This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2019-14444C

Approaches for Quantifying Uncertainties in Computational Modeling for Aerospace Applications

John Schaefer*
The Boeing Company, Saint Louis, Missouri, 63166

Vicente Romero†
Sandia National Laboratories, Albuquerque, New Mexico, 87185

Steven Schafer‡
Textron Aviation, Wichita, Kansas, 67215

Brian Leyde§
SmartUQ, Madison, Wisconsin, 53705

Casey Denham¶
NASA Langley Research Center, Hampton, Virginia, 23681

In the past few decades, advancements in computing hardware and physical modeling capability have allowed computer models such as computational fluid dynamics to accelerate the development cycle of aerospace products. In general, model behavior is well-understood in the heart of the flight envelope, such as the cruise condition for a conventional commercial aircraft. Models have been well validated at these conditions, so the practice of running a single, deterministic solution to assess aircraft performance is sufficient for engineering purposes. However, the aerospace industry is beginning to apply models to configurations at the edge of the flight envelope. In this regime, uncertainty in the model due to its mathematical form, numerical behavior, or model parameters may become important. Uncertainty Quantification is the process of characterizing all major sources of uncertainty in the model and quantifying their effect on analysis outcomes. The goal of this paper is to survey modern uncertainty quantification methodologies and relate them to aerospace applications. Ultimately, uncertainty quantification enables modelers and simulation practitioners to make more informed statements about the uncertainty and associated degree of credibility of model-based predictions.

I. Introduction

This paper was written in concert with the AIAA Certification by Analysis (CbA) Community of Interest (CoI) Recommended Practices document, which was under development in 2019. The purpose of that document is to describe a set of recommended practices (tasks) for an applicant to accomplish when flight modeling is being developed, proposed, and used to reduce flight testing relative to established aircraft certification practices. The specific certification requirements considered when these recommendations were developed include aircraft performance and handling qualities, static loads and aeroelastic stability. However, the framework consisting of the recommended tasks may also be applicable when showing compliance to other requirements.

*Aerodynamics Engineer, The Boeing Company, Member AIAA
†Verification & Validation, Uncertainty Quantification and Credibility Processes Dept, Associate Fellow AIAA. Sandia is managed and operated by NTESS LLC under DOE NNSA contract DE-NA0003525
‡Advanced Design Engineer, Textron Aviation, Member AIAA
§Principle Application Engineer, Member AIAA
¶Pathways, Aeronautics System Analysis Branch, 1 N. Dryden St. MS 442, Student Member AIAA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

For the purposes of this paper, flight modeling includes analysis methods of all types, including analyses based on wind tunnel results, numerical methods such as Computational Fluid Dynamics (CFD) and Computational Structural Dynamics (CSD), and Flight Dynamics Simulation (FDS) (i.e., simulations of aircraft flight dynamics behavior with or without a human pilot). The CbA CoI recommended tasks address subjects pertinent to most flight modeling analyses. These tasks are:

- configuration and process management
- verification of the models
- verification the models were applied correctly
- validation of the models
- justification of analysis adequacy in recognition of potential modeling error and/or uncertainty
- a summary assessment of applicability for showing compliance

These tasks provide a framework for organizing information used to determine whether or not a particular analysis method is appropriate for supporting a compliance showing in a particular application. The CbA CoI document (still in work at the time of this writing) includes an annex on uncertainty quantification (UQ), however it is limited in scope. The purpose of this paper is to expand upon the fourth and fifth recommended tasks of the CbA CoI document by providing a survey of UQ considerations and methodologies as they apply to CbA or other aerospace applications.

An overview of the subject of uncertainty quantification is given in Section II. The identification, characterization, propagation and aggregation, and analysis components of UQ are described in Sections II.A-II.D. Methods described in Section II are applied to two realistic engineering examples in Section III to provide estimates of uncertainty for CFD drag prediction and for wind tunnel correction calculations. Section IV includes descriptions of two more hypothetical examples involving longitudinal stability and control characteristics and FDS for a time-to-roll maneuver. Finally, conclusions are made in Section V.

II. Overview of UQ

Uncertainty Quantification is the process of characterizing all major sources of uncertainty in the model and experiment, and quantifying their effect on the analysis outcomes. UQ is a closely related activity to verification and validation (V&V) and is essential for verifying and validating analysis models. For practical engineering purposes, the goal of UQ, along with V&V, is to enable modelers and analysts to make justifiable statements about the accuracy/uncertainty and associated degree of credibility in their analysis-based predictions. In this context, UQ can be defined[1–3] as the

- Identification (Where are the uncertainties?),
- Characterization (What form are they, and what are their mathematical descriptions?),
- Propagation and Aggregation (How do they combine to determine total uncertainty in the analysis results?), and
- Analysis (What are their impacts and implications?)

of uncertainties in analysis models. The outcome of UQ is the enablement of modelers and analysts to make more informed statements about the uncertainty and associated degree of credibility they have in their analysis-based predictions, as compared to the practice of performing deterministic analysis (i.e., not considering uncertainties).

Estimating the impacts of modeling errors or uncertainties is not new to certification. For example, the EASA generic Equivalent Safety Finding (ESF) certification review item for Requirement 25.251(b)[4] and the corresponding FAA Equivalent Level of Safety Finding (ELOS) issue papers[5] state that applicants may propose to use computational fluid dynamics to show that installing large radome or antenna covered by an aerodynamic fairing does not introduce excessive vibration, provided that the applicant can demonstrate that the CFD tool is "valid for its intended use" and that "CFD Modeling Errors" are understood. The

Table 1. Some sources of uncertainty in experiments and experimental data.

Source of Uncertainty	Description
Measurement uncertainties	Sensor and data acquisition system inaccuracies, etc., in measurement or estimation of: • Experimental conditions and inputs, e.g., initial and boundary conditions • System responses, behaviors, outputs
Extrapolation-bias uncertainty	Occurs when bias corrections and uncertainty characterizations of experimental results are extrapolated from where they were originally quantified
Random variability over replicate tests	Examples are unit-to-unit geometric and physical variability in replicate tests and variability of measurement accuracy in the different tests
Data processing and inference uncertainties	Examples: • Uncertain bias error in temporal and/or spatial field results interpolation, integration, trend extrapolation, etc. • Uncertain bias in inferring the full population of a random quantity (e.g. a frequency distribution or probability density) from limited data samples • Uncertain bias error in uncertainty propagation and aggregation procedures that combine data measurement, processing, and inference uncertainties

public domain records show the EASA ESF and/or the FAA ELOS have been applied to many certification projects (for the FAA, see Ref. 5). Therefore the industry is already addressing the topic of modeling errors for this particular application. However, as flight modeling technologies improve, offering a potential increase in viable applications, UQ tools and processes can serve as a valuable part of the analysis to assess whether a given modeling process is robust enough for a particular application.

This paper introduces some important aspects of UQ that are often encountered in engineering practice and that are important to certification by analysis. A level of understanding of the four elements of uncertainty quantification outlined above, appropriate to the application, is crucial to the success of any uncertainty quantification efforts; each of these elements will be expanded upon in the following sections.

II.A. Identification

The first step in any UQ analysis is to identify the *sources* of uncertainty which may contribute to uncertainty in the output quantity of interest. Table 1 summarizes various commonly encountered sources of uncertainty in experiments and experimental data; Table 2 summarizes sources of uncertainty in models and simulations. The overview in the remainder of this section adds context to many of the entries in the tables.

Table 2. Some sources of uncertainty in models and simulations.

Source of Uncertainty	Description
Phenomenological modeling uncertainty expressed through use of multiple model forms	Examples: • Employing different plausible turbulence models • Multiple equation-of-state models or material constitutive models • Geometry models (model A with bolts explicitly modeled vs. model B without explicit bolts)
Uncertain values for model input parameters, or *multiple* values for input parameters	Examples: • populations of values described by frequency distributions • manufacturing tolerances and variability
Model form and parameter-related prediction-bias uncertainty	Whatever the model form and parameter values, the approximations involved will generally result in some degree of prediction bias even if the mathematical equations are solved exactly and initial and boundary conditions inputs to the model are exact. (Model validation seeks to quantify model prediction bias; model calibration and/or result correction approaches seek to reduce it.)
Numerical solution-bias uncertainty	This is associated with spatial and temporal discretizations and incomplete iterative convergence of the discretized equations being solved. (This is what *Solution* or *Calculation Verification* seeks to quantify.)
Simulation results processing and inference uncertainties	Examples: • Uncertain bias in temporal and/or spatial field results interpolation, integration, trend extrapolation, etc. • Uncertain bias in uncertainty propagation and aggregation procedures for the modeling and simulation uncertainties above

II.B. Characterization

In general, there are two forms of uncertainty – aleatory and epistemic – which can be defined as:

> **Aleatory** – A type of uncertainty which is due to inherent, irreducible chance. Probabilistic in nature, and most commonly described by a probability density function (PDF) or cumulative distribution function (CDF).
>
> **Epistemic** – A type of uncertainty which is due to lack of knowledge and is potentially reducible. These can be represented probabilistically or non-probabilistically. Probabilistic epistemic uncertainty may be represented by a CDF or PDF that reflects subjective estimates or beliefs about the probabilities associated with a quantity taking a particular value. Non-probabilistic epistemic uncertainty is commonly described by an interval with lower and upper bounds, but with no probabilistic information assumed between.

The characterization step includes both the aleatory / epistemic categorization and the mathematical description of the identified sources of uncertainty. In general the mathematical description of uncertainties will be in one of the forms described above in the aleatory and epistemic definitions. Note that all aleatory uncertainties are probabilistic, but not all probabilistic uncertainties are aleatory. In worst-case analysis studies, which are common in engineering practice, choosing a set of worst-case input parameters is equiv-

alent to choosing a very high or very low probability level from a distribution (for aleatory or probabilistic epistemic uncertainties) or to choosing a lower or upper bound of an epistemic interval (for non-probabilistic epistemic uncertainties).

A simple example can help demonstrate the differences between aleatory and epistemic uncertainty. Consider the uncertainty in the value of a six-sided die. If the die is rolled, it has an inherent, irreducible chance of landing on any number one through six, with equal probability if the die is fair; this would be considered an aleatory uncertainty. Contrastingly, if a die were placed on a table non-randomly and covered up (that is, the number facing up is controlled by the placer instead of being random), there is still uncertainty in its value, but the uncertainty is due to lack of knowledge and can be reduced by questioning the placer or uncovering the die; this case would be considered epistemic.

A special case of uncertainty characterization and treatment occurs when both aleatory and epistemic uncertainties are present in an analysis. The combination of aleatory and epistemic uncertainty is called Mixed Uncertainty, and it requires different treatment than purely aleatory or purely epistemic uncertainties. Returning to the die example, if a die is known to be weighted (which changes the probability of rolling 1, 2, 3, 4, 5, or 6), but the exact nature of the weighting is unknown, then there exists a specific, but unknown (epistemically uncertain) probability distribution for the random (aleatory uncertain) result of rolling the die.

If the uncertain result of the die is to be used in another analysis, then the results of the subsequent analysis will also be uncertain. The translation of uncertainty in the die result to uncertainty in the analysis results must be performed using an uncertainty propagation method.

Examples of aleatory sources of uncertainty in the context of aerospace engineering might include atmospheric conditions and manufacturing or flight test instrumentation variability from one aircraft to the next. Examples of epistemic sources uncertainty might include discretization error in computational modeling, model form error such as a turbulence model that results in an inaccurate extent of smooth-surface flow separation in a CFD analysis, or the omission of unsteady aerodynamic terms in a flight simulation model. When both aleatory and epistemic uncertainty are present in one analysis which is used to inform another analysis, then the inputs to the second analysis are mixed. For example, a CFD simulation with uncertain geometry and turbulence model parameters yields mixed uncertain outputs for aerodynamic force and moment coefficients, which may then be used as inputs to a flight dynamics simulation code to assess aircraft maneuvering performance.

II.C. Uncertainty Propagation and Aggregation

During the uncertainty propagation step, methods are employed to translate the characterizations of the identified sources of uncertainty into estimates of uncertainty in output Quantities of Interest (QoI). The resulting uncertainty may then be aggregated at the level of the output results, such as model-discretization-related solution uncertainty, model-form-related prediction-bias uncertainty, etc. Uncertainty roll-up (see Figure 1) consists of all uncertainty propagations and aggregations in quantifying the cumulative uncertainty associated with experimental or simulation results. The figure illustrates the distinctions between uncertainty propagation, aggregation, and roll up. The figure involves a general case where continuous interval and probabilistic uncertainties (PDFs) are propagated along with discretely represented uncertainties.

The uncertainty propagation step for mixed uncertainty problems is complicated due the need to separately treat the aleatory and epistemic sources of uncertainty in analysis results. The literature contains contemporary uncertainty propagation methods for continuous probabilistic uncertainty in computational applications, including sampling methods,[7–9] spectral response surface surrogate-model based methods,[10,11] quadrature methods,[12,13] non-spectral, non-quadrature based methods,[14–17] and optimization-based reliability methods.[18–20] Refs. [21, 22] demonstrate propagation and aggregation of probabilistic discrete uncertainties and non-probabilistic interval uncertainties. If epistemic uncertainty (e.g., interval or subjective probability) is propagated in addition to frequency-based aleatory uncertainty, then various potential PDFs of aleatory response will exist, as indicated at right in Figure 1. This normally requires dual-level propagation approaches.[23–26] Alternative representations of this type of mixed uncertainty are often advantageous for conciseness of the uncertainty representation and to reduce propagation cost (e.g., Probability Box (p-box) representations[27–30]). Figure 2 shows a p-box representation of mixed aleatory-epistemic uncertainty resulting from one common mixed-uncertainty propagation approach called the Second-Order Probability (SOP) method.[29]

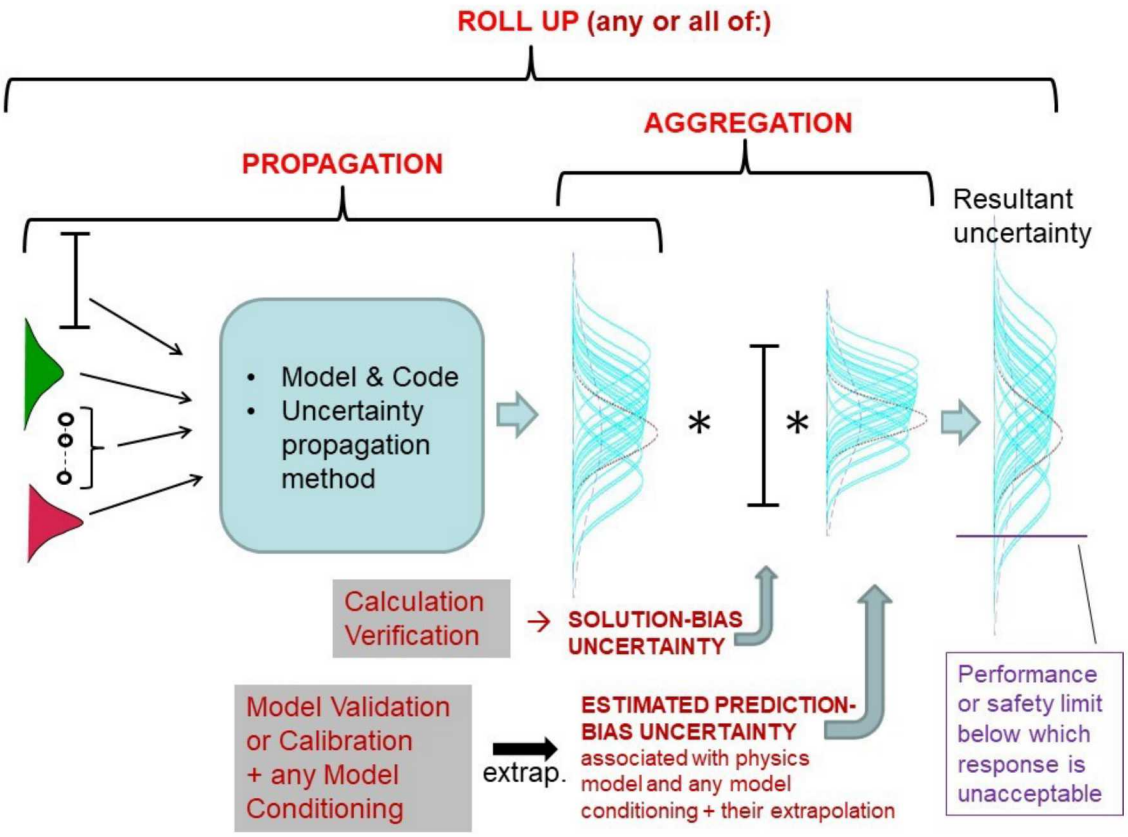


Figure 1. Example of uncertainty roll-up with predicted response assessed against safety or performance requirement (adapted from Ref. 6).

