

LA-UR-13-29365

Approved for public release;
distribution is unlimited.

Title: Robust Decision Making Applied to the NASA Multidisciplinary Uncertainty Quantification Challenge Problem
(Manuscript)

Author(s): Kendra Van Buren, Los Alamos National Laboratory (NSEC)
François Hemez, Los Alamos National Laboratory (XTD-IDA)

Intended for: 16th AIAA Non-Deterministic Approaches Conference, American Institute of Aeronautics and Astronautics (AIAA) Science and Technology Forum and Exposition, National Harbor, Maryland, January 13-17, 2014



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the view point of a publication or guarantee its technical correctness.

This page is left blank intentionally.

Robust Decision Making Applied to the NASA Multidisciplinary Uncertainty Quantification Challenge Problem

Kendra L. Van Buren¹ and François M. Hemez²
Los Alamos National Laboratory, Los Alamos, NM, 87545

This paper addresses the NASA Langley Multidisciplinary Uncertainty Quantification Challenge (MUQC) Problem, which is intended to pose challenges to the uncertainty quantification and robust design communities. The goals of the MUQC problem can be formulated into four main topics that are commonly encountered in the model development process: calibration, sensitivity analysis, uncertainty propagation, and robust design. Our analysis places a particular emphasis on the use of info-gap decision theory (IGDT) to address the goals of the MUQC problem. IGDT provides a convenient framework to treat epistemic uncertainty when using simulation models for decision-making. We utilize a robustness criterion, defined in the context of IGDT, to pursue calibration, uncertainty propagation, and robust design. Herein, our calibration utilizes IGDT to address the situation whereby traditional calibration techniques might result in non-unique results where different sets of calibration variables are able to replicate experiments with comparable fidelity. Uncertainty propagation is performed such that the worst-case and best-case performances of the model output are conditioned on the level of uncertainty that is permitted in the simulations. To pursue robust design, we utilize the robustness criterion to establish whether the amount of uncertainty tolerable in our optimized design is an improvement over the baseline design. We demonstrate that improving the robustness of the model requires different knowledge than improving performance of the model. The main conclusion is that IGDT provides a sound theoretical basis, and practical implementation, to meet the goals of the NASA MUQC problem without formulating simplifying assumptions. (*Manuscript approved for unlimited, public release on December 11, 2013, LA-UR-13-29365.*)

Nomenclature

$\hat{\alpha}$	= Robustness
α	= Horizon-of-uncertainty
$\hat{\beta}$	= Opportuneness
d	= Design parameter
g	= Model output
p	= Calibration variable
R	= Performance criterion
U	= Info-gap uncertainty model
x	= Intermediate variable (or sub-model)

I. Introduction

In recent years, application of modeling and simulation (M&S) techniques has evolved from using simulations to understand experimentally observed behavior to using simulations to predict with *quantifiable confidence* across an application domain. As a result of this paradigm shift, calibration and validation exercises are performed increasingly to explore the underlying physics of a system. Physics-based predictions provided by simulation models can then be implemented to support *decision-making*, such as the feasibility of new design concepts or maintenance needs of existing structures. Herein, we argue that it is important to quantify the robustness of numerical models to sources of uncertainty and lack-of-knowledge for informed decision-making. The basic motivations for pursuing

¹ Post-doctoral Associate, The Engineering Institute, Mail Stop T001, Los Alamos, New Mexico 87545. E-mail: klvan@lanl.gov.

² Technical staff member, "X" Theoretical Design Division, Mail Stop T087, Los Alamos, New Mexico 87545. E-mail: hemez@lanl.gov. AIAA Senior Member.

robustness in the model development process is that the performance of simulation models should remain as consistent as possible as sources of uncertainty are allowed to vary away from their nominal settings.

This manuscript demonstrates the concept of robustness for performance uncertainty management of the NASA Langley Multidisciplinary Uncertainty Quantification Challenge (MUQC) Problem¹. The NASA MUQC problem is a black-box code for aeronautic flight guidance, defined with a total of 21 calibration variables and 14 design parameters. Herein, design parameters are dimensions of the parameter space controlled by the analyst, while calibration variables are introduced by environmental conditions and modeling choices. The overarching goals of the NASA MUQC problem can be formulated into four main questions about the performance of the model:

Question-A: Can experimental evidence be used to reduce uncertainty in calibration variables influential to model output? (Uncertainty characterization, i.e. calibration.)

Question-B: How to determine parameters that are most influential to model output? (Sensitivity analysis.)

Question-C: What are the worst-case and best-case performances of the simulation model? (Uncertainty propagation and extreme-case analysis.)

Question-D: Can the design be improved, given the potential coupling of design parameters and calibration variables on model output? (Robust design.)

Experimental evidence for a sub-problem of the NASA MUQC problem is provided by the challenge problem moderators to pursue Question-A. The intent of the calibration exercise addressed in Question-A is to constrain the uncertainty of the calibration variables. Two forms of calibration are pursued: (i) through deterministic optimization and (ii) through use of a robust-optimal criterion. Our intent is to emphasize that calibration can often result in ambiguous, or non-unique, results. Rather than best-fitting model parameters to reproduce experiments, as is done in the deterministic optimization paradigm, we advocate the use of criteria based on robustness to investigate whether the uncertainty models of calibration variables can be constrained to reduce the overall prediction uncertainty.

Question-B exemplifies a conventional sensitivity analysis question. By answering this question, computational demands of consequential exercises can be alleviated by only investigating the most influential aspects of the model. The sensitivity analysis pursued in Question-B can be applied to determine the influential parameters that should be explored for uncertainty propagation in Question-C. Question-C promotes a better understanding of how robust our simulation predictions are to the uncertain calibration variables. In doing so, it is possible to quantify how reliable our predictions remain as calibration variables are allowed to deviate away from their nominal settings. In our implementation, the worst-case and best-case performances pursued by an extreme-case analysis are conditioned on the total amount of uncertainty that is allowed in the model. Rather than providing one worst-case and one best-case performance, our extreme-case analysis is repeated for different levels of uncertainty. For this reason, uncertainty propagation and extreme-case analysis are handled simultaneously. Examining how the worst-case and best-case performances change, as the amount of uncertainty increases, also defines a non-parametric measure of sensitivity.

Thus far, Questions A-C investigate the calibration variables of the NASA MUQC problem while the design parameters are held constant. The potential coupling between calibration variables and design parameters does not affect the analysis for those questions. Thus, results from Questions B and C, which are useful to reduce the computational demands of the analysis, might no longer apply as the design parameters are varied. Question-D assesses whether the design performance can be improved, which is complicated due to this potential coupling. Our analysis investigates how to improve the design while simultaneously being robust to the uncertainty originating in calibration variables. In doing so, uncertainty originating from the calibration variables is treated separately from the uncertainty originating from design parameters.

Herein, our discussion emphasizes the use of info-gap decision theory² (IGDT) to address the aforementioned goals of the MUQC problem. IGDT is useful to quantify robustness to uncertainties in either a probabilistic or non-probabilistic framework, where the uncertainty is due to ignorance, lack-of-knowledge, or variability. In particular, IGDT provides the robustness criterion for our calibration (Question-A), quantifies the allowable uncertainty for our uncertainty propagation and extreme-case analysis (Question-C), and implements robust design (Question-D). Our robustness criteria are formulated without imposing any simplifying assumption of the problem structure.

