_j) = \left\{F^{x_j}\left(x_j|\boldsymbol{\theta}^{x_j}\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$ of x_j:

$$U_{\boldsymbol{p}}(x_j) = A_{\boldsymbol{p}}(x_j) = \int_0^1 \left(\left[\underline{F}^{x_j}\right]^{-1}(r) - \left[\overline{F}^{x_j}\right]^{-1}(r) \right) dr \,,\ j = 1, 2, \ldots, 5 \tag{11}$$

where $\left[\underline{F}^{x_j}\right]^{-1}(r)$ and $\left[\overline{F}^{x_j}\right]^{-1}(r)$ are the inverse of $\underline{F}^{x_j}(x_j)$ and $\overline{F}^{x_j}(x_j)$, respectively, at cumulative probability level r. Obviously, the larger the area $A_{\boldsymbol{p}}(x_j)$ (i.e., the larger the separation between the bounding CDFs), the larger the imprecision, i.e., the epistemic uncertainty, in the definition of a precise probability model for variable x_j. Notice that the CDFs $\overline{F}^{x_j}(x_j)$ and $\underline{F}^{x_j}(x_j)$ are formally defined as:

$$\overline{F}^{x_j}(x_j) = \max_{\theta^{x_j} \in \Omega^{x_j}} \left\{ F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}\right) \right\} \text{ and } \underline{F}^{x_j}(x_j) = \min_{\theta^{x_j} \in \Omega^{x_j}} \left\{ F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}\right) \right\}, \ \forall x_j \in \Re \text{ , } j = 1, 2, \ldots, 5. \tag{12}$$

In subproblems (B.2) and (B.3) the output quantities of interest Q are represented by the following quantites:

$$J_1 = E_{\boldsymbol{p}}[w(\boldsymbol{p}, \boldsymbol{d})] \tag{13}$$

$$J_2 = 1 - P[w(\boldsymbol{p}, \boldsymbol{d}) < 0], \tag{14}$$

where

$$w(\boldsymbol{p}, \boldsymbol{d}) = \max_{1 \le o \le 8} g_o = \max_{1 \le o \le 8} f_o(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}), \boldsymbol{d}) \tag{15}$$

is the so-called *worst-case requirement* metric. Notice that $J_1 = E_{\boldsymbol{p}}[w(\boldsymbol{p}, \boldsymbol{d})]$ (13) is expected value of the worst-case requirement metric $w(\boldsymbol{p}, \boldsymbol{d})$ (8) and $J_2 = P[w(\boldsymbol{p}, \boldsymbol{d}) > 0]$ (14) is the system failure probability respectively. In these cases, the quantitative indicators $U_{\boldsymbol{p}}(J_1)$ and $U_{\boldsymbol{p}}(J_2)$ of the amount of epistemic uncertainty in J_1 and J_2 are represented by the *lengths* $L_{\boldsymbol{p}}(J_1)$ (16) and $L_{\boldsymbol{p}}(J_2)$ (17) of the corresponding intervals $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_2, \overline{J}_2]$, respectively[§]:

$$U_{\boldsymbol{p}}(J_1) = L_{\boldsymbol{p}}(J_1) = \overline{J}_1 - \underline{J}_1 \tag{16}$$

$$U_{\boldsymbol{p}}(J_2) = L_{\boldsymbol{p}}(J_2) = \overline{J}_2 - \underline{J}_2 . \tag{17}$$

Again, the larger the intervals, the larger the uncertainty in the definition of a precise value for the erformance metrics J_1 (13) and J_2 (14).

A final consideration is in order with respect to the *computational cost* associated to the evaluation of the sensitivity indices $S_i(U_{\boldsymbol{p}}(Q))$, $Q = x_j, J_1, J_2$, $i = 1, 2, \ldots, 21$ (10). For *each* input parameter of interest p_i, $i = 1, 2, \ldots,$

[§] Notice that the choice of intervals to represent the epistemic uncertainty in J_1 and J_2 is a "natural" consequence of the *hybrid representation* of uncertainty adopted in the present paper (i.e., probabilistic/aleatory and interval-based/epistemic). In such a framework, the worst-case requirement metric $w(\boldsymbol{p}, \boldsymbol{d})$ is represented by a distributional p-box, i.e., an ensemble of probability distributions (as its inputs $\boldsymbol{p}$). Thus, a value of the mean and of the failure probability can be computed for *each* element of the p-box: such *ensemble* of values identifies the corresponding *intervals* for J_1 and J_2.

21, a number N_e (e.g., $N_e \approx$ 10-20 in this paper) of realizations $\boldsymbol{\varphi}_i^k$ of the corresponding epistemically-uncertain 'factor' $\boldsymbol{\varphi}_i$ have to be selected. Then, for *each* realization $\boldsymbol{\varphi}_i^k$, k = 1, 2, …, N_e, the quantitative indicator $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^k)$ has to be calculated. The evaluation of indicator $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^k)$ implies: (i) the *propagation* of mixed aleatory and epistemic uncertainty from the input parameters $\boldsymbol{p}$ to the output Q of interest through the corresponding mathematical model (i.e., $h_j(\boldsymbol{p}_j)$ in (2)-(6) for the evaluation of $U_{\boldsymbol{p}}(x_j)$ and $\boldsymbol{g} = \boldsymbol{f}(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}), \boldsymbol{d})$ in (1), for the computation of the worst-case requirement metric $w(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}), \boldsymbol{d})$ and correspondingly of $U_{\boldsymbol{p}}(J_1)$ and $U_{\boldsymbol{p}}(J_2)$); (ii) the identification of the *extreme bounds* of Q (i.e., $\underline{F}^{x_j}(x_j)$ and $\overline{F}^{x_j}(x_j)$, $\underline{J}_1$ and $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$, respectively), which requires the solution of *several optimization problems* (see Section III.C for further details about the uncertainty propagation process). The execution of steps (i) and (ii) above entails the *repeated* evaluation of the output Q (i.e., of the corresponding mathematical model) for *every* possible solution proposed by the optimization algorithm during the search. As a consequence, the total number of system model evaluations can easily reach tens/hundreds millions for *each* realization $\boldsymbol{\varphi}_i^k$ of *each* input parameter p_i analyzed, which makes the proposed approach impractical also in the presence of mathematical system models that take even only few minutes to run. For example, in this case the evaluation of $w(\boldsymbol{p},\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ (1) for N_a = 10000 values $\boldsymbol{p}_s$, s = 1, 2, …, N_a, of the inputs $\boldsymbol{p}$ takes 2125s = 35.4 min.

In the present paper, we address this computational burden by replacing the original model $\boldsymbol{g} = \boldsymbol{f}(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ (1) by a *fast-running*, *surrogate regression model* (also called meta-model): since calculations with the surrogate model can be performed quickly (e.g., in fractions of seconds), the problem of long simulation times is circumvented. The regression model is constructed on the basis of a *finite* (and possibly reduced) set D_{tr} of N_{tr} data representing *examples* of the input/output nonlinear relationships underlying the original system model. The generation of this data set D_{tr} entails running the original system mathematical model $\boldsymbol{f}(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ a predetermined (and possibly reduced) number of times N_{tr} for specified values $\{\boldsymbol{x}_t: t = 1, 2, \ldots, N_{tr}\}$ of the input variables $\boldsymbol{x} = \{x_j: j = 1, 2, \ldots, n_{int} = 5\}$ and collecting the corresponding values $\{\boldsymbol{g}_t: t = 1, 2, \ldots, N_{tr}\}$ of the outputs $\boldsymbol{g} = \{g_o: o = 1, 2, \ldots, n_{out} = 8\}$ of interest; then, statistical techniques (for example, regression error minimization procedures) are employed for calibrating/adapting the internal *parameters/coefficients* of the regression model in order to fit the input/output data $D_{tr} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t): t = 1, 2, \ldots, N_{tr}\}$ generated in the previous step and to capture the underlying (possibly nonlinear and non-monotonic) relationship. Once built, the meta-model can be used for

performing, in an acceptable computational time, the numerous repeated evaluations of the system worst-case requirement metric $w(\boldsymbol{p},\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o\left(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d}\right)$ (1) needed for an accurate estimation of the sensitivity indices above.**

In this work, a three-layered feed-forward Artificial Neural Network (ANN) regression model is considered. ANNs are computing devices inspired by the function of the nerve cells in the brain[20-25]. They are composed of many parallel computing units (called *neurons* or *nodes*) arranged in different *layers* and interconnected by weighed connections (called *synapses*). Each of these computing units performs a few simple operations and communicates the results to its neighbouring units. From a mathematical viewpoint, ANNs consist of a set of nonlinear (e.g., sigmoidal) basis functions with adaptable parameters that are adjusted by a process of *training* (on many different input/output data examples), i.e., an iterative process of regression error minimization[55]. ANNs have been demonstrated to be universal approximants of *continuous* nonlinear functions (under mild mathematical conditions)[21], i.e., in principle, an ANN model with a properly selected architecture can be a consistent estimator of any continuous nonlinear function. Further details about ANN regression models are not reported here for brevity; the interested reader may refer to the cited references and the copious literature in the field.

Notice that the recommendation of using ANN regression models is mainly based on (i) theoretical considerations about the (mathematically) demonstrated capability of ANN regression models of being *universal* approximants of continuous nonlinear functions[21] and (ii) the experience of the authors' in the use of ANN regression models for propagating the uncertainties through mathematical model codes simulating safety systems[56-60]. Since no further comparisons with other types of regression models have been performed by the authors yet, no additional proofs of the superiority of ANNs with respect to other regression models can be provided at present, in general terms.

1.2 Application Results

First, we train a 8-output ANN regression model using a set $D_{train} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t=1,2,...,N_{train}\}$ of input/output data examples of size N_{train} = 30000. A Latin Hypercube Sample (LHS) of the inputs is drawn to give the vectors $\boldsymbol{x}_t$ = $\{x_{1,t}, x_{2,t}, \ldots, x_{j,t}, \ldots, x_{n_{in},t} = x_{5,t}\}$, t = 1, 2, …., N_{train}.[61] Then, the original model (1) is evaluated on the input vectors $\boldsymbol{x}_t$, t = 1, 2, …, N_{train}, to obtain the corresponding output vectors $\boldsymbol{g}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{d})$ = $\{g_{1,t}, g_{2,t}, ..., g_{l,t}, ..., g_{n_{out},t} = g_{8,t}\}$, t = 1,

** Notice that on the contrary the computation of sensitivity indices $S_i(A(x_j))$ does not require the evaluation of $w(\boldsymbol{p},\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o\left(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d}\right)$: thus, no regression model-based approximation is employed in this case.

2, …, N_{train}, and build the data set $D_{train} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, ..., N_{train}\}$. Finally, the adjustable internal parameters of the ANN regression model are calibrated to fit the generated data: in particular, the common error back-propagation algorithm is implemented to *train* the ANN[55]. Note that a *single* ANN can be trained to estimate *all* the eight outputs of the model here of interest.

In the present case study, the number of inputs to the ANN regression model is equal to $n_{int} = 5$ (i.e., the number of intermediate variables $\boldsymbol{x} = \{x_j: j = 1, 2, \ldots, n_{int} = 5\}$ (2)-(6)), whereas the number of outputs is equal to $n_{out} = 8$ (i.e., the number of requirement metrics of interest $\boldsymbol{g} = \{g_l: l = 1, 2, \ldots, n_{out} = 8\}$ (1), as reported in Section II). With respect to that, it is worth pointing out that although the quantity of interest in the present study is the (scalar) worst-case requirement metric $w(\boldsymbol{p}, \boldsymbol{d}) = \max_{1 \le o \le 8} g_o = \max_{1 \le o \le 8} f_o(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}), \boldsymbol{d})$ (1), we choose to reproduce by ANN the relationship between $\boldsymbol{x}$ and the (eight-dimensional) vector $\boldsymbol{g} = \boldsymbol{f}(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}), \boldsymbol{d})$: this is due to the fact that (i) the components of $\boldsymbol{g}$ are *continuous* functions of the inputs $\boldsymbol{x}$ that prescribe them[11] (with benefits for the ANN approximation), and (ii) the behavior of $w(\boldsymbol{p}, \boldsymbol{d})$ (involving a 'max' operator) may be too abrupt for a satisfactory fitting by ANNs. The number of nodes n_h in the hidden layer has been set equal to 27 by trial and error.

A *validation* data set $D_{val} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, ..., N_{val} = 10000\}$ (different from the training set D_{train}) is used to monitor the accuracy of the ANN model during the training procedure: in practice, the Root Mean Squared Error (RMSE) is computed on D_{val} (over all the outputs) at different phases of the training procedure. At the beginning, the RMSE computed on the validation set D_{val} typically decreases together with the RMSE computed on the training set D_{train}; then, when the ANN regression model starts *overfitting* the data, the RMSE calculated on the validation set D_{val} starts increasing: this is the time to stop the training algorithm. The time needed to train the ANN is approximately 20s on a Intel(R) Core(TM) i5-3380M CPU@2.90GHz.

For a realistic measure of the ANN model accuracy, the widely adopted coefficient of determination R^2 and the RMSE are computed for each output $\{g_l: l = 1, 2, \ldots, n_{out} = 8\}$ on a new data set $D_{test} = \{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, ..., N_{test}\}$ also of size $N_{test} = 10000$, not used during training[17]. Table 2 reports the values of the coefficient of determination R^2 and of the RMSE associated to the *final* estimates of the worst-case requirement metric $w(\boldsymbol{p}, \boldsymbol{d}) = \max_{1 \le o \le 8} g_o = \max_{1 \le o \le 8} f_o(\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p}), \boldsymbol{d})$ of interest, computed on the test set D_{test} of size $N_{test} = 10000$ by the ANN model with $n_h = 27$ hidden neurons, built on a data set D_{train} of size $N_{train} = 30000$. For completeness, the values of R^2 and of the RMSE associated to the estimates of each output $\{g_l: l = 1, 2, \ldots, n_{out} = 8\}$ are also reported.

