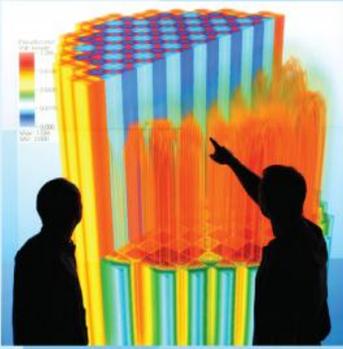




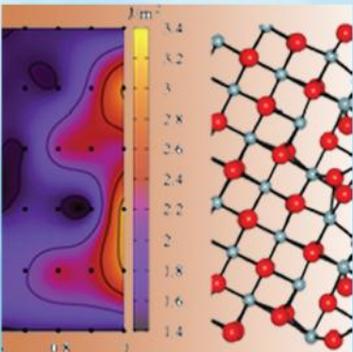
Power uprates
and plant life extension



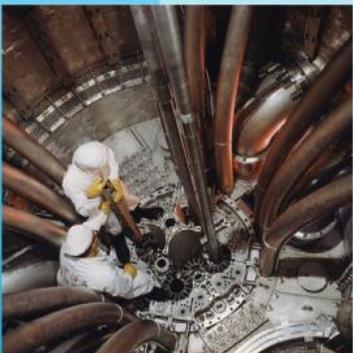
Engineering design
and analysis



Science-enabling
high performance
computing



Fundamental science



Plant operational data

L3:VUQ.VVDA.P2-2.03

William Rider

SNL

Completed: 3/28/11



U.S. DEPARTMENT OF
ENERGY

Nuclear Energy

Verification, Validation and Uncertainty Quantification Workflow in CASL

William J. Rider
Computational Shock and MultiPhysics Department

James R. Kamm and V. Gregory Weirs
Optimization and Uncertainty Quantification Department

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185

Dan G. Cacui
Department of Nuclear Engineering
North Carolina State University
Raleigh, NC 27695-7909

Abstract

The overall conduct of verification, validation and uncertainty quantification (VVUQ) is discussed through the construction of a workflow relevant to CASL. The workflow contained herein is defined at a high level and constitutes an overview of the activity. Nonetheless, the workflow represents an essential activity in predictive simulation and modeling. VVUQ is complex and necessarily hierarchical in nature. The particular characteristics of VVUQ elements depend upon where the VVUQ activity takes place in the overall hierarchy of physics and models. In this document, we focus on the differences between and interplay among validation, calibration and UQ, as well as the difference between UQ and sensitivity analysis. We also discuss the complementary approach of best estimate plus reduced uncertainties (BERU) approach in some detail. The discussion in this document is at a relatively high level and attempts to explain the key issues associated with the overall conduct of VVUQ. The intention is that CASL developers and analysts can refer to this document for guidance regarding how VVUQ analyses fit into their efforts toward conducting truly predictive calculations.

1. OVERVIEW

Verification, Validation, and Uncertainty Quantification (VVUQ) is a complex and often controversial set of procedures for determining the overall quality of a simulation activity. The definitions of what properly constitutes VVUQ are subject to debate and the field continues to evolve. Aspects of the overall process we describe have been addressed by many researchers, e.g., [AIA98, ASM06, Han01, Kam08, Knu03, Obe02, Obe04, Obe07a, Obe10, Ore94, Rid10, Roa04, Roa09, Roy05, Roy10, Sar98, Sar01, Sor07, Sch06, Ste01, Ste05a, Tru03, Tru06]. Despite all that has been written about VVUQ, there remain diverse and often imperfect definitions for each element of the process. Figure 1 contains the Verification & Validation Flowchart from the ASME guide on V&V for Solid Mechanics [Sch06], which provides one high-level overview of the entire VVUQ approach to computational modeling. This document contains a general workflow for verification, validation and uncertainty quantification that is based on the concepts developed by previous investigators and is tailored with the broad CASL objectives in mind.

Cacuci and Ionescu-Bujor [Cac10] presented a new and rigorous mathematical methodology for *BERU*-predictions through data assimilation and *simultaneous calibration of model parameters and responses* for a generic time-dependent physical system. This work has generalized the data assimilation methodologies currently used in geophysical sciences (see, e.g., [Kal03, Lew06]). The methodology of Cacuci and Ionescu-Bujor [Cac10] also provides quantitative indicators (based on uncertainties and sensitivities) for determining the degree of agreement (or disagreement) relevant to the assimilation and best-estimate adjustment of computational and experimental parameters and responses. These quantitative indicators serve as the basic elements for *quantitative model extrapolation* (i.e., prediction of uncertainty in new environments or conditions of interest, including both untested parts of the parameter space and higher levels of system complexity in the validation hierarchy) and *estimation of the validation domain*.

The intention of this presentation is not to be dogmatic regarding a particular VVUQ approach (including the one espoused), but rather to inform CASL analysts of the issues and to suggest approaches by which to enhance the VVUQ content of their work.

2. THE SCIENTIFIC SIMULATION CONTEXT

Scientific simulations are run for a variety of reasons, including scientific discovery (e.g., in astrophysics) and in a capacity to inform decision-making (e.g., in many engineering, national security and manufacturing fields). In this document, we focus on scientific simulation in the latter role. The goal of scientific simulation is to provide

answers to particular questions; these answers must be inferred from numerical values output by the simulations of the specified problems. The numerical values are rarely precise and, as such, contain intrinsic uncertainty, which should be regarded as essential to quantify as part of the simulation activity. We echo the terminology and sentiment of Trucano et al. [Tru06], who focus on “confident prediction” and whose observation below captures the essence of the use of scientific simulations in the context of CASL:

“In computational science, of course, to some degree confidence is correlated with belief in the quantitative numerical, mathematical and physical accuracy of a calculated prediction. Intuitively, confident prediction then implies having some level of confidence or belief in the quantitative accuracy of the prediction. This further implies a willingness to use the prediction in some meaningful way, for example in a decision process.”

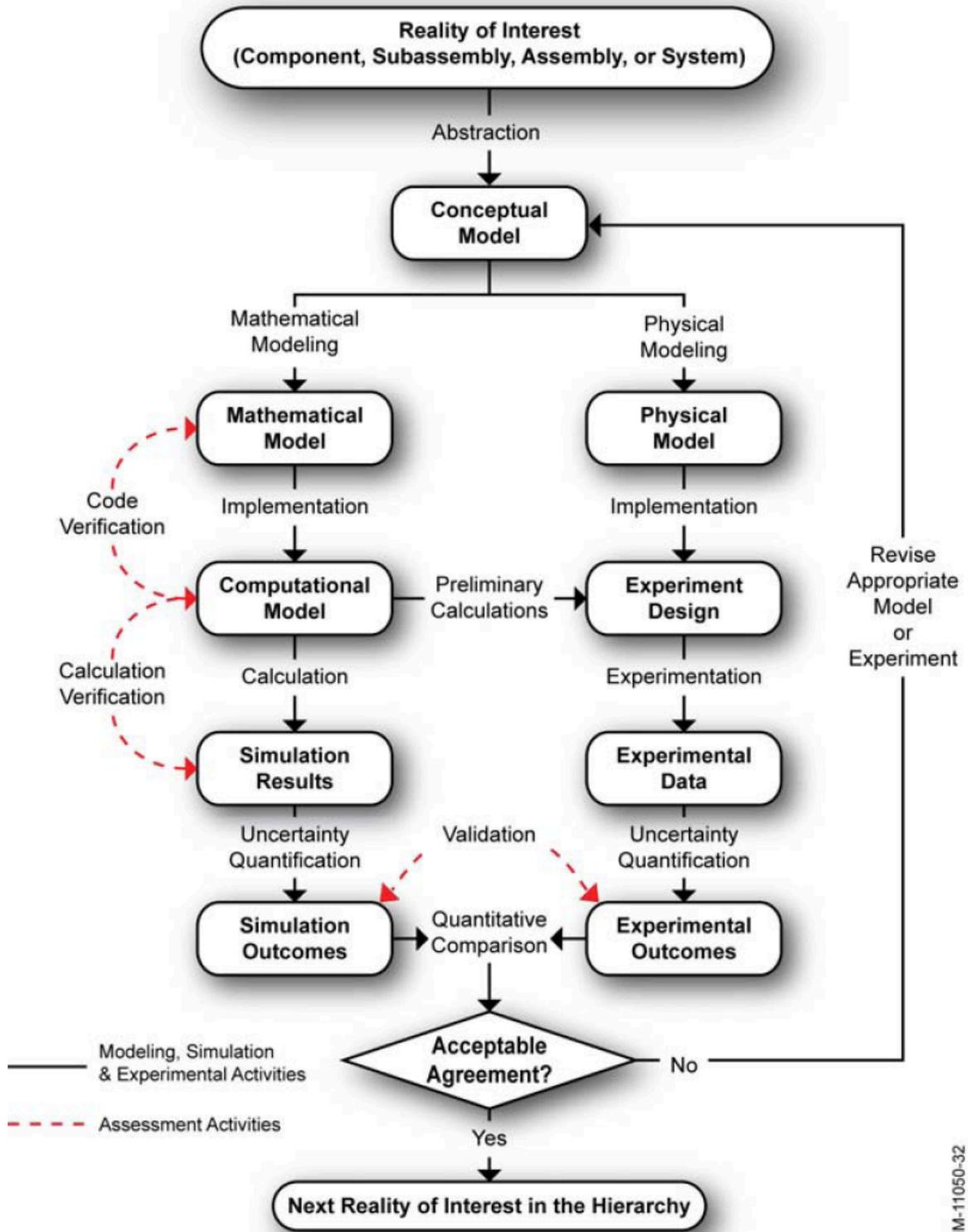


Figure 1. The overall Verification & Validation Flowchart taken from the ASME guide on V&V for Solid Mechanics [Sch06]. This document covers the analyses and boxes at the bottom center of the parallel columns, encompassing Uncertainty Quantification and Validation.

CASL modeling efforts necessarily involve complex, multiphysics simulations, the results of which are unavoidably *approximate* solutions. This immediately raises important questions. How good are the approximations? Is the approximate solution good enough (in the context of the decision)? Or considering the quality of the approximate solution can a decision be comfortably made? The former question is difficult to address because the exact solution is not known (if it were, scientific simulation would not be needed). Differences between the simulation results and reality arise from many sources. Verification, Validation, and Uncertainty Quantification (VVUQ) are different forms of assessment intended to quantify various aspects of these differences; see [AIA98, ASM06, Han01, Kam08, Knu03, Obe02, Obe04, Obe07a, Obe10, Ore94, Rid10, Roa04, Roa09, Roy05, Roy10, Sar98, Sar01, Sor07, Sch06, Ste01, Ste05a, Tru03, Tru06]. These assessments address the first question and help inform the decision maker regarding the second question.