The remainder of the document is organized as follows. First, a brief description of NASA MUQC is provided in Section II. Section III overviews the relevant details of IGDT such that this paper is self-contained; further details of the theory can be found in Reference 2. The goals of the challenge problem are addressed in Sections IV-VII, and the main conclusions from this paper are presented in Section VIII.

II. NASA MUQC Problem Overview

The NASA MUQC problem is used to describe the dynamics of the Generic Transport Model, remotely operated aircraft developed by the NASA Langley Research Center^{3,4}. As described in Equation (1) and illustrated in Figure

1, the formulation of the model relies on 14 design parameters, d , 21 calibration variables, p , and 5 intermediate variables, x to produce 8 model outputs, g . The model inputs describe losses in control effectiveness and time delays, while the model outputs describe vehicle stability and performance requirements for pilot command tracking and handling or riding qualities:

$$g = f(x, d); \quad x = h(p). \quad (1)$$

A visual representation of the problem structure is provided in Figure 1, where the green dashed box is used to highlight the model inputs, which are the design parameters, d , and calibration variables, p . The calibration variables, p , produce the intermediate variables, x , as denoted by the solid blue box. The intermediate variables, x , and design parameters, d , produce the outputs of interest, g . The space in which Equation (1) is defined, is a space of dimension 21 when the design parameters, d , are fixed, and a space of dimension 35 when the design parameters are not fixed. It is noted that exploring such large-dimensional spaces, without formulating simplifying assumptions about the potential interactions between design parameters, calibration variables, and output responses, requires computationally efficient procedures and algorithms.

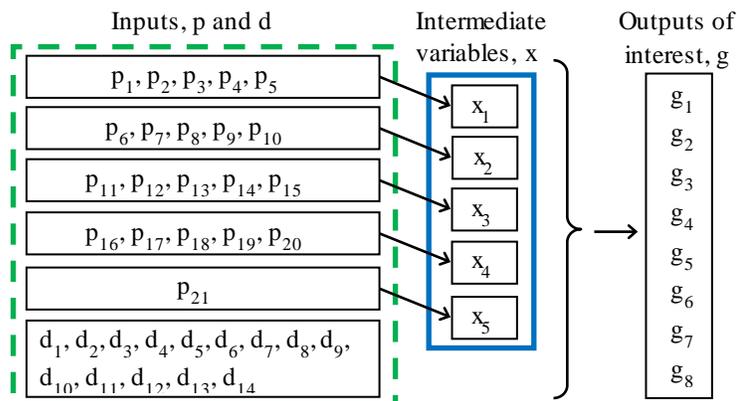


Figure 1. Illustration of the NASA MUQC problem formulation.

Herein, the model is considered *requirement compliant* when the outputs satisfy the inequality, $g < 0$. The set of calibration variables that produces requirement-compliant outputs while the design parameters are held constant is called the *safe domain*, whereas the set of calibration variables that is unable to produce requirement-compliant outputs, that is, $g_k \geq 0$ for at least one output quantity ($1 \leq k \leq 8$), is called the *failure domain*.

III. Info-Gap Decision Theory

This section provides an overview of info-gap decision theory (IGDT) such that the reader has an understanding for how the robustness and opportuneness criteria are formulated for decision-making in our analysis. Herein, robustness describes the worst-case performance predicted by the model whereas opportuneness describes the best-case performance. Further details of the theory and its implementation can be found in References 2 and 5-6. Engineering analysis and design typically uses the worst-case performance to define design constraints, whereas the best-case performance is often disregarded. For this reason, our discussion places a particular emphasis on the use of IGDT to quantify robustness.

Implementation of IGDT necessitates the combination of three attributes: (i) the model, (ii) the performance criterion, and (iii) the info-gap model used to represent what is unknown about the decision. The model represents the relationship between the model inputs and model outputs, as described previously in Equation (1). Performance, denoted herein as R , is a scalar quantity used to determine the ability of the model to satisfy a critical performance value, R_C . The performance can quantify the test-analysis correlation of model output to experimental evidence, or determine whether the model is requirement-compliant, as suggested in Equation (2):

$$R(p) = \max_{1 \leq i \leq 8} g_i \leq R_C. \quad (2)$$

As suggested in Section II, the NASA MUQC problem is requirement-compliant when $g < 0$. However, different critical performance levels, R_C , can be considered for the purpose of decision-making in the context of IGDT, as reflected in Equation (2). Here, a value of R_C that becomes more negative indicates a better-performing design.

The uncertainty that we wish to be robust to originates from the uncertain calibration variables. The info-gap model, $U(\alpha; p_0)$, describes how the uncertain calibration variables, p , vary with respect to their nominal values or settings. The nominal values, p_0 , denote the best-available knowledge that is used as the starting point of the

analysis. The “distance” between the current knowledge, p_0 , and hypothetical settings, p , which could be selected for decision-making, is parameterized using a single scalar α denoted as the horizon-of-uncertainty:

$$U(\alpha; p_0) = \{p : \|p - p_0\| \leq \alpha\}, \alpha \geq 0. \quad (3)$$

When $\alpha = 0$, the uncertain variables remain fixed at their nominal settings, $p = p_0$. As the horizon-of-uncertainty increases, more settings of the calibration variables, p , are taken into consideration. Each setting considered leads to a different model whose prediction of performance is evaluated against the performance criterion (2). Equation (3) shows that, at a fixed horizon-of-uncertainty, α , the set $U(\alpha; p_0)$ defines a family of models that can be explored to search for the worst-possible performance, given this maximum amount of calibration variable uncertainty.

With these three attributes defined, an info-gap analysis searches for the calibration variables that, while they potentially deviate from the best-available knowledge p_0 by the greatest amount α , define a model whose predictions satisfy the performance criterion (2). When searching for the worst-case performance, robustness is achieved if the model can tolerate deviations from its nominal settings, as quantified in (3), while delivering the expected level of accuracy, R_C , as quantified in Equation (2). Conversely, opportuneness is achieved if the best-case performance of the model, as it deviates from its nominal settings (3) improves the critical performance level, R_C , of Equation (2).

Conceptually, the robustness, $\hat{\alpha}$, of an info-gap analysis quantifies the degradation of performance as the horizon-of-uncertainty α increases, while the opportuneness, $\hat{\beta}$, quantifies the potential improvement in performance. Mathematically, these functions depend on the target critical performance, R_C , and are defined as:

$$\hat{\alpha} = \arg \max_{\alpha \geq 0} \left\{ \alpha : \max_{p \in U(p_0; \alpha)} R(p) \leq R_C \right\} \quad \text{and} \quad \hat{\beta} = \arg \min_{\alpha \geq 0} \left\{ \alpha : \min_{p \in U(p_0; \alpha)} R(p) \leq R_C \right\}. \quad (4)$$

Solving the robustness and opportuneness conditions defined in Equation (4) involves two nested optimization problems. To bypass this difficulty, we focus on the inner optimization. First, a horizon-of-uncertainty α is selected, which defines lower and upper bounds for the model parameters, as shown in Equation (3). The worst-case and best-case performances are searched for, given that the uncertainty calibration variables are allowed to vary within the lower and upper bounds. These steps are repeated for increasing levels of the horizon-of-uncertainty α , which provides the information needed to construct the opportuneness and robustness functions. Strictly speaking, a critical level of performance, R_C , needs not be defined when the robustness and opportuneness functions are explored as described previously. Solving Equations (4), on the other hand, requires setting R_C .

