Comput. Methods Appl. Mech. Engrg. 200 (2011) 2131–2144

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

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

A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing

Christopher J. Roy [a,*], William L. Oberkampf [b]

[a] Aerospace and Ocean Engineering Department, Virginia Tech, 215 Randolph Hall, Blacksburg, Virginia 24061, USA
[b] Consulting Engineer, 5112 Hidden Springs Trail, Georgetown, Texas 78633, USA

ARTICLE INFO

Article history:
Received 6 May 2010
Received in revised form 20 January 2011
Accepted 28 March 2011
Available online 1 April 2011

Keywords:
Uncertainty quantification
Verification
Validation
Modeling
Computational simulation

ABSTRACT

An overview of a comprehensive framework is given for estimating the predictive uncertainty of scientific computing applications. The framework is comprehensive in the sense that it treats both types of uncertainty (aleatory and epistemic), incorporates uncertainty due to the mathematical form of the model, and it provides a procedure for including estimates of numerical error in the predictive uncertainty. Aleatory (random) uncertainties in model inputs are treated as random variables, while epistemic (lack of knowledge) uncertainties are treated as intervals with no assumed probability distributions. Approaches for propagating both types of uncertainties through the model to the system response quantities of interest are briefly discussed. Numerical approximation errors (due to discretization, iteration, and computer round off) are estimated using verification techniques, and the conversion of these errors into epistemic uncertainties is discussed. Model form uncertainty is quantified using (a) model validation procedures, i.e., statistical comparisons of model predictions to available experimental data, and (b) extrapolation of this uncertainty structure to points in the application domain where experimental data do not exist. Finally, methods for conveying the total predictive uncertainty to decision makers are presented. The different steps in the predictive uncertainty framework are illustrated using a simple example in computational fluid dynamics applied to a hypersonic wind tunnel.

1. Introduction

Scientific computing plays an ever-growing role in predicting the behavior of natural and engineered systems. In many cases, scientific computing is based on mathematical models that take the form of coupled systems of nonlinear partial differential equations. We will refer to the application of a model to produce a result, often including associated numerical approximation errors, as a *simulation*. While scientific computing has undergone extraordinary increases in sophistication over the years, a fundamental disconnect often exists between simulations and practical applications. Whereas most simulations are deterministic in nature, engineering applications are steeped in uncertainty arising from a number of sources such as those due to manufacturing processes, natural material variability, initial conditions, wear or damaged condition of the system, and the system surroundings. Furthermore, the modeling process itself can introduce large uncertainties due to the assumptions in the model as well as the numerical approximations employed in the simulations. The former is commonly addressed through model validation, while the latter is addressed by code and solution verification. Each of these different sources of uncertainty must be estimated and included in order to estimate the total uncertainty in a simulation. In addition, an understanding of the sources of the uncertainty can provide guidance on how to reduce or manage uncertainty in the simulation in the most efficient and cost-effective manner. Information on the magnitude, composition, and sources of uncertainty in simulations is critical in the decision-making process for natural and engineered systems. Without forthrightly estimating and clearly presenting the total uncertainty in a prediction, decision makers will be ill advised, possibly resulting in inadequate safety, reliability, or performance of the system. Consequently, decision makers could unknowingly put at risk their customers, the public, or the environment.

This paper presents a high level overview of our comprehensive framework for verification, validation, and uncertainty quantification (VV&UQ) in scientific computing. This framework has much in common with previous work in VV&UQ, but it also includes new concepts for estimating and combining various uncertainties. For more details on the approach, see Ref. [1]. The organization of this paper is as follows. First, the two different types of uncertainty are discussed; aleatory and epistemic. Then the key sources of

* Corresponding author. Tel.: +1 540 231 0080.
E-mail addresses: cjroy@vt.edu (C.J. Roy), wloconsulting@gmail.com (W.L. Oberkampf).

0045-7825/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cma.2011.03.016

uncertainty in scientific computing are identified: model inputs, numerical approximations, and model form uncertainty. A comprehensive framework is then described for treating all sources of uncertainty and their effects on the predicted system responses quantities (SRQs) of interest. The framework includes (1) the identification of all sources of uncertainty, (2) characterization of model input uncertainties, (3) elimination or estimation of code and solution verification errors, (4) propagation of input uncertainties through the model to obtain uncertainties in the SRQs, (5) quantification of model form uncertainty, and (6) estimation of model form uncertainty due to extrapolation to application conditions of interest. We conclude with a discussion of the paradigm shift from deterministic to nondeterministic simulations and the impact of this shift on risk-informed decision-making.

2. Types of uncertainty

While there are many different ways to classify uncertainty, we will use the taxonomy prevalent in the risk assessment community which categorizes uncertainties according to their fundamental essence [2–5]. Thus, uncertainty is classified as either (a) *aleatory* – the inherent variation in a quantity that, given sufficient samples of the stochastic process, can be characterized via a probability density distribution, or (b) *epistemic* – uncertainty due to lack of knowledge by the modelers, analysts conducting the analysis, or experimentalists involved in validation. The lack of knowledge can pertain to, for example, modeling of the system of interest or its surroundings, simulation aspects such as numerical solution error and computer round-off error, and lack of experimental data. In scientific computing, there are many sources of uncertainty including the model inputs, the form of the model, and poorly-characterized numerical approximation errors. All of these sources of uncertainty can be classified as either purely aleatory, purely epistemic, or a mixture of the two.

2.1. Aleatory uncertainty

Aleatory uncertainty (also called irreducible uncertainty, stochastic uncertainty, or variability) is uncertainty due to inherent variation or randomness and can occur among members of a population or due to spatial or temporal variations. Aleatory uncertainty is generally characterized by either a probability density function (PDF) or a cumulative distribution function (CDF). A CDF is simply the integral of the PDF from minus infinity up to the value of interest. An example of an aleatory uncertainty is a manufacturing process which produces parts that are nominally 0.5 m long. Measurement of these parts will reveal that the actual length for any given part will be different than 0.5 m. With a sufficiently large number of samples (i.e., information), both the form of the CDF and the parameters describing the distribution of the population can be determined. The aleatory uncertainty in the manufactured part can only be changed by modifying the manufacturing or quality control processes; however, for a given set of processes, the uncertainty due to manufacturing is considered irreducible.

2.2. Epistemic uncertainty

Epistemic uncertainty (also called reducible uncertainty or ignorance uncertainty) is uncertainty that arises due to a lack of knowledge on the part of the analyst, or team of analysts, conducting the modeling and simulation. If knowledge is added (through experiments, improved numerical approximations, expert opinion, higher fidelity physics modeling, etc.) then the uncertainty can be reduced. If sufficient knowledge (which costs time and resources) is added, then the epistemic uncertainty can, in principle, be eliminated. Epistemic uncertainty is traditionally represented as either an interval with *no associated PDF* or a PDF which represents degree of belief of the analyst (as opposed to frequency of occurrence of an event in aleatory uncertainty). We will represent epistemic uncertainty as an interval-valued quantity, meaning that the true (but unknown) value can be any value over the range of the interval, with *no likelihood or belief* that any value is more true than any other value. The Bayesian approach to uncertainty quantification characterizes epistemic uncertainty as a PDF that represents the degree of belief of the true value on the part of the analyst [6–8].

The distinction between aleatory and epistemic uncertainty is not always easily determined during characterization of input quantities or the analysis of a system. For example, consider the manufacturing process mentioned above, where the length of the part is described by a PDF, i.e., it is an aleatory uncertainty. If we are only able to measure a small number of samples (e.g., three) from the population, then we will not be able to accurately characterize the PDF representing the random variable. In this case, the uncertainty in the length of the parts could be characterized as a combination of aleatory and epistemic uncertainty. By adding information, i.e., by measuring more samples of manufactured parts, then the PDF (both its form and its parameters) could be more accurately determined. When one obtains a sufficiently large number of samples, then one can characterize the uncertainty in length as a purely aleatory uncertainty given by a precise PDF, i.e., a PDF with scalar values for all of the parameters that define the chosen distribution.

In addition, the classification of uncertainties as either aleatory or epistemic depends on the question being asked. In the manufacturing example given above, if one asks "What is the length of a *specific* part produced by the manufacturing process?" then the correct answer is a single true value that is not known, unless the specific part is accurately measured. If instead, one asks "What is the length of any part produced by the manufacturing process?" then the correct answer is that the length is a random variable that is given by the PDF determined using the measurement information from a large number of sampled parts.

3. Sources of uncertainty in scientific computing

For a comprehensive VV&UQ framework, all of the possible sources of uncertainty must be identified and characterized. When fixed values are known precisely (or with negligibly small uncertainty), then they can be treated as deterministic. Otherwise, they should be classified as aleatory, epistemic, or mixed uncertainties and characterized with the appropriate mathematical representation. Sources of uncertainty can be broadly categorized as occurring in model inputs, numerical approximations, or in the assumptions embodied in the mathematical model, i.e., model form. We will briefly discuss each of these categories below; see Ref. [1] for a complete description.

3.1. Model inputs

Model inputs include not only parameters used in the model of the system, but also data from the description of the surroundings (see Fig. 1). Model input data includes things such as geometry, constitutive model parameters, and initial conditions, and can come from a range of sources including experimental measurement, theory, other supporting simulations, or expert opinion. Data from the surroundings include boundary conditions and system excitation (e.g., mechanical forces or moments acting on the system, force fields such as gravity and electromagnetism). Uncertainty in model inputs can be aleatory, epistemic, or mixed.

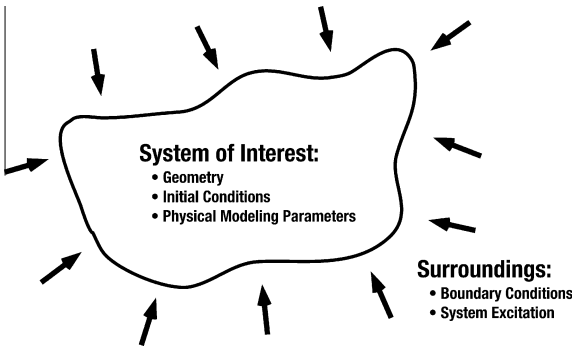

Fig. 1. Schematic of model input uncertainty (from [1]).