The accuracy of the approximate solutions (and their corresponding uncertainties) for a complex, multiphysics problem of interest is directly related to the fidelity of the solutions for simpler “component” problems, with additional errors arising from coupling and the increased complexity. For simpler foundational problems, different (usually better and inherently less error-prone) sources of information can be used to provide the estimation of uncertainty without the complications of the coupled models. These circumstances lead to a hierarchical view of validation problems, with the problem of interest at the top of a notional pyramid and the simplest problems at the base. Scenarios at the base of this conceptual pyramid can be solved (simulated or examined experimentally) with demonstrably greater confidence and more certainty, as these simpler problems are, by definition, less complex and better understood. In this context, “simpler” means, e.g., fewer physical phenomena, less complex submodels, limited physical scale, or reduced geometric complexity. Uncertainty and empiricism naturally increase as the top of the pyramid is approached. At the highest level the problem is, by definition, complicated and not well understood. This hierarchy is a useful paradigm by which to identify where knowledge (e.g., of a particular physics model) is insufficient and where empiricism is a placeholder for this missing knowledge. The purpose of hierarchical validation is not to eliminate empiricism, but to restrict it to the parts of the hierarchy where it truly represents a lack of knowledge. This concept is reflected in the CASL-specific validation hierarchy, shown in Fig. 2.

In consideration of this hierarchical paradigm, validation can be understood as an intrinsically multiscale process. The following conceptual approach can be followed to address the overall validation process:

1. Establish a multiscale hierarchy of physics and phenomena to be validated.
2. Determine simulations and experiments that can be conducted to examine the hierarchy.
3. Execute the validation process on each level of the hierarchy.
4. Assess the overall validation.

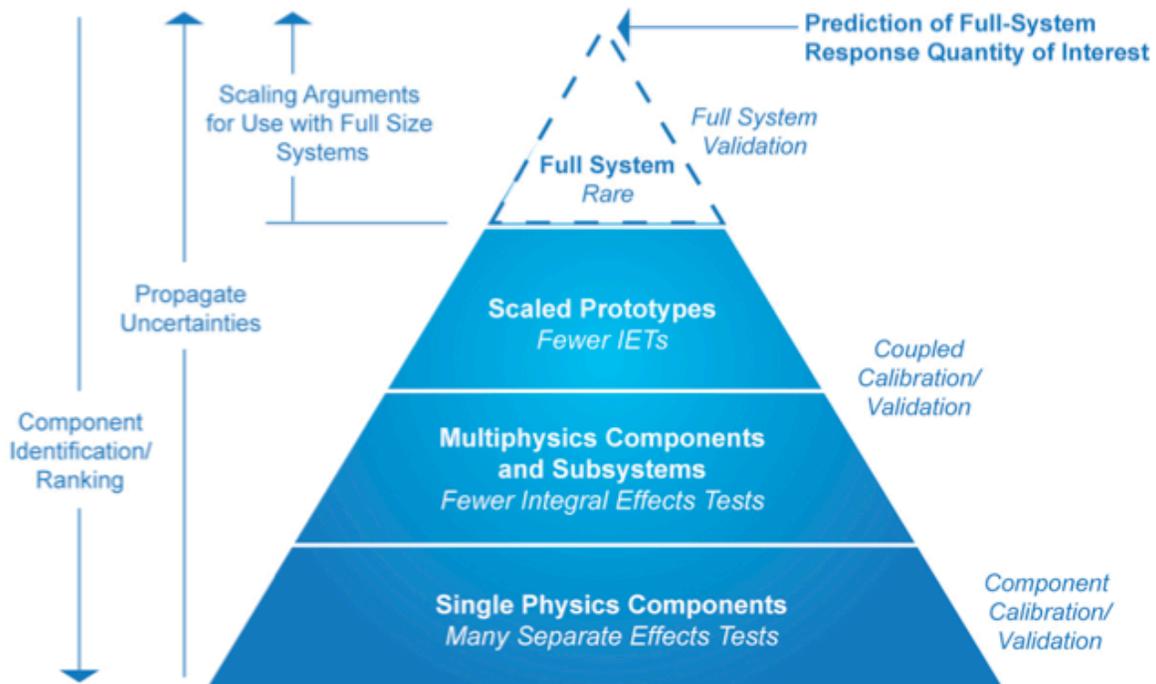


Figure 2. The overall Validation Hierarchy from the CASL proposal.

A goal of this framework is to reduce uncertainties at each conceptual level of the pyramid. The practical purpose of this hierarchical approach is not to eliminate uncertainty, but to identify where uncertainty cannot be reduced. VVUQ can be used to help identify where, in the hierarchical modeling and simulation decomposition, empiricism and uncertainty are the greatest: these are the elements where improvements can have the largest impact. This process is necessarily imperfect and, at times, ambiguous; nevertheless, it can be used to focus efforts, e.g., by identifying where additional model calibration may be particularly useful. It is important to note that the intermediate portions of this pyramid present the most difficult validation activities and, so, are absent from the process. Part of the challenge of VVUQ is to fill in the missing portions of the hierarchy and provide a more complete picture of the nature and sources of uncertainty.

Regarding this observation, we caution that calibration is a potentially dangerous activity when carried out in an ad hoc manner, e.g., as a process by which to determine modeling coefficients where no (or limited) first-principles guidance is available. Calibration can account for inappropriate effects and, so, can mask what is in fact uncertainty and lack of knowledge. Consequently, calibration experiments are appropriate and well suited at the lowest level in a modeling hierarchy. Calibration at the higher levels of the hierarchy may be less rigorous if not applied in a principled fashion. Thus, calibration is an acknowledged necessity, but it is increasingly problematic as the modeling moves up the problem hierarchy.

It is important to acknowledge that empirical/calibrated models may also be appropriate when they replace better-grounded models, but do not increase the overall uncertainty of the problem of interest. This endeavor should be conducted with exceptional care because of its seeming violation of scientific principle and potential for harm when the conditions fall outside the range for the empirical calibration. The validation of the subproblem may demonstrate the sufficiency of the empirical model for replacing the better-grounded model; however, this is different from calibrating an empirical model to experimental data because a better-grounded model does not exist. In this case, the empirically calibrated model actually results in a lower uncertainty, but the risks intrinsic in this approach must be kept clearly in mind.

Below, we present general workflows for applying validation and uncertainty quantification. These workflows become more specific by the context in which they are applied. The validation of the top-level problem in the pyramid will be quite different than that at the lowest level, because the context is different. Likewise, uncertainty quantification will be context dependent: it can be applied as a part of validation at any level of the hierarchy but also across levels.

3. VVUQ WORKFLOW

At the highest level, the workflow for conducting validation at each level of a hierarchy can be described concisely in a short series of steps, each of which will be described in more detail in their own workflows. The conceptual validation workflow we advocate is described at this level in Figure 3. The steps are the following:

1. Begin the validation process.
2. The simulation path of the process:
 - a. The conduct of the simulations defines the baseline simulation model.
 - b. The determination of the simulation uncertainty defines the sources and magnitude of uncertainty in the simulation model.
3. The experimental path of the process:
 - a. The conduct of the experiment provides data to compare the simulation results to reality.
 - b. The determination of the experimental uncertainty provides the quality of the data used for simulation comparison.
4. The validation assessment provides the overall context for the simulation and validation outcomes in relation to their intended application.
5. The determination of the sufficiency of the validation status based on the defined requirements for the application of the simulation: if “yes” then proceed to prediction; if “no” then address the shortcomings in the process, i.e., refine simulations and/or experiments.

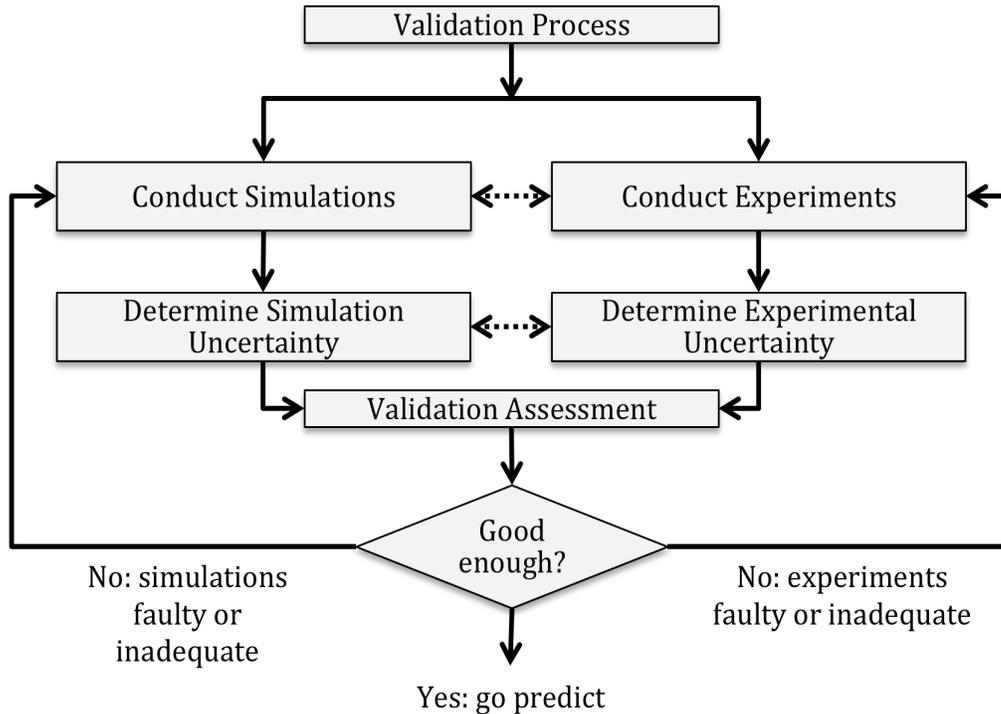


Figure 3. The overall Validation process as described in this document, encompassing both simulations and experiments, the associated Uncertainty Quantification analyses, and the ultimate validation assessment.