A representative info-gap plot is provided in Figure 2, where smaller performance values are more desirable than larger performance values. The nominal performance, R_{NOM} , is obtained when the horizon-of-uncertainty, α , is equal to zero. As the horizon-of-uncertainty increases, the best-case and worst-case performances start to deviate from the nominal performance, as demonstrated by the notional shapes of the opportuneness and robustness functions. The info-gap plot of Figure 2 quantifies, first, by how much model predictions vary and, second, the range of predictions that are obtained at any given horizon-of-uncertainty. In a broad sense, these attributes of the info-gap analysis can be viewed as sensitivity analysis (“do the predictions vary?”) and uncertainty quantification (“by how much?”).

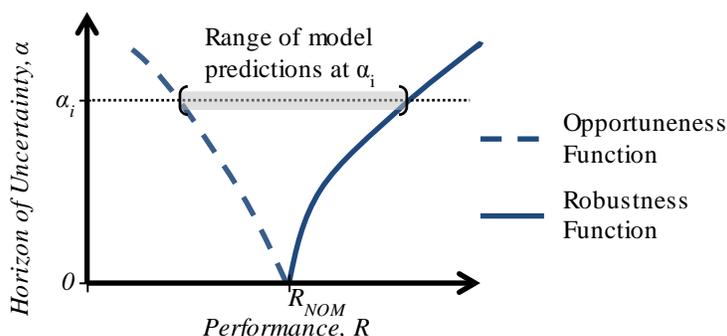


Figure 2. Nominal info-gap plot with opportuneness and robustness functions.

We proceed by using the opportuneness and robustness functions defined in the context of IGDТ to address the goals of the NASA MUQC problem. As discussed previously, the goals are addressed autonomously through four main themes: (i) calibration, (ii) sensitivity analysis, (iii) uncertainty propagation and extreme-case analysis, and (iv) robust design.

IV. Application to the NASA MUQC Problem: Calibration

Herein, the goal of calibration is to investigate whether the uncertainty description of the calibration variables, used to predict the intermediate variable x_1 , can be constrained using experimental evidence. As shown previously in Figure 1, five calibration variables, $p_1 - p_5$, are used to define the intermediate variable x_1 . The definitions of these five variables are given in Table 1. Calibration variable p_1 is a unimodal beta distribution, p_2 is defined over an uncertain range from 0 to 1, p_3 is a uniform distribution between 0 and 1, and p_4 and p_5 are possibly correlated multivariate normal distributions. As shown in Table 1, the uncertainty bounds are reducible for p_1, p_2, p_4 , and p_5 .

Table 1. Definitions of the calibration variables for intermediate variable x_1 .

Parameter	Reducible?	Uncertainty Model
p_1	Yes	Unimodal Beta, $0.6 \leq E[p_1] \leq 0.8$, $0.02 \leq V[p_1] \leq 0.04$
p_2	Yes	$\Delta = [0,1]$
p_3	No	Uniform, $\Delta = [0,1]$
p_4, p_5	Yes	Normal, $-5 \leq E[p_i] \leq 5$, $0.0025 \leq V[p_i] \leq 4$, $ \rho \leq 1$ for $i = 4,5$

The challenge problem moderators provide two sets of experiments, with 25 observations each. The first set of experiments is intended for calibration, and the second set is treated as ‘hold-out’ experiments used to validate the calibration results obtained with the first set of experiments. Figure 3 compares the histograms of the experiments, with the first set shown in the left plot, the second set in the middle plot, and the two sets of experiments combined in the right plot. Our calibration is performed using all of the experimental evidence at once (right-most histogram), rather than keeping them split into two sets as shown in the two left images of Figure 3. Our rationale for this choice is that, given the scarcity of data available for calibration, all of them ought to be used simultaneously.

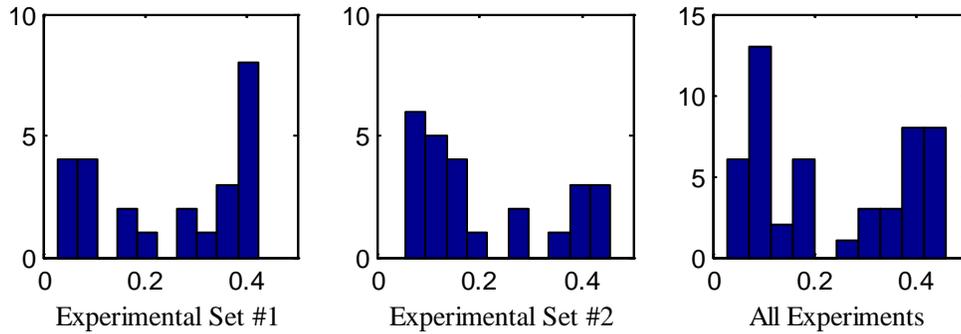


Figure 3. Histograms of the experimental evidence provided by the problem moderators.

Clearly, not enough experimental evidence is provided to determine whether the values of intermediate variable x_1 can be described using a normal, or any other type of, probability distribution. Rather than defining the objective function of our calibration using hyper-parameters of probability distributions, we utilize a coefficient of correlation applied to vectors of the experimental and simulation histograms. Here, bins of the histograms are defined between the lower and upper bounds of experimental evidence, as demonstrated nominally with three bins to describe the experiments in Figure 4. Note that, two more bins may be necessary to describe the predictions if they are lower or higher than those observed experimentally. The histograms of simulated values are obtained using the frequencies of occurrence, expressed on the right side of Figure 4, as percentages of the simulation output.

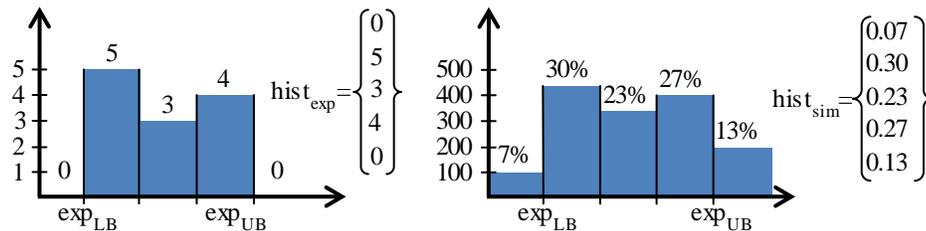


Figure 4. Development of vectors used to describe histograms.

The coefficient of correlation used to define the objective function is adapted from the well-known modal assurance criterion, valued between $[0; +1]$ and subtracted from one, as shown in Equation (5). In our objective function, a value of zero indicates that the two vectors are parallel, whereas a value of one indicates that the vectors are orthogonal. Calibration is performed to search for parameter values that minimize the objective function:

$$\text{Objective} = 1 - \frac{\left(\text{Hist}_{\text{Exp}}^T \cdot \text{Hist}_{\text{Sim}} \right)^2}{\left(\text{Hist}_{\text{Exp}}^T \cdot \text{Hist}_{\text{Exp}} \right) \left(\text{Hist}_{\text{Sim}}^T \cdot \text{Hist}_{\text{Sim}} \right)} \quad (5)$$

where the subscripts $(\cdot)_{\text{Exp}}$ and $(\cdot)_{\text{Sim}}$ identify the histograms of experimental and simulated values, respectively.