II.D. Analysis

The analysis involved in UQ typically represents the majority of the effort, and can be classified into several sub-analyses, such as sensitivity analysis, UQ-based resource allocation, model validation and calibration, and margin assessment.

Once the various uncertainties in a model or analysis have been identified, their form characterized, and then propagated and aggregated, it can be useful to determine which uncertainties contribute the most to the total uncertainty. This is typically referred to as Sensitivity Analysis. Sensitivity analysis (SA) is defined for the present purposes as:

> **Sensitivity Analysis** – the determination of how much uncertainty an individual source contributes to the total uncertainty in a simulated or experimental quantity of interest (QoI). Information on specific methods for sensitivity analysis can be found in representative references such as Refs. [32–36].

SA enables resources to be focused on characterization and/or propagation of the most influential uncertainty sources. Frequently, it arises naturally as an output of the UQ analysis. The parameter variation studies common in engineering practice are a form of SA. A simple means of addressing modeling uncertainty is to use a sufficiently conservative combination of input parameters to determine a conservative output value for the QoI. However, this approach requires the uncertainty in QoI to vary in a known manner with respect to each input parameter in order to ensure the intended level of conservatism is achieved. If the QoI variation with input parameters is in not well understood, then a sensitivity analysis covering the range of uncertainty in the relevant model input parameters should be employed.

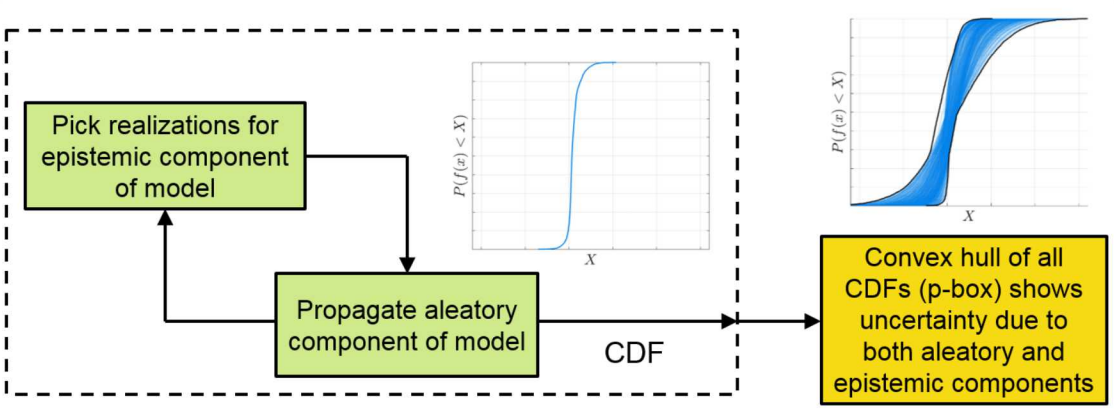


Figure 2. Flowchart describing the various aspects of Second-Order Probability (adapted from Ref. 31).

A related process, UQ-based resource allocation, involves UQ and SA as forward analyses in an inverse problem. The goal of UQ-based resource allocation is as follows:

> **UQ-Based Resource Allocation** – addresses the quantification, management, and optimization of experimental, analysis, and computational resources such that their value-added impact on total QoI uncertainty is adequately estimated and used to prioritize activities and resources to most cost-effectively quantify, control, and reduce uncertainty.

Improved methods for UQ-based resource allocation in engineering practice is an active area of research and development.[37–39] The top row of Figure 3 shows a linking between uncertainty quantification, sensitivity analysis, and UQ-based resource allocation.

Referring to the bottom row of Figure 3: uncertainty categorization, representation, propagation, and aggregation have been discussed in the preceding sections. This section discusses the other topics in the bottom row of the figure: solution bias uncertainty estimation, prediction bias and uncertainty due to model form, extrapolation and performance margin UQ.

For numerical models (such as CFD and FEM), solution bias error and the uncertainty associated with its estimation typically comes from spatial and/or temporal discretization of the governing continuum physics equations and of mathematical descriptions of geometry, and from incomplete convergence in iterative solutions of the discrete equations due to non-zero error tolerances needed for computational affordability. Solution bias error is what "Solution" or "Calculation" Verification attempts to quantify; see Refs. [40–46]. Sources of error which are quantified include discretization error, convergence errors, machine precision or round-off error, and coding errors.[40] Richardson extrapolation is a common method for estimating discretization error.[41,42] A framework for quantifying discretization error based on Richardson extrapolation can be found in Ref. [44], while an example implementation of that framework can be found in Ref. [40]. It is noted that convergence error (also known as iteration error) has an effect on the discretization uncertainty estimation, thus quantifying discretization uncertainty requires iterative convergence.[40] A general rule of thumb is to ensure that the convergence error for a given calculation is negligible when compared to the estimate for the discretization error; Ref. [44] recommends at least one order of magnitude smaller. Ref. [46] discusses assessment and management of discretization related error over a large parameter space of analysis with the model, such as in optimization and/or uncertainty propagation. Round-off, or machine precision error, is typically negligible when compared to discretization error and convergence error. Coding errors are typically "unknown errors," as they are corrected as soon as they are discovered. They are thus very difficult to estimate; see Ref. [1].

Uncertainty often exists regarding the most appropriate way to mathematically represent or model particular physical phenomena, geometry, boundary and initial conditions, materials properties and behaviors, etc. This representational "model form" uncertainty, per the nomenclature in Table 2 and Figure 3, is sometimes explicitly expressed through the presence and use of multiple candidate model forms that are distinct from each other. That is, they are not simple variations of each other given by the same model form but different numerical values for the parameters. In the literature and in this document, the term "model-form

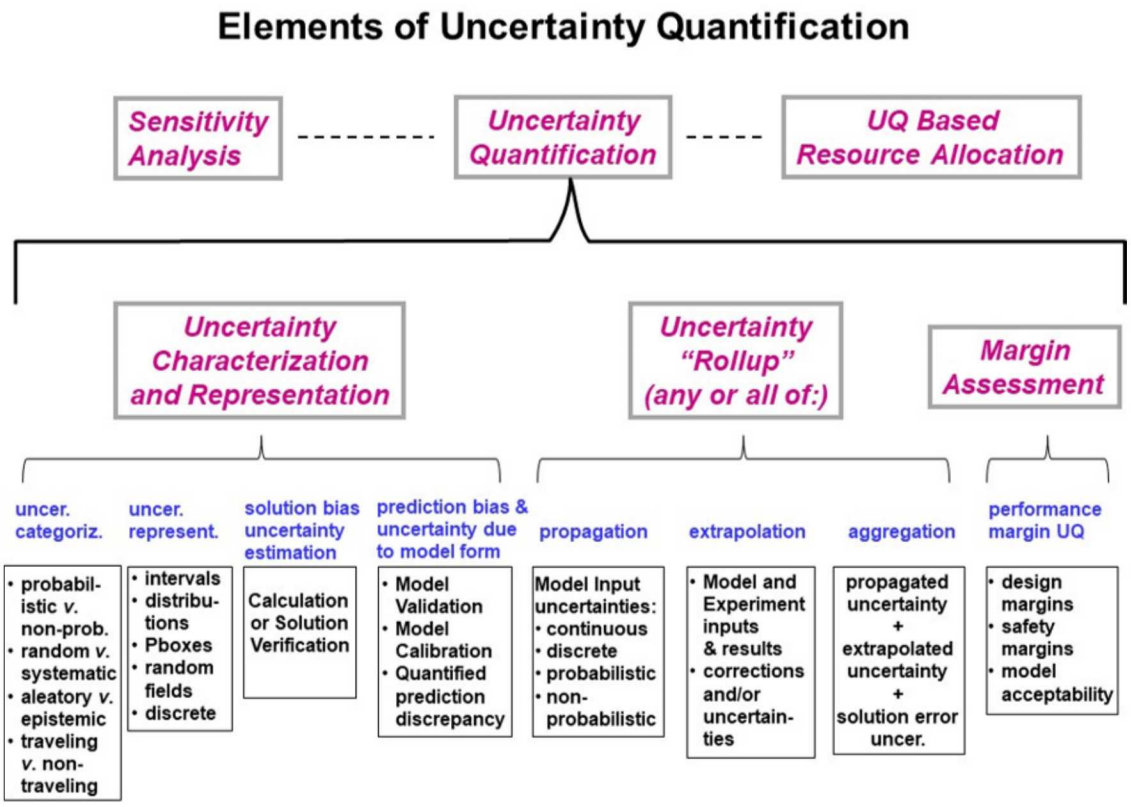


Figure 3. Some important aspects of UQ that can arise in analysis projects (reproduced from Ref. 6).

uncertainty" is sometimes meant to *exclusively* involve the mathematical structure of a model and at other times may include *both* the mathematical structure of a model and the uncertainty in its parameter values (which themselves may be considered models). More information regarding model form uncertainty can be found in Refs. [45, 47–54]. An example of applying model-form uncertainty to an industry-relevant problem is described in Ref. [55]. Expert judgment is required to guide the use of the results from this approach, as there is no general assurance that the results from the multiple model forms selected for analysis encompass the "true" value for the output QoI.

A practical objective of model validation is to quantify prediction bias and uncertainty for an identified QoI. This is accomplished by assessing the model at validation conditions that present relevant and significant extrapolative tests of model predictive capability away from the calibration conditions. Relevant, in this case, means that the extrapolation bias for other analysis cases of interest can be estimated from these validation results. Validation quantification of prediction bias and prediction bias uncertainty should also include any significant errors and uncertainties of measured, processed, modeled, and inferred elements of the validation experiments, such as initial conditions (ICs), boundary conditions (BCs), geometries and output quantities of interest.

The objective of model calibration is to reduce the quantified model prediction bias and uncertainty for an identified QoI at a set of calibration conditions. Calibration is typically accomplished by adjusting model inputs that are not fixed by the calibration experiment conditions and are thus available degrees of freedom. In model calibration, model-form and/or parameter range constraints may prevent the model results from exactly matching the data being calibrated to, whereupon a calibration "matching or fitting gap" occurs. Calibration results and any fitting gap (often termed discrepancy) are functions of the calibration procedure used and of any error or uncertainty in the inputs to, or performance of, the calibration model relative to the experiments that is not addressed as a degree of freedom in the calibration itself. Correction functions are sometimes applied to approximately correct QoI results for any fitting-gap discrepancy left after calibration and before predicting with the model at conditions different from the calibration experimental conditions. Validation seeks to quantify the model prediction-bias uncertainty due to the model form; calibration seeks to reduce it.

How to best use a model's validation- or calibration- characterized prediction bias and uncertainty to potentially adjust or bias-correct the model to mitigate prediction risk beyond the validation or calibration conditions is a very difficult question and an active area of research. But ultimately, estimation of prediction bias error and uncertainty at points where no experimental data exists should include extrapolation of validation-quantified prediction bias/uncertainty (after any model and output QoI corrections for this), and should also include estimated error/uncertainty associated with this extrapolation. Example references [24, 51,52,56–60] address these issues in terms of applied methodology and the presence of various heterogeneous sources and types of uncertainty as itemized in the leftmost two boxes on the bottom row of Figure 3.

Estimation of the margin between system performance and important or critical performance, safety, or failure thresholds is important in engineering design and safety analyses. An example of a margin assessment circumstance for a given requirement or limit is shown at right in Figure 1. Another example using p-boxes is shown in Figure 4. Per the examples, margin analysis is often accompanied by significant uncertainty in model predictions of system performance. Significant uncertainty may also be present in critical threshold limits for system degradation or failure. Engineers are frequently interested in the probability that a design exceeds some performance threshold, or perhaps a threshold with a factor of safety. Therefore, margin analysis is an important and distinct type of UQ analysis, just as solution verification and model validation are. Margin analysis can also have bearing on model acceptability and/or need for refinement, depending on the magnitude of prediction uncertainty relative to the magnitude of the predicted margin. Representative references that describe margin uncertainty analysis include Refs. [61–63].

Finally, in many real engineering applications of UQ, simulation and experimental practitioners are very limited to the true precision or accuracy that can be claimed in UQ characterizations of model and experimental inputs and analysis outputs, especially when many heterogeneous types of uncertainties are involved. Consequently, the outcome of UQ is the enablement of modelers and analysts to make useful and more informed (though perhaps not precise) statements about the uncertainty and associated degree of credibility they have in their analysis-based predictions. As modelling technology continues to advance and become applicable to more and more complicated applications, the concepts and tools of UQ will be critical to guiding its appropriate use.

III. Examples with Results Computed

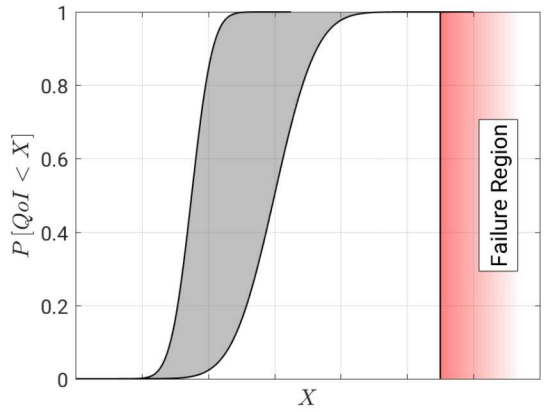
In this section, the methods and considerations discussed in Section II are applied to two realistic engineering problems. Both problems involve the NASA Common Research Model (CRM),[64] an open geometry developed by Boeing and NASA which is representative of a modern commercial transport aircraft. Multiple configurations of the CRM have been designed and experimentally tested for the AIAA Drag Prediction Workshop Series[65,66] and the AIAA High-Lift Prediction Workshop Series.[67,68] The first example problem investigates the uncertainty in CFD drag predictions of the High Speed CRM,[69] which is designed to fly cruise Mach number of 0.85 at a lift coefficient of 0.5. The second example problem assesses the uncertainty in High Speed CRM and High-Lift CRM (HL-CRM)[70] wind tunnel corrections. Images of the High Speed CRM and HL-CRM geometry are included in Figure 5.