3.2. Numerical approximation

Since differential equation-based models rarely admit exact solutions for practical problems, approximate numerical solutions must be used. The characterization of the numerical approximation errors associated with a simulation is called *verification* [9,10]. It includes discretization error, iterative convergence error, round-off error, and also errors due to computer programming mistakes. Discretization error arises due to the fact that the spatial domain is decomposed into a finite number of nodes/elements and, for unsteady problems, time is advanced with a finite time step. Discretization error is difficult to estimate for practical problems and is often the largest of the numerical errors. Iterative convergence errors are present when the discretization of the model results in a simultaneous set of algebraic equations that are solved approximately or when relaxation techniques are used to obtain a steady-state solution. Round-off errors occur due to the fact that only a finite number of significant figures can be used to store floating point numbers in a digital computer. Finally, coding mistakes can occur when numerical algorithms are implemented into a software tool. Since coding mistakes are, by definition, unknown errors (they are generally eliminated when they are identified), their effects on the numerical solution are extremely difficult to estimate.

When nondeterministic methods are used to propagate input uncertainties through the model (as will be discussed in Section 4.4), then the statistical convergence of the propagation technique itself must also be considered. The key issue in *nondeterministic simulations* is that a single solution to the mathematical model is no longer sufficient. A set, or ensemble, of calculations must be performed to map the uncertain input space to the uncertain output space. Sometimes, this is referred to as *ensemble simulation* instead of nondeterministic simulations. Fig. 2 depicts the propagation of input uncertainties through the model to obtain output uncertainties. The number of individual calculations needed to accurately accomplish the mapping depends on four key factors: (a) the nonlinearity of the partial differential equations, (b) the dependency structure between the uncertain input quantities, (c) the nature of the uncertainties, i.e., whether they are aleatory, epistemic, or mixed uncertainties, and (d) the numerical methods used to compute the mapping. The number of mapping evaluations, i.e., individual numerical solutions of the mathematical model, can range from tens to hundreds of thousands. Many techniques exist for propagating input uncertainties through the mathematical model to obtain uncertainties in the SRQs. Sampling techniques (e.g., Monte Carlo and Latin hypercube sampling) are the most common techniques to propagate input uncertainties through the model. Monte Carlo sampling is discussed in more detail in Section 5.4.

For cases where numerical approximation errors can be estimated, their impact on the SRQs of interest can, in principle, be eliminated, given that sufficient computing resources are available. If this is not practical, they should generally be converted to epistemic uncertainties due to the uncertainties associated with the error estimation process itself. Some researchers argue that numerical approximation errors can be treated as random variables and that the variance of the contributors can be summed in order to obtain an estimate of the total uncertainty due to numerical approximations [11,12]. We believe this approach is unfounded and that traditional statistical methods cannot be used. Estimates of numerical approximation errors are analogous to bias (systematic) errors in experimental measurements; *not* random measurement errors. That is, numerical approximation errors do not display a random character, unless the discretization is far from the asymptotic region; in which case one should not be relying on the approximation procedure. As is well known, bias errors are much more difficult to identify and quantify than random errors.

3.3. Model form

The form of the model results from all assumptions, conceptualizations, abstractions, approximations, and mathematical formulations on which the model relies [1]. The characterization of model form uncertainty is commonly estimated using *model validation*. Since the term *validation* can have different meanings in various communities, we expressly define it to be: *assessment of model accuracy by way of comparison of simulation results with experimental measurements*. This definition is a specific, and more restrictive, interpretation of the common definition of validation given in Ref. [10]. Although model validation has been a standard procedure in science and engineering for over a century, our approach takes two additional steps. First, it *statistically quantifies* the disagreement between the simulation results and all of the conditions for which experimental measurements are available. Second, it *extrapolates* this uncertainty structure from the domain of available experimental data to the application conditions of interest where experimental data are *not* available. This extrapolation entails both a regression fit of the computed model form uncertainty in the validation space and the evaluation of prediction

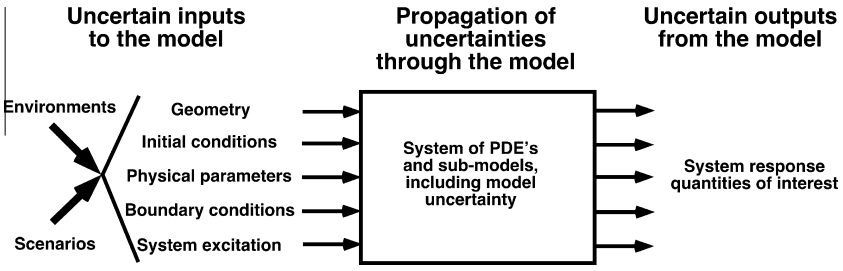

Fig. 2. Propagation of input uncertainties to obtain output uncertainties (from [1]).

intervals (similar to confidence intervals) at the prediction conditions of interest; see Section 4.5.2 for more details. A key aspect of this approach is that model form uncertainty is treated as epistemic [13].

4. Uncertainty framework

The proposed framework for treating predictive uncertainty uses the technique of probability bounds analysis, which was developed by Ferson and his colleagues. [13–15]. Only a high-level description of this predictive uncertainty framework is given here, additional details can be found in Ref. [1]. A novel aspect of the framework is that it addresses all sources of uncertainty in scientific computing including uncertainties due to the form of the model and numerical approximations. The purpose of this framework is to be able to estimate the uncertainty in a SRQ for which no experimental data are available. That is, the mathematical model, which embodies approximations to the relevant physics and characterization of the input uncertainties, is used to predict the uncertain SRQ, which includes numerical approximation uncertainties and an extrapolation of the model form uncertainty to the conditions of interest. The basic steps in the VV&UQ framework are described next, with emphasis on aspects of the framework that are new in the field of predictive uncertainty.

4.1. Identify all sources of uncertainty

All potential sources of uncertainty in model inputs must be identified. If an input is identified as having little or no uncertainty, then it is treated as deterministic, but such assumptions must be justified and understood by the decision maker using the simulation results. The *goals of the analysis* should be the primary determinant for what is considered as fixed versus what is considered as uncertain. The general philosophy that should be used is: consider an aspect as uncertain unless there is strong evidence (e.g., from a sensitivity analysis) that the uncertainty in the aspect will result in minimal uncertainty in *all* of the SRQs of interest in the analysis.

As discussed earlier, sources of uncertainty are categorized as occurring in model inputs, numerical approximations, and in the form of the mathematical model. We point out that there are some varieties of model form uncertainty that are difficult to identify. These commonly deal with assumptions in the conceptual model formulation or cases where there is a very large extrapolation of the model. Some examples are (a) assumptions concerning the environment (normal, abnormal, or hostile) to which the system is exposed, (b) assumptions concerning the particular scenarios the system is operating under, e.g., various types of damage or misuse of the system, and (c) cases where experimental data are not available on any closely related systems, e.g., data are only available on subsystems, but predictions of complete system performance are required. Sometimes, separate simulations are conducted with independent models (using different assumptions) in order to help identify the additional sources of uncertainty in the model inputs or the model form. In this case, care must be taken to avoid the introduction or elimination of important uncertainties in the problem.

4.2. Characterize uncertainties

By *characterizing a source of uncertainty* we mean (a) assigning a mathematical structure to describe the uncertainty and (b) determining the numerical values of all of the needed parameters of the structure. Stated differently, characterizing the uncertainty requires that a mathematical structure be given to the uncertainty and all parameters of the structure be numerically specified such that the structure represents the state of knowledge of every uncertainty considered. The primary decision to be made concerning the mathematical structure for each source is whether it should be represented as a purely aleatory uncertainty, a purely epistemic uncertainty, or a mixture of the two.

For purely aleatory uncertainties, the uncertainty is characterized as a precise distribution, e.g., a CDF is given with scalar quantities for each of the parameters of the chosen distribution. If the model input uncertainties are correlated, then joint PDFs or CDFs should be used; however, for practical engineering problems there may be little or no information concerning the covariance structure of the marginal probability distributions for correlated inputs. For purely epistemic uncertainties, such as numerical approximations and model form, the uncertainty is characterized as an interval. For an uncertainty that is characterized as a mixture of aleatory and epistemic uncertainty, then an *imprecise distribution* is given. This mathematical structure is a distribution where the parameters of the distribution are not scalars, but are either (a) distributions themselves or (b) interval-valued quantities. For this latter case, the structure represents the ensemble of all possible distributions of a single family that exist whose parameters are bounded by the specified intervals. This structure commonly arises in characterization of information from expert opinion. For example, suppose a new manufacturing process is going to be used to produce a component. Before inspection samples can be taken from the new components, a manufacturing expert could characterize its features or performance as an imprecise distribution based on experience with similar processes.

Information for characterizing input uncertainties typically derives from (a) experimentally measured data from the actual system or similar systems of interest, (b) data generated from separate models that support the modeling of the system of interest, and (c) opinions expressed by experts familiar with the system of interest. See, for example, Refs. [2–5] for more information on methods for characterizing input uncertainties.

4.3. Estimate uncertainty due to numerical approximations

Recall that *verification* deals with estimating *numerical errors* which include discretization error, iterative error, round-off error, and coding mistakes. Methods for estimating discretization error can be broadly categorized as either higher-order estimators (Type I) or residual-based estimators (Type II) [1,16]. The Type I methods involve post-processing of the solution (or multiple solutions) and include Richardson extrapolation [1,17], order extrapolation [18], and recovery methods from finite elements [19]. The Type II methods employ additional information about the problem being solved and include discretization error transport equations [20,21], defect correction methods [22], and implicit/explicit residual methods in finite elements [23,24]. Of particular note is the recent emergence of adjoint methods for estimating the error in solution functionals (i.e., SRQs) for both finite element [24] and finite volume methods [25]. However, regardless of the approach used for estimating the discretization error, the reliability of the estimate depends on the solutions being in the asymptotic mesh convergence range [1,16], which is extremely difficult to achieve for complex scientific computing applications.

Various techniques are available for estimating iterative convergence errors (e.g., see [1]). Round-off errors are usually small, but can be reduced if necessary by increasing the number of significant figures used in floating point computations (e.g., by changing from double to quadruple precision arithmetic). Since errors due to the presence of unknown coding mistakes or algorithm inconsistencies are difficult to characterize, their effects should be minimized by employing good software engineering practices and using specific techniques for scientific computing software such as order of

accuracy verification and the method of manufactured solutions (e.g., see [1,17,26]).