One complication to acknowledge at the outset of the process is the difference between concurrent and legacy experiments. Concurrent experiments are conducted in collaboration with the simulations, so that the two processes experience direct feedback from each other. In particular, dedicated validation experiments¹ are those concurrent experiments that are specifically designed to validate code calculations; they are executed and analyzed in concert with the corresponding simulations. [Obe01, Obe02]. The value of dedicated validation experiments cannot be overstated, as the corresponding data are particularly valuable for evaluating the simulation capability of a given code. Moreover, validation experiments are designed to be simulated with relatively less difficulty than traditional experiments, which are often designed for different purposes. Unfortunately, it can be challenging to identify resources for dedicated validation experiment campaigns, as these efforts are sometimes perceived to be less valuable than, say, scientific discovery or design experiments (i.e., traditional experimental science), a perspective that we argue should be resisted in CASL. Legacy data, on the other hand, come from experiments in the past; for the purposes of this discussion, legacy data will be more generally characterized as being from experiments that are not for the express purpose of validation or are not influenced by the simulation activity. Since legacy data may not have been

¹ As pointed out by Oberkampf [Obe01], “validation experiments” comprise a unique class that does not fall into the usual categories of scientific discovery experiments, model improvement experiments, or performance/acceptance tests.

generated in coordination with modern simulation tools and current VVUQ perspectives, such data may be of more limited value, as they often lack sufficient documentation to be used in support of a defensible simulation-validation effort. Lastly, we note that the use of experimental data to conduct calibration of simulation models will have a nearly identical structure as for validation, but the purpose of the activity is distinctly different in character.

Each of the steps in this overall process can be expanded into its own process, which we describe and depict below. This process can be applied at any level of the notional hierarchy presented earlier in the document.

3.1 Conduct Simulations (see Fig. 4)

1. The starting point is a simulation code that have been sufficiently SQAed and code verified.
2. The problem of interest (the experiment) is defined including models, submodels, initial and boundary conditions, as well as any supporting data.
3. The simulation codes' input is defined, incorporating models, data, initial and boundary conditions, and mesh/geometric description; this input should undergo a verification/SQA process.
4. The system response characteristics to be analyzed in the simulation should be defined as well as a means of comparison.
5. The simulation should be run on the computer.
6. The metrics are used to determine the viability and quality of the simulation for the problem of interest.
7. The simulations results are determined to either be or not be sufficient to move to the determination of uncertainty.

It is important that the simulations of the problem of interest be made as faithful as possible to the corresponding experiment. This issue can pose particular difficulties for validation and calibration of complex phenomena. This concept is characterized by Trucano et al. [Tru06] as the “alignment” between a simulation and an intended application. They categorize the code input parameters into two (possibly overlapping) sets, p_A (input parameters that specify alignment with the intended model application) and p_N (numerical parameters and other quantities required to execute a calculation and control its numerical accuracy). A principal virtue of validation experiments is that they may be designed and executed in coordination with the simulation tool to increase the experiment-simulation alignment.

3.2 Determination of Simulation Uncertainty (see Fig. 5)

1. The starting point is a simulation of sufficient quality or importance to examine simulation uncertainty.
2. The types and sources of uncertainty need to be determined. Sources of uncertainty include intrinsic variability, model form, model parameters, simulation choices, and numerical uncertainties. These uncertainties can be broadly categorized as aleatory (“irreducible” or “random”) or epistemic (“reducible” or “lack of knowledge”); see, e.g., [Tru06, Roy10] and references therein.

3. The metrics and methodology for examining for uncertainty determination should be established.
4. The simulation cases for determining uncertainty should be conducted.
5. Process the results of the simulations to determine the uncertainties from the defined sources.
6. Examine the sufficiency of the uncertainty estimation and their magnitude.

The literature on simulation uncertainty quantification is vast (see the citations above and their references) and the challenges associated with it are well recognized in the VVUQ community. It is important to acknowledge that addressing simulation uncertainty increases the technical difficulty of simulation activity (as well as the difficulty of subsequent simulation-experiment comparisons). The intrinsic statistical or probabilistic nature of uncertainty information is the unavoidable source of this issue.

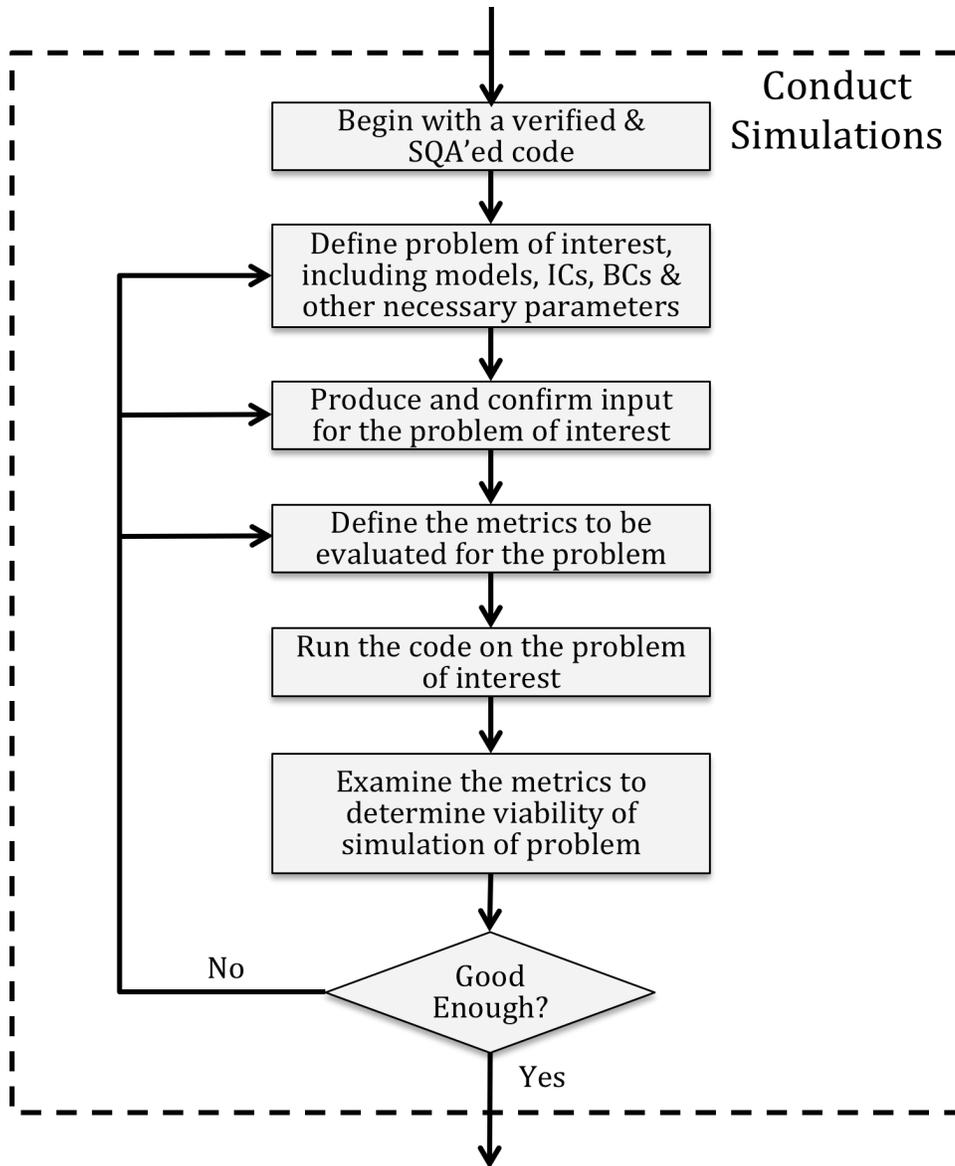


Figure 4. The overall Conduct Simulations process as described in this document is shown. This process clearly identifies three possible aspects that must be investigated if the quality of the simulation does not meet the modeler's expectations.

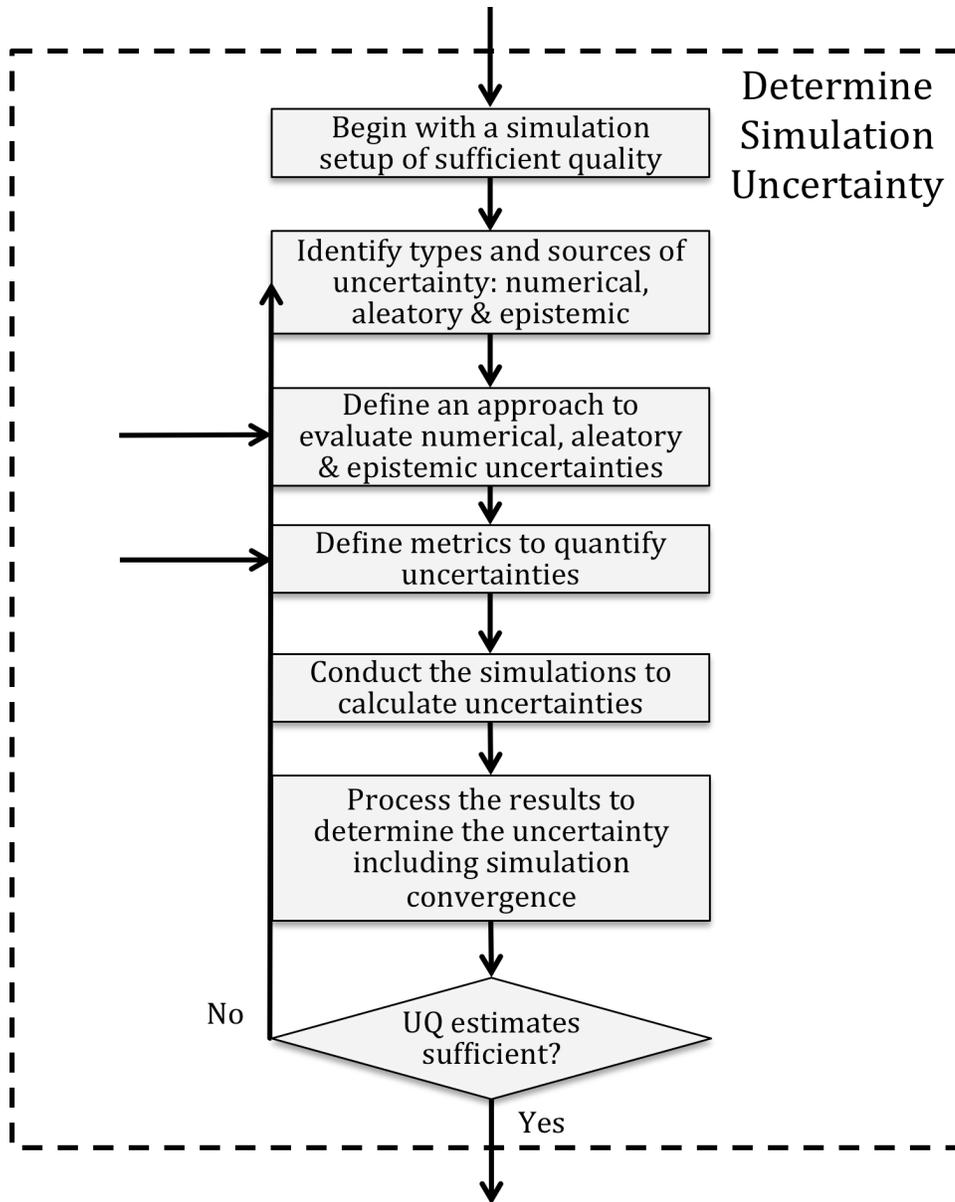


Figure 5. The overall Simulation Uncertainty Quantification process as described in this document is displayed. This process clearly identifies three possible aspects that must be investigated if the quality of the simulation UQ estimates does not meet the modeler’s expectations.