Artificial Neural Network (ANN) - $w(\boldsymbol{p},\boldsymbol{d})$				
Optimal configuration selected: n_{int} = 5, n_h = 27, n_{out} = 8				
N_{train}	N_{val}	N_{test}	R^2 (test)	RMSE (test)
30000	10000	10000	0.9944	0.1468

Artificial Neural Network (ANN) – g_o, o = 1, 2, …, 8		
Optimal configuration selected: n_{int} = 5, n_h = 27, n_{out} = 8		
g_o	R^2 (test)	RMSE (test)
g_1	0.9994	0.1368
g_2	0.9895	0.1633
g_3	0.9976	0.1232
g_4	0.9945	0.1479
g_5	0.9987	0.1498
g_6	0.9937	0.1311
g_7	0.9821	0.1703
g_8	0.9952	0.1488

Table 2. Coefficient of determination R^2 and RMSE associated to the ANN (test) estimates of the worst-case requirement metric $w(\boldsymbol{p},\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$. The same quantities are reported also for the eight output g_o, o = 1, 2, …, 8, separately

The large value of the coefficient of determination R^2, i.e., 0.9944, and the small value of 0.1468 for the RMSE produced lead us to assert that the accuracy of the ANN model can be considered satisfactory for the needs of capturing the global behavior of the highly nonlinear and non-monotonic function $w(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ and, thus, of estimating the corresponding sensitivity indices. This is also pictorially confirmed by a visual inspection of the ANN approximation capabilities. Figure 4, left and right, shows in logarithmic scale the behavior of $w(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ as a function of x_1, when x_2, x_3, x_4 and x_5 are set to 0.6250, 0.4000, 0.7450 and 0.5000, respectively (solid line), and the corresponding ANN fitting (dashed line); instead, Figure 4 right shows $w(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ as a function of x_3, when x_1, x_2, x_4 and x_5 are fixed to 0.4500, 0.6250, 0.7450 and 0.2, respectively (solid line), together with the corresponding ANN approximation (dashed line). In both cases, the ANN estimates are in satisfactory agreement with the real trend of $w(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$. Notice that the evaluation of $w(\boldsymbol{p},\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ (1) for, e.g., N_a = 10000 values $\boldsymbol{p}_s$, s = 1, 2, …, N_a, of the inputs $\boldsymbol{p}$ takes only 1.25s, i.e., 1700 times less than the original model.

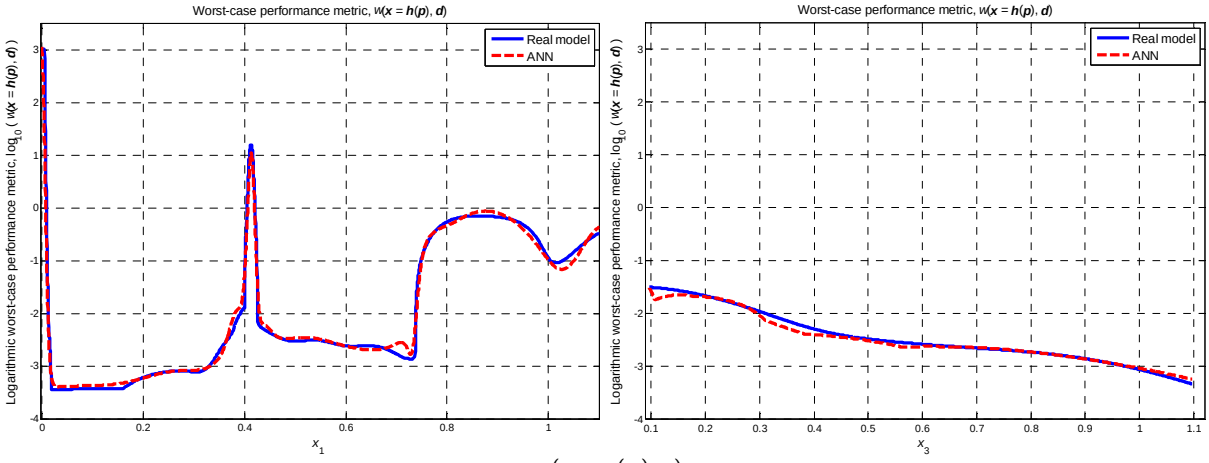

Figure 4. Worst-case requirement metric $w(x = h(p), d)$ as a function of x_1 (with x_2, x_3, x_4 and x_5 set to 0.6250, 0.4000, 0.7450 and 0.5000, respectively) (left) and of x_3 (with x_1, x_2, x_4 and x_5 fixed to 0.4500, 0.6250, 0.7450 and 0.2, respectively) (right) (solid lines), together with the corresponding ANN approximations (dashed lines)

The trained ANN regression model is then used for computing the sensitivity index $S_i(U_{\boldsymbol{p}}(Q))$ (10), i = 1, 2, …, 21, for $Q = x_j$, J_1, J_2. Notice that the propagation of uncertainty needed for the estimation of the indices (see steps 2. and 3. of the algorithm in Section III.B.1.1) is carried out by Monte Carlo Simulation (MCS) with N_a = 10000 random samples. The values are reported in Table 3 together with the corresponding parameters ranking $R_i(Q)$ (in parentheses). The number N_e of epistemic realizations $\boldsymbol{\varphi}_i^k$ used to estimate $E_{\boldsymbol{\varphi}i}[U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i)]$ as $1/N_e \cdot \sum_{k=1}^{N_e} U_{\boldsymbol{p}}\left(Q\middle|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^k\right)$ are also reported for each parameter p_i. Notice that the coefficient of correlation ρ between parameters p_4 and p_5 is not explicitly listed in Table 3 as a 'stand-alone' coefficient, since it is considered part of the corresponding uncertainty models of p_4 and p_5, i.e., $\boldsymbol{\theta}_4 = [\mu_4, \sigma_4, \rho]$ and $\boldsymbol{\theta}_5 = [\mu_5, \sigma_5, \rho]$.

Sensitivity to epistemic uncertainty (factor prioritization setting)							
				Sensitivity index, $S_i(U_p(Q))$ (ranking, $R_i(Q)$)			
Parameter	**Category**	**Intermediate variable**	N_e	$S_i(A_p(x_j))$	$S_i(L_p(J_1))$	$S_i(L_p(J_2))$	**Accumulated ranking, $R_{acc,i}$**
p_1	III	x_1	16	**0.5071 (1)**	**0.4009 (4)**	**0.1978 (2)**	**6**
p_2	II	x_1	10	0.0156 (4)	0.0499 (6)	$6.70 \cdot 10^{-3}$ (9)	15
p_3	I	x_1	/	/	/	/	/
p_4	III	x_1	64	0.0706 (3)	0.1775 (5)	**0.0421 (4)**	9
p_5	III	x_1	64	0.3085 (2)	**0.7727 (2)**	**0.0436 (3)**	**5**
p_6	II	x_2	10	**0.6108 (1)**	$3.40 \cdot 10^{-3}$ (9)	0.0152 (6)	15
p_7	III	x_2	16	0.4773 (2)	$6.30 \cdot 10^{-3}$ (8)	$9.60 \cdot 10^{-3}$ (8)	16
p_8	III	x_2	16	0.1677 (4)	$1.27 \cdot 10^{-4}$ (13)	$6.94 \cdot 10^{-4}$ (13)	26
p_9	I	x_2	/	/	/	/	/
p_{10}	III	x_2	16	0.2232 (3)	$1.49 \cdot 10^{-5}$ (14)	$1.77 \cdot 10^{-4}$ (14)	28
p_{11}	I	x_3	/	/	/	/	/
p_{12}	II	x_3	10	**0.9277 (1)**	**0.4237 (3)**	**0.6852 (1)**	**4**
p_{13}	III	x_3	16	$1.20 \cdot 10^{-4}$ (2-3)	$2.06 \cdot 10^{-7}$ (17)	$1.22 \cdot 10^{-15}$ (17)	34
p_{14}	III	x_3	16	0 (4)	$5.26 \cdot 10^{-7}$ (15)	$1.29 \cdot 10^{-5}$ (16)	31
p_{15}	III	x_3	16	$1.20 \cdot 10^{-4}$ (2-3)	$3.47 \cdot 10^{-7}$ (16)	$2.28 \cdot 10^{-5}$ (15)	31
p_{16}	II	x_4	10	**0.7178 (1)**	0.0125 (7)	0.0174 (5)	12
p_{17}	III	x_4	16	0.1522 (3)	$2.20 \cdot 10^{-3}$ (11)	$5.40 \cdot 10^{-3}$ (10)	21
p_{18}	III	x_4	16	0.2425 (2)	$2.40 \cdot 10^{-3}$ (10)	$2.80 \cdot 10^{-3}$ (11)	21
p_{19}	I	x_4	/	/	/	/	/
p_{20}	III	x_4	16	0.0803 (4)	$5.64 \cdot 10^{-4}$ (12)	$1.00 \cdot 10^{-3}$ (12)	24
p_{21}	III	x_5	/	/	**0.8566 (1)**	$9.90 \cdot 10^{-3}$ (7)	**8**

Table 3. Values of the sensitivity indices $S_i(U_p(Q))$ (10), i = 1, 2, …, 21, for $Q = x_j$, J_1, J_2, together with the corresponding parameters ranking $R_i(Q)$ (in parentheses); the accumulated ranking $R_{acc,i} = R_i(J_1) + R_i(J_2)$ is also reported

It can be seen that the parameters whose epistemic uncertainty contributes more to the epistemic uncertainty 'contained' in the p-boxes of the corresponding intermediate output variables x_j, j = 1, 2, …, 4, are p_1, p_6, p_{12} and p_{16}, respectively (highlighted in bold in Table 3): in detail, refining the uncertainty models of p_1, p_6, p_{12} and p_{16} according to the *particular* strategy proposed (i.e., reducing their epistemic uncertainty from a *set* to a *point*) would lead to an *expected* reduction in the epistemic uncertainty of x_1, x_2, x_3 and x_4 of 50.71%, 61.08%, 92.77% and 71.78%, respectively. This information is of paramount importance for the experts in the disciplines modeled by the relations $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p})$ (2)-(6), because in the light of the results obtained they can focus their efforts primarily on increasing the state-of-knowledge on the identified important parameters and the related physical phenomena. For illustration purposes, Figure 5 left analyzes the different effect that an improvement in the uncertainty models of parameters p_1 (rank $R_1(x_1) = 1$) and p_2 (rank $R_4(x_1) = 4$) has on the p-box of x_1. The upper and lower CDFs, $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$, bounding the distributional p-box $PB^{x_1}(x_1)$ of x_1 obtained by propagating the original uncertainty models of p_1 and p_2 are shown as solid lines, whereas those produced by fixing $\boldsymbol{\theta}_1 = [m, s^2] = [0.63, 0.0207]$ and p_2 =

1.00 are shown as dashed and dot-dashed lines, respectively. The area $A_{\boldsymbol{p}}(x_1)$ contained between $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$ (i.e., the epistemic uncertainty in x_1) is reduced by 56.77% in the first case, whereas a reduction of only 6% is obtained in the second case. In addition, Figure 5 right shows the extreme upper and lower CDFs, $\overline{F}^{x_3}(x_3)$ and $\underline{F}^{x_3}(x_3)$, bounding the distributional p-box $PB^{x_3}(x_3)$ of x_3, obtained by propagating the original uncertainty models of all the corresponding input parameters (solid lines) and those produced by fixing parameter p_{12} (rank $R_{12}(x_3) = 1$) to $p_{12}^* = 0$ (dashed lines), 0.5 (dot-dashed lines) and 1 (dotted lines). It is evident that 'pinching' p_{12} to *different* values within its range of variation produces *extremely different* results: for example, when $p_{12} = 0$, the p-box of x_3 almost collapses into a single CDF (actually, the area $A_{\boldsymbol{p}}(x_3)$ contained is $2.1 \cdot 10^{-7}$); on the contrary, when $p_{12} = 1$, the area $A_{\boldsymbol{p}}(x_3)$ contained is around 0.13. This exemplary situation demonstrates that the sensitivity indicator $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ ($A_{\boldsymbol{p}}(x_3|p_{12} = p_{12}^*)$, in this case) is in general *strongly dependent* on the *position* of the point $\boldsymbol{\varphi}_i^*$ (= p_{12}^*) and confirms the necessity to calculate the *average* of the measure $U_{\boldsymbol{p}}(Q|\boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^*)$ (= $A_{\boldsymbol{p}}(x_3|p_{12} = p_{12}^*)$) over many possible points $\boldsymbol{\varphi}_i^* \in \boldsymbol{\Omega}_i$ ($p_{12}^* \in \Delta_{p_{12}}$) in order to obtain robust and reliable sensitivity rankings.

The four parameters that influence most the uncertainty of J_1 (i.e., the expected value $E_{\boldsymbol{p}}[w(\boldsymbol{p},\boldsymbol{d})]$ of $w(\boldsymbol{p},\boldsymbol{d})$) are p_{21}, p_5, p_{12} and p_1 (highlighted in bold in Table 3), in decreasing order of ranking: actually, refining the corresponding uncertainty models according to the *particular* strategy proposed (i.e., reducing their epistemic uncertainty from a *set* to a *point*) leads to an expected reduction of about 85.66%, 77.21%, 42.37% and 40.09% in the width of the interval of J_1 (i.e., in its epistemic uncertainy); some parameters (e.g., p_4, p_2 and p_{16}) have a non negligible influence on J_1 (in fact, the corresponding indices $S_i(L_{\boldsymbol{p}}(J_1))$, i = 2, 4, 16, range from 0.0125 to 0.1775), whereas some others (in particular, p_{13}, p_{14} and p_{15}) have almost no effect on the uncertainty of J_1 (in fact, the corresponding indices $S_i(L_{\boldsymbol{p}}(J_1))$, i = 13, 14, 15, are around 10^{-7}). Instead, the four parameters that influence most the uncertainty of J_2 (i.e., the system failure probability $J_2 = P[\, w(\boldsymbol{p},\boldsymbol{d}) > 0]$) are p_{12}, p_1, p_5 and p_4 (highlighted in bold in Table 3), in decreasing order of ranking: actually, reducing the corresponding epistemic uncertainty from a *set* to a *point* leads to an expected reduction of about 68%, 20%, 4.4% and 4.2%, respectively, in the width of the interval of J_2 (i.e., in its epistemic uncertainy); again, some parameters (e.g., p_{16}, p_6 and p_{21}) seem to have a non negligible influence on J_2 (in fact, the corresponding indices $S_i(L_{\boldsymbol{p}}(J_2))$ range from 0.0100 to 0.0174), whereas some others (in particular, again p_{13}, p_{14} and p_{15}) have almost no effect on the uncertainty of J_2 (in fact, the corresponding indices $S_i(L_{\boldsymbol{p}}(J_2))$ are lower than or equal to 10^{-5}). As expected, all the parameters that are relevant in the analysis of the

integrated system (i.e., relevant for J_1 and J_2), are also relevant in the analysis of the individual disciplines modeled by the relations $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p})$ (2)-(6). On the contrary, some parameters that are *very relevant* in the models $\boldsymbol{x} = \boldsymbol{h}(\boldsymbol{p})$ (2)-(6) may *not* be *so* important in the analysis of the integrated system (see, e.g., p_6, p_7, p_{10}, p_{16}, p_{17} and p_{18}).