Because of the difficulties in obtaining accurate estimates of the different numerical approximation errors, in most cases they should be converted to and explicitly represented as epistemic uncertainties. The simplest method for converting error estimates to uncertainties is to use the magnitude of the error estimate to apply uncertainty bands above and below the simulation prediction, possibly with an additional factor of safety included, e.g.: $U_{DE} = Fs|\bar{\varepsilon}|$ where U_{DE} is the uncertainty due to discretization error, $Fs \geqslant 1$ is the factor of safety, and $\bar{\varepsilon}$ is the discretization error estimate. Roache's Grid Convergence Index [17,27] provides a specific example of converting the discretization error estimate from Richardson extrapolation to an uncertainty. Uncertainty due to numerical approximation error is epistemic since additional information (e.g., mesh levels, iterations, digits of precision) could be added to reduce it. When treating epistemic uncertainties as intervals, the proper mathematical method for combining uncertainties due to discretization (U_{DE}), incomplete iteration (U_{IT}), and round off (U_{RO}) is to simply sum the intervals using interval arithmetic. Assuming the lower bound on each interval is zero, and the upper bound is given by the associated uncertainty estimate, then one has:

$$U_{NUM} = U_{DE} + U_{IT} + U_{RO}. \tag{1}$$

It can easily be shown that U_{NUM} is a guaranteed bound on the total numerical error, given that each contributor is itself an accurate bound.

Implementing Eq. (1) in practice is a tedious and demanding task, even for relatively simple simulations, because of two features. First, Eq. (1) must be calculated for *each* SRQ of interest in the simulation. For example, if the SRQs are pressure, temperature, and three velocity components in a numerical solution, then U_{NUM} should be calculated for each quantity over the domain of the flow field. Second, each of the SRQs varies as a function of the uncertain input quantities in the simulation. One common technique for limiting the computational effort involved in making all of these estimates is to determine the locations in the domain of the partial differential equations and the conditions for which the input uncertainties are believed to produce the largest values of U_{DE}, U_{IT}, and U_{RO}. The resulting maximum uncertainties at these selected conditions are then applied over the entire physical domain and application conditions. This simplification is a reasonable approach, but it is not always reliable because of potentially large variations in the SRQs over the domain of the differential equation, characteristics of the physics in the mathematical model, and nonlinear interactions between input uncertainties.

4.4. Propagate input uncertainties through the model

This section briefly describes requirements for propagating both aleatory and epistemic uncertainty in model inputs through the model in order to determine the effects on the SRQs. Recall that model inputs can arise from the parameters used in describing either the system of interest or the surroundings. Although both aleatory and epistemic uncertainties can be propagated using a variety of methods, they should each be treated independently because they characterize two different types of uncertainty. For example, consider a simple sampling-based approach for propagating combined aleatory and epistemic uncertainty. Sampling an aleatory uncertainty implies a sample is taken from a random variable and that each sample is associated with a probability. However, sampling an epistemic uncertainty implies a sample is taken from a range of *possible* values. The sample has no probability or frequency of occurrence associated with it based on how the sample was chosen. That is, we only know it is possible, which could be viewed as assigning a probability of unity, given the information available concerning the input quantity. As mentioned above, this characterization of uncertainties, where aleatory and epistemic uncertainties are segregated, is referred to as probability bounds analysis [13–15] and is a fundamental aspect of the proposed uncertainty quantification framework. The treatment of model form uncertainty and numerical approximation uncertainty are addressed in Sections 4.5 and 4.3, respectively.

When a given model input contains a mixture of aleatory and epistemic uncertainty, it can usually be characterized as a PDF or CDF with interval-valued parameters. For example, if an input is assumed to be normally distributed, then the two parameters that describe the normal distribution (the mean and standard deviation) could be treated as intervals. In the uncertainty propagation approaches described below, this model input would result in three uncertainties: the original model input would be an aleatory uncertainty and the mean and standard deviation (used to define the PDF) would be treated as two separate epistemic uncertainties. When there is conflicting evidence for the characterization of uncertain model inputs, the Dempster-Shafer theory of evidence can be used [28], but this case is beyond the scope of the current work.

4.4.1. Aleatory uncertainty

The simplest approach for propagating aleatory uncertainty through a model is sampling. In Monte Carlo sampling, a random number is chosen between zero and one, then the inverse CDF can be used to obtain the corresponding sample for the model input parameter. Although sampling methods are conceptually simple, large numbers of samples are needed in order to accurately characterize low probability events (e.g., when high confidence levels such as 99% or 99.9% are required for the SRQ). More advanced approaches for propagating aleatory uncertainty include polynomial chaos (e.g., [29–33]), stochastic collocation (e.g., [31,33]), and response surface approximation methods (e.g., [34]). For example, in polynomial chaos, the SRQ is expanded as a polynomial in terms of the uncertain model inputs, and the coefficients of this expansion are themselves allowed to be random variables. While initial polynomial chaos implementations were code intrusive, more recently non-intrusive forms of polynomial chaos have been developed. When only a few aleatory uncertain variables are present (e.g., less than five or ten), then polynomial chaos has the potential to significantly reduce the number of samples required for statistical convergence of the CDF. For larger numbers of uncertain variables, or when statistical correlations exist between input quantities, more traditional sampling methods have proven to be more robust.

4.4.2. Combined aleatory and epistemic uncertainty

When aleatory and epistemic uncertainties occur in the input quantities, the propagation of each type of uncertainty must be separated. For example, when employing sampling to propagate uncertainty, each of the samples obtained from an aleatory uncertainty is associated with a probability of occurrence. When a sample is taken from an epistemic uncertainty, however, there is *no probability* associated with the sample. The sample is simply a possible realization over the interval-valued range of the input quantity. Furthermore, if one takes a sample from each of the epistemic uncertainties, and then one computes the aleatory uncertainty as just described, the computed CDF of the SRQ can be viewed as a conditional probability. That is, the computed CDF is for the condition of the given vector of fixed samples of the epistemic uncertainties. This type of segregated sampling between aleatory and epistemic uncertainties is usually referred to as double-loop or nested sampling.

The recommended uncertainty propagation approach can be summarized as follows. For each sample of all of the epistemic

uncertainties, the aleatory uncertainties are propagated through the model (using any of the aleatory uncertainty propagation methods mentioned above) to produce a single CDF of the SRQ. After all of the epistemic samples have been chosen, and the resulting CDFs have been computed, one has an ensemble of M CDFs. The widest extent of the ensemble of CDFs is used to form a probability box (also called a p-box) [13–15]. The p-box is a special type of CDF which contains information on both aleatory and epistemic uncertainties (see Fig. 3). A p-box expresses both epistemic and aleatory uncertainty in a way that does not confound the two. A p-box shows that an SRQ cannot be displayed as a precise probability, but it is now an interval-valued probability. For example, in Fig. 3, for a given value of the SRQ, the probability that that value will occur is given by an interval-valued probability. That is, no single value of probability can describe the uncertainty given the present state of knowledge. Likewise, for a given probability value of the SRQ, there is an interval-valued range predicted for the SRQ of interest. Stated differently, the p-box accurately reflects the system response *given the present state of knowledge* of the input uncertainties.

A p-box is an interval-valued probability, or stated another way, it is a dual-valued probability. For example, suppose that for structural safety a decision maker requires that a predicted beam deflection must be less that 5 mm with a probability of 0.95. Using our approach, one uses the right-hand boundary of the p-box prediction for deflection and then reads the cumulative probability for a deflection of 5 mm. If the cumulative probability is less than 0.95, then the safety requirement is not satisfied. What the p-box structure says is that, given (a) the stochastic nature of the system and its input data, (b) the state of knowledge of the system and its input data, and (c) our knowledge of the predictive uncertainty, one can only assure that the probability will be not be any less than the right-hand boundary of the p-box. An additional graphical example of how to interpret p-boxes is given in the example problem in Section 5.

4.5. Estimate model form uncertainty

Model form uncertainty is estimated through the process of *model validation*. As mentioned above, there are two aspects to estimating model form uncertainty. First, we quantitatively estimate the model form uncertainty at the conditions where experimental data are available using a mathematical operator referred to as a validation metric. During the last ten years there has been a flurry of activity dealing with the construction of validation metrics [13,35–37]. Second, we extrapolate the uncertainty structure expressed by the validation metric to the application conditions of interest. As is common in scientific computing, no experimental data is available for the application conditions of interest. Then the extrapolated model form uncertainty is included in the prediction of the model at the conditions of interest as an epistemic uncertainty. This section discusses both of these topics.

4.5.1. Validation metrics

A validation metric is a mathematical operator that requires two inputs: the experimental measurements of the SRQ of interest and the prediction of the SRQ at the conditions used in the experimental measurements. A flowchart for computing a validation metric is given in Fig. 4. In the ideal case, the validation metric is computed using specially-tailored validation experiments; but more commonly one must use existing experimental data, e.g., from the literature or from an industrial/laboratory database. A key part of computing the SRQ of interest is that the uncertainty in all model input parameters should be carefully measured during the experiment. Once the input uncertainties have been characterized, they are used as input to the model and are propagated through it (as discussed above in Section 4.4) to obtain the SRQ of interest. Depending on the nature of the model input uncertainties (purely aleatory, purely epistemic, or mixed), the SRQ will be a precise CDF, an interval, or a p-box, respectively. The key point that should be stressed with regard to the input uncertainties is that when they are propagated through the model, the model is expected to predict the experimentally measured variability in the SRQs that is due to the experimental variability of the inputs. Any disagreement between the experimentally measured and simulated SRQs (whether they are CDFs, intervals, or p-boxes) is attributed to model form uncertainty, the source of which can be either physics modeling assumptions and/or imprecise knowledge of the input uncertainties.