3.3 Conduct Experiments (see Fig. 6)

1. Begin the experiment at a well-characterized facility with defined quality controls.
2. Define the experimental setup including the physics of interest, initial and boundary conditions.
3. Identify the experimental parameters that can be explicitly controlled as part of the experimental operation.

4. Define the diagnostics used in the experiment including data capture, post-processing and related information.
5. Conduct an instance of the experiment.
 - a. Conduct replicate experiments (if possible) to determine the repeatability of the results, and/or assess the aleatory uncertainty in the experiment.
 - b. Where experiments are single time and unique, the physical phenomena must be carefully scrutinized.
6. Post-process the experimental data and assess the quality of the experimental data providing uncertainty estimates.

Ideally, experimental data used for validation are generated by experiments designed expressly for validation as opposed to scientific discovery (although experiments serving both purposes are inherently desirable). Such experiments are focused on generating high quality data, i.e., with high accuracy, with well-quantified errors and uncertainties, and of a highly repeatable nature. Experiments and simulations should be carried out in concert, with dynamic feedback between the two. This requires close coordination and frequent constructive discussion between computational scientists/engineers and experimentalists in the roles of collaborators. Of course, the larger the scale and greater the complexity of the experiment, the more difficult this objective is to achieve. Nonetheless, the greater the common understanding of techniques, issues, and practice in both experiments and simulations among all parties involved, the higher the quality of the VVUQ results.

3.4 Determination of Experimental Uncertainty (see Fig. 7)

1. Begin with well-defined experimental results
2. Identify the various sources of error in the experiment (measurement, data reduction, etc.) as well as their type (aleatory or epistemic).
3. Define resources for determining the magnitude of uncertainty.
4. Examine the experimental measurements for each uncertainty.
5. Reduce the data associated with the determination of uncertainty.
6. Post-process the uncertainty to separate the different effects, and produce an overall experimental uncertainty budget.
7. Determine the sufficiency of the uncertainty estimates.

The complete definition of experimental uncertainties and any detailed discussion of their determination is beyond the scope of this document. A few broad-based observations regarding the nature of experimental uncertainty are, however, in order. Generally speaking, experimental data have errors in their measurement, processing, inference, and statistics. Each of these errors has a fundamentally different character and has an analog in the computational domain. Measurement error is the most obvious, being the relative inability to produce an exact measurement of the real condition present in the experiment. Processing errors are introduced when raw signals are processed (often, through several steps) into values used in subsequent analysis or inference. Inference is often used to produce a “measurement” from the original or processed data. For example, particle image velocimetry (PIV) is based upon an inference from the actual

measurement. This inference is usually software and algorithm based, and as such is subject to many of the same sources of error as the computations being validated. Finally, the statistical error is the variability of a measurement over time, or with repeated conduct of the experiment. This type of error is strongly associated with aleatory uncertainty.

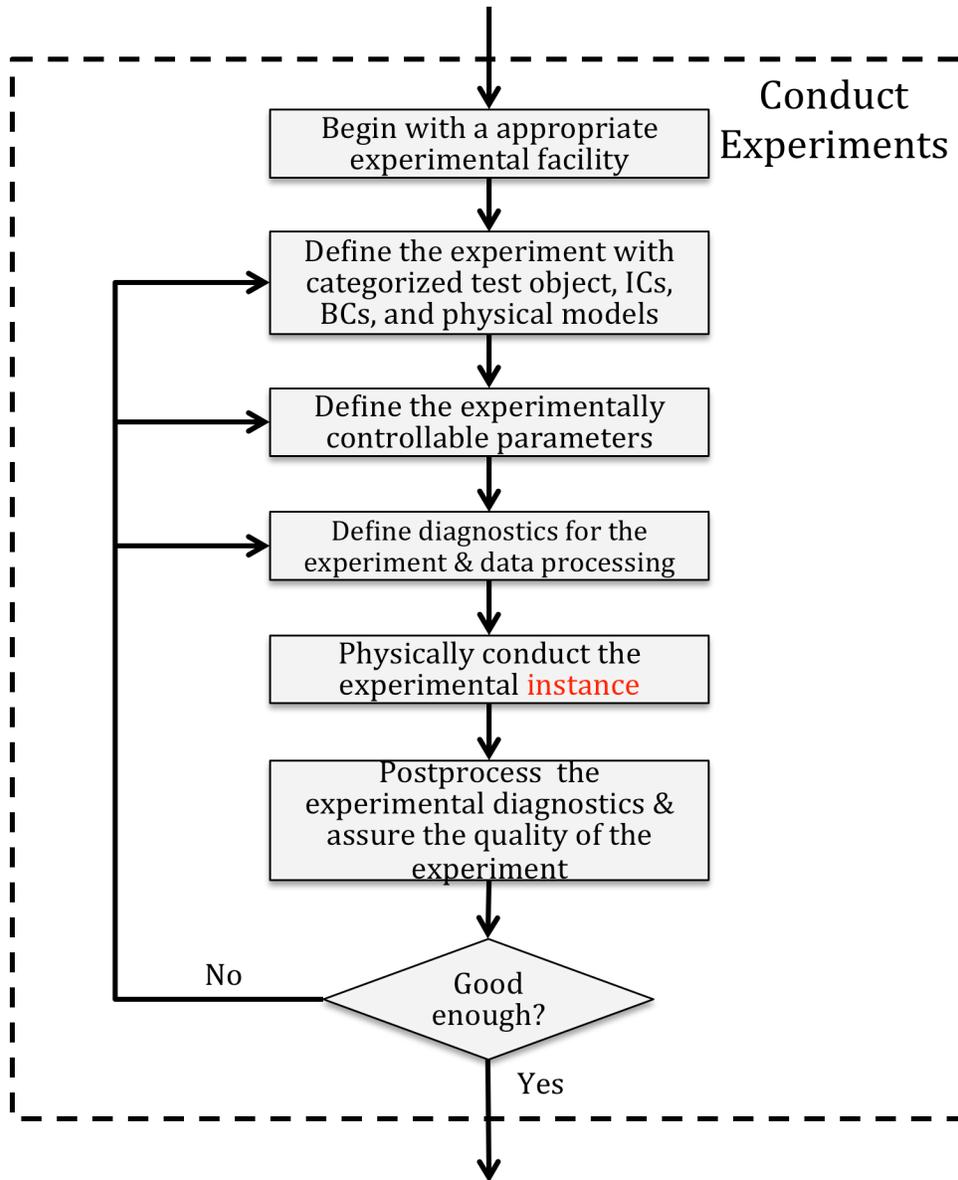


Figure 6. The overall Conduct Experiments process as described in this document is shown above. This process clearly identifies three possible aspects that must be investigated if the quality of the experimental data does not meet the experimentalist’s expectations.

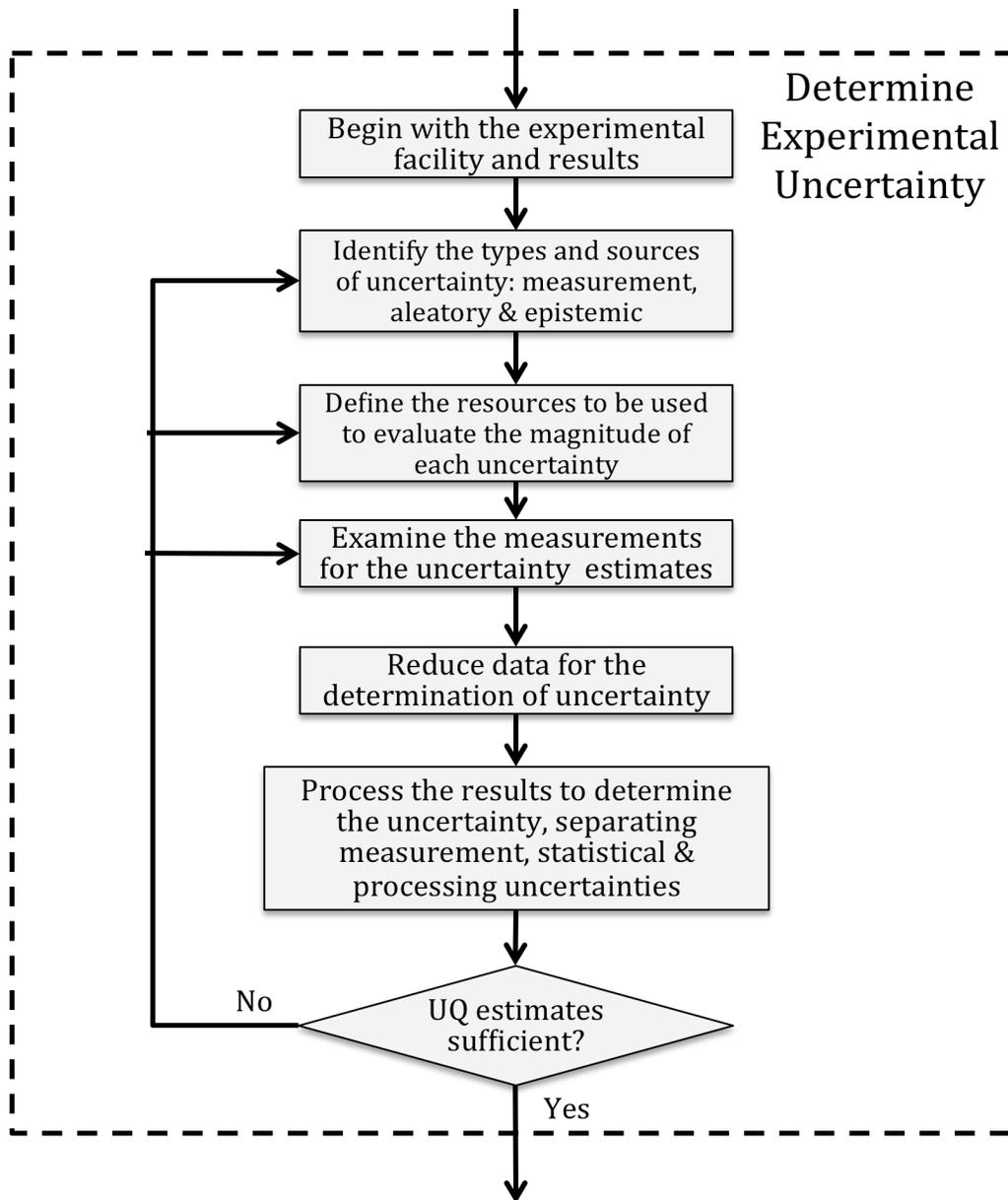


Figure 7. The overall Experimental Uncertainty Quantification process as described in this document is displayed here. This process clearly identifies three possible aspects that must be investigated if the quality of the experimental UQ estimates does not meet the experimentalist's expectations.