Given these considerations, in order to identify the set of the *four* parameters that contribute more to the epistemic uncertainty in *both* J_1 *and* J_2 (see subproblem (C.3)), a joint, accumulated ranking is here introduced: in particular, the accumulated ranking $R_{acc,i}$ of parameter p_i is obtained as the sum of $R_i(J_1)$ (i.e., the ranking based on indicator J_1) and $R_i(J_2)$ (i.e., the ranking based on indicator J_2); the corresponding values are reported in Table 3. The analysis shows that the most relevant parameters are p_{12}, p_5, p_1 and p_{21} that are ranked $R_{acc,12} = R_{12}(J_1) + R_{12}(J_2) = 3+1 = 4$, $R_{acc,5} = R_5(J_1) + R_5(J_2) = 3+2 = 5$, $R_{acc,1} = R_1(J_1) + R_1(J_2) = 4+2 = 6$ and $R_{acc,21} = R_{21}(J_1) + R_{21}(J_2) = 1+7 = 8$, respectively (see Table 3). With respect to that, notice that the probability distribution of parameter p_4 (which has a non negligible influence on both J_1 and J_2, as highlighted above) 'shares' an epistemically-uncertain coefficient (i.e., the Pearson correlation factor ρ) with the uncertainty model of parameter p_5. Thus, an improvement in the uncertainty model of p_5 (i.e., a *reduction* in the *epistemic uncertainty* of ρ) will 'indirectly' lead *also* to a *reduction* in the *epistemic uncertainty* of the uncertainty model of (the relatively important) parameter p_4 (with further beneficial effect on the refinement of the ranges of indicators J_1 and J_2).

In conclusion, we *expect* that a reduction in the epistemic uncertainty of parameters p_1, p_5, p_{12} and p_{21} will lead to a consistent reduction in the uncertainty of *both* J_1 *and* J_2.

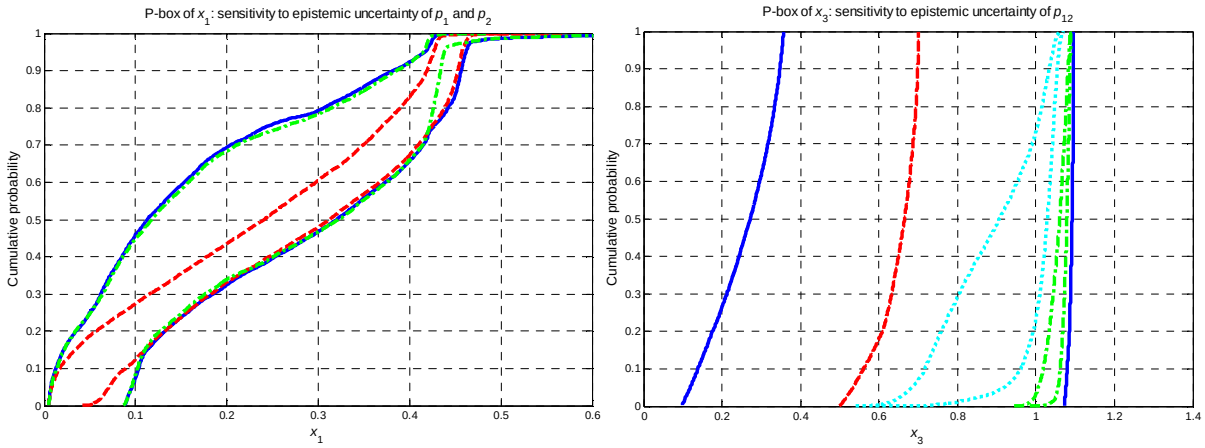

Figure 5. Left: extreme upper and lower CDFs, $\overline{F}^{x_1}(x_1)$ and $\underline{F}^{x_1}(x_1)$, bounding the distributional p-box $PB^{x_1}(x_1)$ of x_1 obtained by propagating the original uncertainty models of p_1 (rank 1) and p_2 (rank 4) (solid lines), together with those produced by fixing $\theta_1 = [m, s^2] = [0.6300, 0.0207]$ (dashed lines) and $p_2 = 1.00$ (dot-dashed lines), respectively. Right: extreme upper and lower CDFs, $\overline{F}^{x_3}(x_3)$ and $\underline{F}^{x_3}(x_3)$, bounding the distributional p-box $PB^{x_3}(x_3)$ of x_3, obtained by propagating the original uncertainty models of all the

corresponding input parameters (solid lines) and those produced by fixing parameter p_{12} (rank 1) to $p_{12}^* = 0$ (dashed lines), 0.5 (dot-dashed lines) and 1 (dotted lines)

2. Sensitivity Analysis in a 'factor fixing' setting

In Section III.B.2.1, the general approach adopted to identify those parameter that can be fixed to a constant value without significantly affecting the outputs of interest is illustrated in detail; in Section III.B.2.2, the results of the application of the method to the tasks of Subproblem (B) are reported.

2.1 The Proposed Approach

In this case, we aim at finding those parameters that *minimally* affect the outputs when they are *fixed* to a given constant value. In particular, the objective is to determine whether the quantities Q of interest analyzed in the previous Section III.B.1, i.e., $Q = x_j$ (subproblem (B.1)), J_1 (subproblem (B.2)) and J_2 (subproblem (B.3)), are sufficiently *insensitive* to the uncertainty in any given parameter such that that parameter can be assumed to take on a fixed constant value without incurring in significant 'error'[11]. In this context, we define the 'error' as the *mismatch* between the results obtained using the original uncertainty models and those produced by a configuration where one of the parameters p_i, i = 1, 2, …, 21, is fixed to the constant p_i^*.

In more detail, in subproblem (B.1) we quantify the 'error' as the relative 'lack of overlapping' between the distributional p-box of x_j obtained using the original uncertainty models, $PB^{x_j}(x_j) = \left\{F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$, and the p-box produced by setting $p_i = p_i^*$, $PB^{x_j}(x_j | p_i^*) = \left\{F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}, p_i^*\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$. The area $A_{\mathbf{p}|p_i^*}^{x_j,over}$ of intersection between the two p-boxes is calculated as

$$A_{\mathbf{p}|p_i^*}^{x_j,over} = \int_0^1 \left(\min\left\{\left[\underline{F}^{x_j}\right]^{-1}(r), \left[\underline{F}^{x_j}\right]^{-1}\left(r \middle| p_i^*\right)\right\} - \max\left\{\left[\overline{F}^{x_j}\right]^{-1}(r), \left[\overline{F}^{x_j}\right]^{-1}\left(r \middle| p_i^*\right)\right\}\right) dr\,, \tag{18}$$

where $\overline{F}^{x_j}(x_j)$ and $\overline{F}^{x_j}(x_j | p_i^*)$ (resp., $\underline{F}^{x_j}(x_j)$ and $\underline{F}^{x_j}(x_j | p_i^*)$) are the extreme bounding upper (resp., lower) CDFs computed as $\max_{\boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}} \left\{F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$ and $\max_{\boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}} \left\{F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}, p_i^*\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$ (resp., $\min_{\boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}} \left\{F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$ and $\min_{\boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}} \left\{F^{x_j}\left(x_j \middle| \boldsymbol{\theta}^{x_j}, p_i^*\right): \boldsymbol{\theta}^{x_j} \in \boldsymbol{\Omega}^{x_j}\right\}$). Thus, the fractional error (i.e., mismatch or 'lack of overlapping') $\varepsilon_{p_i^*}^{x_j}$ can be simply defined as

$$\varepsilon_{p_i^*}^{x_j} = 1 - \frac{A_{\boldsymbol{p}|p_i^*}^{x_j,over}}{\max\left\{A_{\boldsymbol{p}}^{x_j}, A_{\boldsymbol{p}|p_i^*}^{x_j}\right\}}, \tag{19}$$

where $A_{\boldsymbol{p}}^{x_j}$ (11) and $A_{\boldsymbol{p}|p_i^*}^{x_j}$ are the areas of the p-box $PB^{x_j}(x_j)$ obtained by propagating the original uncertainty models and by setting $p_i = p_i^*$, respectively. Notice that when the parameter p_i under consideration (i.e., the one that is fixed to the constant value p_i^*) is of category (I) or (III), then the term $A_{\boldsymbol{p}|p_i^*}^{x_j}$ in (19) may occasionally be larger than $A_{\boldsymbol{p}}^{x_j}$: this explains the necessity of introducing the term max{ $A_{\boldsymbol{p}}^{x_j}$, $A_{\boldsymbol{p}|p_i^*}^{x_j}$ } at the denominator of (19), in order to keep $\varepsilon_{p_i^*}^{x_j} \leq 1$. Obviously, if $\varepsilon_{p_i^*}^{x_j}$ is close to zero, then parameter p_i can be set to the constant value p_i^* without significantly affecting the shape of the p-box of x_j.

Instead, in subproblems (B.2) and (B.3) we quantify the mismatch between the intervals $[\underline{J}_1, \overline{J}_1]$, $[\underline{J}_2, \overline{J}_2]$ (obtained by propagating the original uncertainty models) and $[\underline{J}_1(p_i^*), \overline{J}_1(p_i^*)]$, $[\underline{J}_2(p_i^*), \overline{J}_2(p_i^*)]$ (obtained by setting $p_i = p_i^*$) as the *maximum* between the *relative absolute errors* produced in the estimation of the corresponding upper and lower bounds. In more detail, the relative absolute errors generated in the estimation of $\underline{J}_1$, $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$ are computed as

$$\varepsilon_{p_i^*}^{\underline{J}_1} = \left|\frac{\underline{J}_1(p_i^*) - \underline{J}_1}{\underline{J}_1}\right|, \ \varepsilon_{p_i^*}^{\overline{J}_1} = \left|\frac{\overline{J}_1(p_i^*) - \overline{J}_1}{\overline{J}_1}\right|, \ \varepsilon_{p_i^*}^{\underline{J}_2} = \left|\frac{\underline{J}_2(p_i^*) - \underline{J}_2}{\underline{J}_2}\right| \text{ and } \varepsilon_{p_i^*}^{\overline{J}_2} = \left|\frac{\overline{J}_2(p_i^*) - \overline{J}_2}{\overline{J}_2}\right|, \tag{20}$$

where $\underline{J}_1$, $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$ are different from zero. Then, we take the maximal values among { $\varepsilon_{p_i^*}^{\underline{J}_1}$, $\varepsilon_{p_i^*}^{\overline{J}_1}$ } and { $\varepsilon_{p_i^*}^{\underline{J}_2}$, $\varepsilon_{p_i^*}^{\overline{J}_2}$ } as 'conservative representatives' of the errors $\varepsilon_{p_i^*}^{J_1}$ and $\varepsilon_{p_i^*}^{J_2}$ produced in the estimation of indicators J_1 and J_2, i.e.,

$$\varepsilon_{p_i^*}^{J_1} = \max\{\varepsilon_{p_i^*}^{\underline{J}_1}, \varepsilon_{p_i^*}^{\overline{J}_1}\} \text{ and } \varepsilon_{p_i^*}^{J_2} = \max\{\varepsilon_{p_i^*}^{\underline{J}_2}, \varepsilon_{p_i^*}^{\overline{J}_2}\}, \text{ respectively.} \tag{21}$$

Notice that defining the errors conservatively as in (21) allows treating those cases where the upper and lower bounds of J_1 and J_2 differ by several orders of magnitude (as in the present case study). For example, letting $\underline{J}_1$ = 0.01, $\overline{J}_1$ = 15, $\underline{J}_1(p_i^*)$ = 0.02 and $\overline{J}_1(p_i^*)$ = 15.1, the computation of a 'length of overlap' (similarly to (18) and

(19) for the p-box) would be meaningless, since this length would be always 'dominated' by the large value of the upper bound $\overline{J}_1$. In fact, the original length of the interval of J_1 is (15 – 0.01) = 14.99, whereas the 'length of overlapping' is (15 – 0.02) = 14.98: thus, the corresponding relative error (when computed as the fractional 'lack of overlapping') would be $1 - 14.98/14.99 = 6.67 \cdot 10^{-4} = 0.67\%$. However, such a low mismatch is unrealistic: in fact, while the relative error produced in the estimation of $\overline{J}_1$ is actually very low, i.e., $(15.1 - 15)/15 = 6.67 \cdot 10^{-4} = 0.67\%$, the one generated in the estimation of $\underline{J}_1$ is instead very large, i.e., $(0.02 - 0.01)/0.01 = 1.00 = 100\%$ (notice that this problem is not present in the analysis of the p-boxes, since in the present case the values of the intermediate variables x_j are of the same order of magnitude). Again, if $\varepsilon_{p_i^*}^{J_1}$ and $\varepsilon_{p_i^*}^{J_2}$ are close to zero, then parameter p_i can be set to the constant value p_i^* without significantly affecting the intervals of metrics J_1 (13) and J_2 (14).

Finally, in order to identify those parameters that in general *minimally* affect the output quantity $Q = x_j, J_1, J_2$ of interest, we *exhaustively* explore the *entire* range of variation of all the parameters p_i to find the corresponding (constant) values p_i^* that give rise to the *maximal* mismatch (i.e., maximal error) $\overline{\varepsilon}_{p_i}^{Q} = \max_{p_i^*}\left\{\varepsilon_{p_i^*}^{Q}\right\}$ between the output quantities $Q = x_j, J_1, J_2$, of interest, i.e.:

$$\overline{\varepsilon}_{p_i}^{x_j} = \max_{p_i^*}\left\{\varepsilon_{p_i^*}^{x_j}\right\},\ \overline{\varepsilon}_{p_i}^{J_1} = \max_{p_i^*}\left\{\varepsilon_{p_i^*}^{J_1}\right\} \text{ and } \overline{\varepsilon}_{p_i}^{J_2} = \max_{p_i^*}\left\{\varepsilon_{p_i^*}^{J_2}\right\}. \tag{22}$$

If such maximal error is sufficiently *small* (e.g., lower than 1% in the present paper), then there exists *no* realization within the *entire* domain of parameter p_i that affects appreciably the output Q: in other words, the uncertainty of parameter p_i can be considered *not important* in the analysis of Q and p_i can thus be *fixed* to a constant value in the corresponding mathematical model of Q.