III.A. Uncertainty in CFD Predictions of Transonic Cruise Drag

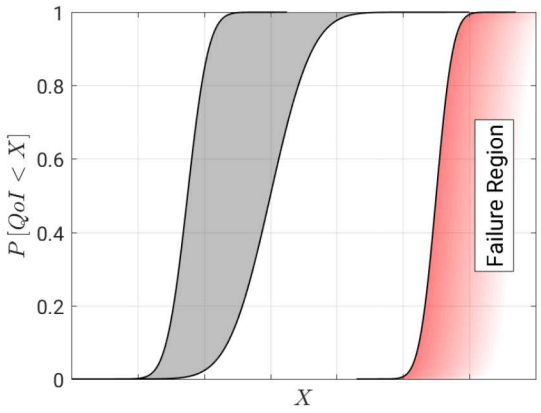
For a large, generic, twin-engine transport aircraft, a 1 count (0.0001) decrease in drag coefficient (C_D) amounts to roughly 200 lbs increase in payload capacity.[71] The average American weighs around 185 lbs.[72] Therefore, accurate predictions of drag are essential in determining the potential passenger capacity (which equals economic value) for commercial aircraft designs.

There are many potential sources of error and uncertainty in CFD simulations. In this example, the following are considered:

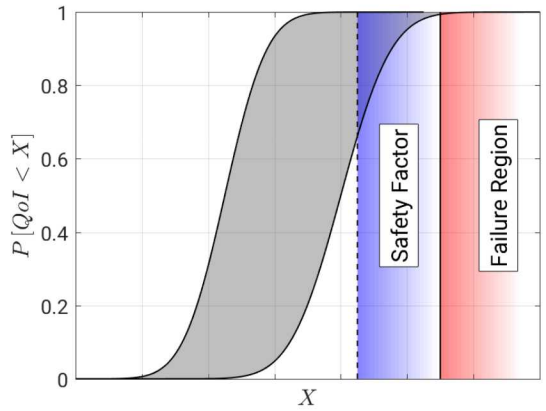
- Freestream conditions. Specifically, Mach number (M), angle of attack (α), and angle of sideslip (β).
- Spalart-Allmaras (S-A)[73] turbulence model coefficients. Specifically, σ, κ, and c_{w3}.
- Grid convergence error



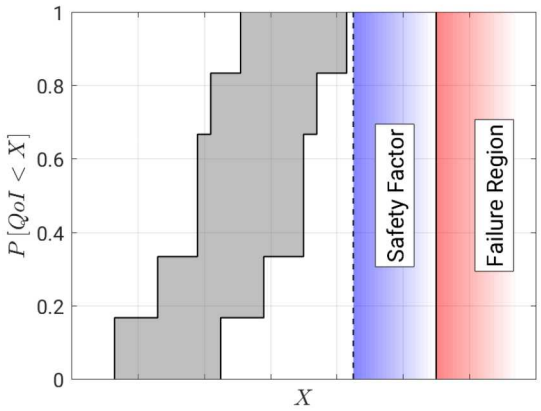
(a) Uncertain simulation and deterministic failure region.



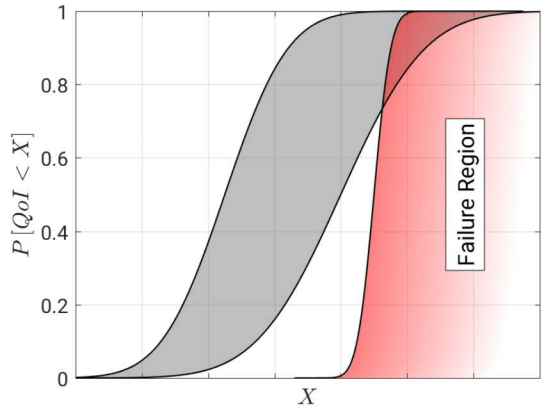
(b) Uncertain simulation and uncertain failure region.



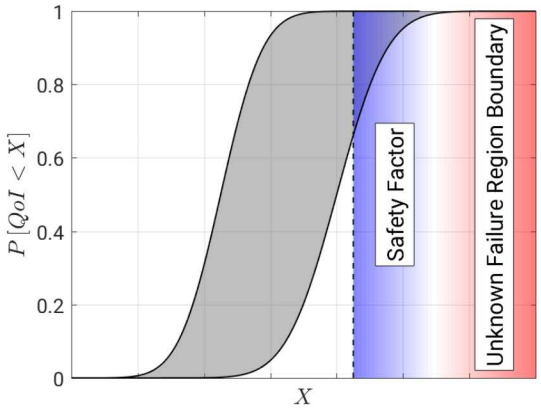
(c) A simulated design which violates a safety factor, but not failure region.



(d) Uncertain experiment or flight test and failure region with safety factor.

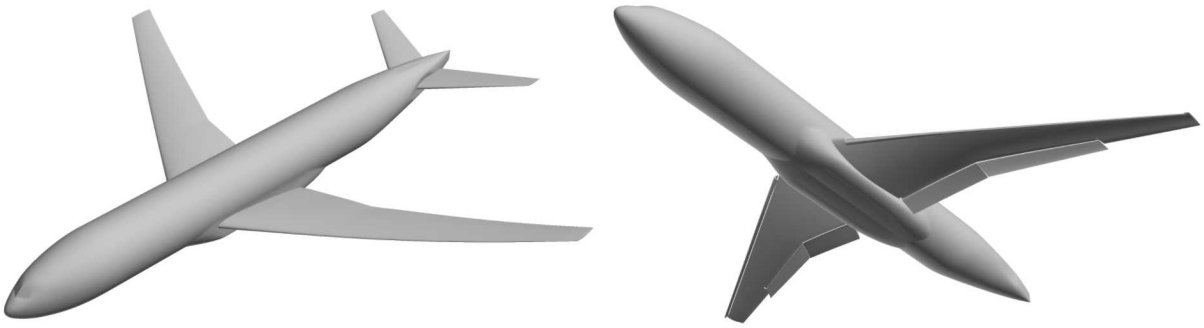


(e) A simulated design which violates an uncertain failure region.



(f) A simulated design which violates a safety factor for an unknown failure region.

Figure 4. Examples of results which might be obtained using margin analysis for simulation or experimental data. Analysis data represented by gray p-boxes.



(a) High Speed CRM geometry. (b) HL-CRM geometry.

Figure 5. High Speed and High-Lift NASA Common Research Model geometries.

The flow solver used in this example is the Boeing-developed BCFD code,[74] version 8.0. BCFD is an unstructured, grid cell-centered, three-dimensional, finite-volume RANS code capable of solving steady and unsteady Euler, laminar, or turbulent flows. The code allows changes to the values of the S-A turbulence model coefficients to support uncertainty analyses. Multiple cases from the NASA Turbulence Model Resource (TMR) webpage[75] have been used to verify the S-A implementation.

All freestream conditions are assumed to be aleatory uncertainties, with normal distributions given by

$$M = \mathcal{N}(0.85, 0.0005) \tag{1}$$

$$\alpha = \mathcal{N}(2.3103^\circ, 0.008^\circ) \tag{2}$$

$$\beta = \mathcal{N}(0.0^\circ, 0.01^\circ) \tag{3}$$

where $\mathcal{N}(m, s)$ is a normal distribution with a mean of m and standard deviation of s. The distributions for M and α are estimates based on the work by Walter et al.,[76] and the distribution for β was provided by test engineers* at the NASA Langley National Transonic Facility (NTF), where a large amount of the High Speed CRM wind tunnel experiments took place. Nominal conditions of $M = 0.85$, $\alpha = 2.3103^\circ$, and $\beta = 0.0^\circ$ correspond to the designed cruise lift coefficient of $C_L = 0.5$ for the wing-body-tail0 configuration.

A characterization of uncertainty in S-A turbulence model coefficients is described by Schaefer et al.[77] Each of the three coefficients is consisdered to be an epistemic interval given by

$$\sigma \in [0.6, 1.0] \tag{4}$$

$$\kappa \in [0.38, 0.42] \tag{5}$$

$$c_{w3} \in [1.75, 2.5] \tag{6}$$

The intervals for σ and c_{w3} were chosen based on the recommendations of Spalart and Allmaras.[73] Bailey et al.[78] determined that $\kappa = 0.40 \pm 0.02$ in their turbulent pipe flow experiments. In private communication with the first author†, Philippe Spalart (a co-author of the S-A model) confirmed that these bounds on turbulence model coefficients are physically realistic. Additional relations between σ, κ, and four other S-A coefficients are enforced and are described by Schaefer et al.[77]

Grid convergence error (G) is characterized as an epistemic uncertainty. Its effects can be modeled as a linear function of some epistemic variable $\zeta \in [-1, 1]$, where

$$G(\zeta) = \zeta * (\text{Error Estimate}) \tag{7}$$

The error estimate is computed using the Roache grid convergence index (GCI) method described by Celik et al.,[44] and the ASME V&V 20 Standard,[40] as discussed in Section II.C. In a sequence of three grids, the GCI method computes an error estimate for the finest grid solution, resulting in an epistemic interval centered at the finest grid solution. An argument could be made that the interval ought to be centered at

*Private communication between Joe Morrison, Melissa Rivers, Andrew Cary, and John Schaefer 2016.

†Email correspondence between Philippe Spalart and John Schaefer, June-December 2014.

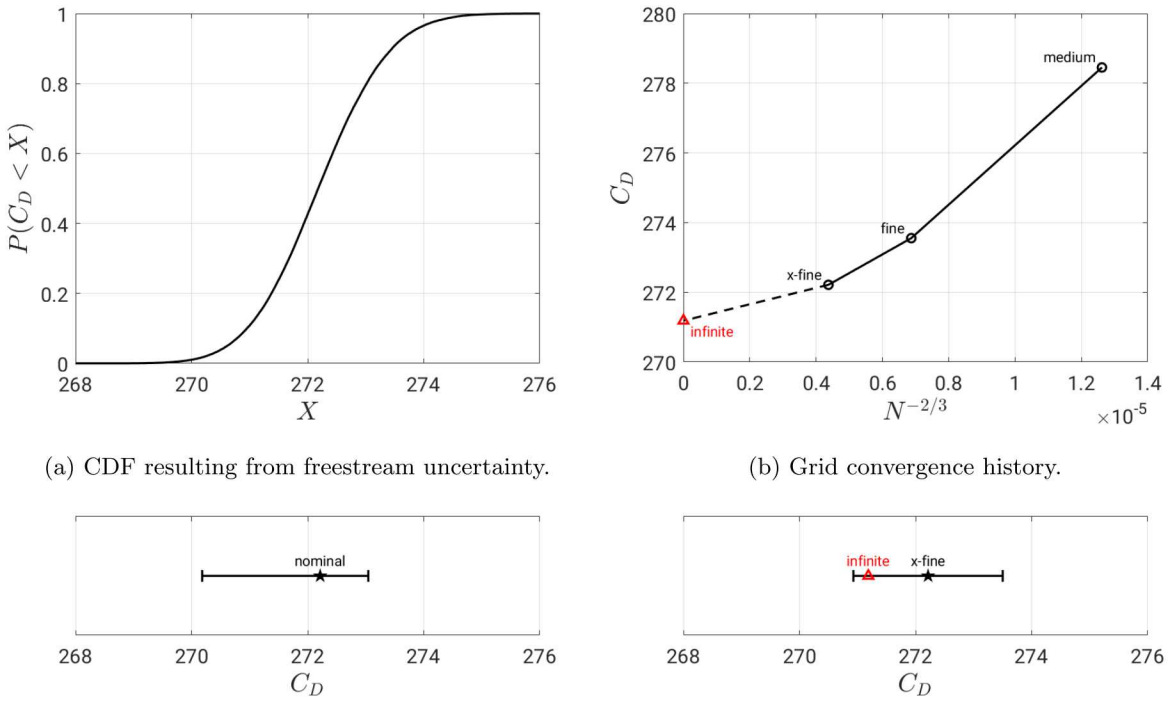
the extrapolated solution for well-behaved grid convergence sequences; however centering the interval at the finest grid solution protects against the possibility that the asymptotic region may not have actually been reached.

To propagate uncertainty, multiple methods were employed. All three components of C_D uncertainty (freestream, S-A, and grid convergence) are assumed to be orthogonal to each other; their combined effects are assumed to be purely additive. This assumption ignores any interactions between the three groups of uncertainty sources, however it reduces the sampling cost required to perform the analysis. Propagation of freestream uncertainty is performed using the point-collocation non-intrusive polynomial chaos (NIPC) method[10,11] with a second order polynomial fit and an oversampling ratio of 2.0. Given these choices, and the three sources of freestream uncertainty ($M,\ \alpha,\ \beta$), NIPC requires 20 CFD solutions for the propagation of freestream uncertainty. NIPC was also used for the propagation of S-A turbulence model uncertainty. Once again using a second order polynomial, oversampling ratio of 2.0, and three uncertain dimensions ($\sigma,\ \kappa,\ c_{w3}$), an additional 20 CFD solutions were required. As described above, uncertainty due to grid convergence error was computed using Richardson extrapolation and the Roache Grid Convergence Index, as described by Celik et al.[44] Medium, fine, and x-fine grid solutions were required to perform this computation. The number of cells in the medium, fine, and x-fine grids is $N_{\mathrm{m}} = 22.3$, $N_{\mathrm{f}} = 55.5$, and $N_{\mathrm{xf}} = 109.4$ million, respectively; a representative length-scale of $N^{-2/3}$ was employed. Although not required for uncertainty propagation, a coarse grid solution was also run to ensure asymptotic behavior of grid convergence across four grids. In total, 43 CFD solutions were required to propagate the uncertainty in the seven dimensions considered in this example (1 medium grid, 1 fine grid, and 41 x-fine grid solutions).

The results of the three CFD uncertainty propagations are shown in Figure 6. For reference, the nominal drag coefficient (from an x-fine solution at the nominal flight conditions with the nominal S-A coefficients) is $C_D = 272.22$ counts. The 95% confidence interval on C_D from the freestream uncertainty propagation is $[270.32, 274.13]$ (Figure 6a); the epistemic interval for the S-A uncertainty propagation is $[270.17, 273.04]$ (Figure 6c); and the epistemic interval for the grid convergence error propagation is $[270.93, 273.50]$ (Figure 6d). The widths of these freestream, S-A, and grid convergence intervals are 3.81, 2.87, and 2.57, respectively, indicating that uncertainty in C_D due to the freestream conditions is slightly larger, but on the same order of magnitude as the uncertainty due to S-A coefficients and grid convergence error.