3.5 Validation Assessment (see Fig. 8)

1. Define the framework used for the assessment (e.g., CSAU [Boy90], PCMM [Obe07b], PCMM++ [Obe07b], QMU [Hel09, Pil06])
2. Conduct a PIRT [Tru02] analysis of the simulation (this could/should be conducted as part of the conduct of simulations or experiment). If this has already been conducted, reassess the previous PIRT.

3. Examine each element of the simulation and experimental work in the context of the assessment framework.
4. Evaluate the validation in terms of the requirements associated with the intended application.

As indicated in the first item above, there are various frameworks available that provide guidelines for validation assessments. CSAU, Code Scaling Applicability and Uncertainty [Boy90], is the framework accepted by the Nuclear Regulatory Committee for validation by dimensionally scaled experiments. PCMM, the Predictive Capability Maturity Model [Obe07b], identifies four qualitative levels of modeling and simulation capability maturity that gauge (1) model representation and geometric fidelity, (2) physics and material model fidelity, (3) code verification, (4) solution verification, (5) model validation, and (6) uncertainty quantification and sensitivity analysis. PCMM++, the enhanced Predictive Capability Maturity Model [Ober07b], is PCMM with additional evaluations of software modularity and extensibility (an evaluation of the ease with which software modifications can be made) and readiness of the software for HPC platforms (e.g., how easily the software is adapted to different machines). QMU, the Quantifications of Margins and Uncertainties [Hel09, Pil06], is a framework used at the Defense Program National Laboratories to support the certification of the nuclear stockpile. The emphasis in QMU is the incorporation of uncertainties into a decision framework.

The PIRT, or Phenomena Identification and Ranking Table, analysis mentioned in the second item is an approach to define the importance of phenomena and their relative importance to the situation at hand. The PIRT should be guided by the relative impact of the identified phenomena on the system response quantities of interest for CASL. Progress toward implementation and UQ of identified phenomena is also identified in the PIRT. The PIRT, because it ranks the relative impact of phenomena and the status of the modeling, can help guide additional model development, V&V, and UQ.

A complexity inherent in any of these assessment frameworks is that the comparison between simulations and experiment cannot be expressed simply as difference of the two quantities, but instead involves comparison of the probability distributions associated with simulations and experiments. The mapping of metrics used in simulations to those used in experiments is problematic. Simulation science and its associated mathematics provide error measures in the form of error norms (a functional measure of a mathematical function), and experiments measure quantities made available through instruments. Fortunately, the error norms typically involve some sort of integration in time or space, which mimics the processes involved in instrumental science. As a result there is a rough correspondence between experimental and simulation measures. Nonetheless, the correspondence is approximate and notional leading to yet another source of uncertainty in the simulation of physical circumstances. The end result of these differences leads to the failure of mathematical theory to provide assurances to simulation quality. This is most acutely felt in the uncertainty estimation due to numerical approximation (calculation verification), but influences all uncertainties arising through numerical approximation.

The determination of metrics requires the close interaction of computational and experimental scientists. The coordination strives to provide a balance between what is possible with experimental measurement science, and metrics that are accurate and well behaved mathematically. As experiments reach the upper echelon of the validation hierarchy this becomes increasingly difficult as measurements become more and more related to the underlying application-specific measures of success (see Section 4.2 below).

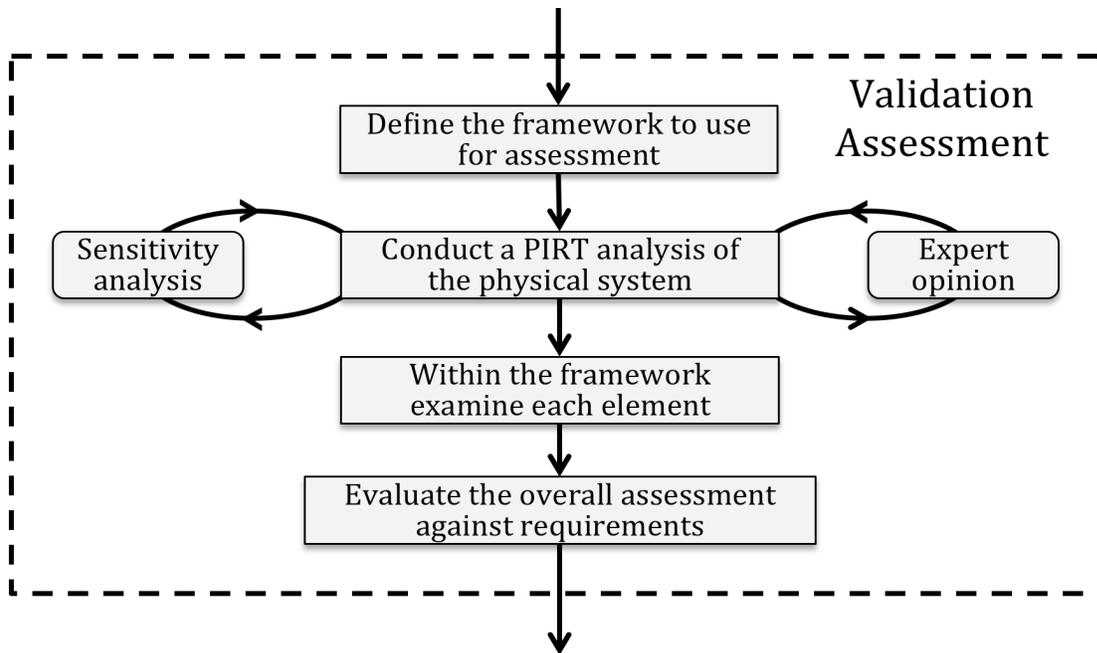


Figure 8. The Validation Assessment process as described in this document. This process clearly defines a course of action for assessment, including a well-defined framework for assessment and a PIRT analysis of the physical system.

4 BEST ESTIMATE PLUS REDUCED UNCERTAINTIES PREDICTIONS

Repeated measurements of the same physical quantity yield values that differ from each other, as well as from the true but unknown value of that quantity. This variation in results is due to experimental errors, imperfect instruments, and imperfectly known calibration standards. Hence, around any reported experimental value, there always exists a range of values that may also be plausibly representative of the true value. In turn, this means that all inferences, predictions, engineering computations, and other applications of measured data are necessarily founded on weighted averages over all the possibly true values, with weights indicating the degree of plausibility of each value. These weights and weighted averages are what we call probabilities and expectation values. Hence, probabilities encode incomplete information, in that persons possessing different information or knowledge would assign different probabilities, and would update the respective probabilities whenever new information became available. Thus, since the true value of physical quantities cannot be measured exactly, nominally measured values are insufficient, by themselves, for applications; the quantitative uncertainties accompanying the measurements are also needed, along with the respective nominal values. Combination of data from different sources involves a weighted propagation (e.g., using sensitivities) of various uncertainties, requiring reasoning from incomplete information for extracting “best” values together with “best” uncertainties from often sparse, incomplete, error-afflicted, and occasionally discrepant experimental data. A wide range of probability-theory concepts and tools are employed in data evaluation and assimilation, from deductive statistics involving mainly frequencies and sample tallies to inductive inference for assimilating non-frequency data and a priori knowledge.

Since the combination of data from different sources involves a weighted propagation (e.g., using sensitivities) of various, data evaluation is intrinsically intertwined with uncertainty analysis, requiring reasoning from incomplete information and using probability theory for extracting “best” values together with “best” uncertainties from often sparse, incomplete, error-afflicted, and occasionally discrepant experimental data. The probabilistic description of possible future computational and experimental outcomes, based on all recognized errors and uncertainties, is the aim of “*best estimates plus reduced uncertainties*” (*BERU*) predictions. *BERU*-predictions relies on the assimilation of experimental data (“data assimilation”) for updating (i.e., “calibrating” or “adjusting”) the parameters characterizing a computational model. The procedures for model calibration must encompass the propagation of all relevant uncertainties, including:

- (i) data uncertainties (input data, model parameters, initial and boundary conditions, forcing functions, etc);
- (ii) numerical discretization errors;
- (iii) discrepancies within the experimental data and/or discrepancies between data and model predictions; and
- (iv) uncertainties in the physics of the modeled processes (e.g., due to incomplete knowledge).

The results of *BERU*-predictions are best-estimated values for parameters and predicted responses, as well as best-estimate reduced uncertainties (i.e., “smaller” values for the variance-covariance matrices) for the predicted best-estimate parameters and responses, provided all elements involved in the calibration process are consistent with each other.

Cacuci and Ionescu-Bujor [Cac10] presented a new and rigorous mathematical methodology for *BERU*-predictions through data assimilation and *simultaneous calibration of model parameters and responses* for a generic time-dependent physical system. This works has generalized the data assimilation methodologies currently used in geophysical sciences (see, e.g., [Kal03,Lew06]). The methodology of Cacuci and Ionescu-Bujor [Cac10] also provides quantitative indicators (based on uncertainties and sensitivities) for determining the degree of agreement (or disagreement) relevant to the assimilation and best-estimate adjustment of computational and experimental parameters and responses. These quantitative indicators serve as the basic elements for *quantitative model extrapolation* (i.e., prediction of uncertainty in new environments or conditions of interest, including both untested parts of the parameter space and higher levels of system complexity in the validation hierarchy) and *estimation of the validation domain*.

The methodology of Cacuci and Ionescu-Bujor [Cac10] for *BERU*-predictions considers that a general physical system and/or the result of an indirect experimental measurement is modeled mathematically in terms of:

- (a) A system of linear and/or nonlinear equations that relate the system's independent variables and parameters to the system's state (i.e., dependent) variables;
- (b) Inequality and/or equality constraints that delimit the ranges of the system's parameters;
- (c) One or several quantities, customarily referred to as system responses (or objective functions, or indices of performance), which are computed using the mathematical model; and
- (d) Experimentally measured responses, with their respective nominal (mean) values and uncertainties (covariance matrices).