2.2 Application Results

A greedy search strategy is applied to tackle the problem. For each input parameter $\{p_i: i = 1, 2, \ldots, 21\}$ a series of N_{p^*} equally spaced values $p_{i,k}^*$, $k = 1, 2, \ldots, N_{p^*}$, is selected deterministically within the corresponding ranges of variation and the associated maximal 'errors' are evaluated as $\overline{\varepsilon}_{p_i}^{Q} = \max_{p_{i,k}^*}\left\{\varepsilon_{p_{i,k}^*}^{Q}\right\}$, $Q = x_j, J_1, J_2$ (notice that $N_{p^*} = 2000$ for p_4 and p_5, whereas $N_{p^*} = 100$ for all the other parameters ranging within [0, 1]). The values of $\overline{\varepsilon}_{p_i}^{Q} = \max_{p_{i,k}^*}\left\{\varepsilon_{p_{i,k}^*}^{Q}\right\}$, $Q = x_j, J_1, J_2$, $i = 1, 2, \ldots, 21$, are reported in Table 4. For illustration purposes, those maximal errors $\overline{\varepsilon}_{p_i}^{Q}$ that are lower than or equal to the (arbitrarily chosen) threshold of 1% (i.e., $\overline{\varepsilon}_{p_i}^{Q} < 1\%$) are indicated in bold to highlight the

corresponding *non influential* parameters. Also, the fixed constant values $p_i^*\left(\bar{\varepsilon}_{p_i}^Q\right)$ that produce such maximal errors $\bar{\varepsilon}_{p_i}^Q$ (i.e., $p_i^*\left(\bar{\varepsilon}_{p_i}^Q\right) = \arg\max_{p_i}\left\{\bar{\varepsilon}_{p_i}^Q\right\}$) are reported in parentheses *only* for these non influential parameters.

		Sensitivity in a factor fixing setting		
		Maximal errors, $\bar{\varepsilon}_{p_i}^Q = \max_{p_i}\left\{\varepsilon_{p_i}^Q\right\}$ ($p_i^*\left(\bar{\varepsilon}_{p_i}^Q\right) = \arg\max_{p_i}\left\{\bar{\varepsilon}_{p_i}^Q\right\}$)		
Param.	**Cat.**	**$Q = x_i$**	**$Q = J_1$**	**$Q = J_2$**
p_1	III	100%	100%	100%
p_2	II	5.87%	7.95%	3.27%
p_3	I	5.38%	1.02%	**0.31% (0.60)**
p_4	III	24.33%	9.62%	13.65%
p_5	III	36.29%	97.43%	29.16%
p_6	II	91.99%	**0.79% (0.08 or 0.92)**	1.81%
p_7	III	100%	1.57%	2.03%
p_8	III	34.08%	**$8.27\cdot10^{-2}$% (1.00)**	**0.63% (1.00)**
p_9	I	8.31%	**$1.36\cdot10^{-2}$% (1.00)**	**0.14% ([0.98, 1.00])**
p_{10}	III	3.55%	**$3.58\cdot10^{-3}$% (0.00)**	**$2.86\cdot10^{-2}$% ([0.00, 0.45])**
p_{11}	I	16.45%	8.71%	18.85%
p_{12}	II	100%	92.07%	73.21%
p_{13}	III	**0.53% (0.70)**	**$3.58\cdot10^{-3}$% (0.00)**	**$1.43\cdot10^{-2}$% ([0.13, 1.00])**
p_{14}	III	**0.53% (0.30)**	**$1.59\cdot10^{-2}$% (1.00)**	**$7.14\cdot10^{-2}$% ([0.29, 0.31])**
p_{15}	III	**0.53% (1.00)**	**$5.53\cdot10^{-2}$% (0.96)**	**0.84% ([0.99, 1.00])**
p_{16}	II	88.93%	4.02%	2.17%
p_{17}	III	89.11%	1.58%	2.74%
p_{18}	III	100%	3.21%	2.41%
p_{19}	I	3.20%	**$4.85\cdot10^{-2}$% (1.00)**	**0.11% ([0.99, 1.00])**
p_{20}	III	28.16%	**0.84% (1.00)**	**0.80% ([0.99, 1.00])**
p_{21}	III	Not applicable	96.73%	5.33%

Table 4. Maximal errors $\bar{\varepsilon}_{p_i}^Q = \max_{p_i}\left\{\varepsilon_{p_i}^Q\right\}$, $Q = x_j, J_1, J_2$, produced by setting p_i to a constant value within its range of variation. Errors $\bar{\varepsilon}_{p_i}^Q$ < 1% are highlighted in bold to indicate those parameters p_i that do not significantly affect the output quantity Q of interest. The fixed constant values $p_i^*\left(\bar{\varepsilon}_{p_i}^Q\right)$ that produce such maximal errors $\bar{\varepsilon}_{p_i}^Q$ (i.e., $p_i^*\left(\bar{\varepsilon}_{p_i}^Q\right) = \arg\max_{p_i}\left\{\bar{\varepsilon}_{p_i}^Q\right\}$) are reported in parentheses only for the non influential parameters

From the analysis of the indicator $Q = x_j$, it can be seen that p_{13}, p_{14} and p_{15} are the only parameters that can be set to *any* constant within their *entire* ranges of variation [0, 1] with almost *no* influence on the p-box of the corresponding intermediate variable x_3: actually, $\bar{\varepsilon}_{p_i}^{x_3}$ = 0.53% << 1%, i = 13, 14, 15. These results suggest that the uncertainty of p_{13}, p_{14} and p_{15} is *not relevant* in determining the characteristics of the p-box of x_3 and could thus be *neglected*. Also notice that this outcome is in agreement with the (very low) importance of these parameters in 'building' the epistemic uncertainty in the p-box of x_3 (see Table 3). With respect to that, only for illustration purposes and by way of example Figure 6 left depicts the upper and lower CDFs of x_3 obtained by propagating the original uncertainty model (solid lines) and those produced by fixing p_{13} to 0 (dashed lines) and to 1 (dot-dashed

lines): in all cases the CDFs completely overlap. All the other parameters are found to have a *non negligible* effect on the p-boxes of the corresponding intermediate variables: actually, the maximal errors $\bar{\varepsilon}_{p_i}^{x_j}$ produced range from 3.20% (p_{19}) to 100% (p_1). This means that there are *at least some parts* of the domain of variation of such parameters whose contribution to the uncertainty of the corresponding intermediate variables x_j is significant. In this respect, for the sake of illustration Figure 6 right reports the upper and lower CDFs of x_2 obtained by propagating the original uncertainty model (solid lines) and by fixing p_8 to 0.5 (dashed lines) and to 1 (dot-dashed lines): when p_8 = 0.5, the two p-boxes completely overlap, whereas when p_8 = 1 the mismatch between the bounding CDFs is significant.

Similar analyses can be carried out with respect to indicators $Q = J_1$ and J_2. The parameters that can be set to *any* constant within their *entire* ranges of variation [0, 1] with almost no influence on the bounds of J_1 are p_6, p_8, p_9, p_{10}, p_{13}, p_{14}, p_{15}, p_{19} and p_{20}: actually, the corresponding maximal errors $\bar{\varepsilon}_{p_i}^{J_1}$ produced range from $3.58 \cdot 10^{-5}$% (for p_{13}) to 0.84% (for p_{20}) (i.e., they are far below 1%). This outcome is in agreement with the relatively low importance of these parameters in 'building' the epistemic uncertainty of J_1 (see Table 3): in facts, the corresponding sensitivity rankings $R_i(J_1)$ vary from 9 (for p_6) to 17 (for p_{13}). Figure 7 left shows the upper and lower bounds of J_1 obtained by propagating the original uncertainty model (solid lines) and those produced by fixing p_{20} to different values within its range of variation (circles): it is evident that these bounds tend to overlap for *all* the possible values of p_{20}. Instead, the other parameters (with particular reference to p_1, p_2, p_4, p_5, p_{11}, p_{12}, p_{16} and p_{21}) have at least a portion of their domain of variation whose contribution to the uncertainty of J_1 is non negligible: actually, the maximal errors $\bar{\varepsilon}_{p_i}^{J_1}$ produced range from 4.02% (for p_{16}) to 100% (for p_1). With respect to that, by way of example Figure 7 right shows the upper and lower bounds of J_1 obtained by propagating the original uncertainty model (solid lines) and those produced by fixing p_4 to different values within its range of variation (circles): it is evident that these bounds tend to overlap *only* when $p_4 \in$ [6.61, 6.63] and $p_4 \approx 8.8$, whereas they differ *significantly* when p_4 lies far from these values (actually, $\bar{\varepsilon}_{p_4}^{J_1}$ = 9.62%). Again, these outcomes are coherent with the results of the sensitivity analysis of the previous Section that highlighted the importance of such parameters in the determination of the uncertainty of J_1. Finally, concerning the remaining parameters (i.e., p_7, p_{17} and p_{18}), the following consideration is in order. Although they are not so relevant in building the epistemic uncertainty of J_1 (actually, their rank $R_i(J_1)$ is 8, 11 and

10, respectively), according to the present analysis they *cannot* be *completely neglected* in the system model: actually, the corresponding maximal errors $\bar{\varepsilon}_{p_i}^{J_1}$ produced are 1.57%, 1.58% and 3.21%, respectively.

Discussions about indicator J_2 are similar and not reported here for brevity (see Table 4 for details): notice that the uncertainty of parameters p_3, p_8, p_9, p_{10}, p_{13}, p_{14}, p_{15}, p_{19} and p_{20} seems to have *very little* or *no effect* on J_2 (actually, they can be set to *any* constant within their *entire* ranges of variation [0, 1] producing errors that do not exceed 0.84%). Figure 8 depicts only two exemplary (and different) situations with reference to parameters p_{15} (left) and p_{21} (right). Fixing p_{15} within its entire range [0, 1] does not lead to significant variations in the bounds of J_2 ($\bar{\varepsilon}_{p_{15}}^{J_2}$ = 0.84%). Instead, p_{21} seems to have insignificant effect on J_2 only within [0.00, 0.35], whereas its contribution becomes quite relevant, e.g., in [0.40, 0.90] ($\bar{\varepsilon}_{p_{21}}^{J_2}$ = 5.33%).

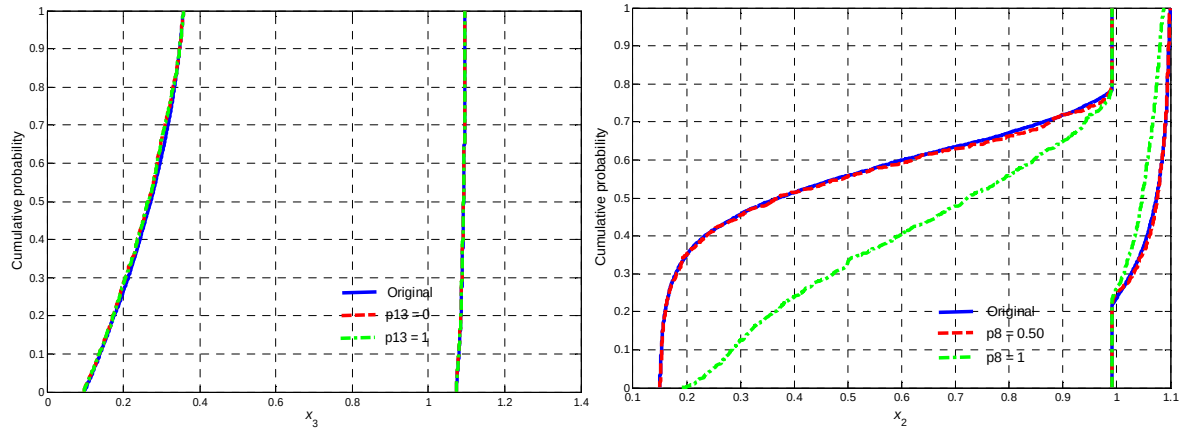

Figure 6. Left: upper and lower CDFs of x_3 obtained by propagating the original uncertainty model (solid lines) and by fixing p_{13} to 0 (dashed lines) and to 1 (dot-dashed lines); right: upper and lower CDFs of x_2 obtained by propagating the original uncertainty model (solid lines) and by fixing p_8 to 0.5 (dashed lines) and to 1 (dot-dashed lines)

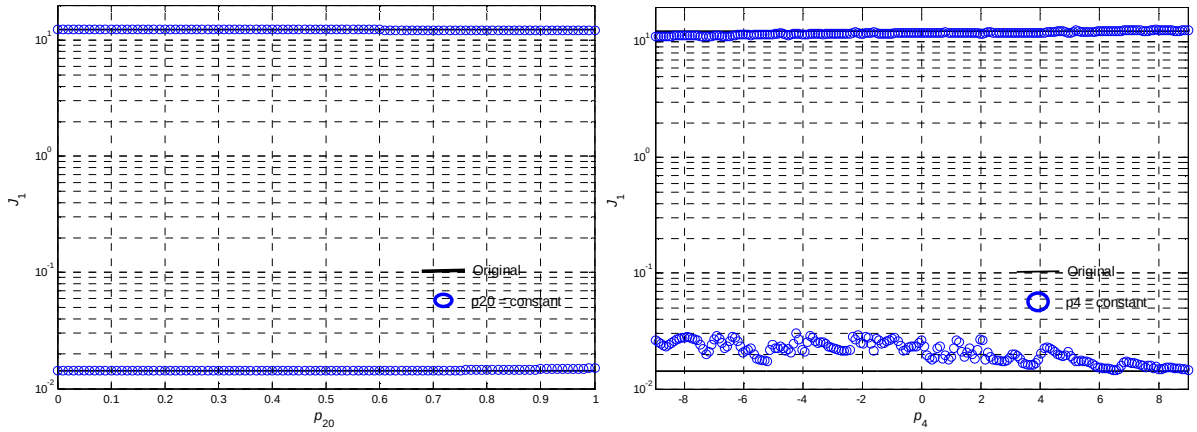

Figure 7. Left: upper and lower bounds of J_1 obtained by propagating the original uncertainty model (solid lines) and by fixing p_{20} to different values within its range of variation (circles); right: upper and lower

bounds of J_1 obtained by propagating the original uncertainty model (solid lines) and by fixing p_4 to different values within its range of variation (circles)

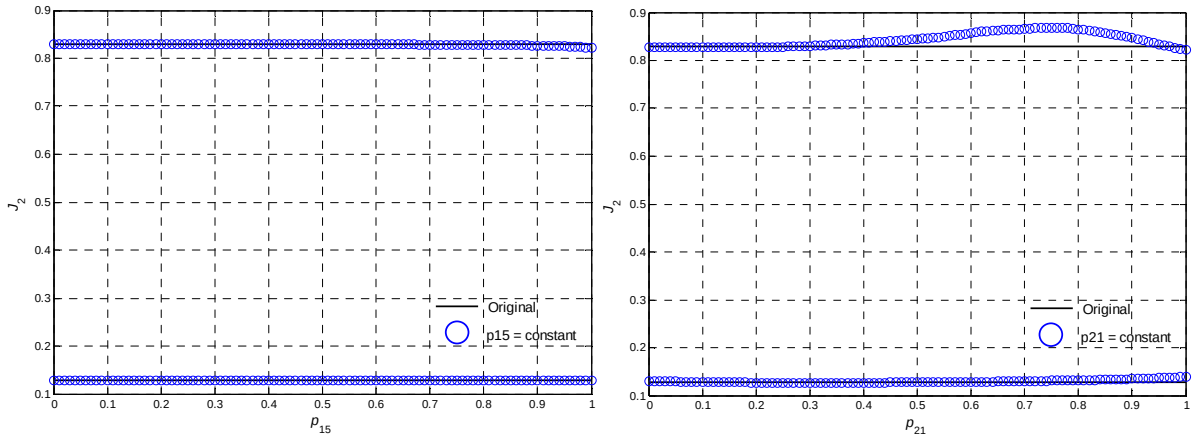

Figure 8. Left: upper and lower bounds of J_2 obtained by propagating the original uncertainty model (solid lines) and by fixing p_{15} to different values within its range of variation (circles); right: upper and lower bounds of J_1 obtained by propagating the original uncertainty model (solid lines) and by fixing p_{21} to different values within its range of variation (circles)

C. Subproblem (C): Uncertainty Propagation

This subproblem aims at finding the range of the metrics J_1 (13) and J_2 (14) that result from propagating both the original uncertainty model and an improved one (provided by the challengers). In Section III.C.1, the general approach adopted to propagate the input uncertainties onto the system performance metrics J_1 (13) and J_2 (14) is illustrated in detail; in Section III.C.2, the results of the application of the method to the tasks of Subproblem (C) are reported.