Calibration is pursued using two approaches: (i) through deterministic optimization, as described in the left image of Figure 5 and (ii) in the context of IGDT, as depicted in the right image of Figure 5. Calibration is applied to the hyper-parameters of the beta distribution used to describe p_1 , the interval used to define p_2 , and the hyper-parameters of the multi-variate distribution used to describe parameters p_4 and p_5 . The expected value and variance for p_1 , as reported in Table 1, are converted to alpha and beta parameters of a beta distribution, described as a and b .

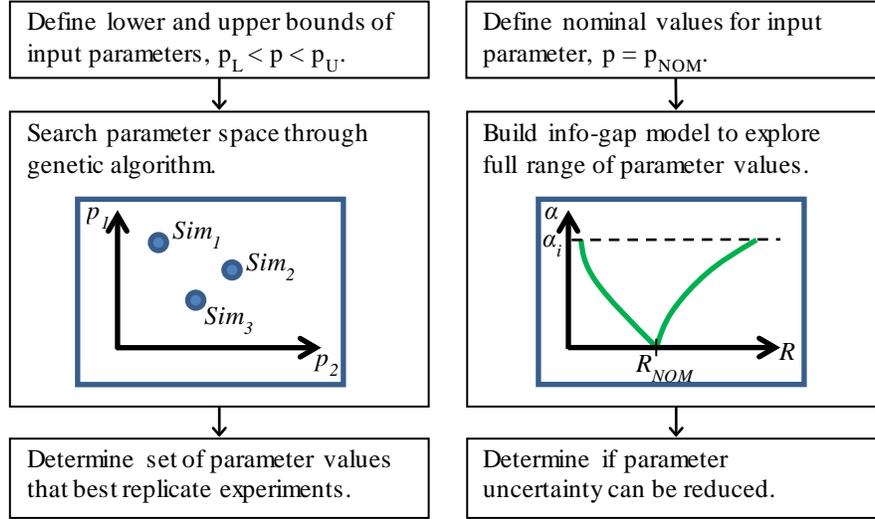


Figure 5. Process for deterministic optimization (left) and calibration through IGDT (right).

Deterministic optimization is performed to demonstrate the ability of the model to replicate experiments. The parameter space is explored, as shown graphically in a two-dimensional space in the left image of Figure 5. Within this optimization, the set of parameters that minimize the objective function (5) is searched for within the lower and upper parameter bounds defined in Table 1.

The deterministic optimization is repeated twice to show the effect of changing the number of bins that define the histograms: once where only one bin is used, and once with ten bins. Conceptually, using a single bin searches for the set of parameters that produces simulation predictions within the bounds of experimental evidence, whereas using ten bins also attempts to capture the frequencies of experimental values. To test for repeatability, the optimization is repeated to observe whether the genetic algorithm converges to the same calibrated parameters. Three sets of parameter values that the optimization converges to, and their corresponding objective functions, are reported in Table 2.

Table 2. Combinations of parameter values and their corresponding objective function values.

Number of Bins	p_1		p_2	p_4 and p_5					Objective
	a_1	b_1		μ_4	σ_4	μ_5	σ_5	$\rho_{4,5}$	
1 bin	0.776	0.020	0.842	2.028	1.673	1.699	0.102	0.311	1.6e-07
	0.694	0.022	0.914	-0.299	0.208	2.828	0.919	0.051	7.3e-07
	0.786	0.022	0.078	2.650	0.208	4.789	0.722	0.311	1.5e-07
10 bins	0.690	0.038	0.1359	-3.631	0.069	-4.758	0.050	0.673	0.11
	0.702	0.032	0.9853	-4.564	0.552	-4.679	0.182	0.961	0.10
	0.690	0.038	0.8859	-2.360	1.655	-4.758	0.050	0.660	0.13

All three sets of parameter values obtained using one bin, and reported in Table 2, provide less than 0.1% simulation output located outside the experimental bounds when sampled 60,000 times using Monte Carlo sampling. The convergence of the optimization approach based on a single bin demonstrates that it is possible to search for parameter values that can provide predictions located within the experimental bounds; however, as made evident by the range of values given in Table 2, the results of this optimization approach are non-unique.

The optimization has difficulty converging when ten bins are used to describe the experimental histograms. As shown in Table 2, different sets of parameter values are able to provide simulations that achieve similar correlation to the experiments. A comparison of histograms of the experiments and simulation predictions is shown in Figure 6, for the set of parameters that produce an objective function equal to 0.11 in Table 2. Clearly, it is more difficult to converge to the experimental evidence when using several bins to describe the frequencies of occurrence.

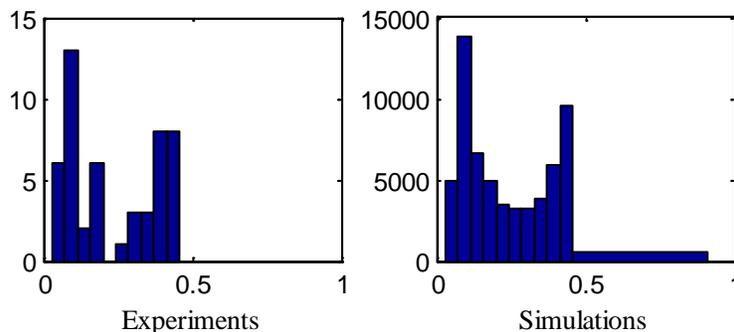


Figure 6. Comparison of histograms for experiments and simulations using ten bins.

Rather than relying on deterministic optimization to answer the question, *can our model parameters be adjusted to replicate experimental evidence*, we also pursue calibration using a robustness criterion to answer the similar question, *can an info-gap representation of uncertainty be used to reduce the uncertainty bounds of calibration variables, given experimental evidence*? As illustrated in Figure 7, an info-gap analysis is performed where we define nominal values for the model parameters, and seek a range of parameter values whose predictions bound the experimental evidence. The nominal values are chosen as the midpoint of calibration variables, which corresponds to $E[p_1] = 0.70$, $V[p_1] = 0.03$, $p_2 = 0.50$, $E[p_4] = 0$, $V[p_4] = 2.00$, $E[p_5] = 0$, $V[p_5] = 2.00$, and $\rho_{4,5} = 0.50$. The definitions of the horizon-of-uncertainty, α , used for info-gap analysis are provided in Equation (6):

$$\alpha = 0 \left\{ \begin{array}{l} p_1 \sim \beta(a_0; b_0) \\ p_2 = 0.5 \\ p_3 \sim U[0;1] \\ p_4, p_5 \sim N(\mu_0; \sigma_0; \rho_0) \end{array} \right. \quad \alpha > 0 \left\{ \begin{array}{l} \left. \begin{array}{l} a_0(1-\alpha) \leq a \leq a_0(1+\alpha) \\ b_0(1-\alpha) \leq b \leq b_0(1+\alpha) \end{array} \right\} p_1 \sim \beta(a; b) \\ 0.5(1-\alpha) \leq p_2 \leq 0.5(1+\alpha) \\ p_3 \sim U[0;1] \\ \left. \begin{array}{l} \mu_0(1-\alpha) \leq \mu \leq \mu_0(1+\alpha) \\ \sigma_0(1-\alpha) \leq \sigma \leq \sigma_0(1+\alpha) \\ \rho_0(1-\alpha) \leq \rho \leq \rho_0(1+\alpha) \end{array} \right\} p_4, p_5 \sim N(\mu; \sigma; \rho) \end{array} \right. \quad (6)$$

The definitions of Equation (6) indicate that the horizon-of-uncertainty, α , can be viewed as a maximum percentage of change relative to the nominal value of a model parameter, or hyper-parameter of its probability law.