The time-dependent generic physical system to be analyzed in the sequel is considered to comprise N_α^ν model parameters and N_r^ν distinct responses, respectively, at every time node $\nu = 1, 2, \dots, N_t$. Hence, at every time node ν , the (column) vector $\boldsymbol{\alpha}^\nu$ of J_α^ν system parameters, and the (column) vector \mathbf{r}^ν of J_r^ν measured responses can be represented in component form as

$$\boldsymbol{\alpha}^\nu = \{\alpha_n^\nu \mid n=1, K, N_\alpha^\nu\}, \quad \mathbf{r}^\nu = \{r_i^\nu \mid i=1, K, N_r^\nu\}, \quad \nu = 1, \dots, N_t. \quad (1)$$

At any time node ν , the system parameters are considered to be variates with mean values $(\boldsymbol{\alpha}^0)^\nu$. Furthermore, the correlations between two parameters α_i^ν and α_j^μ , at two time nodes μ and ν , have the general form

$$c_{\alpha,ij}^{\nu\mu} \equiv \left\langle \left[\alpha_i^\nu - (\alpha_i^\nu)^0 \right] \left[\alpha_j^\mu - (\alpha_j^\mu)^0 \right] \right\rangle, \quad (2)$$

The above covariances constitute the elements of symmetric covariance matrices of the form

$$\mathbf{C}_\alpha^{\mu\nu} @ \left\langle (\mathbf{\alpha} - \mathbf{\alpha}^0)^\mu \left[(\mathbf{\alpha} - \mathbf{\alpha}^0)^\nu \right]^\dagger \right\rangle = (\mathbf{C}_\alpha^{\mu\nu})^\dagger = \mathbf{C}_\alpha^{\nu\mu} = (\mathbf{C}_\alpha)^\dagger. \quad (3)$$

Similarly, the *measured* responses are characterized by mean values $(\mathbf{r}_m)^\nu$ at a time node ν , and by symmetric covariance matrices between two time nodes μ and ν defined as

$$\mathbf{C}_m^{\mu\nu} @ \left\langle (\mathbf{r} - \mathbf{r}_m)^\mu \left[(\mathbf{r} - \mathbf{r}_m)^\nu \right]^\dagger \right\rangle = (\mathbf{C}_m^\nu)^\dagger = \mathbf{C}_m^{\nu\mu} = (\mathbf{C}_m^{\nu\mu})^\dagger. \quad (4)$$

In the most general case, the measured responses may be correlated to the parameters through symmetric *response-parameter uncertainty* matrices of the form

$$\mathbf{C}_{r\alpha}^{\mu\nu} @ \left\langle (\mathbf{r} - \mathbf{r}_m)^\mu \left[(\mathbf{\alpha} - \mathbf{\alpha}^0)^\nu \right]^\dagger \right\rangle = (\mathbf{C}_{r\alpha}^{\mu\nu})^\dagger = \mathbf{C}_{r\alpha}^{\nu\mu} = (\mathbf{C}_{r\alpha})^\dagger. \quad (5)$$

Note that the matrices $\mathbf{C}_{r\alpha}^{\mu\nu}$ are *not* bona-fide variance-covariance matrices, in that they are not necessarily square positive matrices (often, they are rectangular), and the elements on their respective main diagonals (if they happen to be square) are also covariances (or correlations) rather than variances.

At any given time node ν , a response r_i^ν can be a function of not only the system parameters at time node ν , but also of the system parameters at all previous time nodes μ , $1 \leq \mu \leq \nu$; this means that $\mathbf{r}^\nu = \mathbf{R}^\nu(\mathbf{p}^\nu)$, where $\mathbf{p}^\nu @ (\mathbf{\alpha}^1, \mathbf{K}, \mathbf{\alpha}^\mu, \mathbf{K}, \mathbf{\alpha})$. In general, the response computed using the model depends nonlinearly and implicitly (in an analytically intractable form) on the model parameters. The uncertainties in parameters and modeling induce uncertainties in the computed responses. These can be computed either by means of statistical methods (for relatively simple models with few parameters) or deterministically, by using the well-known propagation of moments (errors) method (see, e.g., Ref. 4).

As shown by Cacuci and Ionescu-Bujor [Cac10], the *BERU*-predictions for best-estimate parameters, responses, and their corresponding reduced uncertainties (covariance matrices) are as follows:

(i) The best-estimate predicted nominal values for the calibrated (adjusted) parameters:

$$(\mathbf{\alpha}^{be})^\nu = (\mathbf{\alpha}^0)^\nu + \sum_{\mu=1}^{N_t} \left\{ \left[\mathbf{C}_{\alpha r}^{\nu\mu} - \sum_{\rho=1}^{\mu} \mathbf{C}_{\alpha}^{\nu\rho} (\mathbf{S}^\dagger)^\mu \right] \left[\sum_{\eta=1}^{N_t} \mathbf{K}_d^{\mu\eta} \mathbf{d}^\eta \right] \right\}, \quad \nu = 1, \mathbf{K}, N_t. \quad (6)$$

where $\mathbf{K}_d^{\nu\eta}$ denotes the corresponding (ν, η) -element of the block-matrix \mathbf{C}_d^{-1} , with the block-matrix $\mathbf{C}_d(\mathbf{\alpha}^0)$ defined as follows:

$$\begin{aligned} \mathbf{C}_d(\boldsymbol{\alpha}^0) @ \langle \mathbf{d} \mathbf{d}^\dagger \rangle &= \left\langle (\delta \mathbf{r} - \mathbf{S}(\boldsymbol{\alpha}^0) \delta \boldsymbol{\alpha}) (\delta \mathbf{r}^\dagger - \delta \boldsymbol{\alpha}^\dagger [\mathbf{S}(\boldsymbol{\alpha}^0)]^\dagger \boldsymbol{\alpha}) \right\rangle \\ &= \mathbf{C}_{rc}(\boldsymbol{\alpha}^0) - \mathbf{C}_{r\alpha} [\mathbf{S}(\boldsymbol{\alpha}^0)]^\dagger - [\mathbf{S}(\boldsymbol{\alpha}^0)] \boldsymbol{\epsilon}_{\alpha r} \quad \mathbf{C}_m \end{aligned} \quad (7)$$

where the vector

$$\mathbf{d} @ \mathbf{R}(\boldsymbol{\alpha}^0) - \mathbf{r}_m \quad (8)$$

measures, in the corresponding metric, the deviations between the experimental and nominally computed responses. In component form, the matrix \mathbf{C}_d is expressed as

$$\begin{aligned} \mathbf{C}_d @ \begin{pmatrix} \mathbf{C}_d^{11} & \mathbf{K} & \mathbf{C}_d^{1N_t} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{C}_d^{N_t 1} & \mathbf{L} & \mathbf{C}_d^{N_t N_t} \end{pmatrix} &= \begin{pmatrix} \mathbf{C}_{rc}^{11} + \mathbf{C}_m^{11} & \mathbf{K} & \mathbf{C}_{rc}^{1N_t} + \mathbf{C}_m^{1N_t} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{C}_{rc}^{N_t 1} + \mathbf{C}_m^{N_t 1} & \mathbf{L} & \mathbf{C}_{rc}^{N_t N_t} + \mathbf{C}_m^{N_t N_t} \end{pmatrix} \\ &- \begin{pmatrix} \mathbf{C}_{r\alpha}^{11} (\mathbf{S}^\dagger)^{11} + \mathbf{S}^{11} \mathbf{C}_{\alpha r}^{11} & \mathbf{K} & \mathbf{S}^{11} \mathbf{C}_{\alpha r}^{1N_t} + \sum_{\rho=1}^{N_t} \mathbf{C}_{r\alpha}^{1\rho} (\mathbf{S}^\dagger)^{N_t \rho} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{C}_{r\alpha}^{N_t 1} (\mathbf{S}^\dagger)^{11} + \sum_{\rho=1}^{N_t} \mathbf{S}^{N_t \rho} \mathbf{C}_{\alpha r}^{\rho 1} & \mathbf{L} & \sum_{\rho=1}^{N_t} [\mathbf{C}_{r\alpha}^{N_t \rho} (\mathbf{S}^\dagger)^{N_t \rho} + \mathbf{S}^{N_t \rho} \mathbf{C}_{\alpha r}^{\rho N_t}] \end{pmatrix}. \end{aligned} \quad (9)$$

The $(J_r^\nu \times J_\alpha^\mu)$ -dimensional matrix $\mathbf{S}^{\nu\mu}(\mathbf{p}_0^\mu)$, $1 \leq \mu \leq \nu$, contains the first Gateaux-derivatives of the computed responses with respect to the parameters evaluated at the parameter nominal values \mathbf{p}_0^ν , i.e.,

$$\mathbf{S}^{\nu\mu}(\mathbf{p}_0^\nu) @ \begin{pmatrix} s_{11}^{\nu\mu\nu\mu} & \mathbf{K} & s_{1N} \\ \mathbf{M} & s_{in} & \mathbf{M} \\ s_{I1}^{\nu\mu\nu\mu} & \mathbf{L} & s_{IN} \end{pmatrix} @ \begin{pmatrix} \frac{\partial R_1^\nu(\mathbf{p}_0^\mu)}{\partial \alpha_1^\mu} & \mathbf{K} & \frac{\partial R_1(\mathbf{p}_0)}{\partial \alpha_N^\mu} \\ \mathbf{M} & \frac{\partial R_i^\nu}{\partial \alpha_n^\mu} & \mathbf{M} \\ \frac{\partial R_I^\nu(\mathbf{p}_0^\mu)}{\partial \alpha_1^\mu} & \mathbf{L} & \frac{\partial R_I(\mathbf{p}_0)}{\partial \alpha_N^\mu} \end{pmatrix}, \quad 1 \leq \mu \leq \nu. \quad (10)$$

The above sensitivities can be computed most efficiently by using the *adjoint sensitivity analysis procedure* for nonlinear systems pioneered by Cacuci (see [Cac81a, Cac81b]) and extended by Cacuci and Ionescu-Bujor (see [Cac05a, Cac05b]). Since the response $\mathbf{R}^\nu(\mathbf{p}_0^\nu)$ at time node ν can depend only on parameters $(\boldsymbol{\alpha}^0)^\mu$ which appear up to the current time node ν , it follows that $\mathbf{S}^{\nu\mu} = \mathbf{0}$ when $\mu > \nu$.