1. The Proposed Approach

The objective is to obtain the interval $[\underline{Q}, \overline{Q}]$ for the metrics of interest $Q = J_1 = E_{\boldsymbol{p}}[w(\boldsymbol{p}, \boldsymbol{d})]$ (13) and $Q = J_2 = P[w(\boldsymbol{p}, \boldsymbol{d}) > 0]$ (14) that result from propagating the mixed aleatory and epistemic uncertainty affecting the input parameters $\boldsymbol{p}$. As already mentioned in the previous Section III.B, this amounts in general to solving the following optimization problems:

$$\underline{Q} = \underline{J}_1 = \min_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{J_1\} = \min_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{E_{\boldsymbol{p}}[w(\boldsymbol{p}, \boldsymbol{d})]\}, \quad \overline{Q} = \overline{J}_1 = \max_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{J_1\} = \max_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{E_{\boldsymbol{p}}[w(\boldsymbol{p}, \boldsymbol{d})]\}, \tag{23}$$

$$\underline{Q} = \underline{J}_2 = \min_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{J_2\} = \min_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{P[w(\boldsymbol{p}, \boldsymbol{d}) > 0]\}, \quad \overline{Q} = \overline{J}_2 = \max_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{J_2\} = \max_{\boldsymbol{\theta}^{ali} \in \boldsymbol{\Omega}^{ali}} \{P[w(\boldsymbol{p}, \boldsymbol{d}) > 0]\}. \tag{24}$$

It is worth remembering that $\boldsymbol{\theta}^{all}$ is the vector containing: (i) *all* the epistemically-uncertain category (II) parameters and (ii) *all* the epistemically-uncertain internal coefficients of the probability distributions $q^{p_i}\left(p_i|\boldsymbol{\theta}_i\right)$ of the inputs p_i, i = 1, 2, …, 21; also, $\boldsymbol{\Omega}^{all}$ is the corresponding space of variation.

In this paper, problems (23) and (24) are tackled by embedding the Monte Carlo Simulation (MCS) technique for uncertainty propagation within a Genetic Algorithms (GAs) search for the extreme values $\underline{Q}$ and $\overline{Q}$. In more detail, for obtaining $\underline{Q}$ (resp., $\overline{Q}$), the following conceptual steps have to be performed:

1) the GA conducts its search using a population of *candidate* solutions $\{\boldsymbol{\theta}^{all,c}: c = 1, 2, \ldots, N_{pop}\}$ 'sampled' within the corresponding space of variation $\boldsymbol{\Omega}^{all}$;
2) for *each* candidate solution $\boldsymbol{\theta}^{all,c}$, the (aleatory) uncertainty in the input parameters $\boldsymbol{p}$ is *propagated* to the output metric Q of interest by MCS.[30-31] In more detail:
   a. N_a realizations $\{\boldsymbol{p}_{i_a}: i_a = 1, 2, \ldots, N_a\}$ of the input parameters $\boldsymbol{p}$ are randomly sampled from the corresponding probability distributions $\boldsymbol{q}^{\boldsymbol{p}}\left(\boldsymbol{p}|\boldsymbol{\theta}^{all,c}\right)$;
   b. using the realization $\boldsymbol{p}_{i_a}$, the value $Q(\boldsymbol{\theta}^{all,c})$ of the output metric Q of interest is estimated. In particular, if $Q = J_1$, then $Q(\boldsymbol{\theta}^{all,c}) \approx 1\Big/ N_a \cdot \sum_{i_a=1}^{N_a} w\left(\boldsymbol{p}_{i_a}, \boldsymbol{d}\right)$; instead, if $Q = J_2$, then $Q(\boldsymbol{\theta}^{all,c}) \approx 1\Big/ N_a \cdot \sum_{i_a=1}^{N_a} I_w\left(\boldsymbol{p}_{i_a}, \boldsymbol{d}\right)$, where $I_w\left(\boldsymbol{p}_{i_a}, \boldsymbol{d}\right) = 1$, when $w\left(\boldsymbol{p}_{i_a}, \boldsymbol{d}\right) > 0$, and 0, otherwise.
3) on the basis of the estimates of $Q(\boldsymbol{\theta}^{all,c})$ computed at step (2.b) above, the GA 'intelligently' drives the population of possible solutions $\{\boldsymbol{\theta}^{all,c}: c = 1, 2, \ldots, N_{pop}\}$ towards the (near) optimal region of the search space $\boldsymbol{\Omega}^{all}$; at the end of the search, $\underline{Q}$ (resp., $\overline{Q}$) $\approx \min_{c=1,2,\ldots,N_{pop}} \left\{Q\left(\boldsymbol{\theta}^{all,c}\right)\right\}$ (resp., $\max_{c=1,2,\ldots,N_{pop}} \left\{Q\left(\boldsymbol{\theta}^{all,c}\right)\right\}$).

Finally, notice that the sequence of steps (1)-(3) above have to be done *once* for *each* extreme bound of interest $\underline{J}_1$, $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$.

2. Application Results

Two different analyses are performed. In the first (subproblems (C.1) and (C.2)), the space of variation $\boldsymbol{\Omega}^{all}$ in (23) and (24) is the original one, i.e., that based on Table 1 and on the answer given to subproblem (A.3): then, $\boldsymbol{\Omega}^{all}$

= $\boldsymbol{\Omega}_{n_{d3}}^{all} = \boldsymbol{\Omega}_{n_{d3}}^{x_1} \text{x} \prod_{j=1}^{4} \boldsymbol{\Omega}^{x_j}$ (see Section III.A). In the second (subproblems (C.3) and (C.4)), parameters p_1, p_5, p_{12} and p_{21} are selected by the authors on the basis of the sensitivity rankings obtained in subproblems (B.2) and (B.3) (see Section III.B) and improved uncertainty models $\boldsymbol{\Omega}_{1,red}$, $\boldsymbol{\Omega}_{5,red}$, $\boldsymbol{\Omega}_{12,red}$ and $\boldsymbol{\Omega}_{21,red}$, respectively, are provided for them by the challengers: then, in this case $\boldsymbol{\Omega}^{all} = \boldsymbol{\Omega}_{red}^{all} = \boldsymbol{\Omega}_{red}^{x_1} \text{ x } \boldsymbol{\Omega}_{12,red} \text{ x } \boldsymbol{\Omega}_{21,red} \text{ x } \prod_{i=6,i\neq12}^{20} \boldsymbol{\Omega}_i = [(\boldsymbol{\Omega}_{1,red} \text{ x } \boldsymbol{\Omega}_{5,red}) \cap \boldsymbol{\Omega}_{n_{d3}}^{x_1}] \text{ x } \boldsymbol{\Omega}_{12,red} \text{ x } \boldsymbol{\Omega}_{21,red} \text{ x } \prod_{i=6,i\neq12}^{20} \boldsymbol{\Omega}_i$ (in passing, notice that $\boldsymbol{\Omega}_{red}^{x_1}$ is the *new*, *reduced* joint space of variation of the epistemically-uncertain parameters/coefficients of category (II) and (III) inputs p_1, p_2, p_4 and p_5 to intermediate variable x_1: in particular, $\boldsymbol{\Omega}_{red}^{x_1}$ is given by the *intersection* between space $\boldsymbol{\Omega}_{n_{d3}}^{x_1}$ – improved in Section III.A by means of n_{d3} = 50 data – and the further improved ranges $\boldsymbol{\Omega}_{1,red}$ and $\boldsymbol{\Omega}_{5,red}$ of p_1 and p_5 provided by the challengers). As highlighted in Section III.A, the identification of those solutions that belong to set $\boldsymbol{\Omega}_{n_{d3}}^{all}$ is guaranteed by introducing the property in (8) as a *hard constraint* in the GA: only those candidates that satisfy such property are retained in the genetic evolution, whereas the others are discarded.

As in Section III.B, the original system model $w(\boldsymbol{p},\boldsymbol{d}) = \max_{1\le o\le 8} g_o = \max_{1\le o\le 8} f_o(\boldsymbol{x}=\boldsymbol{h}(\boldsymbol{p}),\boldsymbol{d})$ (1) is replaced by an ANN regression model to reduce the computational burden associated to the solution of (23) and (24). Notice that N_a = 10000 random realizations $\boldsymbol{p}_{i_a}$ of $\boldsymbol{p}$ are sampled to propagate uncertainty by MCS for both J_1 and J_2 (step 2.a of Section III.C.1). The intervals $[\underline{J}_1, \overline{J}_1]^{ANN}$ and $[\underline{J}_2, \overline{J}_2]^{ANN}$ resulting from the optimization searches within both $\boldsymbol{\Omega}_{n_{d3}}^{all}$ and $\boldsymbol{\Omega}_{red}^{all}$ are reported in Table 5; the relative reduction in the length of these intervals due the improvement of the uncertainty models of the selected parameters p_1, p_5, p_{12} and p_{21} is shown in parentheses. It can be seen that the width of the intervals $[\underline{J}_1, \overline{J}_1]^{ANN}$ and $[\underline{J}_2, \overline{J}_2]^{ANN}$ has been reduced by 90.92% and 74.20%, respectively, after the refinement of the uncertainty models of parameters p_1, p_5, p_{12} and p_{21} selected according to our sensitivity analysis (Section III.B.1). In order to validate *a posteriori* the results obtained using the ANN meta-model, the optimal solutions $\boldsymbol{\theta}^{all}$ thereby found are sent in input to the *real* system model and the corresponding intervals $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_2, \overline{J}_2]$ are re-calculated (highlighted in bold in Table 5). It can be seen that the results are in satisfactory agreement, confirming the effectiveness of ANNs in mapping complicated nonlinear and non-monotonic functions. Also, it can be further verified that the width of the intervals $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_2, \overline{J}_2]$ have been significantly reduced

(i.e., by 99.01% and 72.26%, respectively) after the refinement of the uncertainty models of parameters p_1, p_5, p_{12} and p_{21} selected by our sensitivity analysis (Section III.B.1).

Uncertainty Propagation, N_a = 10000			
		Original uncertainty model, $\Omega^{all}_{n_{d3}}$ (after Section III.A)	**Improved uncertainty model, Ω^{all}_{red} (after Section III.B)**
J_1	$[\underline{J}_1, \overline{J}_1]^{ANN}$	[0.0142, 12.2756]	[0.0583, 0.1721] (90.92%)
	$[\underline{J}_1, \overline{J}_1]$	**[0.0129, 13.1552]**	**[0.0316, 0.1612] (99.01%)**
J_2	$[\underline{J}_2, \overline{J}_2]^{ANN}$	[0.1285, 0.8288]	[0.2778, 0.4585] (74.20%)
	$[\underline{J}_2, \overline{J}_2]$	**[0.0900, 0.8142]**	**[0.2389, 0.4398] (72.26%)**

Table 5. Intervals $[\underline{J}_1, \overline{J}_1]^{ANN}$ (23) and $[\underline{J}_2, \overline{J}_2]^{ANN}$ (24) of performance metrics J_1 (13) and J_2 (14) obtained by embedding ANNs regression models and MCS (with N_a = 10000 samples) within a GA optimization search; the relative reduction in the length of these intervals due the improvement of the uncertainty models of the selected parameters p_1, p_5, p_{12} and p_{21} is shown in parentheses. The intervals $[\underline{J}_1, \overline{J}_1]$ (23) and $[\underline{J}_2, \overline{J}_2]$ (24) resulting from the a posteriori validation of the optima found on the real system model are also reported

Finally, in order to take into account the *statistical variability* in the estimates of J_1 and J_2 (obtained by plain random sampling), the upper and lower bounds of the corresponding intervals (see Table 5) are 'extended' above and below, respectively, of an amount equal to two standard deviations: the 'conservative' estimates thereby obtained (i.e., $[\underline{J}_1, \overline{J}_1]^{cons} = [\underline{J}_1 - 2\sigma_{\underline{J}_1}, \overline{J}_1 + 2\sigma_{\overline{J}_1}]$ and $[\underline{J}_2, \overline{J}_2]^{cons} = [\underline{J}_2 - 2\sigma_{\underline{J}_2}, \overline{J}_2 + 2\sigma_{\overline{J}_2}]$, respectively) are reported in Table 6.