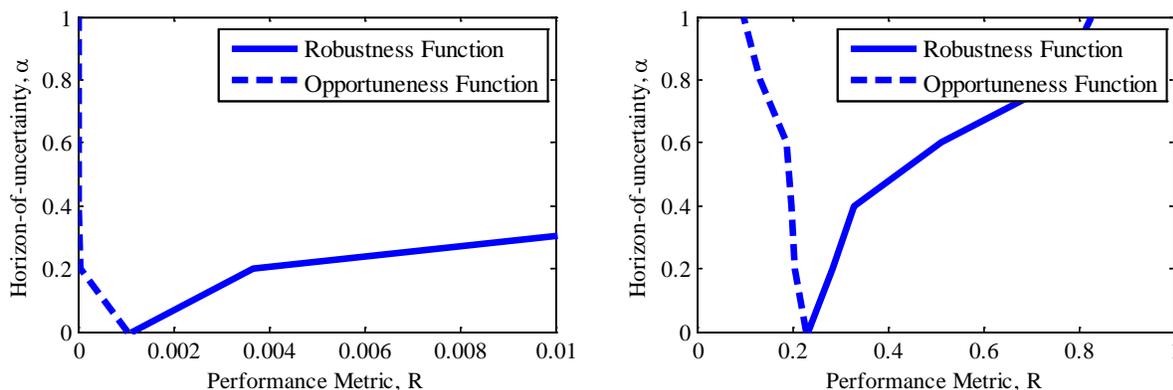


Figure 7. Info-gap calibration performed with one bin (left) and ten bins (right).

Calibration utilizing the robustness criterion is performed twice using the performance defined by the objective function of Equation (5): once with only one bin and again with ten bins. The results of the corresponding info-gap analyses are provided in Figure 7. Note the different limits on the x-axis. For the case where the histogram has only one bin, as shown in the left image of Figure 7, the objective function is minimized to $4.8e-05$ when 20% variation in the calibration variables is allowed ($\alpha = 0.2$). An objective function of $4.8e-05$ provides less than 0.1% simulation

output outside of the range of experimental evidence. This result suggests that uncertainty of the calibration variables can be constrained to $\pm 20\%$ variation if we wish to bound the experimental evidence. In comparison, the objective function is minimized to 0.09 at $\alpha = 1$ when utilizing ten bins. The parameter values corresponding to an objective function of 0.09 provides comparable test-analysis correlation results to the parameter values that produce an objective function of 0.11, illustrated in Figure 6. As discussed previously, it is difficult to converge to the experimental evidence using a histogram defined with ten bins. The fact that the best-case performance of the model is unable to replicate experiments when the objective function utilizes ten bins suggests that the parameter uncertainty cannot be reduced.

V. Application to the NASA MUQC Problem: Sensitivity Analysis

A sensitivity analysis is performed to identify the uncertain calibration variables, p , that most significantly influence the model outputs. Conversely, the least influential calibration variables can be eliminated from further consideration. Doing so is advantageous to reduce the computational demands of subsequent exercises, for example, the uncertainty propagation that is pursued in the next section.

One observation of the NASA MUQC problem formulation is that the calibration variables, p , and intermediate variables, x are decoupled. In other words, intermediate variable x_1 is not dependent on the same inputs as x_2 , x_3 , x_4 , or x_5 (refer to Figure 1). Our approach for sensitivity analysis and uncertainty propagation takes advantage of this decoupled formulation. Rather than explore the 21-dimensional space defined by the calibration variables, we propose to follow a two-step procedure whereby we first study the effect of calibration variables on the intermediate variables, and then study the effect of intermediate variables on the model output, g . Doing so reduces the 21-dimensional space of calibration variables to a 5-dimensional space defined by the intermediate variables.

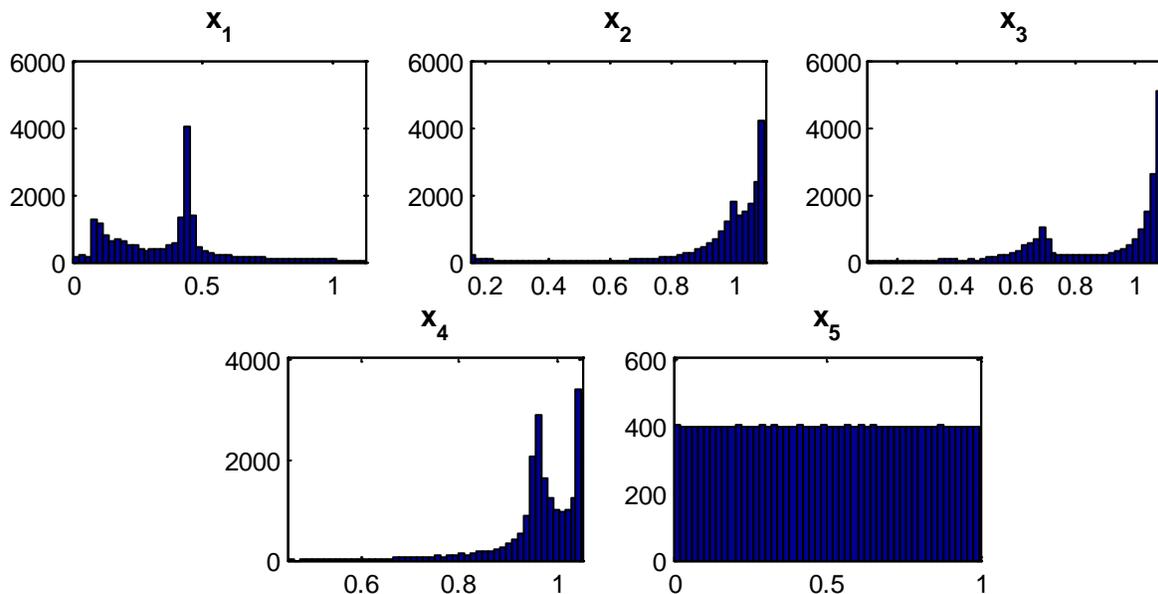


Figure 8. Histograms of intermediate variable outputs, x .

A Latin Hypercube exploration is performed to verify that the intermediate variables, x are continuous functions of the input parameters, p , as confirmed by the histograms of sampled results illustrated in Figure 8. Each parameter varies in the interval $[0; +1]$, except for parameters p_4 and p_5 that are bounded within $[-11; +11]$. These bounds are chosen based on information learned from the problem formulation. Establishing the continuity of intermediate variables is important to define their values between minimum and maximum bounds. Figure 8 also illustrates that, with the exception of intermediate variable x_5 , some values of x are more likely than others; these frequencies of occurrence are ignored in our analysis to focus only on lower and upper bounds. This approach, however, is justified because sensitivity analysis is used to support the extreme-case analysis, which seeks the best-case and worst-case performances, regardless of the likelihood of intermediate variables, x .

A five-level, full-factorial design-of-experiments (DOE), in which the calibration variables are bounded by their lower and upper bounds, is performed to analyze the effect of each variable, p_k , on the intermediate variables, x . This DOE provides the data necessary to perform an analysis-of-variance (ANOVA) that determines the influential input parameters for each intermediate variable. The input parameters that contribute over 95% of the main effect for each intermediate variable are identified. Table 3 summarizes the results of this main-effect ANOVA. The fifth

intermediate variable, x_5 , is not considered because it depends on a single input parameter. The table indicates that variations of intermediate variables x_3 and x_4 are controlled, for the most part, by only three input parameters each. Even fewer input parameters influence the values of intermediate variables x_1 and x_2 . Table 3 indicates that, due to the low influence that they exercise, 11 of the 21 input parameters can be eliminated from further consideration.