One additional important point should be made concerning Fig. 4. The concept behind Fig. 4 is that the experimental processes of nature, symbolized on the left, are expected to be reproduced to some degree by the mathematical model on the right. However, the existence of experimental uncertainty in the measurement of the SRQ can potentially interfere with this seemingly reasonable expectation of the model. Even if bias errors in the measurements have been reduced to a negligible level, there is always random

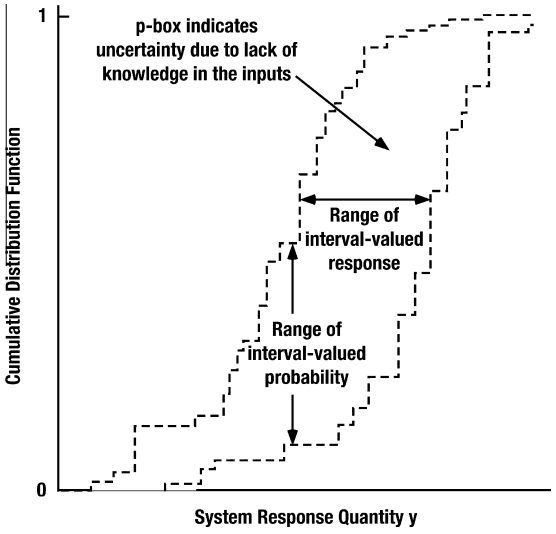

Fig. 3. Example p-box resulting from the propagation of combined aleatory and epistemic uncertainties through a model (from [1]).

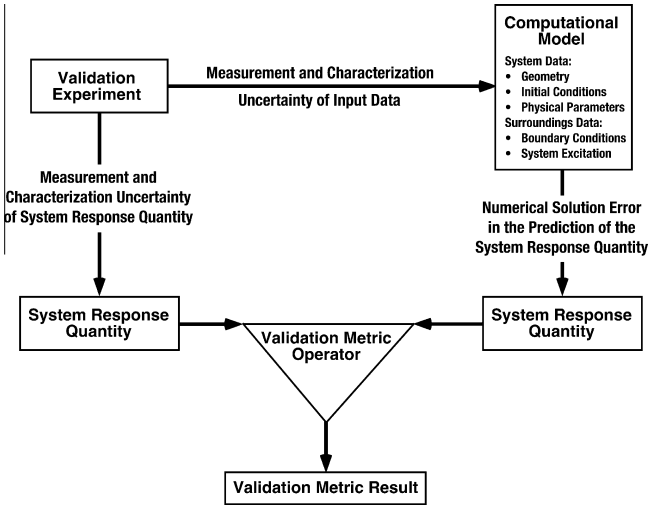

Fig. 4. Flowchart detailing the key steps in computing a validation metric (from [1]).

measurement uncertainty in the SRQ. However, measurement uncertainty is *not* expected to be modeled on the right. As a result, there will be increased variance in the experimental results due to measurement uncertainty. Since this will not be exhibited in the model predictions, the model will be (incorrectly) assessed to be in error due to this effect.

While there are many possible validation metrics, we will focus on one implementation called the area validation metric. When only aleatory uncertainties are present in the model inputs, then propagating these uncertainties through the model produces a CDF in the SRQ. Experimental measurements are then used to construct another empirical CDF of the SRQ. The area between these two CDFs is referred to as the area validation metric d (also called the Minkowski L_1 norm) and is given by

$$d(F, S_n) = \int_{-\infty}^{\infty} |F(x) - S_n(x)| dx, \tag{2}$$

where $F(x)$ is the CDF from the simulation, $S_n(x)$ the CDF from the experiment, and x is the SRQ. The area validation metric d has the same units as the SRQ and effectively provides a *measure of the evidence for disagreement between the simulation and the experiment* [13]. It should be noted that the area validation metric satisfies all the conditions for a distance function on a metric space (see Ref. [1] for details).

An example of this area validation metric for a case with only aleatory uncertainty occurring in the model input parameters is given in Fig. 5. In this figure it is assumed that a large number of simulations, i.e., Monte Carlo samples, are performed, but only four experimental measurements are available. With the large number of simulation samples, one is able to construct an accurate CDF for the simulation. The stair-steps in the experimental measurements are the locations of each of the four experimental measurements. As a result, each measurement is assigned a probability of 0.25. The *area validation metric* is the smallest area between the simulation CDF and the experimental CDF. It can be seen in the figure that the model is assessed an error (i.e., an area) whether the model prediction is greater than or less than the individual measurements.

Fig. 6 is an example of the area metric for the case when there are only three samples from the simulation and five experimental measurements. This situation is not uncommon for computationally intensive scientific computing applications such as computational fluid dynamics or nonlinear computational structural mechanics. A significant advantage of the area validation metric is that it can be computed even when very few samples are available from the simulation or from the experiment. In addition, when only a few simulation or experimental samples are available, *no assumptions* are required concerning the statistical nature of the simulation or the experiment. The validation metric deals directly with the samples computed and the measurements obtained.

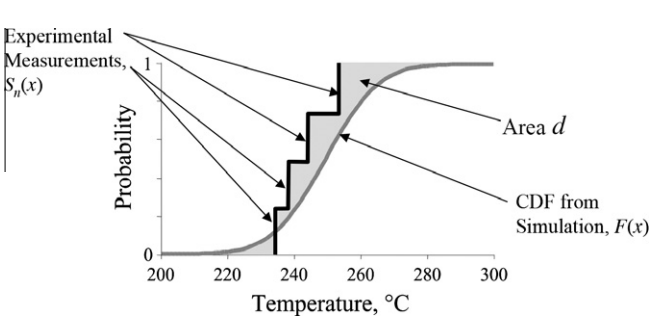

Fig. 5. Area validation metric example with four experimental measurements of the SRQ available (from [13]).

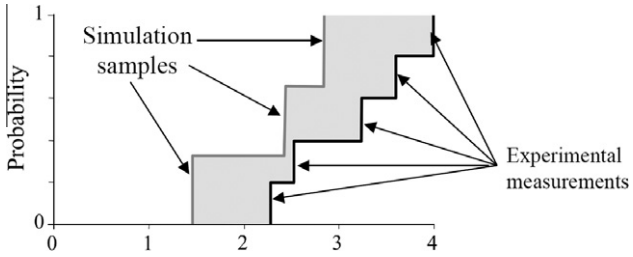

Fig. 6. Example of the area validation metric when only a small number of simulations and experimental measurements are available (from [13]).

When little experimental information is available on needed model input parameters, these parameters should be characterized as an interval, i.e., epistemic uncertainties. This situation occurs very commonly with published experimental data and experiments that were not designed to be validation experiments. As a result, when these intervals are propagated through the model, the predicted SRQ of interest is represented as a p-box. The validation metric can also deal with this situation. Fig. 7 shows the case where the model prediction is a p-box and only a single experimental measurement is available. Fig. 7a occurs when the measurement falls entirely within the p-box and Fig. 7b occurs when the measurement is slightly larger than the p-box. When the experimental measurement falls entirely within the simulation's p-box (Fig. 7a), then the area validation metric is zero. When a portion of the experimental measurement falls outside of the p-box (Fig. 7b), then the area validation metric is nonzero. Recall that the area validation metric is the *smallest area* between the p-box and the experimental CDF. That is, the validation metric reflects the *evidence for disagreement* between the model and the experiment. When the simulation is a p-box due to insufficient information provided by the validation experiment, the model is given more leeway in comparing with the experiment, as is appropriate.

4.5.2. Model extrapolation

Numerous validation experiments would normally be required in order to estimate the area validation metric over the entire space of model input parameters for the application of interest. In many cases, however, it is not possible to obtain experimental data at the application conditions of interest. As a result, the more common situation is that the model must be applied at conditions where there are no experimental data. Consider a simple example when there are only two input parameters for the model: α and β (Fig. 8). The validation domain consists of the set of points in this parameter space where experiments have been conducted and the validation metric has been computed (denoted by a "V" in Fig. 8). In this example, the application domain (sometimes referred to as the operating envelope of the system) is larger than

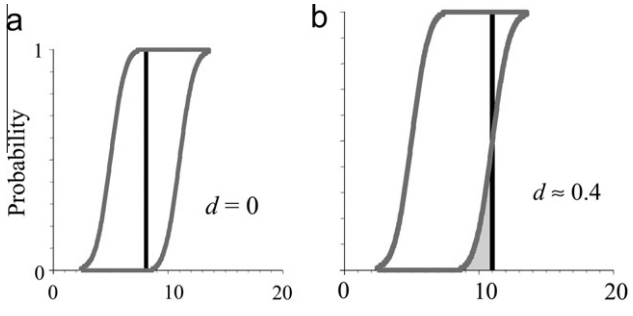

Fig. 7. Example of area validation metric computed at one location in the model input parameter space: (a) aleatory uncertainties only and (b) combined aleatory and epistemic uncertainties (from [38]).

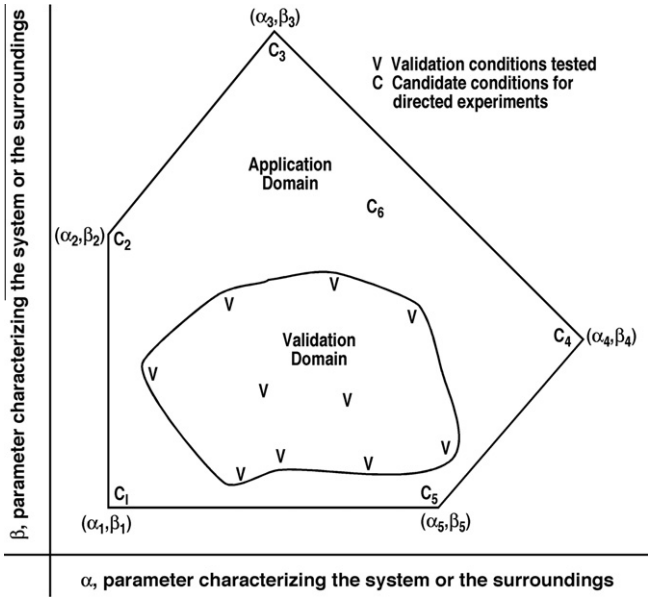

Fig. 8. Model validation relative to experimental data (adapted from [40]).

the validation domain. Thus, one must choose between either extrapolating the validation metric outside of the validation domain or performing additional validation experiments. In Fig. 8, some conditions for candidate validation experiments are given by a "C". The key point here is that the validation domain is generally not coincident with the application domain, thus either interpolation or extrapolation to the conditions of interest is needed. Another point is that when one makes a prediction with the model by extrapolating beyond the validation conditions, one must conceptually abandon the notion of assessment of model accuracy because we use the term "prediction" to mean "no experimental data is presently available." Thus, extrapolation of the validation metric deals only with the estimation of model form uncertainty due to the fact that there are no experimental data available for the conditions of interest.