Since the groups of C_D sources of uncertainty (freestream, S-A, and grid convergence) are assumed to be independent of one another, their effects can be combined in an additive manner. First, the CDF resulting from freestream uncertainty is shifted left and right by amounts equal to the differences between the S-A epistemic interval bounds minus the nominal S-A solution. Next, the left and right bounds of the resulting p-box are further shifted left and right by the differences between the grid convergence epistemic interval bounds and the x-fine solution (which is the same as the nominal S-A solution). An aggregated p-box resulting from this build-up of uncertainty is shown in Figure 7. This aggregated p-box has a 95% conservative confidence interval (CCI) of $[267.00, 276.26]$, which has a width of 9.26 counts (as described by Cary et al.,[31] a CCI is defined by taking the lower probability level from the left-bounding CDF of a p-box and the upper probability level from the right-bounding CDF of a p-box). The solid CDF in the middle of the p-box is the same as the freestream CDF shown in Figure 6a; the blue-shaded region enclosed by the dash-dot line is the aggregation of freestream and S-A uncertainty; and the orange-shaded region enclosed by the long-dash line is the aggregation of freestream, S-A, and grid convergence uncertainty.

To analyze the results of the uncertainty propagation and aggregation, comparisons are made to experimental data. Wind tunnel testing was performed on the High Speed CRM at both the NASA Langley NTF facility and the NASA Ames 11-foot facility. These test campaigns utilized the exact same test article, and both tunnels were operated at identical Reynolds numbers. Limited replicate data points were collected at the $M = 0.85$, $C_L = 0.5$, $Re = 5 \times 10^6$ condition during testing at both facilities (NTF Test 197, Runs 92, 97, 99 and Ames 11-foot Test 216, Runs 76, 77, 80, 83). Data collected during these tests is availabile for download at the NASA CRM website.[64] Acheson and Balakrishna[79] describe the repeatability of individual measurands collected during both tests; at the NTF, the $2\sigma_{\mathrm{avg}}$ C_D repeatability is reported to be $\pm$ 3.4 drag counts, and in the Ames 11-foot, the $2\sigma_{\mathrm{avg}}$ C_D repeatability is reported to be $\pm$ 2.2 drag counts. These repeatability figures are attractive because of their relatively easy interpretability, but understanding the full experimental data UQ problem may require wind tunnel expertise; for this reason, wind-tunnel experimentalists and UQ specialists such Vassberg et al.,[66] Walter et al.,[76] and Acheson and Balakrishna[79] should be directly involved in order to make the most appropriate validation comparisons to simulation results.



(a) CDF resulting from freestream uncertainty.

(b) Grid convergence history.

(c) Epistemic interval from S-A uncertainty.

(d) Epistemic interval from grid convergence.

Figure 6. CFD Drag Prediction: Uncertainty propagation results for C_D (shown in counts).

According to Walter et al.,[76] the NTF uncertainty model at the time of the High Speed CRM experiments required a better description of systematic uncertainties so that data could be properly compared to data from other test facilities. Furthermore, there appears to be a systematic bias between the NTF and Ames 11-foot results. With these observations in mind, the NTF and Ames 11-foot C_D repeatability data are treated as separate uncertain experimental results. The experimental data points from each facility (three from NTF, four from Ames 11-foot) are sorted in ascending order and used to contruct an empirical CDF for C_D. Since Walter et al. indicate that the repeatability figures do not include systematic uncertainty, the individual data points are interpreted to be samples from a distribution with no epistemic uncertainty. Note that the difference between the maximum and minimum C_D at the NTF and Ames 11-foot are 1.9817 and 1.7419 counts, respectively; these are well within the repeatability figures reported for the test facilities at which each data set was collected. Once the empirical experimental CDFs have been contructed, they may then be used to compute validation metrics for the CFD drag prediction. Plots showing the experimental CDFs and the area validation metrics[1,24] for each facility are included in Figure 8. In this example, the area validation metric for the NTF data has a value of 0.6146 drag counts and the area validation metric for the Ames 11-foot data has a value of 9.9866 counts.

An alternative to the area validation metric is to use the Real Space (RS) model validation approach,[21,50] which focuses on comparing percentiles of experimental and predicted behaviors of stochastic systems possessing inherent random variability (aleatory uncertainty). RS also treats deterministic systems, as a reduction of the more general stochastic case. Comparison of relevant percentiles is especially pertinent for models to be used in the analysis of performance and safety margins. These types of validation problems also usually involve substantial epistemic uncertainty due to lack of knowledge. The RS methodology handles these and many other categories of uncertainty itemized in the leftmost two boxes in the bottom row of Figure 3.

Results of the RS method applied to the CFD drag prediction problem are shown in Figure 9. The uncertainty bars for the 2.5 and 97.5 percentiles of the model-predicted drag coefficient in the figure are formed from the CFD prediction results in Figure 7. These are used for expediency in this paper, although the RS method would segregate, propagate, and represent various uncertainty categories somewhat differently (but with effectively the same results as in Figure 7 for this particular problem). Although the uncertainty bars for the predicted 2.5 and 97.5 percentiles of drag overlap graphically, it must be kept in mind that

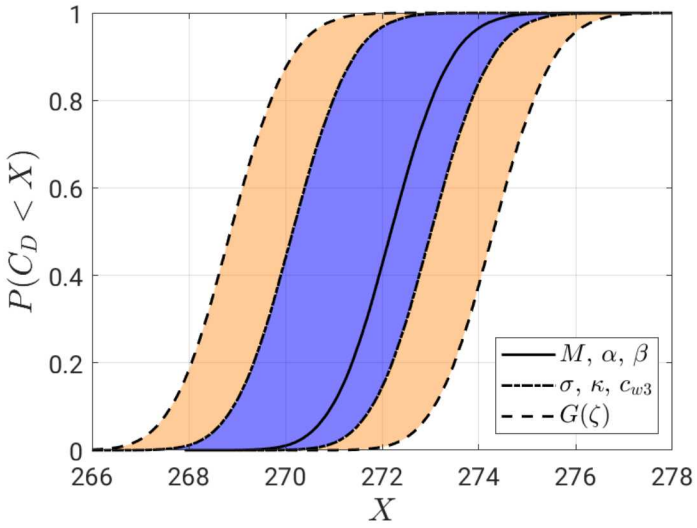


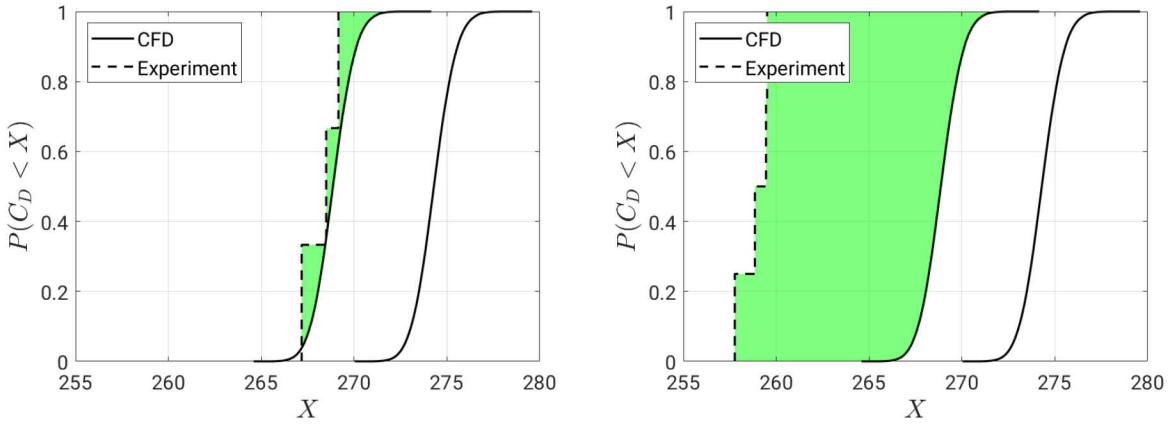
Figure 7. CFD Drag Prediction: Aggregated p-box for C_D resulting from combination of all sources of uncertainty.

these uncertainties are perfectly correlated; if the epistemic uncertainties related to numerical/discretization effects and turbulence modeling parameters could be eliminated, then point values for the predicted 2.5 and 97.5 percentiles would result at some equal percentage of the way between the lower and upper bounds of each percentile's interval. Thus, despite the graphical appearance, it is not the case that the predicted 2.5 percentile of C_D could, for example, lie above the predicted 97.5 percentile.

For the experimental results, ∼90% confident[‡] statistical tolerance intervals (TIs) for the central 95% of behavior between the 2.5 and 97.5 percentiles of an asymptotically large population of C_D test results are inferred for the AMES data points and separately for the NTF data points. The TI constructed from the four AMES data points has endpoints [254.9, 262.9]. These endpoints correspond to 90% confidence on the lowest value of the 2.5 percentile and 90% confidence on the highest value of the 97.5 percentile, respectively. This range is about 81% larger than the AMES facility's reported repeatability uncertainty of ± 2.2 drag counts; the 81% larger range is not unreasonable given the very few data points (only four) it was constructed from, and the fact that TIs from very few samples are necessarily conservatively broad in order that they contain the true central 95% of behavior with a high ∼90% confidence. The three NTF data points give a 95/90 TI of [261.3, 275.3], which is about 100% larger than the range of ± 3.4 drag counts stated repeatability of that facility.

For the 2.5 percentile of C_D, the inferred ∼90% confident lower bounds on the experimental results from both facilities (bottoms of their TIs) are far below the model predicted uncertainty range (left half of Fig 9). Thus, the model prediction uncertainty does not likely contain the lower-end responses (as represented by the 2.5 percentile) of the asymptotically large populations of facility test results. Nor do the model prediction results bound this lower-end of the population from below. The TI 97.5 percentile results at right in Figure 9 can also be used as 100% confident *upper* bounds on the 2.5 percentiles of the experimental populations. Then the model predictions are shown to bound the upper-most possible 2.5 percentile value for C_D in the AMES facility, but not for the NTF facility.

[‡]This cited confidence level is a nominal value that is only accurate if the distribution being sampled is a Normal distribution. We do not know the distribution shape from which the AMES or NTF data samples come. Although derived for Normal populations, 95/90 TIs will span the central 95% ranges of many other sparsely sampled PDF types with reasonably/usefully high odds or confidence. For instance, 89% of 144 PDFs (including highly skewed and multi-modal highly non-Normal distributions) studied in Refs. 80 and 81 had empirical confidence levels of 75% or greater with 95/90 TIs and $N = 4$ random samples. From studies in Ref. 81 on several representative PDFs it is projected that 90% of the 144 PDFs considered would have confidence levels > 85% with 95/95 TIs and $N = 4$ random samples. Reliability rates of 75% or 85% are often adequate to sufficiently manage risk, especially if conservatism from other sources exists in the analysis or results – such as applied factors of safety, and/or large indicated design, safety, or performance margins from high-quality analysis, and/or when several sources of uncertainty are present where each involves sparse data conservatively treated with the TI method.



(a) NTF, area = 0.6146 counts.

(b) Ames 11-foot, area = 9.9866 counts.

Figure 8. CFD Drag Prediction: Comparison of experimental and CFD p-boxes for C_D, with area validation metric shaded green.

For the 97.5 percentile of drag, the model predicts larger drag than inferred from the AMES facility (see right half of Figure 9). This could be considered an error on the conservative side of things for design and analysis purposes because the model would not under-estimate the higher-end drag coefficient values (according to this facility). Conversely, an over-prediction in C_D may result in a less optimal airplane due to design decisions based on the inaccurate information. Within the model prediction uncertainty range, an over-prediction margin of about 7.9 to 13.4 counts exists in C_D. The validation conclusion is provisionally the same for the case with NTF data, although qualitative differences exist; the NTF TI upper end falls within the model-predicted range (with a small margin of 1.0 drag counts). If the analyst uses the upper end of the model-predicted range to be at the conservative end of the prediction uncertainty in design or safety analysis, then this conservatively bounds (from above) the NTF's conservatively inferred 97.5 percentile of drag.

Given the area validation metric results in Figure 8 and the real-space validation results in Figure 9, what should the CFD practitioner do next? Presumably, the CFD practitioner wishes to improve the accuracy and reduce the uncertainty of the predictions (i.e., reduce the area validation metric or the differences between experimental and CFD 2.5 and 97.5 percentiles). One option is to use a finer grid; from Figure 6b, it is apparent that continuing to refine the grid will reduce C_D further, trending it closer towards the experimental data in Figure 8 or the 2.5 percentiles of the experimental TIs in Figure 9. Output-based grid adaptation may also be employed to reduce the grid convergence error. Another option may be to perform calibration of the S-A turbulence model coefficients; perhaps a set of coefficients exists within their physically realistic bounds which yields lower, more accurate predictions of C_D for transonic commercial aircraft geometries. Sobol indices[36] are easily computed from the NIPC response surfaces generated for the freestream and S-A uncertainty propagations, and these may be used to guide further refinement of the CFD model. If updating or improving the model is not possible due to time or resource constraints of a project, then expert judgments, analysis conservatisms, safety factors, etc. may be employed to argue the credibility of model results for their intended use. Whatever route is taken, the inevitable issue of extrapolation will be involved to go forward, and it will be necessary to factor in the considerations above into methods for extrapolation with UQ (as discussed near the end of Section II).