The covariance matrix of the computed responses, denoted as \mathbf{C}_{rc} in the above expressions, has the symmetric structure

$$\mathbf{C}_{rc} @ \begin{pmatrix} \mathbf{C}_{rc}^{11} & \mathbf{K} & \mathbf{C}_{rc}^{1N_t} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ \mathbf{C}_{rc}^{N_t 1} & \mathbf{L} & \mathbf{C}_{rc}^{N_t N_t} \end{pmatrix} \mathbf{C}_{rc}^{v\mu} = \sum_{\eta=1}^v \sum_{\rho=1}^{\mu} \mathbf{S}^{v\eta} \mathbf{C}_{\alpha}^{\eta\rho} (\mathbf{S}^{\mu\rho})^{\dagger} = (\mathbf{C}_{rc}^{\mu v})^{\dagger}; \quad v, \mu = 1, \dots, N_t. \quad (11)$$

(ii) The best-estimate predicted nominal values for the calibrated (adjusted) responses:

$$(\mathbf{r}^{be})^v = (\mathbf{r}_m)^v + \sum_{\mu=1}^{N_t} \left\{ \left[\mathbf{C}_m^{v\mu} - \sum_{\rho=1}^{\mu} \mathbf{C}_{r\alpha}^{v\rho} (\mathbf{S}^{\dagger})^{\mu} \right] \left[\sum_{\eta=1}^{N_t} \mathbf{K}_d^{\mu\eta} \mathbf{d}^{\eta} \right] \right\}, \quad v = 1, K, N_t. \quad (12)$$

(iii) The expressions for the best-estimate predicted covariances \mathbf{C}_{α}^{be} and \mathbf{C}_r^{be} , corresponding to the best-estimate parameters $\boldsymbol{\alpha}^{be}$ and responses $\mathbf{r}(\boldsymbol{\alpha}^{be})$, together with the predicted best-estimate parameter-response covariance matrix $\mathbf{C}_{r\alpha}^{be}$ are as follows:

$$(\mathbf{C}_{\alpha}^{be})^{v\mu} = \mathbf{C}_{\alpha}^{v\mu} - \sum_{\eta=1}^{N_t} \sum_{\rho=1}^{N_t} \left[\mathbf{C}_{\alpha r}^{v\rho} - \sum_{\pi=1}^{\rho} \mathbf{C}_{\alpha}^{v\pi} (\mathbf{S}^{\dagger})^{\rho\pi} \right] \mathbf{K}_d^{\rho\eta} \left[\mathbf{C}_{r\alpha}^{\eta\mu} - \sum_{\pi=1}^{\eta} \mathbf{S}^{\eta\pi} \mathbf{C}_{\alpha}^{\pi\mu} \right]. \quad (13)$$

$$(\mathbf{C}_r^{be})^{v\mu} = \mathbf{C}_m^{v\mu} - \sum_{\eta=1}^{N_t} \sum_{\rho=1}^{N_t} \left[\mathbf{C}_m^{v\rho} - \sum_{\pi=1}^{\rho} \mathbf{C}_{r\alpha}^{v\pi} (\mathbf{S}^{\dagger})^{\rho\pi} \right] \mathbf{K}_d^{\rho\eta} \left[\mathbf{C}_m^{\eta\mu} - \sum_{\pi=1}^{\eta} \mathbf{S}^{\eta\pi} \mathbf{C}_{\alpha r}^{\pi\mu} \right]. \quad (14)$$

$$(\mathbf{C}_{r\alpha}^{be})^{v\mu} = \mathbf{C}_{r\alpha}^{v\mu} - \sum_{\eta=1}^{N_t} \sum_{\rho=1}^{N_t} \left[\mathbf{C}_m^{v\rho} - \sum_{\pi=1}^{\rho} \mathbf{C}_{r\alpha}^{v\pi} (\mathbf{S}^{\dagger})^{\rho\pi} \right] \mathbf{K}_d^{\rho\eta} \left[\mathbf{C}_{\alpha r}^{\eta\mu} - \sum_{\pi=1}^{\eta} \mathbf{S}^{\eta\pi} \mathbf{C}_{\alpha}^{\pi\mu} \right]. \quad (15)$$

Note in Eq. (13) that a symmetric positive matrix is subtracted from the initial parameter covariance matrix \mathbf{C}_{α} ; hence, in this sense, *the best-estimate predicted parameter uncertainty matrix \mathbf{C}_{α}^{be} has been reduced by the calibration (adjustment) procedure, through the introduction of new information from experiments.* Similarly in Eq. (14), a symmetric positive matrix is subtracted from the initial covariance matrix \mathbf{C}_m of the experimental-responses; hence, *the best-estimate predicted response covariance matrix \mathbf{C}_r^{be} has been improved (reduced) through the introduction of new experimental information.* Furthermore, Eq. (15) indicates that the calibration (adjustment) procedure will introduce correlations between the calibrated (adjusted) parameters and responses even if the parameters and response were initially uncorrelated, since $\mathbf{C}_{r\alpha}^{be} \neq 0$ even if $\mathbf{C}_{r\alpha} = 0$, i.e.,

$$\mathbf{C}_{r\alpha}^{be} = \mathbf{C}_m \left[\mathbf{C}_{rc}(\boldsymbol{\alpha}^0) + \mathbf{C}_m \right]^{-1} \left[\mathbf{S}(\boldsymbol{\alpha}^0) \right] \mathbf{C}_{\alpha}, \quad \text{when } \mathbf{C}_{r\alpha} = 0. \quad (16)$$

As the above expression indicates, the adjustment (calibration) modifies the correlations among the parameters through couplings introduced by the sensitivities of

the participating responses. In the calibration procedure, the sensitivities play the role of weighting functions for propagating the initial parameter-covariances and experimental-response covariances to the adjusted best-estimate predicted quantities. Thus, as indicated by Eqs. (13) through (15), the incorporation of additional (experimental) information in the adjustment (calibration) process reduces the variances of the adjusted parameters and responses while also modifying their correlations.

It is important to note that the computation of the best estimate parameter and response values, together with their corresponding best-estimate uncertainties, cf. Eqs. (6) and (12) through (15) require the inversion of a single matrix, namely the matrix $\mathbf{C}_d(\boldsymbol{\alpha}^0)$ defined in Eq. (7). This is usually advantageous in practice, since the order of the matrix $\mathbf{C}_d(\boldsymbol{\alpha}^0)$ is given by the number of measured (or computed responses), which is most often considerably smaller than the number of model parameters under consideration.

On the other hand, when the number of parameter exceeds the number of responses, it is possible to derive alternative expressions for the best-estimate calibrated parameters and their corresponding best-estimate covariances, by performing all derivations in the “parameter space” rather than in “response space”. The expression below provides the bridge between the “response-space” and “parameter-space” formulations.

$$\begin{aligned} \mathbf{C}_d^{-1} @(\mathbf{C}_{rc} - \mathbf{C}_{r\alpha} \mathbf{S}^\dagger - \mathbf{S} \mathbf{C}_{\alpha r} + \mathbf{C}_m)^{-1} \\ = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{S} (\mathbf{C}_\alpha^{-1} + \mathbf{S}^\dagger \mathbf{A}^{-1} \mathbf{S})^{-1} \mathbf{S}^\dagger \mathbf{A}^{-1}; \quad \mathbf{A} @\mathbf{C}_m - \mathbf{C}_{r\alpha} \mathbf{S}^\dagger - \mathbf{S} \mathbf{C}_{\alpha r}. \end{aligned} \quad (17)$$

The above expression also highlights the fact that the response-space formulation requires a single inversion of a square symmetric matrix (namely, the matrix \mathbf{C}_d) of the same dimensions as the number of responses. In contradistinction, the “parameter space” formulation requires the inversion of three symmetric matrices, two of which have dimensions equal to the number of parameters and one of dimensions equal to the number of responses. Hence, from a computational standpoint, the “response-space” formulations should be used whenever possible.

In view of Eq. (7), it is important to note that the inverse matrix, \mathbf{C}_d^{-1} , incorporates simultaneously all of the available information about the system parameters and responses, at all time nodes [i.e., $\nu = 1, 2, \dots, N_t$]. Specifically, at any time node ν , \mathbf{C}_d^{-1} incorporates information not only from time nodes prior in time to ν (i.e., information regarding the “past” and “present” states of the system) but also from time nodes posterior in time to ν (i.e., information about the “future” states of the system). Through the matrix \mathbf{C}_d^{-1} , at any specified time node ν , the calibrated best-estimate parameters $(\boldsymbol{\alpha}^{be})^\nu$ and responses $\mathbf{r}(\boldsymbol{\alpha}^{be}) @\mathbf{r}^{be}$, together with the corresponding calibrated best-estimate covariance matrices $(\mathbf{C}_\alpha^{be})^{\nu\mu}$, $(\mathbf{C}_r^{be})^{\nu\mu}$, and $(\mathbf{C}_{\alpha r}^{be})^{\nu\mu}$ will also incorporate

automatically all of the available information about the system parameters and responses at all time nodes [i.e., $\nu = 1, 2, \dots, N_t$].

In this respect, the methodology presented in this section is conceptually related to the "foresight" aspects encountered in decision analysis. It is also important to note that, in practice, the application of the methodology developed in this section involves two distinct computational stages. A complete sensitivity data base (i.e., sensitivities $s_{ni}^{\nu\mu}$ at all times nodes $\nu, \mu = 1, K, N_t$) needs to be generated prior to performing the "data assimilation" and "model calibration" (or data adjustment) stage, since all of the sensitivities are needed in order to perform data assimilation and model calibration. These stages are performed subsequently, together with the computation of calibrated best-estimate covariance matrices (the "uncertainty analysis" stage), by combining the sensitivities with covariance matrices.

Because of the "foresight" and "off-line" characteristics, the methodology presented in this Section can be called the "off-line with foresight" data assimilation and adjustment (model calibration) methodology, underscoring that all sensitivities are generated separately, prior to performing the uncertainty analysis, and that foresight characteristics are included in the procedure. Since the incorporation of foresight effects involves the inversion of the matrix \mathbf{C}_d , this methodology is best suited for problems involving relatively few time nodes. For large-scale highly nonlinear problems involving many time nodes, the matrix \mathbf{C}_d becomes very large, requiring large amounts of computer storage; the inversion of \mathbf{C}_d may become prohibitively expensive in such cases. These difficulties can be reduced at the expense of using less than the complete information available at any specific time node. For example, even in time-dependent problems in which the entire time history is known (e.g., transient behavior of reactor systems), one may nevertheless choose to use only information up to the current time index, and disregard the information about "future" system states.

On the other hand, in dynamical problems such as climate or weather prediction, in which the time variable advances continuously and states beyond the current time are not known, information about future states cannot be reliably accounted for anyway. Thus, the most common way of reducing the dimensionality of the data assimilation and model calibration problem is to disregard information about future states and limit the amount of information assimilated about "past states". Data assimilation and model calibration procedure using such a limited amount of information can be performed either off-line or on-line, assimilating the new data as the time index advances.

For *time-independent* problems, the time-dependent results derived in the forgoing reduce to expressions that are formally identical to Eqs. (6) and (12) through (15); hence, these expressions can be used directly to obtain the best-estimate predicted values for parameters, responses, and their respective covariances. It is also important to note that modeling errors can be treated in a manner similar to parameter uncertainties by including the discretization intervals in the vector of model parameters, as shown in [Cac03, Cac81a, Cac81b, Cac05a, Cac05b].