Uncertainty Propagation, N_a = 10000: 'conservative' estimates			
		Original uncertainty model, $\Omega^{all}_{n_{d3}}$ (after Section III.A)	**Improved uncertainty model, Ω^{all}_{red} (after Section III.B)**
J_1	$[\underline{J}_1, \overline{J}_1]^{cons}$	[0.0108, 15.4437]	[0.0278, 0.1693]
J_2	$[\underline{J}_2, \overline{J}_2]^{cons}$	[0.0843, 0.8220]	[0.2304, 0.4497]

Table 6. Intervals $[\underline{J}_1, \overline{J}_1]^{cons} = [\underline{J}_1 - 2\sigma_{\underline{J}_1}, \overline{J}_1 + 2\sigma_{\underline{J}_1}]$ and $[\underline{J}_2, \overline{J}_2]^{cons} = [\underline{J}_2 - 2\sigma_{\underline{J}_2}, \overline{J}_2 + 2\sigma_{\underline{J}_2}]$ of performance metrics J_1 (13) and J_2 (14) obtained by 'extending' the upper and lower bounds $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_2, \overline{J}_2]$ of J_1 and J_2 (see Table 5) above and below of two standard deviations, respectively

A consideration is in order with respect to the use of GA for the identification of the extreme values $\underline{Q}$ and $\overline{Q}$ of a safety variable Q of interest (i.e., J_1 or J_2 in this case). Although GA is a *global* optimizer, in some problems (characterized by massive multimodality of the objective function to be optimized), it may converge to *local* optima. In such a case, the lower bound $\underline{Q}$ would be overestimated, whereas the upper bound $\overline{Q}$ would be underestimated:

in other words, the analyst would *underestimate* the range of epistemic uncertainty in Q. Such a situation can lead the decision maker to the wrong decision (e.g., to an estimate of the largest failure probability that is much lower than the actual value): this is particularly dangerous in the risk assessments of safety-critical systems, such as the aerospace, nuclear and chemical ones. In this view, the performance of the GA is a key issue to avoid such underestimations. The performance of GA depends largely on its ability to thoroughly explore the search space (i.e., to maintain a sufficient "genetic diversity" in the population of candidate solutions), while attempting to efficiently and intelligently drive the search towards the "interesting region" of the search space, i.e., towards the global optimum. On one hand, a thorough exploration of the search space (i.e., a sufficient "genetic diversity") is guaranteed by the following strategies[49,50]: (i) GA is *repeated* several times (say, ten times) with different random seeds (i.e., *different* random *initial populations*) and only the best result over all the simulation is retained; (ii) some of the GA parameters are properly set: for example, a relatively high population size (i.e., $N_{pop} = 100$) is employed. On the other hand, an efficient identification of the global optimum is favoured (but *not guaranteed*) by the following techniques[49,50]: (i) *fitness-guided* candidate selection procedures are adopted: in other words, the probability that a candidate solution survives during the GA evolution is proportional to its objective function (i.e., to its 'quality' or 'fitness'); (ii) *elitism* is implemented: at each generation some of the individuals of the current population (e.g., the *best* $0.1 \cdot N_{pop}$) are deterministically selected to be part of the next population, so that the best genetic code is guaranteed to be propagated; (iii) differently from subproblem A, the algorithm stops only if the average relative change in the best fitness function value over a given number of generations (e.g., 50) is less than or equal to a given tolerance (e.g., 10^{-6}).

D. Subproblem (D): Extreme Case Analysis

In Section III.D.1, the general approach adopted to identify the realizations of the epistemically-uncertain parameters/coefficients $\boldsymbol{\theta}^{all}$ leading to the extreme bounds of the ranges $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_2, \overline{J}_2]$ of J_1 and J_2 is described; in Section IV.D.2, the results of the application of the method to the tasks of Subproblem (D) are reported.

1. The Proposed Approach

The realizations of the epistemically-uncertain parameters/coeffieints leading to the extreme bounds of the ranges $[\underline{J}_1, \overline{J}_1]$ and $[\underline{J}_2, \overline{J}_2]$ of J_1 and J_2 are formally defined as follows:

$$\boldsymbol{\theta}_{1,low}^{all} = \arg\min_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{J_1\} = \arg\min_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{E_{\boldsymbol{p}}[w(\boldsymbol{p},\boldsymbol{d})]\},\ \boldsymbol{\theta}_{1,up}^{all} = \arg\max_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{J_1\} = \arg\max_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{E_{\boldsymbol{p}}[w(\boldsymbol{p},\boldsymbol{d})]\} \tag{25}$$

$$\boldsymbol{\theta}_{2,low}^{all} = \arg\min_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{J_2\} = \arg\min_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{P[w(\boldsymbol{p},\boldsymbol{d})>0]\},\ \boldsymbol{\theta}_{2,up}^{all} = \arg\max_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{J_2\} = \arg\max_{\boldsymbol{\theta}^{all}\in\boldsymbol{\Omega}^{all}}\{P[w(\boldsymbol{p},\boldsymbol{d})>0]\}. \tag{26}$$

Notice that in our approach, solutions to (25) and (26) are obtained in the *same* GA optimization searches carried out to identify the extreme bounds of the ranges [$\underline{J}_1$, $\overline{J}_1$] and [$\underline{J}_2$, $\overline{J}_2$] (see the previous Section III.C.1). Notice that $\boldsymbol{\theta}_{1,low}^{all}$ and $\boldsymbol{\theta}_{2,low}^{all}$ correspond to extreme 'best-case' configurations (i.e., to parameters settings that produce *lower* – i.e., *safer* – values of metrics J_1 and J_2); on the contrary, $\boldsymbol{\theta}_{1,up}^{all}$ and $\boldsymbol{\theta}_{2,up}^{all}$ correspond to extreme 'worst-case' scenarios (i.e., to parameters settings that produce *higher* – i.e., *more risky* – values of metrics J_1 and J_2).

2. Application Results

As before, the solutions are searched for in two different spaces of variation, i.e., the original one $\boldsymbol{\Omega}_{n_{d3}}^{all}$ and the improved one $\boldsymbol{\Omega}_{red}^{all}$. The corresponding values of $\boldsymbol{\theta}_{1,low}^{all}$, $\boldsymbol{\theta}_{1,up}^{all}$, $\boldsymbol{\theta}_{2,low}^{all}$ and $\boldsymbol{\theta}_{2,up}^{all}$ are reported in Table 7 for both cases. The corresponding extreme CDFs of x_1 (top left), x_2 (top right), x_3 (middle left), x_4 (middle right) and x_5 (bottom) are shown in Figures 9 and 10 for the original ($\boldsymbol{\Omega}_{n_{d3}}^{all}$) and reduced ($\boldsymbol{\Omega}_{red}^{all}$) models, respectively; the CDFs producing the extreme values $\underline{J}_1$, $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$ for J_1 and J_2 are depicted in solid, dashed, dot-dashed and dotted lines, respectively.

Extreme Case Analysis, N_a = 10000									
		Realizations of the epistemically-uncertain parameters/coefficients							
		J_1				J_2			
		Model $\Omega^{all}_{n_{d3}}$		Model Ω^{all}_{red}		Model $\Omega^{all}_{n_{d3}}$		Model Ω^{all}_{red}	
Parameter	**Category**	$\theta^{all}_{1,low}$	$\theta^{all}_{1,up}$	$\theta^{all}_{1,low}$	$\theta^{all}_{1,up}$	$\theta^{all}_{2,low}$	$\theta^{all}_{2,up}$	$\theta^{all}_{2,low}$	$\theta^{all}_{2,up}$
p_1	III	0.7791	0.6248	0.6399	0.6203	0.7438	0.6455	0.6400	0.6244
		0.0400	0.0337	0.0322	0.0330	0.0205	0.0266	0.0321	0.0322
p_2	II	0.6000	1.0000	0.8266	0.9999	0.8611	0.6656	0.1682	1.0000
p_3	I	/	/	/	/	/	/	/	/
p_4	III	3.0646	5.0000	4.8554	2.2933	1.4203	0.4386	4.7448	-4.8727
		0.6437	2.0000	1.5948	0.0556	0.6353	1.4342	0.0512	0.2724
		-0.1941	-0.9639	0.3332	0.4987	0.0294	-0.0417	0.1511	-0.3830
p_5	III	-0.4393	-2.9709	-1.3485	-1.5259	1.7777	2.0209	-1.5755	-1.3440
		1.3203	0.0502	0.5756	0.5493	0.5173	0.0574	0.6201	0.5677
		-0.1941	-0.9639	0.3332	0.4987	0.0294	-0.0417	0.1511	-0.3830
p_6	II	1.0000	0.5003	0.0025	0.5106	0.9989	0.9953	0.0098	0.8916
p_7	III	0.9820	0.9820	0.9835	0.9855	0.9876	1.0146	1.0266	1.3275
		1.0800	1.0800	1.0800	1.0799	1.0113	1.0523	1.0773	0.9712
p_8	III	7.4501	7.4501	7.4593	7.4600	12.7776	8.7347	10.4424	11.9942
		7.8640	7.8640	7.8589	7.8531	6.1472	7.7475	7.8057	6.7817
p_9	I	/	/	/	/	/	/	/	/
p_{10}	III	4.5130	4.5130	4.5115	4.5111	4.0797	4.4119	1.9878	4.4368
		1.5360	1.5361	1.5476	1.5425	4.0802	1.8005	2.0735	4.2122
p_{11}	I	/	/	/	/	/	/	/	/
p_{12}	II	0.1209	0.6330	0.9981	0.9675	0.6368	0.0997	0.9623	0.9968
p_{13}	III	0.4120	0.7370	0.4138	0.7355	0.6790	0.6146	0.7253	0.4162
		2.0680	1.0000	2.0680	1.0054	1.6002	1.6874	1.0958	2.0288
p_{14}	III	2.1690	0.9310	2.1675	0.9518	2.1283	1.2083	1.0167	2.1564
		1.7491	2.4070	1.0021	2.4067	1.1463	1.9519	2.4068	1.0014
p_{15}	III	7.0950	7.0950	5.4495	7.0937	7.0432	6.0305	6.5092	5.4787
		5.2871	5.2870	6.9383	5.2880	5.9964	6.2385	5.3701	6.8334
p_{16}	II	1.0000	0.7404	1.0000	0.2545	0.2492	0.9966	0.1298	1.0000
p_{17}	III	1.0600	1.0600	1.0607	1.0723	1.5096	1.0619	1.3060	1.0603
		1.4880	1.4880	1.4880	1.4864	1.1103	1.4866	1.1271	1.4862
p_{18}	III	4.2660	4.2660	4.2625	1.0032	2.1948	4.1990	2.3304	4.0693
		0.5531	0.5530	0.5680	0.9981	0.9529	0.5907	0.6932	0.6122
p_{19}	I	/	/	/	/	/	/	/	/
p_{20}	III	7.5300	7.5300	7.5389	13.4889	9.3426	8.5107	11.8411	12.6506
		8.1480	8.1480	8.1468	4.7143	6.8836	8.0741	6.7160	7.9054
p_{21}	III	0.4211	1.0000	0.8815	0.8707	0.9808	0.9621	0.9995	0.8959
		29.6210	7.7720	22.9982	22.9990	12.4134	7.8048	19.0219	19.8172

Table 6. Realizations $\theta^{all}_{1,low}$, $\theta^{all}_{1,up}$, $\theta^{all}_{2,low}$ and $\theta^{all}_{2,up}$ of the epistemically-uncertain parameters/coefficients θ^{all} leading to the extreme bounds $\underline{J}_1$, $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$, respectively, of the performance metrics J_1 and J_2 in the original and improved uncertainty models $\Omega^{all}_{n_{d3}}$ and Ω^{all}_{red}, respectively

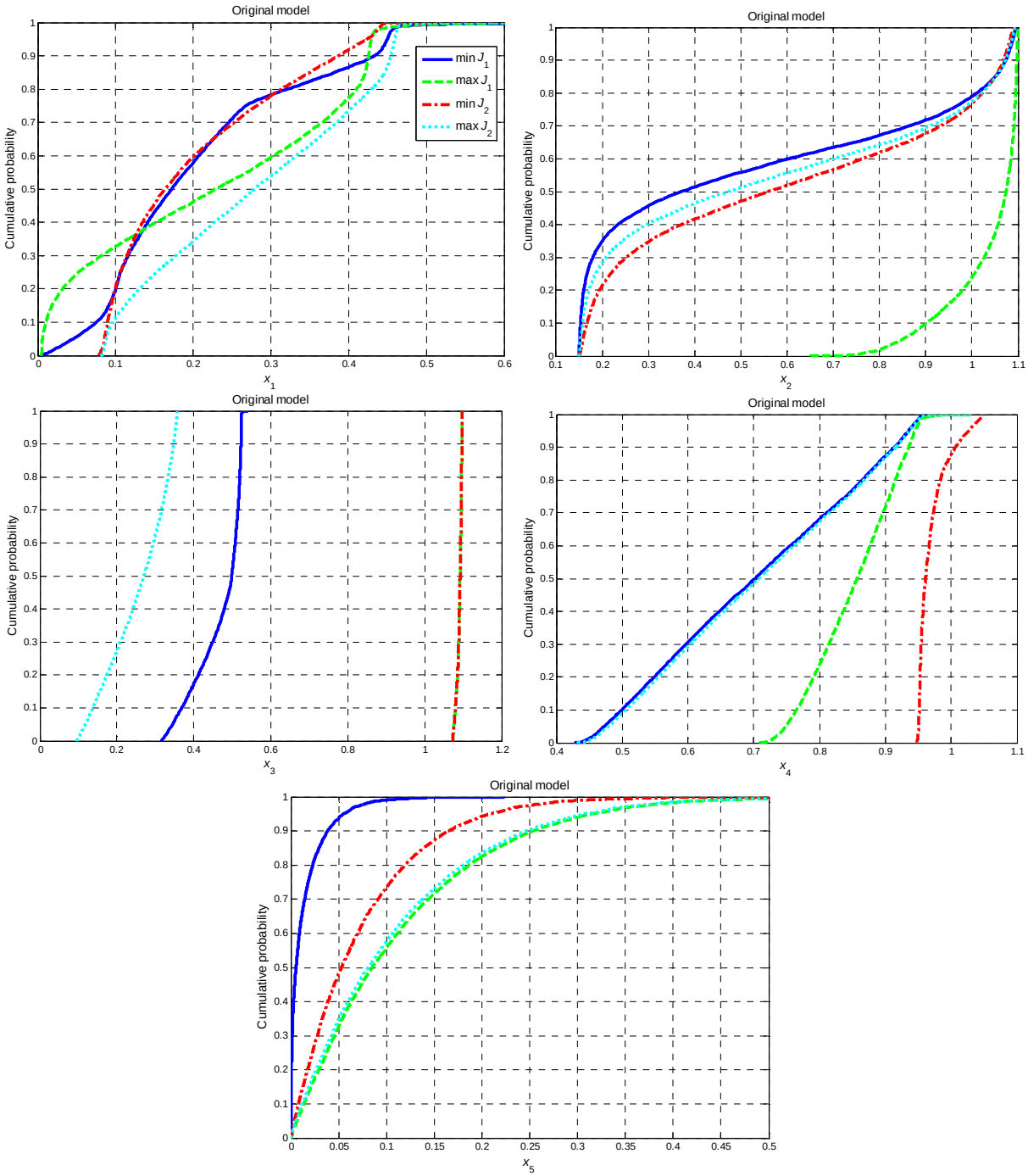

Figure 9. CDFs of x_1 (top left), x_2 (top right), x_3 (middle left), x_4 (middle right) and x_5 (bottom) producing the extreme values $\underline{J}_1$ (solid lines), $\overline{J}_1$ (dashed lines), $\underline{J}_2$ (dot-dashed lines) and $\overline{J}_2$ (dotted lines) for J_1 and J_2 for the original ($\Omega^{all}_{n_{g3}}$) model