Table 3. Composite R^2 values for main-effect statistical screening of input parameters, p.

Intermediate Variable	Parameter	R^2 Value	Keep?	Intermediate Variable	Parameter	R^2 Value	Keep?
x_1	p ₁	99.15%	Yes	x_3	p ₁₁	3.30%	Yes
	p ₂	0.01%	No		p ₁₂	90.78%	Yes
	p ₃	0.70%	No		p ₁₃	0.07%	No
	p ₄	0.11%	No		p ₁₄	4.23%	Yes
	p ₅	0.04%	No		p ₁₅	1.62%	No
x_2	p ₆	78.25%	Yes	x_4	p ₁₆	42.12%	Yes
	p ₇	19.78%	Yes		p ₁₇	14.94%	Yes
	p ₈	1.77%	No		p ₁₈	42.18%	Yes
	p ₉	0.18%	No		p ₁₉	0.02%	No
	p ₁₀	0.01%	No		p ₂₀	0.75%	No

Next, the influence of intermediate variables, x , on the model output, g , is studied. The lower and upper bounds determined from the histograms of Figure 8 are used to define the bounds for a five-level, full-factorial analysis. The normalized R^2 statistics are provided in Table 4. These statistics derive, as before, from a main-effect ANOVA and are used to identify the intermediate variables influential to each output. For clarity, the calibration parameters that are influential to each intermediate variable are shown in parenthesis within the table. The results indicate that the first intermediate variable, x_1 , is influential for all of the model outputs, whereas x_2 is influential only to g_3 and g_5 , x_3 and x_4 are influential to g_4 , g_6 , g_7 , and g_8 , and the x_5 provides minimal effect on all of the outputs.

Table 4. Composite R^2 values for statistical effect screening of intermediate variables, x.

Intermediate Variable	Model Output							
	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
x_1 (p ₁)	95.58%	95.70%	82.00%	69.69%	58.81%	50.74%	68.56%	49.44%
x_2 (p ₆ ; p ₇)	4.52%	0.0%	14.74%	0.0%	37.92%	0.0%	0.0%	0.0%
x_3 (p ₁₁ ; p ₁₂ ; p ₁₄)	0.0%	0.36%	0.0%	14.28%	0.0%	22.74%	16.04%	25.35%
x_4 (p ₁₆ ; p ₁₇ ; p ₁₈)	0.0%	3.94%	0.0%	16.03%	0.0%	26.44%	15.40%	24.95%
x_5 (p ₂₁)	0.0%	0.0%	3.26%	0.0%	3.28%	0.08%	0.01%	0.26%

As mentioned in the introduction, the performance enters the failure domain when the output exceeds zero, $g > 0$. The full-factorial DOE used for Table 4 is analyzed to gain a better understanding of which output, g , controls the maximum value, as shown in Table 5. It is further determined whether the output is in the failure domain ($g > 0$), or the safe domain ($g < 0$). It is shown in Table 5 that output g_7 , followed by g_4 , are most influential to performance values that are requirement-compliant, whereas the other outputs (g_1 , g_2 , g_3 , g_5 , g_6 , g_8) are all in the failure domain.

Table 5. Frequencies of occurrence of the maximum output.

Criterion	Model Output								Cumulated Statistic
	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	
Max. Output	0.4%	29.6%	3.3%	28.0%	0.1%	15.5%	21.1%	2.1%	100%
Max. Output > 0	0.4%	29.6%	3.3%	26.2%	0.1%	15.5%	2.4%	2.1%	79.5%
Max. Output < 0	0%	0%	0%	1.8%	0%	0%	18.7%	0%	20.5%

It is emphasized that the results given in Tables 4-5 depend on the set of baseline design parameters, d , used to define the simulations, whereas the results in Table 3 are independent of the design parameters. It is possible that when the design parameters are varied, output g_7 will no longer be the most influential to the requirement-compliant output, due to the coupling between calibration variables and design parameters. In such a case, the intermediate variables influential to the model output, as reported in Table 5, would need to be reconsidered.

VI. Application to the NASA MUQC Problem: Uncertainty Propagation and Extreme-Case Analysis

Here, the goal is to better understand how much our predictions “break down,” by entering the failure domain, as the calibration variables vary away from their nominal settings. Rather than provide an overall worst-case or best-

case performance, our definition of performance is conditioned on a particular horizon-of-uncertainty that can be tolerated, as discussed in Section III. The horizon-of-uncertainty defines by how much the calibration variables are allowed to deviate from their nominal values while guaranteeing that the predicted performance remains compliant.

The discussion of robust design in the following section is provided to suggest how a robustness criterion can be used to improve the design. To stay consistent with our discussion of robust design, the nominal calibration variables, denoted by p_0 in Equation (3), are chosen from a subset of values that provide requirement-compliant outputs. These nominal variables yield the smallest value of performance, that is, the best performance, amongst those obtained from a 1,000 Monte Carlo runs that explore different combinations of the 21 calibration variables.

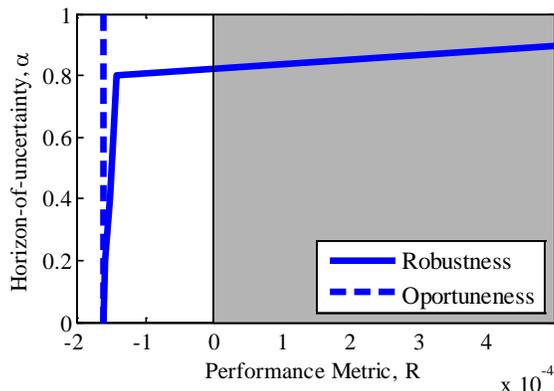


Figure 9. Info-gap analysis using the baseline set of design parameters.

The info-gap robustness and opportuneness functions are calculated by allowing the intermediate variables to vary from their nominal settings. Using the intermediate variables bypasses the difficulty of searching for the worst-case and best-case performances within the 21-dimensional space of calibration variables. The best-case and worst-case performances are determined using the influential calibration variables identified in Table 3. It is noted that the info-gap analysis performed in the current section is applied to the maximum output obtained by the entire model, whereas the info-gap analysis provided previously in Figure 7 is applied to the performance of intermediate variable x_i for calibration purposes. Further, the current info-gap analysis is applied directly to the calibration variables whereas the analysis in Figure 7 is applied to the hyper-parameters of the probability distributions used to define the calibration variables.

Figure 9 shows the info-gap results, in which a horizon-of-uncertainty of $\alpha = 1$ corresponds to $\pm 27\%$ deviation of calibration variables from their nominal settings. (For variables p_1 , p_{13} , p_{14} , and p_{16} , the level of $\alpha = 1$ corresponds to their upper bounds.) The white area of the figure indicates the safe domain, whereas the grey shaded area identifies the failure domain. The opportuneness function is plotted with dashed line; it suggests that, given the baseline set of design parameters used in these simulations, varying the calibration variables does not significantly improve the performance. Conversely, the robustness function, which is plotted with a solid line, suggests that the performance degrades progressively as the calibration variables vary. The performance enters the failure domain when $\alpha = 0.8$, which represents 22% uncertainty. Said differently, the performance is guaranteed to remain requirement-compliant as long as the calibration variables do not deviate by more than 22%, or $\alpha = 0.8$, from their nominal settings.