III.B. Uncertainty in Wind Tunnel Wall Corrections

An important consideration in transonic wind tunnel testing is the wall interference correction, which accounts for the presence of the test section boundary. For NASA Langley Research Center's National Transonic Facility (NTF), one of the facilities in which the CRM has undergone testing, the wall interference correction is calculated using the Transonic Wall Interference Correction System (TWICS).[82] This system uses measurements from a 360 wall-mounted pressure sensors placed throughout the ceiling, floor, and sides of the tunnel

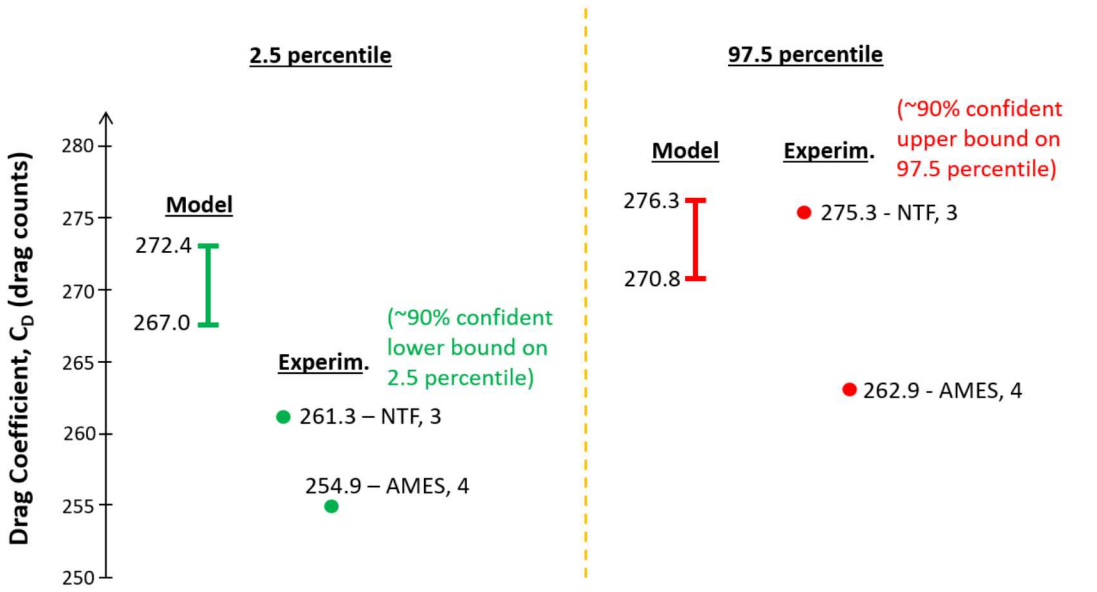


Figure 9. Real Space model validation comparisons of model predicted and experimentally inferred 2.5 and 97.5 percentiles of C_D.

to estimate the blockage interference factor ϵ and make corrections to the angle of attack, sideslip angle, and aerodynamic forces and moments.[82] Although this system has been studied and validated extensively, it assumes that there is no uncertainty in the pressure measurements used in the calculations.[82–86]

Many factors can affect the accuracy of wind tunnel wall pressure measurements including instrument accuracy, surface and orifice quality, leakage in the test section, and the operating dynamic pressure.[83] Uncertainty in these measurements leads to uncertainty in the wall interference correction, which then leads to uncertainty in the aerodynamic forces and moments.[87] For the pressure sensors located in the NTF, the error in accuracy is estimated to be within 0.0025 psi, or 0.1% of the sensor range.[§]

To study the effect of uncertainty in the wall pressure sensors on the wall interference correction, the error in each of the 360 pressure sensors, δ_i is treated as an independent uniformly distributed variable given by

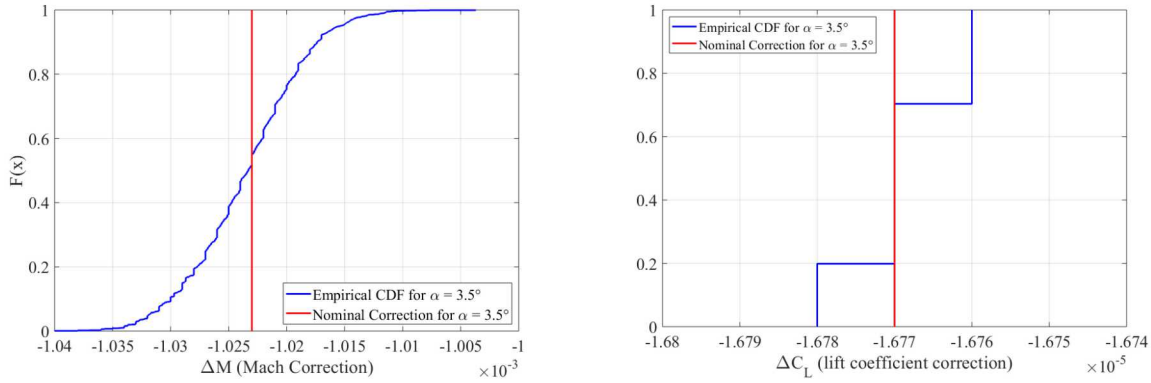
$$\delta_i = \mathcal{U}(-0.0025, 0.0025) \text{ psi} \tag{8}$$

Using data from the NTF Test 197 Runs 92, 97, and 99 and the TWICS software, a Monte Carlo simulation was conducted with 500 realizations of the wall pressure sensor errors for a range of angle of attack. This allows for quantification of the uncertainty in the wall interference correction for Mach number and angle of attack, which is then be used to calculate the uncertainty in the aerodynamic forces and moments.

The variation in the Mach number correction term, shown in Figure 10a is relatively small, with 95% confidence bounds of [-0.001033,-0.001015]. When the corrections in Mach number and angle of attack are used to calculate the corrections for lift coefficient, C_L, drag coefficient, C_D, and pitching moment coefficient, C_M, the effect is minimal. The cumulative distribution function for C_L, Figure 10b, shows that the lift coefficient correction is not significantly influenced by uncertainty in the wall pressure measurement. Although there were 500 realizations of the uncertainty in the wall pressure sensors, there was no change to the lift coefficient correction in approximately half of these cases and only minor effects (to an output file printing accuracy of 10^{-8}) in the remaining cases, resulting in a step-like cumulative distribution function.

Although the lift coefficient correction term is known to vary as a function of angle of attack, the effect of wall pressure measurement uncertainty does not cause significant differences in the standard deviation as a function of angle of attack. Figure 11 shows a comparison of the standard deviation of the lift coefficient correction term calculated as a function of angle of attack to the standard deviation of the correction term calculated across all angles of attack. This standard deviation is especially small when compared to the value of the correction term, which is on the order of 10^{-5}. If accurate estimations of the uncertainty in the

[§]Private communication between Matthew Bailey and Casey Denham September 2019.



(a) Mach number correction at $\alpha = 3.5\,\text{deg}$ (b) Lift coefficient correction at $\alpha = 3.5\,\text{deg}$

Figure 10. Cumulative distribution functions for wind tunnel corrections.

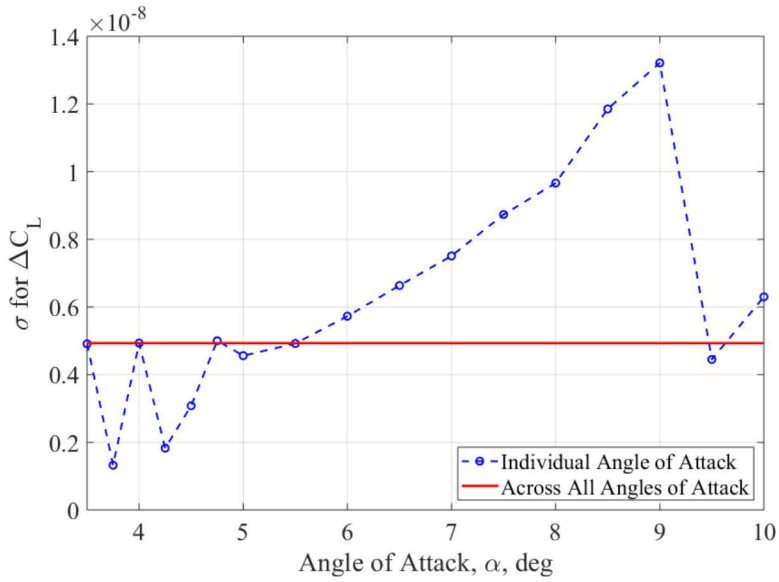


Figure 11. Standard deviation of lift coefficient correction term, showing weak dependence on α.

force and moment coefficient terms at each angle of attack were required, additional simulation runs would be necessary to ensure that results are properly converged. The sharp oscillations in the standard deviation at low angles of attack are due to the output file accuracy and the number of uncertainty realizations.

Overall the wall interference correction is not sensitive to the estimated uncertainty in the wall pressure measurement sensors, with only minor uncertainty in the correction terms. This indicates that the number and accuracy of the sensors is sufficient to provide consistent results in the wall interference corrections. Because the correction terms are known to vary as a function of Mach number, additional analysis would be required to determine if the uncertainty in the correction terms is also dependent on Mach number. This analysis demonstrates how uncertainty within a measurement can propagate through calculations and generate results with different uncertainty distributions than the original uncertainty. It is an example of how UQ can be used to confirm whether or not the accuracy of instrumentation is sufficient for intended use.

IV. Further Hypothetical Examples

In this section, two more example problems are described which are hypothetical in nature and do not have results fully calculated. The purpose of these examples is to demonstrate how the methods and considerations discussed in Section II may be applied to more complex types of analysis than those with computed results in Section III. As before, both problems involve the NASA Common Research Model (CRM).[64] The first hypothetical example presents the first steps towards estimating the uncertainty in the longitudinal stability and control characteristics of an aircraft. The second involes a FDS of roll maneuver performance.

IV.A. Uncertainty in Longitudinal Static Stability Characteristics

Some of the most fundamental characteristics of any aircraft are its longitudinal stability and control (S&C) characteristics. Longitudinal S&C behavior determines a number of critical aircraft criteria; for example, the extents of the weight and center of gravity (c.g.) envelope, the maneuvering characteristics of the aircraft in a given configuration, the loads on the aircraft, takeoff performance, etc. Understanding the uncertainty in a potential aircraft's S&C characteristics can be helpful in assessing the potential programmatic risk associated with that design, whether those risks be to internal, customer, or regulatory requirements.

The example in this section will explore how accounting for uncertainty in the wind tunnel wall corrections, the determination of which was outlined in Section III.B, can be used to determine the impact of those uncertainties in the longitudinal static stability characteristics of an aircraft. Similarly to the TWICS wall correction uncertainty analysis, this example will use data from the NASA CRM NTF 197 wind tunnel test. Information on that test can be found in Ref. 88. In this example we examine three primary longitudinal static stability characteristics and how they are influenced by the TWICS uncertainties: horizontal stabilizer angle and angle of attack required to trim the aircraft at a specific C_L, the most aft c.g. which can be trimmed using -2° to +2° of stabilizer travel, and the determination of the aircraft stick-fixed neutral point. These are basic longitudinal static stability characteristics which will be examined by performing a bounding analysis using the average standard deviations of the angle of attack, pitching moment coefficient, and lift coefficient wall corrections determined from the tail = 0° CRM wind tunnel run at $M = 0.85$, $Re = 5 \times 10^6$. It is the hope of the authors that this simple example can provide engineers who may not specialize in UQ methods with a foundation upon which to begin analyzing uncertainties and their effects within their own analyses. Only the uncertainties due to the wall correction method are studied in this simple example, however a discussion of additional sources of uncertainty which may influence longitudinal static stability characteristics is presented at the conclusion of this section.

IV.A.1. Trim Orientation of Aircraft

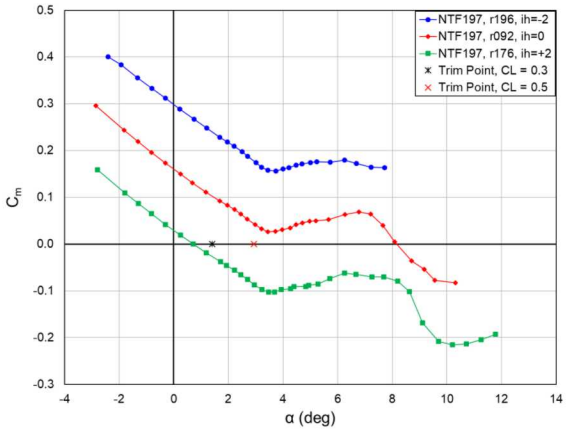
This example will outline the determination of the stabilizer angle (elevator faired, as the CRM does not have an elevator surface defined or tested) and angle of attack for an aircraft at a given lift coefficient (which can also be thought of as an aircraft at a given weight and load factor). This is a classical S&C problem which is the starting point for determining a number of aircraft flight characteristics; it determines whether a given loading can be trimmed (and thus flown), it is the starting point for maneuvering characteristic analysis in a given loading, and it is a starting point in the determination of wing and balancing tail loads in static loads analyses. The trim characteristics of an aircraft can be solved with a textbook style simplification of the longitudinal equations of motion.[89] One can arrive at the following simplified longitudinal equations:

$$C_L = C_{L_{(i_h=0)}} + \frac{dC_L}{di_h} i_h \tag{9}$$

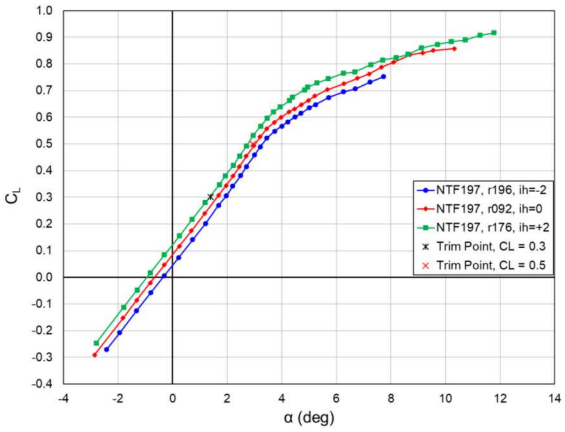
$$0 = C_{m_{(i_h=0)}} + \frac{dC_m}{di_h} i_h \tag{10}$$

which assume that the aircraft is in steady, level, unaccelerated flight. These equations also ignore the force in the x-axis, as the CRM does not have engines, making a determination of the thrust required to balance drag inconsequential.

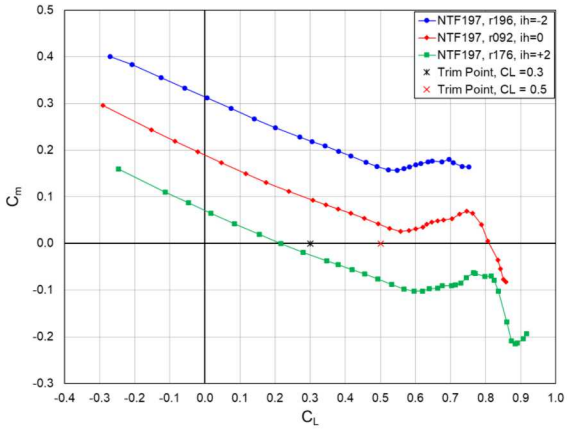
Eqn's (9) and (10) are thus a function of two unknowns: the horizontal stabilizer angle, i_h, and the angle of attack, α (Note that C_L, $C_{L_{i_h=0}}$, $C_{L_{i_h}}$, $C_{m_{i_h=0}}$, and $C_{m_{i_h}}$ are all functions of angle of attack. In many textbook examples the C_L and C_m terms are linearized with respect to α). In practice, it may be necessary



(a) C_m versus α.



(b) C_L versus α.



(c) C_m versus C_L.

Figure 12. CRM pitch and lift behavior at $i_h = -2°$, $0°$, and $2°$ ($M = 0.85$, $Re = 5 \times 10^6$), with trim points calculated for a c.g. at the model reference point and C_L of 0.30 and 0.50.

to include drag terms in order to ensure that the aircraft's engines can produce enough thrust at a given flight condition to balance drag; however, the CRM model is not a real aircraft, and thus does not have engines, necessitating a thrust-off analysis in this example. Note that the data used in this analysis does not include the effects of the nacelles on the aircraft pitch and lift behavior.