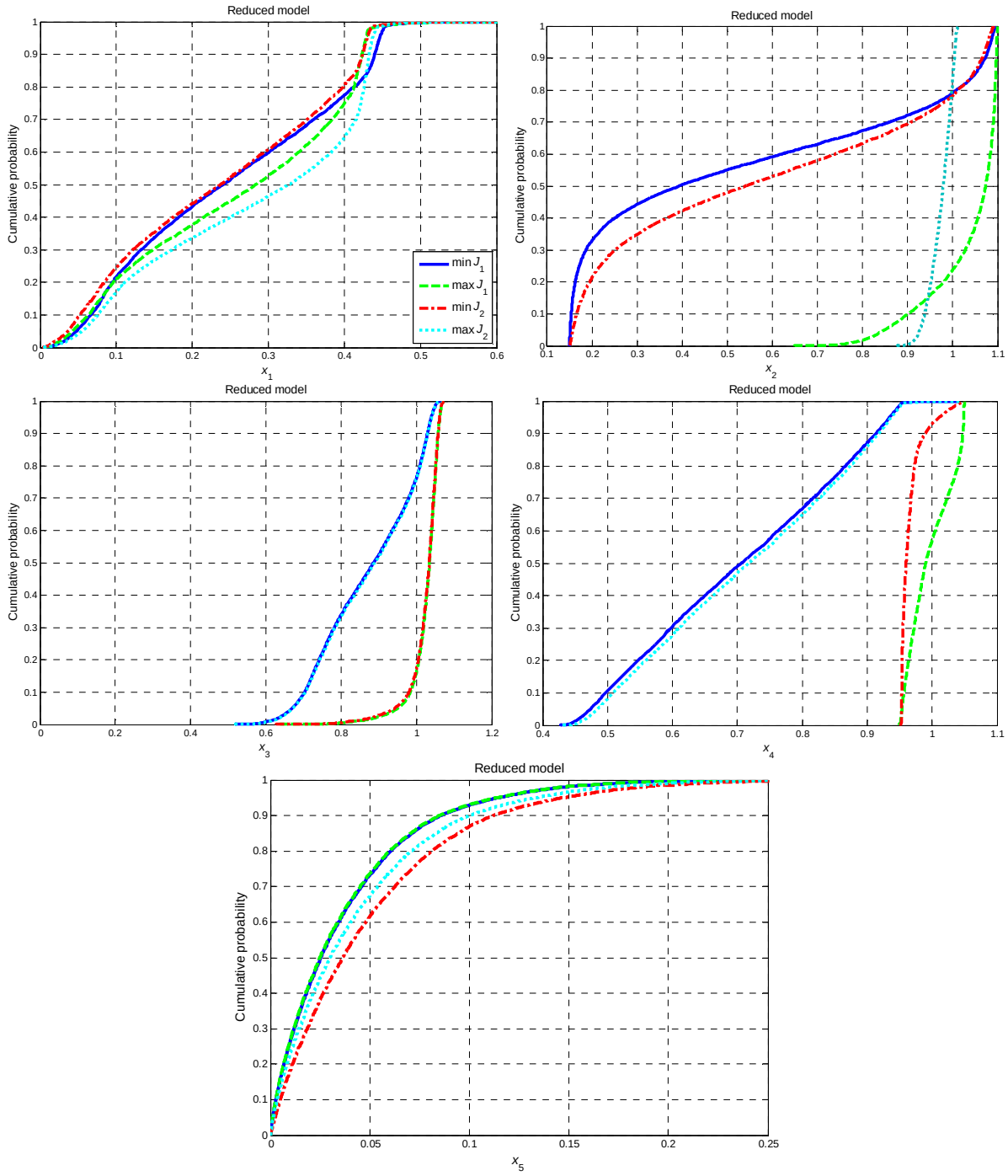

Figure 10. CDFs of x_1 (top left), x_2 (top right), x_3 (middle left), x_4 (middle right) and x_5 (bottom) producing the extreme values $\underline{J}_1$ (solid lines), $\overline{J}_1$ (dashed lines), $\underline{J}_2$ (dot-dashed lines) and $\overline{J}_2$ (dotted lines) for J_1 and J_2 for the improved (Ω_{red}^{all}) model

It can be seen that: (i) as expected, moving from the original ($\Omega_{n_{d3}}^{all}$) to the reduced (Ω_{red}^{all}) uncertainty model the CDFs of intermediate variables x_1, x_3 and x_5 (i.e., those variables containing the selected improved parameters p_1, p_5,

p_{12} and p_{21}) significantly change their shape (see, e.g., x_1 and x_3) and/or their location (see, e.g., x_3 and x_5); (ii) when the reduced uncertainty model ($\boldsymbol{\Omega}_{red}^{all}$) is used, then the parameters ranked as 'less relevant' in Table 2 come to play a non negligible role in 'building up' the epistemic uncertainty of J_1 and J_2. For example, moving from $\boldsymbol{\Omega}_{n_{d3}}^{all}$ (Figure 9) to $\boldsymbol{\Omega}_{red}^{all}$ (Figure 10) one can observe a significant modification in the CDF of x_2 that leads to the maximum value of J_2 (dotted lines in Figures 9 and 10, top right). Also, it is worth noting the change in the CDF of x_4 leading to the maximum value of J_1 (dashed lines in Figures 9 and 10, middle right); this is due, e.g., to the significant difference in the (optimal) values that the epistemically-uncertain parameters/coefficients of p_{16}, p_{18} and p_{20} (ranked 7, 11 and 10 in Table 3) assume when the improved model ($\boldsymbol{\Omega}_{red}^{all}$) is adopted instead of the original one ($\boldsymbol{\Omega}_{n_{d3}}^{all}$).

The different characteristics of the CDFs reported in Figures 9 and 10 allow exploring different parts of the space of variation of the intermediate variables x_j, j = 1, 2, …, 5, and consequently allow probing different areas of the system failure domain. To this aim, all the N_a = 80000 samples *randomly* generated in the uncertainty propagation phase of Section 4.C are taken into account (i.e., all the 8·10000 samples produced to estimate $\underline{J}_1$, $\overline{J}_1$, $\underline{J}_2$ and $\overline{J}_2$ using both the original – $\boldsymbol{\Omega}_{n_{d3}}^{all}$ – and the reduced – $\boldsymbol{\Omega}_{red}^{all}$ – uncertainty model); moreover, additional N = 20000 patterns *deterministically* selected previously to train and test the ANN (see Section VI.B) are added to provide a better covering of the intermediate variable space. Thus, a total of $N_{tot} = N_a + N$ = 100000 points $\{(\boldsymbol{x}_t, \boldsymbol{g}_t), t = 1, 2, ..., N_{tot} = 100000\}$ are analyzed with the objective of identifying those realizations of $\boldsymbol{x}$ leading to $J_2 > 0$; then, among all these failure configurations we identify few representative realizations that typify different possible failure scenarios (in terms of relationship between $\boldsymbol{x}$ and $\boldsymbol{g}$). In more detail, we proceed as follows:

i. we group those configurations of $\boldsymbol{x}$ that lead to the violation of the *same* requirements (i.e., that lead to the same failure scenarios): for example, those patterns $\boldsymbol{x}_t$ that lead to the violation of requirement g_4 alone are separated from those that cause the violation of requirements g_4 and g_6 together, and so on. By so doing, in the present case we identify NS = 63 different failure scenarios S_i, i = 1, 2, …, NS = 63;

ii. for *each* failure scenario S_i, i = 1, 2, …, NS = 63, we characterize the corresponding relations between $\boldsymbol{x}$ and $\boldsymbol{g}$ in order to identify all the combinations of intermediate variable values that lead to a given failure scenario. In order to do that automatically and to capture the complicated dependences between the variables x_j, j = 1, 2, …, 5, we perform k-means clustering on the configurations $\boldsymbol{x}_t$ belonging to a given scenario S_i. Without

going into technical details, *k*-means *clustering* is a partitioning method that separates a set of data $\boldsymbol{x}_t$ into into *k* mutually exclusive clusters; the partitions are such that the objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible (the classical Euclidean distance can be used to measure such distance). Each cluster in the partition is defined by its member objects and by its *centroid*, or *center*: the centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized. An iterative algorithm is employed that minimizes the sum of distances from each object to its cluster centroid, over all clusters; this algorithm moves objects between clusters until the sum cannot be decreased further. The result is a set of clusters that are as compact and well-separated as possible[62]. With respect to that, notice that in the present case for *each* scenario S_i many different clusters (and the corresponding centroids) may be identified, each one corresponding to one representative, archetypical combination of x_j-values that leads to the failure scenario S_i considered.

Table 7 reports a selection of 8 (out of 63) representative failure scenarios: three of them (indices 1-3) lead to the violation of only one requirement (g_4, g_6 and g_8), four of them (indices 4-7) to the violation of two constraints at the same time (g_4, g_7; g_3, g_4; g_5, g_8 and g_1, g_4) and one (index 8) to the violation of three constraints at the same time (g_2, g_4, g_6). This set has been selected because it represents a 'minimal' list of scenarios that contains examples of violations of all the requirements of interest (for example, we have not found any scenario where requirements g_1, g_2, g_3 or g_7 are violated alone; also, we have not found any scenario where requirement g_2 is violated in a group of less than three requirements): more complex scenarios involving the violation of 4, 5, …, 8 constraints at the same time can be obtained as intersections/unions of those reported in Table 7. The centroids of the corresponding clusters are also reported in the Table. Based on the values of these centroids, qualitative descriptions of the relationships between the intermediate variables x_j is given in parentheses: letters 'L', 'H' and 'A' mean that in order to generate the scenario of interest, the variable need to take low, high or any value within its range of variation. Notice that (i) obviously, the fact that a variable may take any value within its range suggests that it is not important in the definition of the scenario of interest; (ii) when a variable is not important, the k-means algorithm locates the corresponding coordinate of the centroid in the middle point of the range of variation of the variable.

N. of violated constraints	Scenario Index	Violated constraints	Cluster number	Intermediate variable values (centroids, representative realizations)				
				x_1	x_2	x_3	x_4	x_5
1	1	g_4	1	0.2466 ($0.05<x_1<0.34$)	0.6310 (A)	0.2904 (L)	0.7422 (A)	0.4072 (A)
			2	0.4300 ($0.425<x_1<0.53$)	0.6148 (A)	0.8517 (H)	0.7733 (A)	0.2453 (L)
			3	0.3750 ($0.34<x_1<0.4$)	0.6148 (A)	0.8517 (H)	0.7422 (A)	0.4072 (A)
	2	g_6	1	0.0288 (L)	0.6191 (A)	0.8832 (H)	0.7124 (A)	0.5609 (H)
	3	g_8	1	0.0245 (L)	0.6200 (A)	0.4566 (A)	0.8468 (H)	0.5640 (H)
2	4	g_4, g_7	1	0.7277 ($0.68<x_1<0.76$)	0.6191 (A)	0.1203 (L)	0.8114 (A-H)	0.3452 (A)
			2	0.7282 ($0.68<x_1<0.76$)	0.6185 (A)	0.2952 (L)	0.8144 (A-H)	0.3252 (A)
	5	g_1, g_4	1	0.4157 ($0.3985<x_1<0.42$)	0.8281 (H)	1.0166 (H)	0.8183 (A-H)	0.2106 (L)
	6	g_3, g_4	1	1.0546 (H)	0.5578 (L-A)	0.1503 (L)	0.8975 (H)	0.3499 (A)
	7	g_5, g_8	1	0.0107 (L)	1.0339 (H)	0.2593 (L)	0.4822 (L)	0.7179 (H)
3	8	g_2, g_4, g_6	1	0.3850 ($0.34<x_1<0.4$)	0.6148 (A)	0.2500 (L)	0.5500 (L-A)	0.2987 (A)
			2	0.4400 ($0.425<x_1<0.53$)	0.6431 (A)	0.1911 (L)	0.6112 (L-A)	0.3251 (A)
			3	0.6849 ($0.68<x_1<0.76$)	0.6344 (A)	0.9559 (H)	0.7322 (A)	0.3331 (A)

Table 7. Representative failure scenarios obtained by *k*-means clustering of system failure configurations

IV. Conclusion

In this work, we have considered the model of a twin-jet aircraft including twenty-one inputs and eight outputs (affected by mixed aleatory and epistemic uncertainties that are represented by probability distributions and intervals, respectively). Within this context, we have addressed and solved the following issues:

A. on the basis of a finite number of *empirical realizations* of one of the model outputs, the uncertainty models of five input parameters have been *improved* (i.e., the *epistemic uncertainty* in the corresponding internal coefficients/parameters has been *reduced*). In particular, Genetic Algorithms have been efficiently devised to identify *some* of the possible combinations of the epistemically-uncertain input parameters/coefficients leading to a distributional p-box for the output that is coherent with the available data (i.e., with the corresponding empirical CDF and the related Kolmogorov-Smirnov confidence bounds). A reduction of about 40% in the epistemic uncertainty has been obtained by means of 50 data;

B. sensitivity analysis has been carried out to study systematically how the inputs to the model influence the outputs. In particular, two tasks have been performed. In the first (namely, 'factor prioritization'), we

have *ranked* the input parameters according to *degree of reduction* in the *output epistemic uncertainty* which one could hope to obtain by refining their uncertainty models. To this aim, a novel *global sensitivity index* has been introduced, which has led to the identification of *four* (out of twenty-one) relevant parameters (p_1, p_5, p_{12} and p_{21}). This information is of paramount importance since it allows the analyst to focus his/her future empirical studies *mainly* on the refinement of these parameters. In the second analysis (namely, 'factor fixing'), on the contrary we have identified those parameters that *minimally* affect the outputs, i.e., those that can be assumed to take on a *fixed constant value* without producing significant errors. The analysis has led to find out *at least four* (out of twenty-one) parameters (p_{13}, p_{14}, p_{15} and p_{19}) that have practically *no influence* on the uncertainty of the system failure probability and of the mean of the worst case requirement metric (other four parameters, p_8, p_9, p_{10} and p_{20} could be also considered negligible in the analysis of the integrated system). This information suggests assigning *constant* values to these inputs from the mathematical system model, which produces a consistent *simplification* of the analysis.