Table 6. Statistics of the robustness function.

Horizon-of-uncertainty, α	Performance, R	Maximum Model Output, g
$\alpha = 0$	-1.62e-04	g_7
$\alpha = 0.2$	-1.59e-04	g_7
$\alpha = 0.4$	-1.54e-04	g_7
$\alpha = 0.6$	-1.48e-04	g_7
$\alpha = 0.8$	-1.42e-04	g_7
$\alpha = 1.0$	+1.20e-03	g_4

An explanation for the “bifurcation” of the robustness function, where the slope decreases significantly between $\alpha = 0.8$ and $\alpha = 1.0$, is given in Table 6. As shown in the table, the model output that controls the performance starts off as the seventh model output, g_7 , then switches to the fourth model output, g_4 , between $\alpha = 0.8$ and $\alpha = 1.0$. Thus, the behavior of the robustness curve changes significantly as the output controlling the performance changes. Lastly, the results of Table 6 are consistent with the results of Table 5, in that the maximum output that is requirement-compliant is controlled by the same model outputs, g_7 and g_4 . Such consistency is expected in our analysis.

VII. Application to the NASA MUQC Problem: Robust Design

The info-gap results provided in the previous section suggest that the model performance can change considerably as the calibration variables are allowed to vary. The objective of our robust design is to improve on the baseline design, defined by the problem moderators and used in the analysis thus far, such that the robustness to calibration variable uncertainty is increased. Doing so ensures that the performance remains requirement-compliant even as these uncertain calibration variables vary away from their nominal settings.

A robust design analysis is complicated by the potential coupling between calibration variables and design parameters. Further, the NASA MUQC problem is defined in a 35-dimensional space, which can be computationally demanding to explore. To reduce the number of design parameters, an effect screening is performed using the Morris one-at-a-time sensitivity analysis. The Morris method relies on calculating elementary effects (EE) of each parameter using “trajectories” where one parameter is changed at a time. The mean, $\mu_k(\text{EE})$, and standard deviation, $\sigma_k(\text{EE})$, of elementary effects can then be used to assess the overall influence and higher-order effects, respectively, of the k^{th} parameter. “Large” statistics of μ_k and σ_k indicate that the k^{th} parameter is significant to explain how predictions change, whereas “small” statistics suggest the parameter is non-influential. The Morris method is adopted because it offers computational advantages discussed in detail in References 7-9.

Table 7. Design parameter ranking using the Morris one-at-a-time screening method.

Ranking	Performance-based Sensitivity Analysis			Robustness-based Sensitivity Analysis		
	Design Parameter	Normalized $\mu_k(\text{EE})$	Cumulated Statistic	Design Parameter	Normalized $\mu_k(\text{EE})$	Cumulated Statistic
1	d_{12}	33.1%	33.1%	d_1	28.7%	28.7%
2	d_{13}	22.6%	55.7%	d_{11}	23.8%	52.5%
3	d_{11}	11.5%	67.2%	d_3	11.1%	63.6%
4	d_{10}	11.1%	78.3%	d_6	10.3%	73.9%
5	d_9	10.3%	88.6%	d_{12}	10.1%	84.0%
6	d_{14}	7.8%	96.4%	d_9	6.0%	90.0%

Herein, the Morris method is implemented to rank the influence of the design parameters and is evaluated using two criteria: (i) performance and (ii) robustness. Evaluating the performance of the model is straightforward: the maximum output is used for each point in the trajectory. The robustness criterion, however, is computationally more demanding to perform. The robustness is obtained by performing the info-gap analysis, as shown previously in Figure 9. Each design point is assessed for the worst-case performance when the calibration variables can vary by up to 27%, or $\alpha = 1$. A maximum variation of 27% is defined because we seek a design that improves upon the robustness-optimal performance of the baseline design pursued in Figure 9.

Table 8. Frequencies of occurrence of the maximum output for two levels of horizon-of-uncertainty.

Horizon-of-uncertainty	Performance Outputs ^(a)							
	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
$\alpha = 0.0$	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	98.0%	1.5%
$\alpha = 1.0$	0.1%	0.0%	1.4%	81.8%	0.6%	16.2%	0.0%	0.0%

^(a) Note that all percentages over 10% are highlighted to emphasize their contributions to the performance output.

The design parameters that contribute to 90%, or more, of the influence for the two criteria, as identified by the Morris method, are reported in Table 7. Six parameters are identified as influential for each criterion, a significant reduction from the fourteen design parameters of the design space. Even for efficient search algorithms, performing an optimization within a 14-dimensional space can be computational demanding and can lead to convergence difficulties. Three of the four parameters identified as the most influential ones using the robustness criterion, d_1 , d_3 , and d_6 , are not found influential when the performance criterion is considered. The simulation runs performed for the Morris screening are evaluated to determine which parameters control the model output, and reported in Table 8. The results provided in Table 8 are consistent with those given in Table 6: as the calibration variables vary, the outputs, and thus, the design parameters, that control the maximum output also change. This provides a plausible explanation for why the design parameter rankings of Table 7 are different when evaluating the model against performance or robustness.

Improvement of the baseline design is pursued using the design parameters identified in Table 7. The parameters identified using the performance criterion are optimized, through genetic algorithm, to determine the performance-optimal design. Similarly, the parameters identified using the robustness criterion are optimized to determine the

robust-optimal design. The two designs are compared in Figure 10, where the white area indicates that the design is requirement-compliant whereas the grey shaded area identifies the failure region. The performance-optimal design improves upon the performance of the baseline design with a performance of $-1.9\text{e-}04$ versus $-1.6\text{e-}04$ previously; however, only 8.7% uncertainty in the calibration variables can be tolerated before the design enters the failure region. In contrast, the robust-optimal design features a performance that degrades slightly at the nominal settings of calibration variables, but it is able to withstand more than 27% calibration variable uncertainty, as indicated by the fact that the robustness function on the right side of Figure 10 does not cross into the failure region before $\alpha = 1.0$ is reached. For this application, the findings obtained demonstrate the superiority of a robust-optimal criterion over a conventional performance-optimal criterion, both for sensitivity analysis and design optimization.

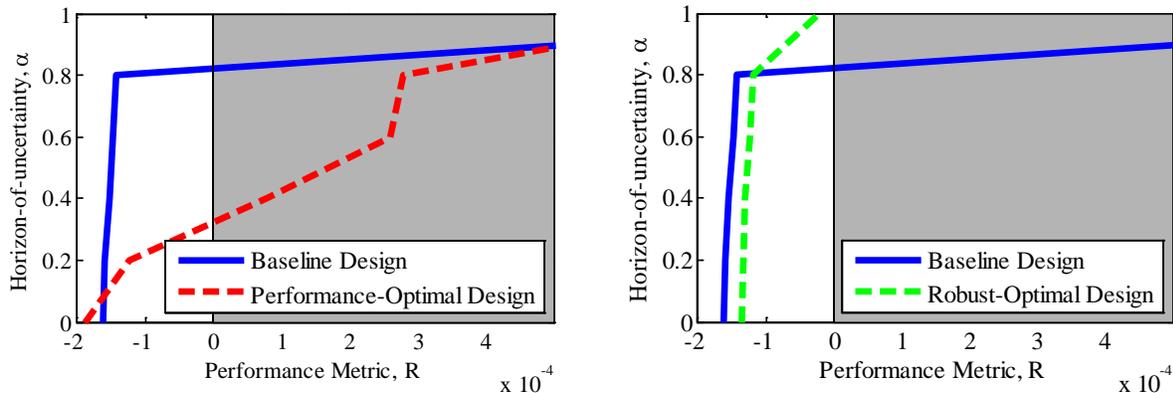


Figure 10. Comparison of info-gap robustness functions obtained performance (left) and robustness (right).