The pitching moment and lift coefficient curves for three difference tail incidence angles, $i_h = -2°$ (NTF197 Run 196), $i_h = 0°$ (NTF197 Run 92), and $i_h = 2°$ (NTF197 Run 176) are shown in Figure 12. The pitching moment curves shown are for a c.g. at the model reference point, which for the 2.7% scale CRM model tested in the NTF is 35.8 inches back from the nose and 2.04 inches below the fuselage centerline.[88] All of the values shown and used in this analysis are corrected using the TWICS method described in Section III.B. Two trim points are shown, calculated using the two simplified equations previously shown. For the wind tunnel model at the model reference point, a stabilizer angle of 1.54° and an angle of attack of 1.41° are needed to trim the aircraft at a C_L of 0.30, and a stabilizer angle of 0.66° and an angle of attack of 2.93° are needed to trim the aircraft at a C_L of 0.50 (which is the CRM design lift coefficient).[69]

Next, the uncertainty values determined in the Section III.B are used to determine whether the uncertainty in the wall correction terms have an appreciable effect on the trim orientation of the aircraft. The following standard deviations are employed:

$$\sigma_\alpha = 3.03 \times 10^{-6} \tag{11}$$

$$\sigma_{C_L} = 4.93 \times 10^{-9} \tag{12}$$

$$\sigma_{C_m} = 4.70 \times 10^{-7} \tag{13}$$

It is worth noting that these are average values of the standard deviation calculated for the entire range of angle of attacks in the NTF Run 92 α sweep. However, the largest individual standard deviation was 2.68 times larger than the average used (in C_L), which is not significant given the small magnitude of the standard deviations. It is also worth noting that the standard deviation values used are for Run 92, with the CRM in the tail=0° configuration only, and these values are applied to the tail=-2° and tail=2° configurations as well. The use of consistent uncertainty characterization (i.e., the same values for σ_α, σ_{C_L}, and σ_{C_m}) across all tail incidences is a reasonable assumption since the tail configuration likely does not affect the scatter of corrections due to measurement uncertainty in the taps; however the nominal value of the corrections (before applying uncertainty) should be applied on a per-tail-incidence basis, as the nominal corrections will be configuration-dependent.

Returning to Figure 12, it is clear that a conservative bounding analysis (sometimes referred to as a "worst-case" analysis) will occur when positive C_L and C_m uncertainties are superimposed on the data, as this will reduce the available range of C_L that can be trimmed at the model reference point. However, the values of the uncertainties in the wall corrections are so small that even adding $+6\sigma$ and -6σ to the baseline curves does not affect the solution – the largest change seen is 0.001° in the stabilizer angle.

IV.A.2. Most Aft c.g. that can be Trimmed

In an analysis similar to the previous trim orientation example, the most aft c.g. which can be trimmed by the -2° to +2° stabilizer travel limits tested in the NTF wind tunnel can be determined. For the purposes of this example we will define the most aft c.g. limit as the point where there are no longer any valid trim points that can be achieved using -2° to +2° stabilizer travel. Only the x-coordinate of the c.g. will be examined. For the CRM model with the same assumptions as those in the previous example, the most aft c.g. that can be trimmed with +2° of horizontal stabilizer is approximately 17% c_{ref} aft of the model reference point. This trim point is at a stabilizer angle of +2° and an angle of attack of approximately 3.22°, corresponding to a C_L of 0.57. As with the previous example, the $+6\sigma$ and -6σ uncertainty bands are so small that they do not influence the calculation of the most aft c.g. which can be trimmed with +2° of stabilizer.

IV.A.3. Determination of Stick Fixed Neutral Point

For an aircraft to possess longitudinal static stability, the change in pitching moment of the aircraft with regard to a change in angle of attack must be negative:[90]

$$\frac{dC_m}{d\alpha} < 0 \tag{14}$$

The stick-fixed neutral point is defined as the point where

$$\frac{dC_m}{d\alpha} = 0 \tag{15}$$

which is the aft most limit of the c.g. for the aircraft to possess longitudinal static stability.

For the CRM model analyzed using NTF runs 92, 176, and 196 in this example, the stick-fixed neutral point lies approximately 32% c_{ref} aft of the model reference point. This point was determined using the average of the $dC_m/d\alpha$ slopes for all three tail incidences, with the slope calculated between α of around -1.3° to +1.7°. Unlike the previous two examples, which examined the trim capability of the aircraft, a conservative analysis regarding the longitudinal static stability and location of the stick-fixed neutral point would be one in which the uncertainties result in a shallower $dC_m/d\alpha$ slope. That being said, as with the previous two examples in this section, an uncertainty band of $+6\sigma$ and -6σ did not produce any appreciable change in the location of the stick-fixed neutral point.

IV.A.4. Other Sources of Uncertainty in Longitudinal Static Stability

In this example a single source of uncertainty, the uncertainty in the wall corrections factors applied to the raw wind tunnel data, was applied to three longitudinal static stability criteria in a simple bounding analysis. The magnitude of the uncertainties, however, was so small that even at a $\pm 6\sigma$ interval there was no impact on the longitudinal static stability characteristics analyzed. In fact, it is much more likely that the modeling

assumptions (i.e., interpolating the tunnel data onto a common α basis, or linearizing the $C_{L_{i_h}}$ and $C_{m_{i_h}}$ terms) impact the results more than the wall correction terms could, although even this impact is likely negligible. The calculation of this impact is not the purpose of this example.

In light of that, it is important to consider that this example is a simple, hypothetical example of characterizing the impact of uncertainty on aircraft-level metrics. In practice there are a number of additional uncertainties that may show up in an aerodynamic model, including, but not limited to:

- Instrumentation error and uncertainty (tunnel balance, repeatability of multiple runs)
- Differences in anticipated and actual flight wing shape in the tunnel
- Wall corrections
- Tare, interference, and mounting corrections
- Buildup terms
- Modeling simplifications, such as interpolating data to a common basis so that buildup terms can be superimposed

In particular, buildup terms may introduce significant uncertainty. Each incremental effect (for example: gear, nacelle, or fairing contributions) added to a model will carry with it its own associated uncertainty, and the aggregation and propagation of these uncertainties through the model may become significant. The uncertainties in each buildup term may be due to measurement uncertainty (i.e., each increment determined from a tunnel run carries with it an uncertainty due to the measurement uncertainty of the runs it was calculated from) or uncertainties due to the analysis used to determine the incremental effect (such as CFD, which was explored at depth in Section III.A). Additionally, terms required for a more complete S&C analysis, such as thrust for power-on analysis, will have their own unique uncertainties determined by the particular model used to determine those effects. The four steps of uncertainty quantification outlined in the beginning of this paper should be exercised for each uncertainty term that is to be included in the analysis.

The conclusion that the the uncertainty in the pressure measurements used for the wall corrections do not significantly influence the simple longitudinal static stability characteristics examined in this example should not be taken to mean that there is not significance in accounting for all uncertainties in a model, but rather that the wall correction method appears to be robust against the uncertainty in the pressure tap measurements that it is based on. An analysis of other uncertainty terms may yield different results, for example, the standard deviation in the lift coefficient due to data repeatability reported by Rivers for the NTF 197 and Ames 216 tests is on the order of 10^{-3}, which, while still small, is significantly larger than the standard deviation in the wall correction terms.[88] Other sources of error and uncertainty have presented themselves in the CRM model as well, such as the pitching moment contributions due to the mounting sting and wing twist, and will have a noticeable influence the longitudinal static stability characteristics of the aircraft.[91,92]

The example in the following section outlines how uncertainties may be characterized and propagated through a dynamic simulation of a roll maneuver, and many of the processes outlined in that example may be appropriate for more complex longitudinal static stability and S&C maneuver analyses.

IV.B. Uncertainty in FDS of Roll Maneuver Performance

In this example, methods of characterizing the uncertainty in the time required for an airplane to execute a specified roll maneuver (t_r) are explored. It is assumed that a flight dynamics simulation will be used to assess the roll rate capability of an aircraft, and for the purposes of this hypothetical example, the problem is simplified by assuming zero sideslip. The primary input parameters of interest for this analysis are:

- Rolling moment coefficient due to lateral control deflection ($C_{M,LCD}$)
- Aircraft moment of inertia about the roll axis (I_{RA})
- Roll damping coefficient (C_{RD})

Table 3. Descriptions of terms used in FDS example approaches.

Term	Description
Probabilistic parameters	Parameters for which the uncertainty may be described using a probability distribution
Interval parameters	Parameters for which the uncertainty is described using an interval or range of values
Extreme output case	A set of input parameter values for a simulation or test that yields a minimum or maximum value of an output parameter. Intervals are bounded by the minimum and maximum extreme output cases.
Input parameter interactions	Input parameters are considered to interact when changes in one input parameter alter the sensitivity of an output to changes in another input parameter. If inputs act independently on the output then there are no input parameter interactions.
Surrogate model	A mathematical model that predicts the responses of another system. Often referred to as response surfaces, reduced order models, statistical models, emulators, or machine learning models.

For this example, it is assumed that the uncertainties in the input parameters are all epistemic in nature and that the goal is to describe an epistemic interval for the output. Further, it is assumed that $C_{M,LCD}$ and I_{RA} may be described either as intervals or probability distributions based on prior data and that C_{RD} may be described as an interval based on some subjectively determined level of confidence.

The context of this example is predicting the behavior of an airplane that has not yet flown, based on various analysis methods including use of empirical data from past designs. Due to the simple nature of this example it is better seen as part of routine preparation for flight testing rather than as an analysis to be used for certification, though the uncertainty quantification analysis could be similar for either case.

This example describes six general approaches for performing the analysis in order to illustrate differences in prerequisite assumptions, simulation resource requirements, and results. Additional uncertainty propagation methods for different scenarios may be found in literature.[34,93] Descriptions for some of the terms used in these approaches can be found in Table 3.

1. Traditional interval-only analysis assuming known extreme cases.
2. Interval-only analysis assuming unknown extreme cases.
3. Uncertainty propagation across mixed probabilistic and interval input parameters assuming known extreme output cases for the interval inputs.
4. Uncertainty propagation across mixed probabilistic and interval parameters assuming unknown extreme output cases for the interval inputs and no input parameter interactions.
5. Uncertainty propagation across mixed probabilistic and interval parameters assuming unknown extreme output cases for the interval inputs and unknown parameter interactions.
6. Surrogate model-based uncertainty propagation across mixed probabilistic and interval parameters assuming unknown extreme output cases for the interval inputs and unknown parameter interactions.

The following sections walk through the approaches listed above and provide graphs visualizing the inputs and output result. The first two approaches treat interval only scenarios while the later four treat mixed probabilistic and interval scenarios.

IV.B.1. Approach 1

A traditional uncertainty analysis assuming known extreme cases would consist of two simulations estimating the upper and lower bounds of the output (t_r) based on known extreme combinations of the inputs (I_{RA}, $C_{M,LCD}$, C_{RD}). This is typically done when it is known that the input-output relationships are monotonic,

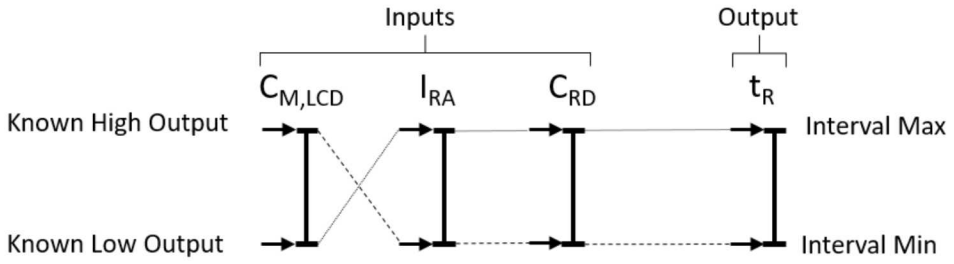


Figure 13. Schematic of FDS example problem Approach 1.

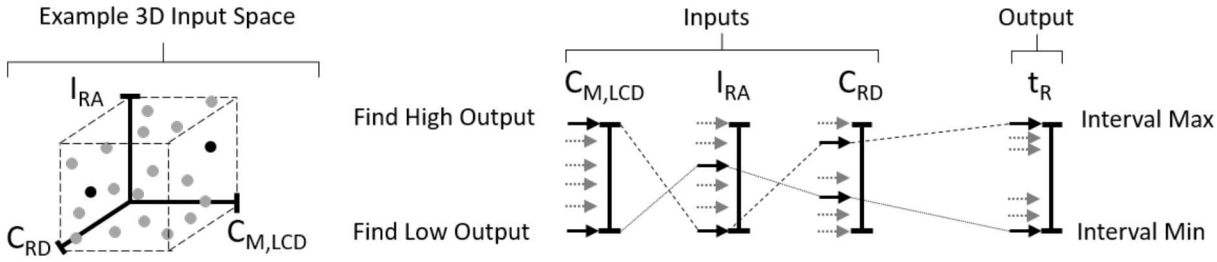


Figure 14. Schematic of FDS example problem Approach 2.

that there are no complicating input parameter interactions, and that the input uncertainties are epistemic intervals (non-probabilistic). The results of this method are an expected interval range of the output with no information about confidence bounds. While this method uses few simulations, the reliability of its results requires a high degree of *a priori* confidence in the behavior of the system over the uncertain ranges. Furthermore, when compared to more advanced UQ treatments, worst-case interval UQ analysis can lead to unrealistically large uncertainty ranges and unnecessarily severe bounds on output quantities when more than two or three uncertain factors are involved. A schematic of Approach 1 is shown in Figure 13.

IV.B.2. Approach 2

In instances where the inputs are considered epistemic intervals and the extreme cases are unknown, a number of options exist for determining the output interval. A common method is to apply upper and lower bounds to each of the three input parameters and analyze all combinations as in a factorial design. Alternatively, the analyst could determine the sensitivity of the output quantity of interest to each of the input variables separately, and based on those 3 initial calculations, determine the two combinations of input parameter extremes that are expected to give the high and low output extremes. This results in only 5 calculations instead of the 8 required to test all combinations. The computational savings of this approach grow as the number of input parameters increases. Approach 2 has potential pitfalls; to be used reliably, this method relies on monotonic behavior and lack of interactions and reduces to Approach 1 given the availability of slope/trend information in each input variable direction. If any of these conditions do not apply, then it becomes necessary to identify the minimum and maximum outputs over the input interval space. In these cases, design of experiments (i.e. multi-level factorial, central composite, LHD), surrogate modeling, or optimization techniques must be employed – see Ref. 94 for examples. A schematic of Approach 2 is shown in Figure 14.

In some situations, it is possible to apply probability distributions to some of the inputs and not others such that the uncertainty propagation must be performed across mixed probabilistic and interval input parameters. For the remaining four approaches, it is assumed that probability distributions for I_{RA} and $C_{M,LCD}$ can be meaningfully approximated based on a combination of expert knowledge, data from previous aircraft programs and analytical tools. The C_{RD} is still assumed to be an interval uncertainty.

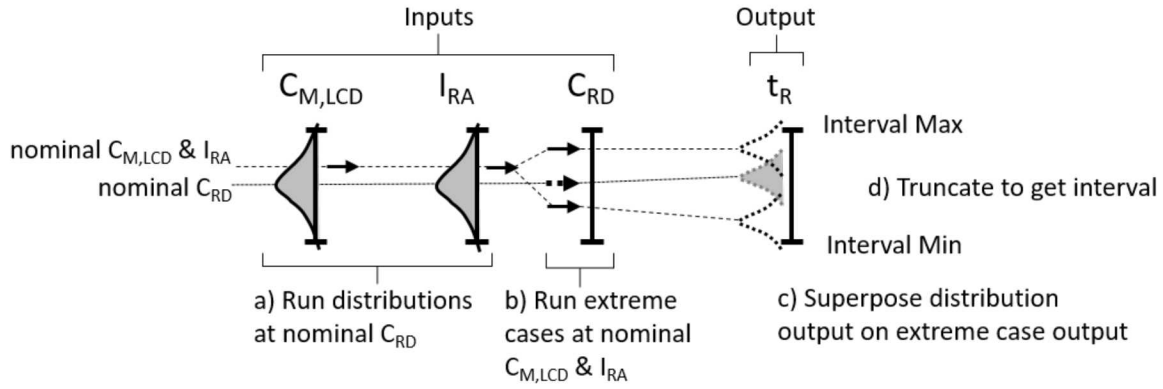


Figure 15. Schematic of FDS example problem Approach 3.