The general process for determining the model form uncertainty at the conditions of interest (i.e., the prediction location) is as follows. First, a regression fit of the validation metric is performed in the space of the validation domain. Next, a statistical analysis is performed to compute the prediction interval at the conditions of interest. This prediction interval is similar to a confidence interval, but it will be larger because we are interested in a future random deviate predicted by the regression fit of the validation metric data, i.e., the uncertainty due to the regression fit and due to the variability of the validation metric evaluated at multiple conditions. The computation of the prediction interval requires a level of confidence to be specified (e.g., 95% confidence). The model form uncertainty at the prediction conditions is then found by taking the maximum of zero and the value found from the regression fit of the validation metric and adding in the upper value of the prediction interval. For example, if the value of the regression fit at the prediction conditions is $\hat{d}$ and the 95% confidence interval is $\hat{d} \pm P$, then the model form uncertainty at that location is given by

$$\max(\hat{d}, 0) + P. \tag{3}$$

In the past, it has been common practice to either (a) ignore the model form uncertainty in the predictions for the application conditions or (b) calibrate adjustable parameters in the mathematical model so that improved agreement could be obtained with the available experimental data at conditions "V". However, we believe a more forthright approach is to explicitly account for the mismatch between the model and the experiments and include this mismatch as additional uncertainty in the predictions of the SRQ in the application domain. When the validation metric results are extrapolated to new conditions, there is a subsequent increase in the uncertainty due to the extrapolation process itself, i.e., the standard error of the regression fit increases as one moves away from the central region of the available data.

It should be stressed that the extrapolation we use is based on a regression fit of the model form uncertainty structure exhibited within the validation domain, along with the statistical prediction interval associated with the regression fit. In our method of extrapolation, the regression function for the validation metric is completely separate from any extrapolation of the prediction results (based on the physics in the model) to conditions where experimental data are not available. A common example of the latter is the use of an aerodynamics model to extrapolate from wind tunnel conditions (where data exist) to flight Reynolds numbers. As one reviewer correctly pointed out, our framework does not address (a) cases where some system response quantities are measured, but these are not the SRQs that are of interest for the application condition and (b) the case where most (or all) experimental data are available for systems that are distantly related to the system of interest, e.g., lower tiers in a validation hierarchy. The only validation framework that we are aware of that attempts to deal with these difficult issues is the recent work of Refs. [13,38,39].

4.6. Determine total uncertainty in the SRQ

The total uncertainty in the SRQ at the application conditions of interest is computed as follows. First, one begins with the *p*-box that was generated by propagating the aleatory and epistemic uncertainties in the model input parameters through the model (recall Figs. 2 and 3). Next, the area validation metric is appended to the sides of the *p*-box, thus showing that the epistemic uncertainty in the SRQ has increased due to model form uncertainty. This is certainly not the only way that the area validation metric could be included in the predictive uncertainty, but it is consistent with the *p*-box representation of epistemically uncertain inputs that are propagated through the model. As discussed above, if extrapolation of the model form uncertainty is required, then it is the extrapolated d values that are appended to the SRQ *p*-box. This process is shown graphically in Fig. 9. Finally, the uncertainty due to numerical approximations is treated as an additional episte-

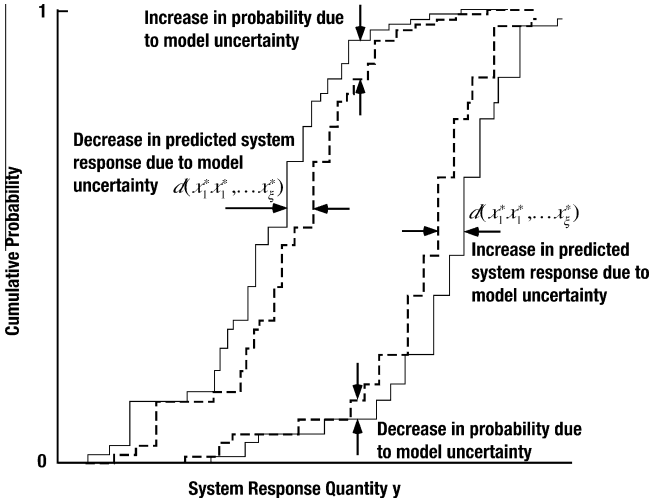

Fig. 9. Increase in predictive uncertainty due to the addition of model form uncertainty (from [1]).

mic uncertainty and, in a similar fashion to the model form uncertainty, is appended to the p-box of the SRQ. A more detailed example of this total uncertainty representation is given in Section 5.

This approach for representing total uncertainty in the predicted SRQ of interest provides valuable information to decision makers using the results from the simulation. The width of the original *p*-box provides information on the effects of epistemic uncertainties in the model inputs on the predicted SRQ. The shape, or the range, of the two bounding CDFs of the *p*-box provides information on the effects of aleatory uncertainties in the model inputs. The validation metric d that is appended to the *p*-box of the SRQ explicitly informs the decision maker of the estimated magnitude of the uncertainty that is due to model form uncertainty. Finally, U_{NUM} that is appended to the *p*-box informs the decision maker of the magnitude of the uncertainty due to numerical approximations. In limited testing of this approach, it has been found that the model form uncertainty has been a dominant contributor to the total uncertainty in the predicted SRQ, even for relatively accurate models of the system of interest [1,13].

5. Example: Hypersonic nozzle flow

As a simple example of the proposed VV&UQ framework, consider the flow of air through a converging–diverging hypersonic wind tunnel. This nozzle is designed to achieve a nominal test section Mach number of 8. The flow is modeled with the quasi-one-dimensional Euler equations which account for cross-sectional area variations in the nozzle. While the inviscid flow through a nozzle could be determined by simpler means (e.g., formulated as a set of boundary-value ordinary differential equations or even algebraic equations resulting from isentropic flow assumptions), we retain the more complex formulation in order to demonstrate all phases of the proposed framework. In order to cluster the mesh cells in regions of high gradients (i.e., near the nozzle throat), a (steady) global transformation $\xi = \xi(x)$ is employed where x is the axial location in the nozzle. The quasi-one-dimensional Euler equations in transformed coordinates can thus be written as:

$$\begin{aligned} &\frac{A}{\xi_x}\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial \xi}(\rho u A) = 0, \\ &\frac{A}{\xi_x}\frac{\partial (\rho u)}{\partial t} + \frac{\partial}{\partial \xi}\left[A(\rho u^2 + p)\right] = p\frac{\partial A}{\partial x}, \\ &\frac{A}{\xi_x}\frac{\partial (\rho e_t)}{\partial t} + \frac{\partial}{\partial \xi}(\rho u h_t A) = 0, \end{aligned} \tag{4}$$

where $A(x)$ is the area distribution of the nozzle, ρ is the density, u the velocity, p the pressure, e_t the total energy (internal plus kinetic), h_t the total enthalpy, and ξ_x the metric of the transformation. For a calorically perfect gas, one has

$$\begin{aligned} p &= \rho R T, \\ e_t &= \frac{RT}{\gamma - 1} + \frac{u^2}{2}, \\ h_t &= \frac{\gamma R T}{\gamma - 1} + \frac{u^2}{2}, \end{aligned} \tag{5}$$

where R is the specific gas constant and γ is the ratio of specific heats.

The weak form of these equations is solved with a cell-centered finite volume formulation using Roe's approximate Riemann solver [41] to determine the interface fluxes. MUSCL extrapolation [42] is used to determine the left and right states at each cell interface (i.e., grid point) in a formally second-order accurate manner. The grid points x_i ($1 \leqslant i \leqslant imax$) define the cell interfaces and are nonuniformly distributed in physical space; however, these points are uniformly distributed in transformed space from $\xi = 1$ to $\xi = imax$. All simulations are performed with 128 cells in the spatial direction unless otherwise noted.

The nozzle has a circular cross section with an effective radius (in meters) given by

$$r(x) = \begin{cases} 0.025 + 0.015\cos(\pi x/x_t) & x \leqslant x_t, \\ \frac{r_{ts}+0.01}{2} + \frac{0.01-r_{ts}}{2}\cos\left(\pi\frac{x-x_t}{2.5-x_t}\right) & x > x_t, \end{cases} \tag{6}$$

where the location of the throat is $x_t = 0.2$ m. The effective tunnel radius at the test section r_{ts} depends on both the wind tunnel contour and the presence of the tunnel wall boundary layer. Thus the effective test section radius can vary depending on whether the tunnel wall boundary layer is laminar, transitional, or turbulent. The nominal laminar value for this effective radius at the test section is $r_{ts} = 0.14$ m, and the corresponding area variation of the tunnel is given in Fig. 10.

The SRQ of interest is the static temperature at the beginning of the test section ($x = 2.44$ m). From prior experience, it is known that condensation can occur in the test section when the static temperature falls below 80 K. In addition to degrading the flow quality in the test section, condensation at such high speeds can also damage any aerodynamic models inserted into the tunnel. The goal of the analysis is to determine whether one can have 95% confidence that the test section temperature will be greater than or equal to 80 K. Fig. 10 shows the static temperature distribution in the nozzle predicted by a deterministic simulation assuming a stagnation temperature of $T_o = 1{,}200$ K and a laminar boundary layer (i.e., $r_{ts} = 0.14$ m). This deterministic simulation results in a test section static temperature of 85.3 K, which is 6% larger than the specified minimum temperature of 80 K. The five steps in the proposed VV&UQ framework will now be applied to this example problem in order to quantify the total uncertainty in the simulation predictions.

5.1. Identify all sources of uncertainty

The two primary sources of uncertainty in the model inputs are the wind tunnel stagnation temperature and the effective area downstream of the tunnel throat. The other model inputs include wind tunnel stagnation pressure, the specific gas constant, the ratio of specific heats, and the nozzle throat radius. These model inputs are assumed to be known to sufficient accuracy so as to justify treating them as deterministic values as listed in Table 1.