VIII. Conclusion

This paper proposes a discussion of uncertainty quantification and robust design whereby a particular emphasis is placed on the use of info-gap decision theory (IGDT) to manage uncertainty in decision-making. Included in the analysis is the discussion of some of our basic assumptions and justifications. Four specific questions are posed to guide the uncertainty assessment of the NASA Multidisciplinary Uncertainty Quantification Challenge (MUQC) Problem.

Question-A: Can experimental evidence be used to reduce uncertainty in calibration variables influential to model output? (Uncertainty characterization, i.e. calibration.)

Answer: Calibration is performed using two methods: (i) deterministic optimization and (ii) through the use of a robustness criterion. The deterministic optimization highlights the non-uniqueness that often confronts calibration exercises, in that different sets of calibration variables are capable of reproducing the experimental evidence with comparable fidelity. The authors posit that reducing the uncertainty of calibration variables without accounting for this non-uniqueness might eliminate sets of calibration variables that should otherwise have been kept. As an alternative method, calibration is pursued utilizing a robustness criterion, defined in the context of IGDT. Through IGDT, the calibration variables can be constrained to $\pm 20\%$ variation from their nominal values to bound the experimental evidence. However, the full range of parameter values is necessary when trying to replicate the frequency of experiments. This leaves us with somewhat of a modeling dilemma: the parameter uncertainty can be reduced to bound experimental evidence, as long as the frequency of experiments is disregarded.

Question-B: How to determine parameters that are most influential to model output? (Sensitivity analysis.)

Answer: Sensitivity analysis is used to identify parameters that are most influential to model output. A design-of-experiments is analyzed to provide information needed to perform an analysis-of-variance (ANOVA) decomposition. For computational efficiency, sensitivity analysis is performed in a two-step procedure: first, for the calibration variables used to evaluate the intermediate variables; and, second, for the intermediate variables used to evaluate the model output. This procedure reduces the analysis to five 5-dimensional problems, rather than one 21-dimensional problem. It is found that the first intermediate variable, x_1 is overwhelmingly the most influential to the model output. ANOVA results further suggest that the first calibration variable, p_1 , is most influential to the intermediate variable, x_1 . It is noted, however, that these results are dependent upon the baseline set of design parameters used to define the problem, and that these sensitivities can change as the design parameters are varied to pursue alternate designs.

Question-C: What are the worst-case and best-case performances of the simulation model? (Uncertainty propagation and extreme-case analysis.)

Answer: Our approach is to define worst-case and best-case performances that are conditioned on the amount of uncertainty introduced in the model, rather than provide absolute worst-case and best-case performances. This is performed in the context of IGDT, where the uncertainty quantifies the percentage amount that calibration variables are allowed to deviate from their nominal settings. The opportuneness function suggests that not much improvement in the performance can be expected, whereas the robustness function demonstrates that the model enters the failure region once the calibration variables are allowed to deviate 22% from their nominal settings. Statistics of the robustness function suggest that, as the calibration variables are allowed to vary, the model outputs that control performance might change, demonstrating the coupling between calibration variables and design parameters.

Question-D: Can the design be improved, given the potential coupling of the design parameters and calibration variables on model output? (Robust design.)

Answer: Here, a particular emphasis is placed on improving the design by choosing design parameters that are robust to the calibration variable uncertainty. The Morris one-at-a-time sensitivity analysis is utilized to rank the design parameters, using two criteria: (i) performance and (ii) robustness. Due to the coupling between calibration variables and design parameters, it is found that the ranking of design parameters changes depending on the criteria used. The results of the Morris sensitivity analysis are taken advantage of to efficiently search for the performance-optimal and robust-optimal designs. It is found that the performance-optimal design improves performance when the calibration variables are held constant at their nominal settings, but it is no longer requirement-compliant when the calibration variables vary only by 8.7% from their nominal settings. The robust-optimal design is unable to improve the performance when the calibration variables are nominal. On the other hand, the robust-optimal design remains requirement-compliant as the calibration variables vary over 27% from their nominal settings. Such a finding suggests that the robust-optimal design is preferable, due to its ability to remain requirement-compliant as more uncertainty is tolerated. It is emphasized that determining the robust-optimal design comes with the added computational expense of performing the info-gap analysis for the robustness criterion.

Our study concludes that relying on IGDT to address the goals of the MUQC problem defines an approach whereby simplifying assumptions about the problem formulation are kept to a minimum, thus, providing a convenient non-probabilistic alternative for uncertainty quantification. Doing so enables rigorous assessment of prediction uncertainties at a time when numerical simulations are used increasingly to support decision making.

Acknowledgments

The first author acknowledges support from the Advanced Scientific Computing program at the Los Alamos National Laboratory (LANL). The second author is grateful for support provided by the Advanced Certification Campaign at LANL. LANL is operated by the Los Alamos National Security, L.L.C., for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396.

References

- ¹Crespo L.G., Kenny, S.P., Giesy, D.P., The NASA Langley Multidisciplinary Uncertainty Quantification Challenge, *16th AIAA Non-Deterministic Approaches Conference*, National Harbor, MD, January 13-17, 2014.
- ²Ben-Haim, Y., *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd ed., Academic Press, Oxford, 2006.
- ³Jordan, T.L., Bailey, R.M., NASA Langley's AirSTAR testbed – a subscale flight test capability for flight dynamics and control system experiments. *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Honolulu, HI, August 18-21, 2008.
- ⁴Foster, J.V., Cunningham, K., Fremaux, C.M., Shah, G.H., Stewart, E.C., Rivers, R.A., Wilborn, J.E., Gato, W., Dynamics modeling and simulation of large transport airplanes in upset conditions, *AIAA Guidance, Navigation, and Control Conference and Exhibit*, San Francisco, CA, August 15-18, 2005.
- ⁵Van Buren, K.L., Atamturktur, S., Hemez, F.M., "Model selection through robustness and fidelity criteria: Modeling the dynamics of the CX-100 wind turbine blade," *Mechanical Systems and Signal Processing*, to appear, doi: 10.1016/j.ymsp.2013.10.010.
- ⁶Hemez, F.M., Ben-Haim, Y., "Info-gap robustness for the correlation of tests and simulations of a non-linear transient," *Mechanical Systems and Signal Processing*, Vol. 18, 2004, pp. 1443-1467.
- ⁷Morris, M.D., "Factorial sampling plans for preliminary computational experiments," *Technometrics*, Vol. 33, No. 2, 1991, pp. 161-174.
- ⁸Campolongo, F., Cariboni, J., Saltelli, A., "An effective screening design for sensitivity analysis of large models," *Environmental Modeling and Software*, Vol. 22, No. 10, 2007, pp. 1509-1518.
- ⁹Hemez, F.M., "Performance of the Morris one-at-a-time sensitivity analysis to explore large-dimensional functions," Technical Report LA-UR-10-0069, Los Alamos National Laboratory, Los Alamos, NM, 2010.