IV.B.3. Approach 3

The simplest scenario for a mixed distribution/interval problem is when the extreme output cases for the interval inputs are known and there are no interactions between parameters. For this example, the behavior of t_r with respect to C_{RD} would be known such that extreme case analysis is viable for C_{RD}. Given these assumptions, the most computationally efficient method is to decouple the propagation:

a) Propagate the probabilistic uncertainties given some nominal value of the interval uncertainty using direct Monte Carlo sampling or another technique (see Ref. 34 for examples) depending on the resources and information available.

b) Propagate the interval uncertainty given some nominal value of the probabilistic uncertainties by running the known extreme cases for the interval parameters.

c) Superpose the probabilistic uncertainty distribution on the results of the interval propagation.

d) Truncate the probabilistic component at reasonable PDF/CDF tail values to form an equivalent interval uncertainty for the output. This is recommended as the inputs are all epistemic in this example problem, yielding an epistemic output.

A schematic of Approach 3 is shown in Figure 15.

IV.B.4. Approach 4

The fourth approach relaxes the assumption in Approach 3 regarding knowledge of the extreme cases of the interval parameters but still assumes no parameter interaction. This is the equivalent of the change in assumptions between Approaches 1 and 2 and results in similar changes in the analysis. The change in assumptions alters the four-step process outlined for Approach 3 by introducing the need to identify extreme cases for the interval uncertainties using the same methods mentioned in Approach 2. The interval analysis would be done at some nominal values of the probabilistic uncertainties prior to superposing the probabilistic uncertainty distributions. A schematic of Approach 4 is shown in Figure 16.

IV.B.5. Approach 5

The fifth approach removes the assumptions that there is no parameter interaction. Thus, no information is known about any of the effects of the input parameters or their interactions with each other. Propagating mixed probabilistic/interval uncertainties must be performed using a dual-loop analysis either with purely Monte Carlo methods, an optimization, or a mixture depending on whether the desired outputs are pure extreme case intervals or include information about the confidence.[34,94] These methods all involve running a sampling of the probabilistic parameters at a number of different combinations of the interval parameters in order to capture interactions that impact the output. A schematic of Approach 5 is shown in Figure 17.

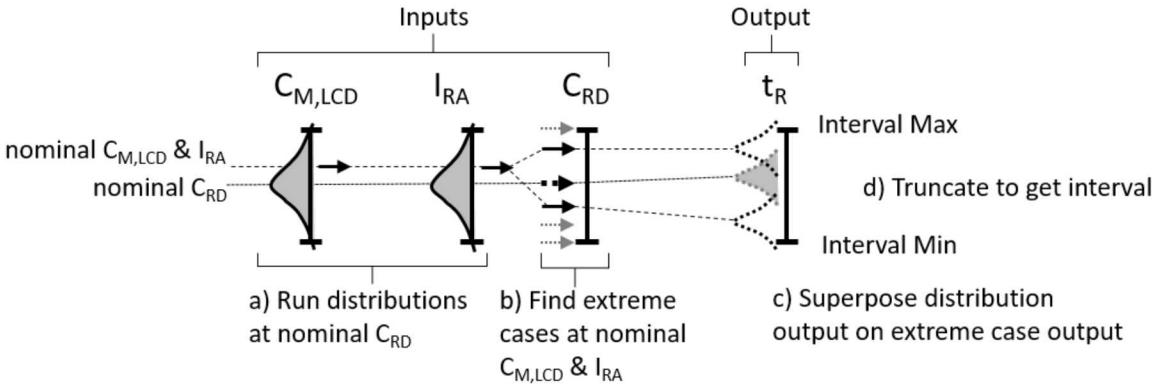


Figure 16. Schematic of FDS example problem Approach 4.

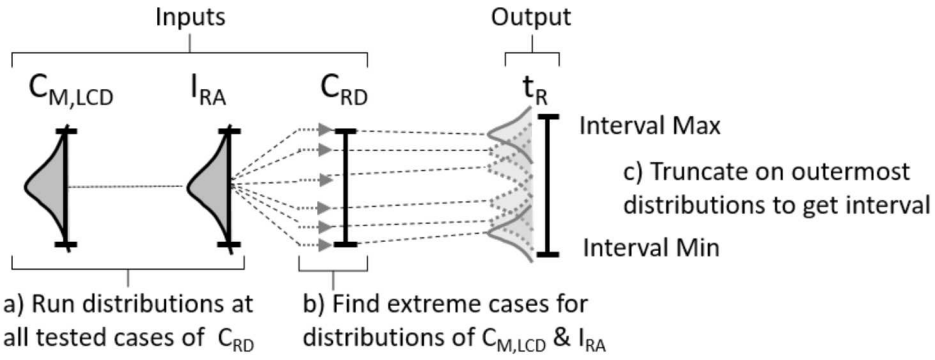


Figure 17. Schematic of FDS example problem Approach 5.

IV.B.6. Approach 6

Surrogate model-based methods (aka response surface, meta model, emulation, machine learning) take a smaller number of samples of a physics-based simulation or test data set and use those to build a predictive model for use in analytics. As knowledge of the system decreases and fewer assumptions are made, more sampling intensive approaches must be used to propagate the uncertainties. For the example outlined here, the simulation may be computationally light enough to run these samples directly for some of the propagation methods discussed but could still take days to run the dual-loop processes necessary for Approach 5. The situation becomes worse as the cost of the individual simulations rises. A typical surrogate modeling process requires five steps:

a) Build and run a sampling design of experiment or collect existing data from the simulation.

b) Fit a surrogate model to the simulation data.

c) Check the accuracy of the surrogate model relative to the underlying simulation.

d) If the model has achieved the desired level of accuracy, use for analysis, otherwise change model fitting, model type, or add additional data using DOEs or adaptive sampling techniques.

e) Run the uncertainty propagation cases, optimizations, etc. on the surrogate model.

Additional details on model types, error metrics, and use cases are available in the literature.[95] A schematic of Approach 6 is shown in Figure 18.

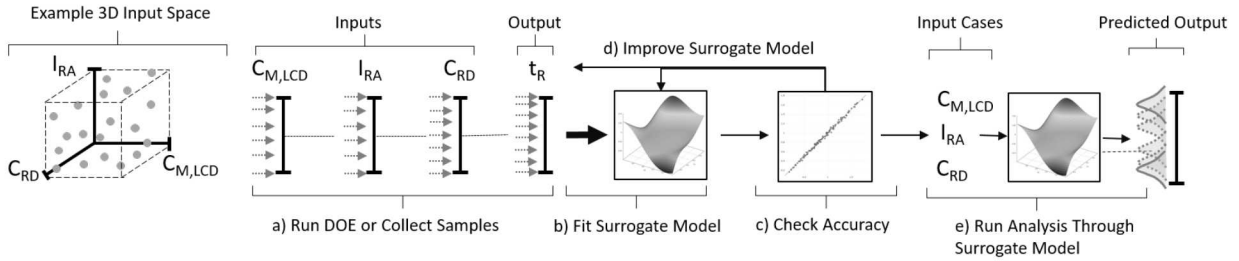


Figure 18. Schematic of FDS example problem Approach 6.

V. Conclusions

This paper was written to serve as a reference on uncertainty quantification for the AIAA Certification by Analysis Community of Interest Recommended Practices Document; specifically, it is meant to aid in the understanding of two of the CbA CoI recommended tasks: (1) validation of models and (2) justification of analysis adequacy in recognition of potential modeling error and/or uncertainty. An overview of uncertainty quantifcation was given, which includes four major components: the identification of sources of uncertainty, the characterization of their statistical form, the propagation and aggregation of uncertainty through models, and finally the analysis of uncertain results. Two realistic engineering problems with results computed were presented, followed by two hypothetical examples for more complex analyses. The outcome of these example problems and UQ in general is the enablement of modelers and analysts to make informed statements about the uncertainty and associated degree of credibility in analysis-based predictions.

Acknowledgements

The authors would like to thank Byram Bays-Muchmore (The Boeing Company) for his leadership in the AIAA Certification by Analysis Community of Interest and his commendable advisorship, support, and review of the content in this paper. Many other members of the CbA CoI have also provided excellent food for thought regarding the presentation of the four steps of uncertainty. Thanks also to Melissa Rivers (NASA LaRC) for her contributions to several example problems, and to Matthew Bailey (NASA LaRC) for his contributions to the wind tunnel uncertainty analysis. Additionally, thank you to Steve Klausmeyer (Textron Aviation) for his inputs and review of the longitudinal static stability example. Furthermore, the authors thank Dave Shikany (The Boeing Company) and Paul Bolds-Moorehead (The Boeing Company) for their contributions to the flight dynamics simulation example problem. The CFD results presented in Section III.A used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

References

[1]Lee, H. B., Ghia, U., Bayyuk, S., Oberkampf, W. L., Roy, C. J., Benek, J. A., Rumsey, C. R., Powers, J. M., Bush, R. H., and Mani, M., "Development and Use of Engineering Standards for Computational Fluid Dynamics for Complex Aerospace Systems," AIAA Paper 2016-3811, June 2016.

[2]National Aeronautics and Space Administration, Standard for Models and Simulations, NASA-STD-7009, 2008.

[3]American Society of Mechanical Engineers (ASME), *Guide for Verification and Validation in Computational Solid Mechanics*, ASME V&V 10-2006.

[4]European Aviation Safety Agency (EASA). "EASA Proposed Equivalent Safety Finding on CS 25.251(b) - "Vibration / buffeting" - Applicable to Large Aeroplanes category fitted with large radome or antenna fairing on the Fuselage." https://www.easa.europa.eu/document-library/product-certification-consultations/proposed-equivalent-safety-finding-cs-25251b-, accessed November 2019.

[5]Federal Aviation Administration. "FAA Regulatory and Guidance Library," https://rgl.faa.gov/ (Select "Equivalent Levels of Safety" and then search for "25.251(b)"), accessed November 2019.

[6]Romero, V.J., "Uncertainty Quantification and Sensitivity Analysis—Some Fundamental Concepts, Terminology, Definitions, and Relationships," Chapter 5 of Joint Army/Navy/NASA/Air Force (JANNAF) e-book: *Simulation Credibility—Advances in Verification, Validation, and Uncertainty Quantification*, U. Mehta (Ed.), D. Eklund, V. Romero, J. Pearce, N. Keim, document NASA/TP-2016-219422 and JANNAF/GL-2016-0001, Nov. 2016.

[7]Santner, T., B. Williams, and W. Notz, *The Design and Analysis of Computer Experiments,* New York, NY: Springer, 2003.

[8]Romero, V.J., Burkardt J.S., Gunzburger M.D., Peterson J.S., "Comparison of Pure and "Latinized" Centroidal Voronoi Tessellation Against Various Other Statistical Sampling Methods," *Reliability Engineering and System Safety,* 91 (2006) 1266 - 1280.

[9]Helton, J.C., J.D. Johnson, C.J. Sallaberry, C.B. Storlie, "Survey of Sampling-Based Methods for Uncertainty and Sensitivity Analysis," Sandia National Laboratories report SAND 2006-2901, June 2006.

[10]Hosder, S., Walters, R. W., and Balch, M., "Point-Collocation Nonintrusive Polynomial Chaos Method for Stochastic Computational Fluid Dynamics," *AIAA Journal*, Vol. 48, No. 12, December 2010.

[11]Eldred, M. S., "Recent Advances in Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Analysis and Design," AIAA Paper 2009-2274, May 2009.

[12]Gerstner, T., and Griebel, M., "Numerical Integration using Sparse Grids," *Numerical Algorithms*, 18(3-4), pp. 209-323, 1998.

[13]Smolyak, S. A., "Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions," *Soviet Mathematics Doklady*, Vol. 4, pp. 240-243, 1963.

[14]Myers, R.H., and Montgomery, D.C., *Response Surface Methodology*, Wiley & Sons, Inc., 1995.

[15]Jin, R., Chen, W., Simpson, T.W., "Comparative Studies of MetaModeling Techniques under Multiple Modeling Criteria," AIAA Paper 2000-4801, 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 6-8 Sept. 2000, Long Beach, CA.

[16]Wang, L., D. Beeson, G. Wiggs, M. Rayasam, "A Comparison of Meta-modeling Methods Using Practical Industry Requirements," AIAA Paper 2006-1811, 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conf., 1 - 4 May 2006, Newport, Rhode Island.

[17]Viana, F., R. T. Haftka, V. Steffen Jr., S. Butkewitsch, M. F. Leal, "Ensemble of Surrogates: a Framework based on Minimization of the Mean Integrated Square Error," AIAA Paper 2008-1885, 49th IAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials, 7 - 10 April 2008, Schaumburg, IL.

[18]Haldar, A., and S. Mahadevan, *Probability, Reliability, and Statistical Methods in Engineering Design,* John Wiley and Sons, 2000.

[19]Bichon, B. J., M. S. Eldred, L. P. Swiler, S. Mahadevan, and J. M. McFarland. "Efficient global reliability analysis for nonlinear implicit performance functions." *AIAA Journal*, 46(10):2459–2468, 2008.

[20]Schueller, G.I., H.J. Pradlwarter, P.S. Koutsourelakis, "A critical appraisal of reliability estimation procedures for high dimensions," *Probabilistic Engineering Mechanics* 19 (2004) 463 – 474.

[21]Romero, V., F. Dempsey, B. Antoun, "Application of UQ and V&V to Experiments and Simulations of Heated Pipes Pressurized to Failure," Chapter 11 of Joint Army/Navy/NASA/Air Force (JANNAF) e-book: *Simulation Credibility—Advances in Verification, Validation, and Uncertainty Quantification*, U. Mehta (Ed.), D. Eklund, V. Romero, J. Pearce, N. Keim, document NASA/TP-2016-219422 and JANNAF/GL-2016-0001, Nov. 2016.

[22]Romero, V., "Discrete Direct Model Calibration and Propagation Approach addressing Sparse Replicate Tests and Material, Geometric, and Measurement Uncertainties," Soc. Auto. Engrs. 2018 World Congress (WCX18) paper 2018-01-1101 (doi:10.4271/2018-01-1101), April 10-12, Detroit, MI.

[23]Ferson, S. and Ginzburg, L. R., "Different Methods are Needed to Propagate Ignorance and Variability," *Reliability Engineering and System Safety*, Vol. 54, 1996, pp. 133-144.

[24]Roy, C. J. and Oberkampf, W. L., "A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing," *Computer Methods in Applied Mechanics and Engineering*, Vol. 200, Issues 25-28, 2011, pp. 2131-2144.

[25]Hoffman, F. O., and Hammonds, J. S., "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish between Uncertainty due to Lack of Knowledge and Uncertainty due to Variability," *Risk Analysis*, 14(5): 707-712, 1994.

[26]Helton, J.C., "Treatment of Uncertainty in Performance Assessments for Complex Systems," *Risk Analysis*, Vol. 14, No. 4, 1994, pp. 483 – 511.