5.2. Characterize uncertainties

After extensive measurements of the temperature in the stagnation plenum (i.e., the low speed region upstream of the nozzle

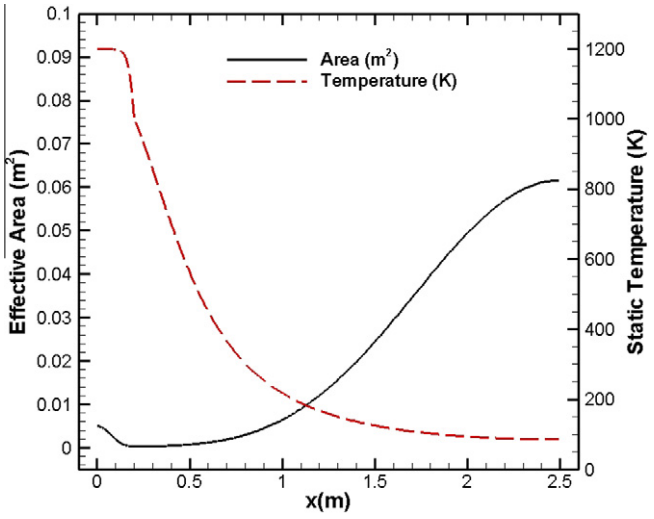

Fig. 10. Effective area and static temperature distributions through the hypersonic wind tunnel.

Table 1
Deterministic model inputs for hypersonic wind tunnel simulations.

Stagnation pressure, p_o	20 MPa
Specific gas constant, R	287 J/(kg K)
Ratio of specific heats, γ	1.4
Tunnel throat radius, r_t	0.01 m

throat), run-to-run variations are observed to be normally distributed about their mean value with a coefficient of variation of 3.33%. For the nominal mean stagnation temperature of T_o = 1200 K, the stagnation temperature is thus treated as an aleatory uncertainty characterized by a normal distribution with a standard deviation of σ = 40 K.

The thickness of the wind tunnel side-wall boundary layer is not measured, and the state of the boundary layer (laminar, transitional, or turbulent) over the length of the wind tunnel is not known. Separate boundary layer simulations are performed assuming fully laminar and fully turbulent tunnel wall boundary layers. After accounting for the displacement thickness of the boundary layer, the effective radius of the wind tunnel test section is found to be 0.14 m for the laminar boundary layer and 0.13 m for the turbulent one. Since the effective wind tunnel test section radius (accounting for the boundary layer displacement thickness) is expected to be somewhere within this range (depending on where boundary layer transition occurs in the nozzle), the test section radius is treated as an epistemic uncertainty using the interval r_{ts} = [0.13, 0.14] m.

5.3. Estimate uncertainty due to numerical approximation

The uncertainties arising due to the numerical solution to the quasi-one-dimensional Euler equations are now estimated. As a preliminary step, code verification studies are first performed on non-uniform meshes in order to ensure that there are no programming mistakes (i.e., bugs) in the code. For this simple case, an exact solution is available and is used to verify that the numerical solutions approach this exact solution at a second-order rate (i.e., as $\Delta\xi^2$).

All simulations are performed using double precision floating point numbers which provide approximately 14 significant digits in the computations. Furthermore, the simulations are advanced in time until a steady state is achieved. This steady-state is confirmed by inserting the current solution into the steady-state portion of the discrete equations and evaluating the non-zero remainder, i.e., the steady-state iterative residuals. These iterative residuals are converged 12 orders of magnitude from their initial levels, thus allowing both round-off and iterative errors to be neglected in the simulations.

The remaining numerical errors occur due to the spatial discretization of the domain. These discretization errors in the SRQ, test section temperature, are estimated by running simulations on three systematically-refined meshes (i.e., by the same refinement factor over the entire domain) for a set of nominal conditions with T_o = 1200 K and r_{ts} = 0.14 m. For systematically-refined meshes of 128, 256, and 512 cells, the test section static temperature was found to be 85.307, 85.824, and 85.954 K, respectively. The observed order of accuracy can then be computed as

$$\hat{p} = \frac{\ln\left(\frac{T_{coarse}-T_{med}}{T_{med}-T_{fine}}\right)}{\ln(r)}, \tag{7}$$

where r = 2 is the grid refinement factor. The observed order of accuracy for this case is $\hat{p} = 1.99$, which is very close to the formal order of accuracy of p = 2, thus Richardson extrapolation [16,26] can be used to reliably estimate the exact solution to the differential equations $\overline{T}$ as

$$\overline{T} = T_{fine} + \frac{T_{fine} - T_{med}}{r^p - 1} = 85.998 \text{ K}.$$

The uncertainty due to numerical discretization is estimated with a procedure similar to Roache's Grid Convergence Index [17,27], but applied to each mesh level. A factor of safety of 1.25 is used because the solutions appear to be in the asymptotic range (i.e., the observed order of accuracy equals the formal order). The resulting uncertainty estimate due to discretization on the coarsest mesh of 128 cells (the mesh used for uncertainty propagation and prediction) is

$$U_{NUM} = U_{DE} \cong 1.25\left|T_{coarse} - \overline{T}\right| = 0.86 \text{ K}.$$

5.4. Propagate input uncertainties through the model

Recall that aleatory uncertainties are characterized as a normally distributed CDF. For Monte Carlo sampling, a sample is chosen between 0 and 1 based on a uniform PDF. This probability is then mapped, using the CDF characterizing the input uncertainty, to determine the corresponding value of the input quantity (see top of Fig. 11). When more than one aleatory uncertain input is present (e.g., x_1, x_2, and x_3), Monte Carlo sampling randomly (and independently) picks probabilities for each of the input parameters as shown in Fig. 11. Once the input parameter samples are chosen, the model is used to compute a SRQ (y) for each sample. This sequence of SRQs is then ordered from smallest to largest, making up the abscissa of the CDF of the SRQ. The ordinate is found by separating the corresponding probabilities into equally-spaced divisions, where the incremental change in cumulative probability is $1/N$ for each Monte Carlo sample and N is the total number of samples (see bottom of Fig. 11). The CDF of the SRQ is the mapping of the uncertain inputs through the model to obtain the uncertainty in the model outputs.

For *epistemic uncertainties*, the issue is one of determining what combination of possible values over the range of the epistemic interval (in combination with all of the other uncertainties in the analysis) produces the largest range (i.e., p-box) for the SRQ of interest. As a result, the propagation of epistemic uncertainty can actually be formulated as a constrained optimization problem. The simplest way to formulate this optimization problem is through sampling, with Latin hypercube sampling (LHS) being the recommended approach [43,44]. For LHS over a single uncertain input, the probabilities are separated into a number of equally-sized divisions and one sample is randomly chosen in each division. Since there is absolutely no structure over the range of the interval, an appropriate structure for sampling over epistemic uncertainties would be a combinatorial design. The number of samples, M, of the epistemic uncertainties must be sufficiently large to ensure satisfactory coverage of the combinations of all of the epistemic uncertainties in the mapping to the SRQs. Based on the work of Refs. [44–46], we recommend that a minimum of three LHS samples be taken for each epistemic uncertainty, in *combination* with all of the remaining epistemic uncertainties [1]. For example, if m is the number of epistemic uncertainties, the minimum number of samples would increase as m^3. Recall that for *each* of these combinations, one must compute all of the samples for the aleatory uncertainties. For more than a handful of epistemic uncertainties, the total number of samples required for convergence becomes extraordinarily large, and other approaches should be considered [44–46].

For this example problem, both aleatory and epistemic uncertainties exist in the model inputs. The epistemic uncertainty in the effective wind tunnel test section radius is propagated using

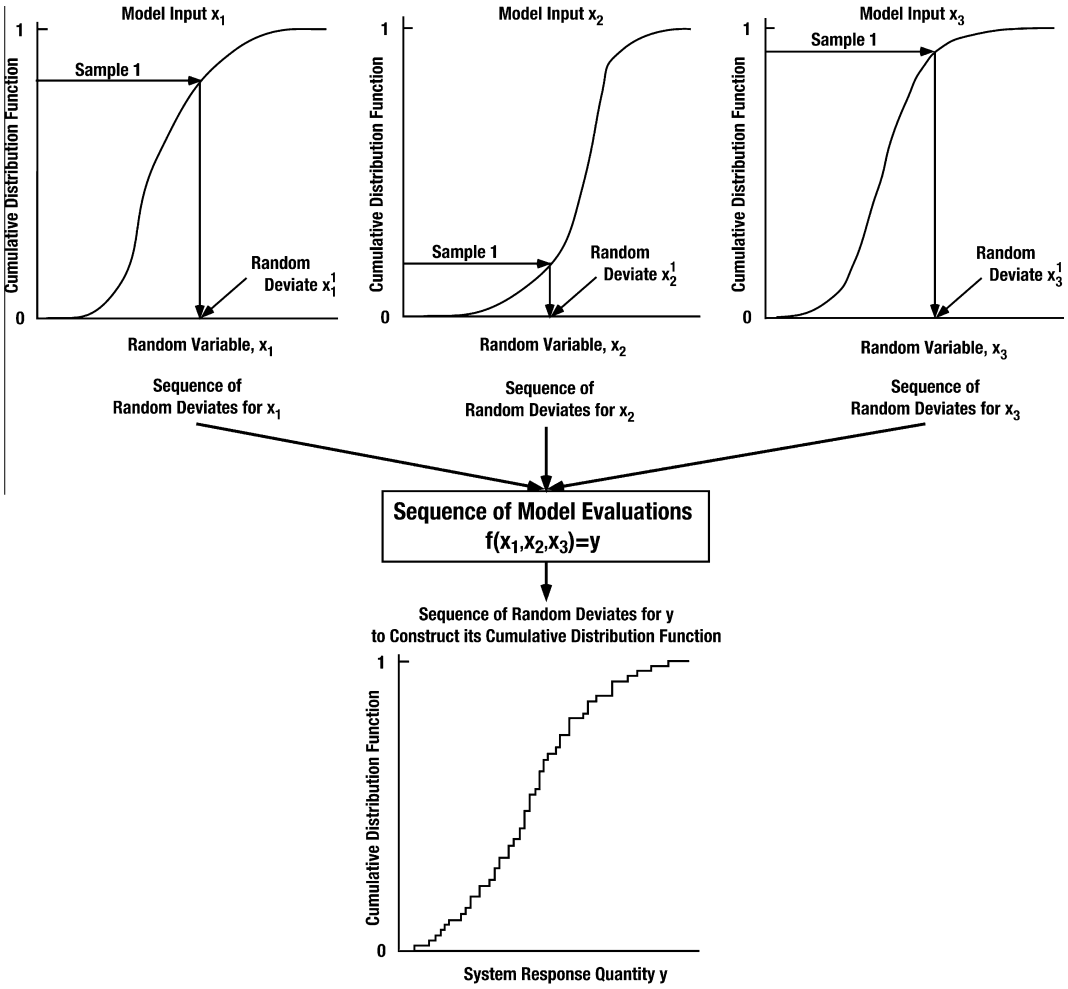

Fig. 11. Monte Carlo sampling to propagate aleatory uncertainties through a model (from [1]).