In all the tasks related to sensitivity analysis, the original mathematical model of the system has been replaced by a *fast-running, surrogate regression model* based on Artificial Neural Networks (ANNs): this has allowed to *reduce* the associated *computational time* by about three orders of magnitude.

C. uncertainty has been propagated from the inputs to the outputs of the system model in order to identify the *extreme bounds* (i.e., the range) of two performance metrics of interest (i.e., the expected value of the worst-case requirement metric and the system failure probability). We have employed (i) standard MCS to propagate the aleatory uncertainty described by probability distributions and (ii) GAs to solve the numerous optimization problems related to the propagation of epistemic uncertainty by interval analysis. The uncertainty propagation phase has been carried out in two different 'system configurations': in the first, the original input uncertainty models were used; in the second, improved (i.e., less uncertain) models were adopted for the four input parameters identified in task (B). The use of the improved models has led to a *reduction* of 99% and 72% in the length of the intervals of the two metrics, confirming the relevance of the four parameters selected by sensitivity analysis;

D. within the uncertainty propagation phase (C), we have also identified the realizations of the epistemically-uncertain coefficients/parameters that yield the *extreme bounding* values of the two

performance metrics defined above. This information can be used to typify *best-* and *worst-case scenarios*, i.e., to identify which combinations of values of the epistemically-uncertain coefficients/parameters lead to the *smallest* and *largest* values, respectively, of the system performance indicators of interest.

References

[1]EPA, "Guidance on the Development, Evaluation, and Application of Environmental Models, Council for Regulatory Environmental Modeling", U.S. Environmental Protection Agency, Washington, DC, 20460, 2009.

[2]USNRC, "Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making", NUREG-1855, US Nuclear Regulatory Commission, Washington, DC, 2009.

[3]NASA, "Risk-Informed Decision Making Handbook", NASA/SP-2010-576 – Version 1.0, April 2010.

[4]Apostolakis, G. E., "The concept of probability in safety assessment of technological systems", *Science*, Vol. 250, No. 4986, 1990, pp. 1359-1364.

[5]Helton, J. C. and Oberkampf, W., "Alternative representations of epistemic uncertainties", *Reliability Engineering and System Safety*, Vol. 85, No. 1-3, 2004, pp. 1-10.

[6]Helton, J. C., Johnson, J. D., Sallaberry, C. J. and Storlie, C. B., "Survey of sampling-based methods for uncertainty and sensitivity analysis", *Reliability Engineering & System Safety*, Vol. 91, 2006, pp. 1175-1209.

[7]Borgonovo E., Castaings, W., and Tarantola, S., "Moment Independent Importance Measures: New Results and Analytical Test Cases", *Risk Analysis*, Vol. 31, No. 3, 2011, pp. 404-428.

[8]Plischke E., Borgonovo E., and Smith, C. L., "Global Sensitivity Measures from Given Data," *European Journal of Operational Research*, Vol. 226, 2013, pp. 536–550.

[9]Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global sensitivity analysis, The Primer*, John Wiley and Sons Ltd., New York, 2008.

[10]Storlie C. B., Swiler L. P., Helton J.C., Sallaberry C. J., "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models", *Reliability Engineering and System Safety*, Vol. 94, 2009, pp. 1735-1763.

[11]Crespo, L. G., Kenny, S. P., Giesy, D. P., "The NASA Langley Multidisciplinary Uncertainty Quantification Challenge", *16th AIAA Non-Deterministic Approaches Conference*, pp. 1-10, 2013, doi: 10.2514/6.2014-1347.

[12]Ferson, S., Kreinovich, V., Ginzburg, L., Sentz, K., Myers, D.S., "Constructing probability boxes and Dempster-Shafer structures, Sandia National Laboratories", Technical Report SAND2002-4015, Albuquerque, New Mexico, 2003.

[13]Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W., and Ginzburg, L., "Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty", Setauket, New York 11733, Technical Report SAND2007-0939, 2007.

[14]Ferson, S., Tucker, W.T., "Sensitivity analysis using probability bounding", *Reliability Engineering and System Safety*, Vol. 91, 2006, pp. 1435–1442.

[15]Ferson, S., Van den Brink, P., Estes, T.L., Gallagher, K., O'Connor, R., Verdonck, F., "Bounding uncertainty analyses", *Application of uncertainty analysis to ecological risks of pesticides*, edited by Warren-Hicks, W.J., and Hart, A., Pensacola and Boca Raton, FL.: SETAC and CRC Press, 2010.

[16]Borgonovo, E. and Tarantola, S., "Moment independent and variance-based sensitivity analysis with correlations: An application to the stability of a chemical reactor", *International Journal of Chemical Kinetics*, Vol. 40, No. 11, 2008, pp. 687-698.

[17]Marrel, A., Iooss, B., Laurent, B., and Roustant, O., "Calculations of Sobol indices for the Gaussian process metamodel", *Reliability Engineering and System Safety*, Vol. 94, 2009, pp. 742-751.

[18]Sobol, I. M., "Sensitivity analysis for nonlinear mathematical model", *Math. Modelling Comput. Exp.*, Vol. 1, 1993, pp. 407-414.

[19]Volkova, E., Iooss, B., and Van Dorpe, F., "Global sensitivity analysis for a numerical model of radionuclide migration from the RRC "Kurchatov Institute" redwaste disposal site", *Stoch Environ Res Assess*, Vol. 22, 2008, pp. 17-31.

[20]Bishop, C. M., *Neural Networks for pattern recognition*, Oxford University Press, New York, NY, USA, 1995.

[21]Cybenko, G., "Approximation by superpositions of a sigmoidal function", *Mathematics of Control Signals Systems*, Vol. 2, 1989, pp. 303-314.

[22]Deng, J., "Structural reliability analysis for implicit performance function using radial basis functions", *International Journal of Solids and Structures*, Vol. 43, 2006, pp. 3255-3291.

[23]Hurtado, J. E., "Filtered importance sampling with support vector margin: a powerful method for structural reliability analysis", *Structural Safety*, Vol. 29, 2007, pp. 2-15.

[24]Cardoso, J. B., De Almeida, J. R., Dias, J. M., and Coelho, P. G., "Structural reliability analysis using Monte Carlo simulation and neural networks", *Advances in Engineering Software*, 2008ol 39, 2008, pp. 505-513.

[25]Cheng, J., Li, Q. S., Xiao, R. C., "A new artificial neural network-based response surface method for structural reliability analysis", *Probabilistic Engineering Mechanics*, Vol. 23, 2008, pp. 51-63.
[26]Crespo, L. G., Giesy, D. P., and Kenny, S. P., "Reliability-based analysis and design via failure domain bounding", *Structural Safety*, Vol. 31, 2009, pp. 306-315.
[27]Crespo, L. G., Kenny, S. P., Giesy, D. P., "Uncertainty analysis via failure domain characterization: unrestricted requirement functions", *Advances in Safety, Reliability and Risk Management*, Proceedings of the ESREL 2011 Conference, 18-22 September 2011, Troyes, France, CRC Press, 2012, pp. 2205–2212.
[28]Crespo, L. G., Munoz, C. A., Narkawicz, A., Kenny, S. P., Giesy, D. P., "Uncertainty analysis via failure domain characterization: polynomial requirement functions", *Advances in Safety, Reliability and Risk Management*, Proceedings of the ESREL 2011 Conference, 18-22 September 2011, Troyes, France, CRC Press, 2012, pp. 1153–1160.
[29]Crespo, L. G., Giesy, D., Kenny, S. P., "Bounding of the failure probability range of polynomial systems subject to p-box uncertainties", *Mechanical Systems and Signal Processing*, Vol. 37, No. 1-2, 2013, p. 121-136.
[30]Kalos, M. H., Whitlock, P.A., *Monte Carlo methods. Volume I: Basics*, Wiley, New York, NY, 1986.
[31]Zio, E., *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, Springer Series in Reliability Engineering, Springer, London, UK, 2013.
[32]Konak, A., Coit, D. W., and Smith A.E., "Multi-Objective Optimization Using Genetic Algorithms: A Tutorial", *Reliability Engineering and System Safety*, Vol. 91, 2006, pp. 992-1007.
[33]Yang, X. S., *Engineering optimization: an introduction with metaheuristic applications*, Wiley, New York, NY, USA, 2010.
[34]Apostolakis, G. E., and Kaplan, S., "Pitfalls in risk calculations", *Reliability Engineering*, Vol. 2, No. 2, 1981, pp. 135-145.
[35]Beer, M., "Engineering quantification of inconsistent information", *Int J Reliability and Safety*, Vol. 3, No. (1/2/3), 2009, pp. 174–197.
[36]Couso, I., and Moral, S., "Independence concepts in evidence theory", *International Journal of Approximate Reasoning*, Vol. 51, 2010, pp. 748–758.
[37]Fetz, M., Oberguggenberger, W., "Propagation of uncertainty through multivariate functions in the framework of sets of probability measures", *Reliability Engineering and System Safety*, Vol. 85, 2004, pp. 73-87.
[38]Pedroni, N., and Zio, E., "Empirical comparison of methods for the hierarchical propagation of hybrid uncertainty in risk assessment, in presence of dependences", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 20, No. 4, 2012, pp. 509-557.
[39]Pedroni, N., and Zio, E., "Uncertainty analysis in fault tree models with dependent basic events", *Risk Analysis, an International Journal*, Vol. 33, No. 6, 2013, pp. 1146–1173.
[40]Pedroni, N., Zio, E., Ferrario, E., Pasanisi, A., Couplet, M., "Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model", *Computers and Structures*, Vol. 126, 2013, pp. 199–213.
[41]Stein, M., and Beer, M., "Bayesian quantification of inconsistent information", *Applications of Statistics and Probability in Civil Engineering* – Faber, Köhler & Nishijima (eds), Taylor & Francis Group, London, 2011, pp. 463–470.
[42]Stein M., Beer, M., and Kreinovich, V., "Bayesian Approach for Inconsistent Information", *Information Sciences*, Vol. 245, 2013, pp. 96–111.
[43]Kolmogorov, A., "Confidence limits for an unknown distribution function", *Annals of Mathematical Statistics*, Vol. 12, 1941, pp. 461–463.
[44]Smirnov, N., "On the estimation of the discrepancy between empirical curves of distribution for two independent samples", Bulletin de l'Université de Moscou, Série internationale (Mathématiques), Vol. 2: (fasc. 2), 1939.
[45]Miller, L.H., "Table of percentage points of Kolmogorov statistics", *Journal of the American Statistical Association*, 51, 1956, pp. 111–121.
[46]Feller, W., "On the Kolmogorov-Smirnov limit theorems for empirical distributions", *Annals of Mathematical Statistics*, 19, 1948, pp. 177–189.
[47]Wan, W., and Birch, J.B., "Using a modified genetic algorithm to find feasible regions of a desirability function", *Quality and Reliability Engineering International*, Volume 27, Issue 8, 2011, pp. 1173–1182.
[48]Limbourg, P., and de Rocquigny, E., "Uncertainty analysis using evidence theory – confronting level-1 and level-2 approaches with data availability and computational constraints", *Reliability Engineering and System Safety*, Vol. 95, No. 5, 2010, pp. 550-564.
[49]Zio, E., Baraldi, P., Pedroni, N., "Selecting features for nuclear transients classification by means of genetic algorithms", *IEEE Transactions on Nuclear Science*, Vol. 53, No. 3, 2006, pp. 1479-1493.
[50]Zio, E., Baraldi, P., Pedroni, N., "Optimal power system generation scheduling by multi-objective genetic algorithms with preferences", *Reliability Engineering & System Safety*, Volume 94, Issue 2, 2009, pp. 432–444.
[51]Saltelli, A., "Making best use of model evaluations to compute sensitivity indices", *Comput. Phys. Commun.*, Vol. 145, 2002, pp. 280-297.
[52]Hall, J. W., "Uncertainty-based sensitivity indices for imprecise probability distributions", *Reliability Engineering and System Safety*, Vol. 91, 2006, pp. 1443–1451.
[53]Aughenbaugh, J. M., and Paredis, C.J.J., "Probability Bounds Analysis as a general approach to sensitivity analysis in decision making under uncertainty", *SAE 2007 Transactions Journal of Passenger Cars: Mechanical Systems (Section 6)*, Vol. 116, 2007, pp. 1325–1339.

[54]Oberguggenberger, M., King, J., and Schmelzer, B., "Classical and imprecise probability methods for sensitivity analysis in engineering: A case study", *International Journal of Approximate Reasoning*, Vol. 50, 2009, pp. 680–693.

[55]Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning internal representations by error back-propagation", *Parallel distributed processing: exploration in the microstructure of cognition* (vol. 1), edited by Rumelhart, D. E. and McClelland, J. L., MIT Press, Cambridge (MA), USA, 1986.

[56]Pedroni, N., Zio, E., and Apostolakis, G. E., "Comparison of bootstrapped Artificial Neural Networks and quadratic Response Surfaces for the estimation of the functional failure probability of a thermal-hydraulic passive system", *Reliability Engineering and System Safety*, Vol. 95, No. 4, 2010, pp. 386-395.

[57]Zio, E., Pedroni, N., "How to effectively compute the reliability of a thermal-hydraulic passive system", *Nuclear Engineering and Design*, Vol. 241, No. 1, 2011, pp. 310-327.

[58]Zio, E., Apostolakis, G. E., and Pedroni, N., "Quantitative functional failure analysis of a thermal–hydraulic passive system by means of bootstrapped Artificial Neural Networks", *Annals of Nuclear Energy*, Vol. 37, No. 5, 2010, pp. 639-649.

[59]Zio, E., and Pedroni, N., "An optimized Line Sampling method for the estimation of the failure probability of nuclear passive systems", *Reliability Engineering and System Safety*, Vol. 95, No. 12, 2010, pp. 1300-1313.

[60]Zio, E., and Pedroni, N., "Monte Carlo Simulation-based Sensitivity Analysis of the model of a Thermal-Hydraulic Passive System", *Reliability Engineering and System Safety*, Vol. 107, 2012, pp. 90-106.

[61]Zhang, J. and Foschi, R. O., "Performance-based design and seismic reliability analysis using designed experiments and neural networks", *Probabilistic Engineering Mechanics*, Vol. 19, 2004, pp. 259-267.

[62]Seber, G.A.F., *Multivariate Observations*, Hoboken, NJ: John Wiley & Sons, Inc.,1984.