[27]Williamson, R.C., and T. Downs, (1990), "Probabilistic Arithmetic I: Numerical Methods for Calculating Convolutions and Dependency Bounds," *International Journal of Approximate Reasoning* 4:89-158.

[28]Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. S., and Sentz, K., "Constructing Probability Boxes and Dempster-Shafer Structures," Sandia Report SAND2002-4015, January 2003.

[29]Ferson, S., and Tucker, W. T., "Sensitivity in Risk Analyses with Uncertain Numbers," Sandia Report SAND2006-2801, July 2006.

[30]Romero, V., "Approximate Probability Boxes and Other Shortcuts in a Broad-before-Deep approach to Balanced UQ," Sandia National Laboratories document SAND2015-3605C and also archived in proceedings of ASME 2015 Verification and Validation Symposium, May 13-15, Las Vegas, NV.

[31]Cary, A. W., Schaefer, J. A., Duque, E. P. N., and Lawrence, S. S. "Application of a CFD Uncertainty Quantification Framework for Industrial-Scale Aerodynamic Analysis," AIAA SciTech 2019 Forum Vol. AIAA 2019-1492, San Diego, California, 2019.

[32]Saltelli, A., Chan, K., Scott, E.M. *Sensitivity Analysis,* New York: Wiley; 2000.

[33]Storlie, C.B., L.P. Swiler, J.C. Helton, and C.J. Sallaberry. "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models." *Reliability Engineering and System Safety*, 94 (2009) 1735–176.

[34]Swiler, L.P., V.J. Romero, "A Survey of Advanced Probabilistic Uncertainty Propagation and Sensitivity Analysis Methods," Chapter 6 of Joint Army/Navy/NASA/Air Force (JANNAF) e-book: *Simulation Credibility—Advances in Verification, Validation, and Uncertainty Quantification*, U. Mehta (Ed.), D. Eklund, V. Romero, J. Pearce, N. Keim, document NASA/TP-2016-219422 and JANNAF/GL-2016-0001, Nov. 2016.

[35]Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S., "Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index," *Computer Physics Communication*, Vol. 181, pp. 259-270, 2010.

[36]Sudret, B., "Global sensitivity analysis using polynomial chaos expansion," *Reliability Engineering and System Safety*, Vol. 93, No. 7, 2008, pp. 964-979.

[37]Drignei, D., Z. P. Mourelatos, M. Kokkolaras, V. Pandey, "Reallocation of Testing Resources in Validating Optimal Designs Using Local Domains," *Structural and Multidisciplinary Optimization*, 50(5), 825-838, 2014.

[38]Park, C. Y., N. H. Kim, R. T. Haftka, "How coupon and element tests reduce conservativeness in element failure prediction," *Reliability Engineering and System Safety*, 123 (2014) 123–136.

[39]Quintana, C., H. Millwater, G. Singh, P. Golden, "Optimal Allocation of Testing Resources for Statistical Simulations," *Engineering Optimization*, Vol. 48, No. 7, pp. 979-993, 2014, https://doi.org/10.1080/0305215X.2014.933824.

[40]American Society of Mechanical Engineers, ASME V&V 20-2009 "Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer," American Society of Mechanical Engineers, 2009.

[41]Richardson, L. F., "The Approximate Arithmetic Solution by Finite Differences of Physical Problems Involving Differential Equations, with an Application to the Stresses in a Masonry Dam," Philos. Trans. Royal Society of London, Series A, 210, pp. 307-357, 1910.

[42]Richardson, L. F. and Gaunt, J. A. "The Deferred Approach to the Limit," Philos. Trans. Royal Society of London, Series A, 226, pp. 299-361. 1927.

[43]Roache, P.J., *Fundmentals of Verification and Validation*, Hermosa Publishers, 2009.

[44]Celik, I.B., Ghia, U., Roache, P. J., FREITAS, C. J., COLEMAN, H., Raad, P.E., "Procedure for Estimation and Reporting of Uncertainty Due to Discretization in CFD applications," ASME Journal of Fluid Engineering, Vol 130, July 2008.

[45]Oberkampf, W. L., and Roy, C. J., *Verification and Validation in Scientific Computing*, Cambridge University Press, New York, 2010.

[46]Romero, V.J., "Model-Discretization Sizing and Calculation Verification for Multipoint Simulations over Large Parameter Spaces," AIAA Paper 2007-1937, 9th AIAA Non-Deterministic Methods Conference, April 23 - 26, 2007, Honolulu, HI.

[47]Oberkampf, W. L., Helton, J. C., and Sentz, K., "Mathematical Representation of Uncertainty," *AIAA Non-Deterministic Approaches Forum*, Apr 2001.

[48]Romero, V.J., "Validated Model? Not So Fast. The Need for Model 'Conditioning' as an Essential Addendum to Model Validation," AIAA Paper 2007-1953, 9th AIAA Non-Deterministic Methods Conference, April 23 - 26, 2007, Honolulu, HI.

[49]Romero, V.J., "Type X and Y Errors and Data & Model Conditioning for Systematic Uncertainty in Model Calibration, Validation, and Extrapolation," SAE Paper 2008-01-1368 for Society of Automotive Engineers 2008 World Congress, April 14-17, 2008, Detroit, MI.

[50]Romero, V., "Real-Space Model Validation and Predictor-Corrector Extrapolation applied to the Sandia Cantilever Beam End-to-End UQ Problem," AIAA Paper 2019-1488, 21st AIAA Non-Deterministic Approaches Conference, AIAA SciTech 2019, Jan. 7-11, San Diego, CA.

[51]Chen, X., S. Zhanpeng, H. Qinshu, X. Liu, "Prediction Considering Multi-Model and Model Form Uncertainty in the Parameter Space," paper 15IDM-0015, Soc. Auto. Engrs. 2015 World Congress, April 21–23, 2015, Detroit, MI.

[52]Chen, W., Y. Xiong, K.-L. Tsui, S. Wang, "A Design-Driven Validation Approach using Bayesian Prediction Models," *ASME J. Mechanical Design*, Vol. 130, No. 2., 2008.

[53]Xi, Z., Y. Fu, R.-J. Yang, "An Ensemble Approach for Model Bias Prediction," *SAE Journal of Materials and Manufacturing*, 6(3):2013, doi:10.4271/2013-01-1387.

[54]Oreskes, N., K. Shrader-Frechette, and K. Belitz, "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," *Science*, Vol. 263, pp. 641-646, Feb. 1994.

[55]Hale, L. E., Patil, M., and Roy, C. J., "Nondeterministic Simulation for Probability of Loss of Control Prediction for Unmanned Aircraft Systems," AIAA Paper 2015-2329, AIAA Modeling and Simulation Technologies Conference, Jun 2015.

[56]Kennedy, M.C., and A. O'Hagan, "Bayesian calibration of computer models," J. Royal Statistical Society Series B, Statistical Methodology, 63(3) pp. 425-464, 2002.

[57]Youn, B.D., B.C. Jung, Z. Xi, S.B. Kim, W. Lee, "A Hierarchical Framework for Statistical Model Calibration in Engineering Product Development," Computer Methods in Applied Mechanics and Engineering, 200:1421-1431, 2011.

[58]Sankararaman, S., and S. Mahadevan, "Comprehensive Framework for Integration of Calibration, Verification, and Validation," AIAA Paper 2012-1367, 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, April 23-26, 2012, Honolulu, HA.

[59]Zhan, Z., Y. Fu., R-J. Yang, "On Stochastic Model Interpolation and Extrapolation Methods for Vehicle Design," *SAE Journal of Materials and Manufacturing*, 6(3):2013, doi:10.4271/2013-01-1386.

[60]Oliver, T., G. Terejanu, C. Simmons, R. Moser, "Validating Predictions of Unobserved Quantities," *Computer Methods in Applied Mechanics and Engineering*, Vol. 283, Jan. 2015, pp. 1310-1335.

[61]Romero V.J., "Issues and Needs in Quantification of Margins and Uncertainty (QMU) for Phenomenologically Complex Coupled Systems," AIAA Paper 2006-1989, 8th AIAA Non-Deterministic Analysis Conference, Newport, RI, May 1-4, 2006.

[62]Helton, J.C. "Conceptual and Computational Basis for the Quantification of Margins and Uncertainty," Sandia Technical Report 2009-3005.

[63]Special Issue: Quantification of Margins and Uncertainties, *Reliability Engineering & System Safety*, Edited by Helton, J.C., and Pilch, M., Vol. 96, No. 9, pp. 959-1256, September 2011.

[64]"NASA Common Research Model," https://commonresearchmodel.larc.nasa.gov/ [retrieved August 2019].

[65]"Announcing the 6th AIAA CFD Drag Prediction Workshop," http://aiaa-dpw.larc.nasa.gov/ [retrieved August 2019].

[66]Vassberg, J. C., Tinoco, E. N., Mani, M., Rider, B., Zickuhr, T., Levy, D. W., Brodersen, O. P., Eisfeld, B., Crippa, S., Wahls, R. A., Morrison, J. H., Mavriplis, D. J., Murayama, M., "Summary of the Fourth AIAA CFD Drag Prediction Workshop," AIAA Paper 2010-4547, June-July 2010.

[67]"The 3rd AIAA CFD High Lift Prediction Workshop (HiLiftPW-3)," https://hiliftpw.larc.nasa.gov/index.html [retrieved August 2019].

[68]Rumsey, C. L., Slotnick, J. P., and Sclafani, A. J., "Overview and Summary of the Third AIAA High Lift Prediction Workshop," AIAA Paper 2018-1258, January 2018.

[69]Vassberg, J. C., DeHaan, M. A., Rivers, S. M., and Wahls, R. A., "Development of a Common Research Model for Applied CFD Validation Studies," AIAA Paper 2008-6919, August 2008.

[70]Lacy, D., S., and Sclafani, A. J., "Development of the High Lift Common Research Model (HL-CRM): A Representative High Lift Configuration for Transonic Transports," AIAA Paper 2016-0308, January 2016.

[71]Payne, F. M., "Low Speed Wind Tunnel Testing Facility Requirements: A Customer's Perspective," AIAA Paper 1999-0306, January 1999.

[72]Fryar, C. D., Kruszon-Moran, D., Gu, Q., and Ogden, C. L., "Mean Body Weight, Height, Waist Circumference, and Body Mass Index Among Adults: United States, 1999–2000 Through 2015–2016," *National Health Statistics Reports*, No. 122, December 2018.

[73]Spalart, P. R. and Allmaras, S. R., "A One-Equation Turbulence Model for Aerodynamic Flows," *Recherche Aerospatiale*, No. 1, 1994, pp. 5-21.

[74]Mani, M., Cary, A., Ramakrishnan, S. V., "A Structured and Hybrid-Unstructured Grid Euler and Navier-Stokes Solver for General Geometry," AIAA Paper 2004-524, January 2004.

[75]"NASA Langley Research Center Turbulence Modeling Resource," http://turbmodels.larc.nasa.gov [retrieved August 2019].

[76]Walter, J., Lawrence, W., Elder, D., and Treece, M., "Development of an Uncertainty Model for the National Transonic Facility (Invited)," AIAA Paper 2010-4925, June-July 2010.

[77]Schaefer, J., Cary, A., Mani, M., and Spalart, P., "Uncertainty Quantification and Sensitivity Analysis of SA Turbulence Model Coefficients in Two and Three Dimensions," AIAA Paper 2017-1710, January 2017.

[78]Bailey, S. C. C., Vallikivi, M., Hultmark, M., and Smits, A. J., "Estimating the value of von Kármán's constant in turbulent pipe flow," *Journal of Fluid Mechanics*, Vol. 749, 2014, pp. 79-98.

[79]Acheson, M. J. and Balakrishna, S., "Effects of Active Sting Damping on Common Research Model Data Quality," AIAA Paper 2011-0878, January 2011.

[80]Romero, V. J., Schroeder, B. B., Dempsey, J. F., Breivik, N. L., Orient G. E., Antoun, B.R., Lewis, J. R., and Winokur J. G., "Simple Effective Conservative Treatment of Uncertainty from Sparse Samples of Random Variables and Functions," *ASCE-ASME Journal of Uncertainty and Risk in Engineering Systems: Part B. Mechanical Engineering*, DOI 10.1115/1.4039558, December 2018, vol. 4, pp. 041006-1 – 041006-17.

[81]Romero, V., M. Bonney, B. Schroeder, V.G. Weirs, "Evaluation of a Class of Simple and Effective Uncertainty Methods for Sparse Samples of Random Variables and Functions," Sandia National Laboratories report SAND2017-12349, November 2017.

[82]Walker, E., "Measuring Wall Interference Correction Accuracy: An Overview of the NTF Program," AIAA Paper 2004-770, January 2004.

[83]Iyer, V., Kuhl, D. D., and Walker, E. L., "Improvements to Wall Corrections at the NASA Langley 14X22-FT Subsonic Tunnel," AIAA Paper 2003-3950, June 2003.

[84]Ulbrich, N., and Boone, A. R., "Direct Validation of the Wall Interference Correction System of the Ames 11-Foot Transonic Wind Tunnel," NASA TM-2003-212268, 2003.

[85]Walker, E. L., "Statistical Calibration and Validation of a Homogeneous Ventilated Wall-Interference Correction Method for the National Transonic Facility," NASA TM-2005-213947, 2005.

[86]Walker, E. L., Everhart, J. L., and Iyer, V., "Parameter Sensitivity Study of the Wall Interference Correction System (WICS)," AIAA Paper 2001-2421, June 2001.

[87]Walker, E., L, "Sensitivity of the Wall Interference Correction System to Measurement Uncertainty," M.S. Thesis, George Washington University, Washington, D.C., 2000.

[88]Rivers, M. B., and Dittberner, A., "Experimental Investigations of the NASA Common Research Model in the NASA Langley National Transonic Facility and NASA Ames 11-Ft Transonic Wind Tunnel (Invited)," AIAA Paper 2011-1126, January 2011.

[89]Roskam, J., *Airplane Flight Dynamics and Automatic Flight Controls, Part 1*, 6th ed., DARcorporation, 2011.

[90]Nelson, R. C., *Flight Stability and Automatic Control*, 2nd ed., McGraw Hill, 1998.

[91]Rivers, M. B., Hunter, C. A., "Support System Effects on the NASA Common Research Model," AIAA Paper 2012-0707, January 2012.

[92]Rivers, M. B., Hunter, C. A., Campbell, R. L., "Further Investigation of the Support System Effects and Wing Twist on the NASA Common Research Model," AIAA Paper 2012-3209, June 2012.

[93]Stephens, J. A., "Algorithms for Design Exploration and Simulation Credibility," Optimization and Uncertainty Quantification 2017 SIAM Computational Science and Engineering, Atlanta, GA, February-March 2017.

[94]Eldred, M. S., Swiler, L. P., and Tang, G., "Mixed Aleatory-Epistemic Uncertainty Quantification with Stochastic Expansions and Optimization-Based Interval Estimation," *Reliability Engineering and System Safety (RESS)*, Vol. 96, No. 9, September 2011, pp. 1092-1113.

[95]Ray, D., Ramirez-Marquez, J., "A framework for probabilistic model-based engineering and data synthesis," *Reliability Engineering & System Safety*, Vol. 193, 2020.