Latin hypercube sampling as discussed above. The interval for the test section radius is decomposed into 10 sub-intervals, and a random sample is selected from each of these sub-intervals using a uniform distribution (i.e., a pseudo-random number). Probabilities of 0 and 1 for the first and last sub-intervals, respectively, are used to ensure that the endpoints of the intervals are included. For each of these 10 samples of the epistemic uncertainty, the aleatory uncertainty in the stagnation temperature is propagated through the model using simple Monte Carlo sampling with 100 samples. After each of these inner loops over the aleatory uncertainty variable, a conditional CDF is generated which is conditioned on the value of the epistemic uncertain variable (r_{ts}). The complete ensemble of 10 CDFs is given in Fig. 12 below, and the widest extent of this ensemble in the SRQ coordinate is used to construct the *p*-box.

5.5. Estimate model form uncertainty

In order to estimate the model form uncertainty, three simple validation experiments are designed at stagnation pressures of 7, 10, and 12 MPa. Since the Reynolds number is proportional to the pressure, these values are sufficiently low that the boundary layer can be safely assumed to be laminar (i.e., $r_{ts} = 0.14$ m); note that the stagnation pressure for the conditions where the uncertainty is to be estimated is 20 MPa where the laminar/turbulent state of the boundary layer is unknown. The stagnation temperature measured as part of the validation experiments is again found to be normally distributed from run-to-run with a mean and standard deviation of 1200 K and 40 K, respectively. All other model inputs are assumed to be deterministic. Since an actual experiment is

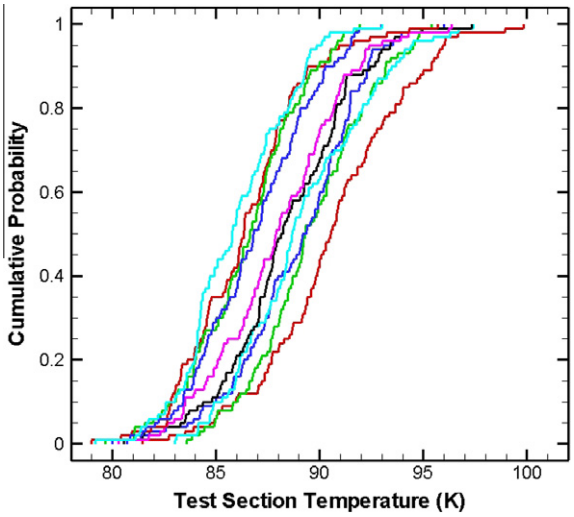

Fig. 12. Ensemble of Cumulative Distribution Functions (CDFs) generated by propagating the aleatory uncertainty in the stagnation temperature and the epistemic uncertainty in the effective test section radius.

not performed for the purposes of this illustrative example, a hypothetical "synthetic" set of experimental data is chosen. For the validation experiment at a stagnation pressure of p_o = 10 MPa, the ten synthetic measurements of the SRQ (test section static temperature) are chosen to be: SRQ_{EXP} = [78.5, 80.2, 81.6, 81.8, 81.9, 82.5, 82.7, 83.6, 84.7, 86.4] K. The synthetic experimental data were obtained using a combination of the present mathematical model, an assumed model bias error, and a random experimental error in the SRQ. While performing an actual validation assessment, these values would come directly from experimental observations. The CDF formed from these experimental observations is given in Fig. 13. Also shown in the figure is the CDF found by propagating the aleatory uncertainty in the stagnation temperature through the model. The area validation metric is then computed as the area between these two CDFs and is found to be d = 2.89 K. Similar area validation metrics are computed for the validation experiments at stagnation pressures of 7 and 12 MPa resulting in values of d = 3.1 and 2.8 K, respectively.

Now that the area validation metric d has been computed for the three validation experiments, it must be extrapolated to the conditions of interest (i.e., p_o = 20 MPa). First, a linear regression fit of the area validation metric is computed as a function of the stagnation pressure. While a simple linear regression was chosen to illustrate the concepts in this example, other regressions could be used as dictated by the amount of data available and its character. For notational simplicity, let the independent variable (stagnation pressure, p_o) be represented by x and the dependent variable (the area validation metric, d) be represented by y. The values used for the statistical analysis are thus given in Table 2.

A linear (least squares) regression fit of the data from Table 2 results in

$$\hat{y} = 3.518 - 0.0608x \text{ K}.$$

Next, a prediction interval [47] is computed at the stagnation pressure where we are interested in estimating the model form uncertainty (i.e., x = 20 MPa). This prediction interval provides the range of possible expected values of the area validation metric for an experiment conducted at a new condition and is given by

$$\hat{y} \pm t_{\alpha/2,N-d} s \sqrt{1 + \frac{1}{N} + \frac{N(x-\bar{x})^2}{N\sum_N x_i^2 - \left(\sum_N x_i\right)^2}}. \quad (8)$$

Table 2
Stagnation pressure (x) and area validation metric (y) used for statistical analysis.

x_i (MPa)	y_i (K)
7	3.1
10	2.89
12	2.8

Here N is the number of validation experiments conducted (N = 3), x is the stagnation pressure where we wish to estimate the prediction interval (x = 20 MPa), $t_{\alpha/2,N-d}$ is the $1-\alpha/2$ quantile of the student's t distribution for the degrees of freedom d (here d = 2 since there are two parameters in the regression fit and for the two-sided 95% interval $t_{95\%,1}$ = 12.71), and s is the square root of the mean square error of the regression fit

$$s = \left[\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N-2}\right]^{1/2} = 0.02433 \text{ K}.$$

The resulting 95% prediction interval for the area validation metric at p_o = 20 MPa is d = 2.30 ± 0.97 K and is shown graphically in Fig. 14. Note that these prediction intervals will always be larger than the corresponding confidence intervals of the regression fit because they represent a predicted difference between two random variables, i.e., the measured and the predicted SRQ. For the final estimation of model form uncertainty at the conditions of interest, the largest value of the 95% prediction interval is used, or d = 2.30 + 0.97 K = 3.27 K.

Note that if some SRQ other than test section static temperature is of interest, then one could not use the approach discussed here. One would instead have to use a different approach such as described in Refs. [13,38], or [39].

5.6. Determine total uncertainty in the SRQ

In order to determine the total predictive uncertainty in the SRQ (test section static temperature), one begins with the p-box determined by propagating aleatory and epistemic uncertainties in the model inputs through the model for the conditions of interest (p_o = 20 MPa). The bounding limits from the ensemble of CDFs from Fig. 12 are shown graphically by the blue region of Fig. 15. Then, one appends the estimated model form uncertainty found by

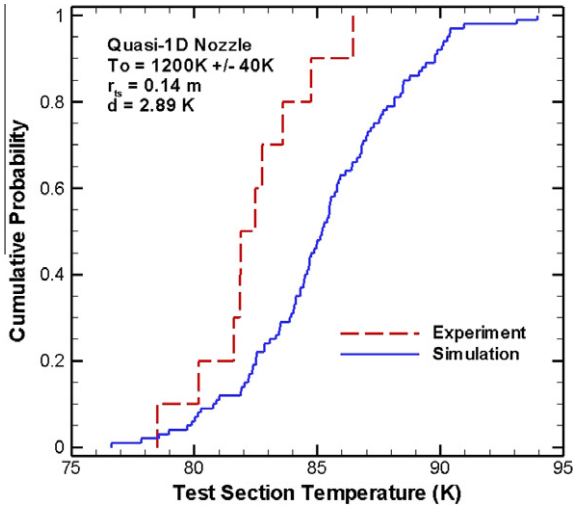

Fig. 13. Area validation metric d computed as the area between the experimental and simulation CDFs for a stagnation pressure of 10 MPa.

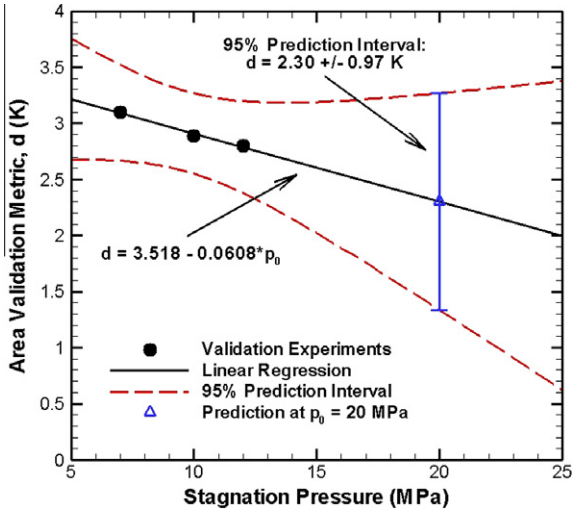

Fig. 14. Extrapolation of the area validation metric d from the conditions of the validation experiments to the condition where the model form uncertainty is to be estimated (p_o = 20 MPa).

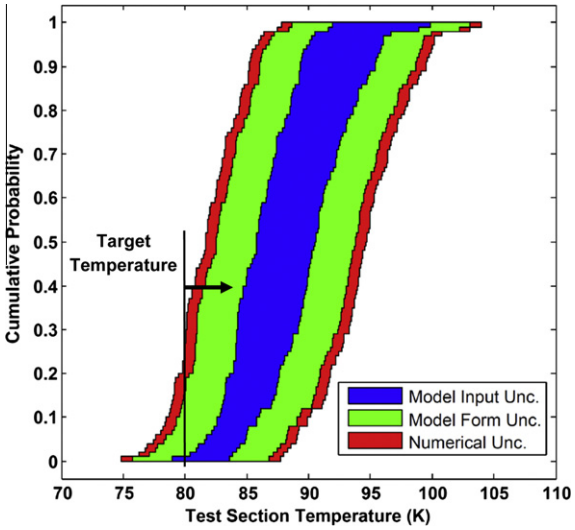

Fig. 15. Total uncertainty in the test section static temperature accounting for the propagation of model input uncertainty (blue), model form uncertainty (green), and uncertainty due to numerical approximations (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

extrapolating the area validation metric to the conditions of interest, i.e., $d = 3.27$ K, to the left and right sides of the p-box. The regions denoting model form uncertainty are shown in green in Fig. 15. Finally, the estimated uncertainty due to numerical approximation $U_{NUM} = 0.86$ K is appended to the left and right sides of the p-box, as shown by the red regions in Fig. 15.

The resulting nondeterministic prediction of uncertainty in the test section static temperature is given in Fig. 15. Note that the total predictive uncertainty is primarily due to the aleatory uncertainty in the model inputs (i.e., the stagnation temperature) and the estimated model form uncertainty. No probability structure is assumed inside the probability box since the various sources of epistemic uncertainty were treated as interval-valued quantities. Since the stated requirement was that the test section temperature should not fall below 80 K due to the possibility of condensation, the results of applying this VV&UQ framework can be interpreted in the following ways: (a) there is as high as a 25% chance that the test section temperature will fall below 80 K with 95% confidence and (b) the interval valued probability of $T < 80$ K is $[0, 0.25]$ with 95% confidence. The result of this total uncertainty quantification exercise is that if one desires 95% confidence that no condensation will occur in the test section (i.e., $T \geqslant 80$ K), then the nominal stagnation temperature should be increased, since currently there is only a 75% likelihood that this requirement will be met.

6. Conclusions

The framework for verification, validation, and uncertainty quantification (VV&UQ) in scientific computing presented here represents a conceptual shift in the way that scientific and engineering predictions are performed and presented to decision makers. The philosophy of the present approach is to rigorously segregate aleatory and epistemic uncertainties in input quantities, and to explicitly account for numerical solution errors and model form uncertainty directly in terms of the predicted system response quantities of interest. In this way the decision maker is clearly and unambiguously shown the uncertainty in the predicted quantities of interest. For example, if the model has been found to be inaccurate in comparisons with relevant experimental data, the decision maker will starkly see this in any new predictions; as opposed to the approach of immediately incorporating newly obtained experimental data for system responses into the model by way of re-calibration of model parameters. We believe our proposed approach to presenting predictive uncertainty to decision makers is needed to reduce the tendency of under estimating predictive uncertainty, especially when large extrapolation of models is required. We believe that with this clearer picture of the uncertainties, the decision maker is better served. This approach is particularly important for predictions of high-consequence systems, such as those where human life, the public safety, national security, or the future of a company is at stake.

Acknowledgments

The authors thank Scott Ferson of Applied Biomathematics, Jon Helton (consultant), and Tim Trucano of Sandia National Laboratories who were instrumental in developing many of the ideas behind the predictive uncertainty framework described here.

References

[1] W.L. Oberkampf, C.J. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, Cambridge, 2010.
[2] M.G. Morgan, M. Henrion, Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis, first ed., Cambridge University Press, Cambridge, UK, 1990.
[3] A.C. Cullen, H.C. Frey, Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs, Plenum Press, New York, 1999.
[4] D. Vose, Risk Analysis: A Quantitative Guide, third ed., Wiley, New York, 2008.
[5] Y.Y. Haimes, Risk Modeling, Assessment, and Management, third ed., John Wiley, New York, 2009.
[6] T. Bedford, R. Cooke, Probabilistic Risk Analysis: Foundations and Methods, Cambridge University Press, Cambridge, UK, 2001.
[7] J.K. Ghosh, M. Delampady, T. Samanta, An Introduction to Bayesian Analysis: Theory and Methods, Springer, Berlin, 2006.
[8] D. Sivia, J. Skilling, Data Analysis: A Bayesian Tutorial, second ed., Oxford University Press, Oxford, 2006.
[9] AIAA, Guide for the verification and validation of computational fluid dynamics simulations, American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA, 1998.
[10] ASME, Guide for Verification and Validation in Computational Solid Mechanics, American Society of Mechanical Engineers, ASME Standard V&V 10-2006, New York, NY, 2006.
[11] H.W. Coleman, F. Stern, Uncertainties and CFD code validation, J. Fluids Engrg. 119 (1997) 795–803.
[12] F. Stern, R.V. Wilson, H.W. Coleman, E.G. Paterson, Comprehensive approach to verification and validation of CFD simulations – Part 1: methodology and procedures, J. Fluids Engrg. 123 (2001) 793–802.
[13] S. Ferson, W.L. Oberkampf, L. Ginzburg, Model validation and predictive capability for the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (2008) 2408–2430.
[14] S. Ferson, L.R. Ginzburg, Different methods are needed to propagate ignorance and variability, Reliab. Engrg. Syst. Safety 54 (1996) 133–144.
[15] S. Ferson, J.G. Hajagos, Arithmetic with uncertain numbers: rigorous and (often) best possible answers, Reliab. Engrg. Syst. Safety 85 (2004) 135–152.
[16] C.J. Roy, Review of discretization error estimators in scientific computing, in: AIAA Paper 2010-0126, 2010.
[17] P.J. Roache, Verification and Validation in Computational Science and Engineering, Hermosa Publishers, Albuquerque, New Mexico, 1998.
[18] R.E. Bank, Hierarchical bases and the finite element method, Acta Numer. 5 (1996) 1–45.
[19] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates, part 2: error estimates and adaptivity, Int. J. Num. Meth. Engrg. 33 (1992) 1365–1382.
[20] P.A. Cavallo, N. Sinha, Error quantification for computational aerodynamics using an error transport equation, J. Aircraft 44 (2007) 1954–1963.
[21] T.I-P. Shih, B.R. Williams, Development and evaluation of an a posteriori method for estimating and correcting grid-induced errors in solutions of the Navier–Stokes equations, in: AIAA Paper 2009-1499, 2009.
[22] R.D. Skeel, Thirteen ways to estimate global error, Numer. Math. 48 (1986) 1–20.
[23] M. Ainsworth, J.T. Oden, A posteriori error estimation in finite element analysis, Comput. Methods Appl. Mech. Engrg. 142 (1997) 1–88.
[24] M. Ainsworth, J.T. Oden, A Posteriori Error Estimation in Finite Element Analysis, Wiley Interscience, New York, 2000.
[25] N.A. Pierce, M.B. Giles, Adjoint recovery of superconvergent functionals from PDE approximations, SIAM Review 42 (2000) 247–264.
[26] C.J. Roy, Review of code and solution verification procedures for computational simulation, J. Comput. Phys. 205 (2005) 131–156.

[27] P.J. Roache, Perspective: a method for uniform reporting of grid refinement studies, J. Fluids Engrg. 116 (1994) 405–413.
[28] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[29] R. Ghanem, P. Spanos, Stochastic Finite Elements: A Spectral Approach, Springer-Verlag, 1991.
[30] H. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, Ann. Rev. Fluid Mech. 41 (2009) 35–52.
[31] M.S. Eldred, L.P. Swiler, Efficient algorithms for mixed aleatory-epistemic uncertainty quantification with application to radiation-hardened electronics. Part I: Algorithms and Benchmark Results, Sandia National Laboratories, SAND2009-5805, Albuquerque, NM, 2009.
[32] M.S. Eldred, C.G. Webster, P. Constantine, Evaluation of non-intrusive approaches for Wiener-Askey generalized polynomial chaos, in: AIAA Paper 2008-1892, 2008.
[33] M.S. Eldred, Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design, in: AIAA Paper 2009-2274, 2009.
[34] A.A. Giunta, J.M. McFarland, L.P. Swiler, M.S. Eldred, The promise and peril of uncertainty quantification using response surface approximations, Struct. Infrastruct. Engrg. 2 (2006) 175–189.
[35] H.W. Coleman, F. Stern, Uncertainties and CFD code validation, J. Fluids Engrg. 119 (1997) 795–803.
[36] W.L. Oberkampf, T.G. Trucano, Verification and validation in computational fluid dynamics, Prog. Aerospace Sci. 38 (2002) 209–272.
[37] W.L. Oberkampf, M.F. Barone, Measures of agreement between computation and experiment: validation metrics, J. Comput. Phys. 217 (2006) 5–36.
[38] W.L. Oberkampf, S. Ferson, Model validation under both aleatory and epistemic uncertainty, in: NATO/RTO Symposium on Computational Uncertainty in Military Vehicle Design, NATO, AVT-147/RSY-022, Athens, Greece, 2007.
[39] I. Babuska, F. Nobile, R. Tempone, A systematic approach to model validation based on Bayesian updates and prediction related rejection criteria, Comput. Methods Appl. Mech. Engrg. 197 (2008) 2517–2539.
[40] T.G. Trucano, M. Pilch, W.L. Oberkampf, General Concepts for Experimental Validation of ASCI Code Applications, Sandia National Laboratories, SAND2002-0341, Albuquerque, NM, 2002.
[41] P.L. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, J. Comput. Phys. 43 (1981) 357–372.
[42] B. van Leer, Towards the ultimate conservative difference scheme, V.A second order sequel to Godunov's method, J. Comput. Phys. 32 (1979) 101–136.
[43] J.C. Helton, Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty, J. Statist. Comput. Simul. 57 (1997) 3–76.
[44] C.J. Sallaberry, J.C. Helton, S.C. Hora, Extension of Latin hypercube samples with correlated variables, Reliab. Engrg. Syst. Safety 93 (2008) 1047–1059.
[45] S. Ferson, W.T. Tucker, Sensitivity in Risk Analyses with Uncertain Numbers, Sandia National Laboratories, SAND2006-2801, Albuquerque, NM, 2006.
[46] V. Kreinovich, J. Beck, C. Ferregut, A. Sanchez, G.R. Keller, M. Averill, S.A. Starks, Monte-Carlo-type techniques for processing interval uncertainty, and their potential engineering applications, Reliab. Comput. 13 (2007) 25–69.
[47] J. Devore, Probability and Statistics for Engineering and the Sciences, seventh ed., Brooks/Cole, Belmont, CA, 2009.