The actual application of the model calibration (adjustment) algorithms, cf. Eqs. (6) and (12) through (15), to the a physical system characterized by nominal values and uncertainties for model parameters together with the computed and measured responses is straightforward, in principle, although it can become computationally very demanding on both data handling and computational speed. It is also important to note that the indiscriminate incorporation of all (seemingly relevant) experimental-response data could produce a set of calibrated (adjusted) parameter values that might differ unreasonably much from the corresponding original nominal values. Worse yet, the indiscriminate use of information might even fail to improve the agreement between the calculated and measured values of some of the very responses by which the library was calibrated (adjusted). When calibrating (adjusting) a library of model parameters, it is tacitly assumed that the given parameters are basically “correct,” except that they are not sufficiently accurate for the objective at hand. The calibration procedure uses additional data (e.g., experimental responses) for improving the parameter values while reducing their uncertainties. Although such additional information induces modifications of the original parameter values, the adjusted parameters are still generally expected to remain consistent with their original nominal values, within the range of their original uncertainties. However, calibration of model parameters by experimental responses, which significantly deviate from their respective computed values would significantly modify the resulting adjusted parameters. Such unlikely adjustments would most probably lead to failure of even reproducing the original experimental responses.

On the other hand, calibrating a parameter library by using measured responses that are very close to their respective computed values would cause minimal parameter modifications and a nearly perfect reproduction of the given responses by the adjusted library (as would be expected). In such a case, the given responses would be considered as being consistent with the parameter library, in contradistinction to adjustment by inconsistent experimental information, in which case the adjustment could fail because of inconsistencies. These considerations clearly underscore the need for using a quantitative indicator to measure the mutual and joint consistency of the information available for model calibration.

As shown by Cacuci and Ionescu-Bujor [Cac10], the expression

$Q_{\min} \equiv Q(\mathbf{z}^{be}) = \mathbf{d}^\dagger [\mathbf{C}_d(\boldsymbol{\alpha}^0)]^{-1} \mathbf{d}, \mathbf{d} @ \mathbf{R}(\boldsymbol{\alpha}^0) - \mathbf{r}_m$ represents the square of the length of the vector \mathbf{d} , measuring (in the corresponding metric) the deviations between the experimental and nominally computed responses. Note that $Q_{\min} \equiv Q(\mathbf{z}^{be})$ can be evaluated directly from the given data (i.e., given parameters and responses, together with their original uncertainties) after having inverted the deviation-vector uncertainty matrix $\mathbf{C}_d(\boldsymbol{\alpha}^0)$. It is also very important to note that $Q_{\min} \equiv Q(\mathbf{z}^{be})$ is independent of calibrating (or adjusting) the original data. As the dimension of \mathbf{d} indicates, the number of degrees of freedom characteristic of the calibration under consideration is equal to the number of experimental responses. In the extreme case of absence of experimental responses, no actual calibration takes place since $\mathbf{d} = \mathbf{R}(\boldsymbol{\alpha}^0)$, so that the best-estimate parameter values

are just the original nominal values, i.e., $(\mathbf{\alpha}^{be})^k = (\mathbf{\alpha}^0)^k$; an actual adjustment occurs only when at least one experimental response is included.

The above relation indicates that experimental responses can be considered as random variables approximately described by a multivariate Gaussian distribution with means located at the nominal values of the computed responses, and with a covariance matrix $\mathbf{C}_d(\mathbf{\alpha}^0)$. In turn, the random variable $Q_{\min} \equiv Q(\mathbf{z}^{be})$ obeys a χ^2 -distribution with n degrees of freedom, where n denotes the total number of experimental responses considered in the calibration (adjustment) procedure. Since $Q_{\min} \equiv Q(\mathbf{z}^{be})$ is the “ χ^2 of the calibration (adjustment) at hand”, it can be used as an indicator of the agreement between the computed and experimental responses, measuring essentially the consistency of the experimental responses with the model parameters. Recall that the χ^2 (chi-square) distribution with n degrees of freedom of the continuous variable ($0 \leq x < \infty$) is defined as

$$P(x < \chi^2 < x + dx) @ k_n(x) dx = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2} dx, \quad x > 0, \quad (n = 1, 2, \dots, \infty).$$

The χ^2 -distribution is a measure of the deviation of a “true distribution” (in this case – the distribution of experimental responses) from the hypothetic one (in this case – a Gaussian). The mean and variance of x are $\langle x \rangle = n$ and $\text{var}(x) = 2n$. Further practically useful asymptotic properties of the χ^2 -distribution for $n \rightarrow \infty$ are as follows: (i) x is asymptotically normal with mean n and variance $2n$; (ii) x/n is asymptotically normal with mean 1 and variance $2/n$; (iii) $\sqrt{2x}$ is asymptotically normal with mean $\sqrt{2n-1}$ and variance 1. Although the χ^2 -distribution is extensively tabulated, the notation is not uniform in the literature for the various derived quantities (in particular, for the corresponding cumulative distribution functions and fractiles). The cumulative distributions, denoted here by $P_n(\chi^2)$ and $Q_n(\chi^2)$, are defined as

$$P_n(\chi_0^2) @ P(\chi^2 \leq \chi_0^2) @ \int_0^{\chi_0^2} k_n(t) dt; \quad Q_n(\chi_0^2) @ P(\chi^2 \geq \chi_0^2) @ \int_{\chi_0^2}^{\infty} k_n(t) dt = 1 - P_n(\chi_0^2).$$

In practice, one rejects a hypothesis using the χ^2 -distribution when, for a given significance level α and number of degrees of freedom n , the value of $Q_{\min} \equiv \chi^2$ exceeds a chosen critical fractile value $\chi_{\alpha}^2(n)$. Published tables often show $\chi_{1-\alpha}^2(n)$ versus α . When the number of degrees of freedom is large ($n > 30$), a useful asymptotic approximation is $\chi_{\alpha}^2(n) \approx 1/2 \left(\sqrt{2n-1} - z_{2\alpha} \right)^2$, with $z_{2\alpha}$ denoting the corresponding fractile of the standard normal distribution $\Phi_0(z)$, computed by solving the equation

$2\Phi_0(z_{2\alpha})=1-2\alpha$, using the tabulated tables for $\Phi_0(z)$. For large or small values of α , a more accurate approximation is $\chi^2_\alpha(n) \approx n \left(1 - \frac{2}{9m} + z_{2\alpha} \sqrt{\frac{2}{9m}} \right)^3$. It may be often more convenient to transform χ^2 to the variate $t = \chi^2/n$ (i.e., “ χ^2 per degree of freedom”), in which case the transformed distribution, $g_n(t)$, becomes $g_n(t) = nk_n(nt)$, with mean value $\langle t \rangle = 1$ and variance $2/n$.

For model calibration (adjustment), it is important to assess if: (i) the response and data measurements are free of gross errors (blunders such as wrong settings, mistaken readings, etc), and (ii) the measurements are consistent with the assumptions regarding the respective means, variances, and covariances. For example, if $\chi^2/n \approx 1$, then that the measurements are very likely to be both free of gross errors and consistent with the assumptions. However, if $\chi^2/n \gg 1$ or $\chi^2/n \ll 1$, the measurements (or at least some measurements), the assumptions, or both are suspect. In particular, unusually large values $\chi^2/n \gg 1$ could be obtained when the original variances are underestimated; increasing them beyond their assumed nominal values would cause the adjusted values of χ^2/n and $P_n(\chi^2)$ to decrease accordingly. The reverse argument would apply if the a priori values of χ^2/n and $P_n(\chi^2)$ were unusually small (e.g., $\chi^2/n \approx 1$, $P_n(\chi^2) \approx 10^{-4}$), which could be the result of a priori overestimated variances. A practical quantitative criterion for the “acceptance” (or “rejection”) of experimental results in conjunction with a given “theoretical” model (i.e., in conjunction with the assumptions regarding the variates underlying the model) is to accept the value of χ^2/n whenever $0.15 < P_n(\chi^2) < 0.85$, in analogy to the “ 1σ ” range of normal distributions. Note that, when setting an acceptance criterion for χ^2/n of the general form $\alpha < P_n(\chi^2) < 1 - \alpha$, the exact value of α is not essential and is subject to personal judgment. This is because the probability $P_n(\chi^2)$ is still sensitive to the value of χ^2/n due to the fact that $\chi^2/n \approx 1 \pm \sqrt{(2/n)}$ (except for few degrees of freedom, e.g., for $n \leq 5$), so the acceptable range of χ^2/n narrows as $1/\sqrt{n}$ (see also the previously noted asymptotic forms for χ^2_α/n). In other words, moderate changes in χ^2/n lead to significant relative changes in $P_n(\chi^2)$. For example, the central 50%-range of $\chi^2/20$ is (0.77, 1.19), and the corresponding 90%-range is (0.54, 1.57), implying that values of $\chi^2/20$ below 0.4 or above 2.0 would be definitely unacceptable.

In addition to measuring the overall consistency of a given set of parameters and responses, the quantity χ^2/n also measures the consistency among the measured responses. Hence, an entire data set (model parameters and/or experimental responses)

should not be indiscriminately disqualified because of too high or too low value of χ^2 / n , since even a single “outlying” response could significantly degrade the set’s overall consistency. Note that a simple-minded assessment and ranking of “questionable responses” according to the values of the “individual consistencies” (i.e., the values of χ^2 obtained for each response as if it were the only response available for calibrating the entire set of parameters), would be very likely misleading. This is because the sum of the respective “individual consistencies” [which would numerically be obtained by dividing the squares of the deviations, d_i^2 , through the sum of the respective variances of the computed and measured responses $\text{var}(r_i^{\text{comp}}) + \text{var}(r_i^{\text{exp}})$], would not be equal to the “joint consistency” (i.e., the joint χ^2) of the entire set of experimental responses. This is because the deviation-vector uncertainty matrix $\mathbf{C}_d(\boldsymbol{\alpha}^0) = \mathbf{C}_{rc}(\boldsymbol{\alpha}^0) - \mathbf{C}_{ra}[\mathbf{S}(\boldsymbol{\alpha}^0)]^\dagger - [\mathbf{S}(\boldsymbol{\alpha}^0)]\mathbf{C}_{ar} - \mathbf{C}_m$ is generally non-diagonal, even if both $\mathbf{C}_{rc}(\boldsymbol{\alpha}^0)$ and \mathbf{C}_m are diagonal. On the other hand, verifying the consistency of *all* partial sets of the array of n responses with respect to their consistency with the given library is usually impractical, since the number of partial sets of an array of n responses is $2^n - 1$; hence, such a verification would be practically feasible only when the number of measured responses is very small.

The model calibration methodology presented in the foregoing provides a rigorous mathematical foundation for similar methodologies used in geophysical sciences, while extending these methodologies in several directions:

- (i) Simultaneous calibration of all parameters and responses (as opposed to just initial conditions, as usually performed in geophysical sciences);
- (ii) Treatment of systems involving correlated parameters and responses (as opposed to no parameter-response correlations, e.g., in geophysical sciences);
- (iii) Simultaneous calibration over all time intervals; the usual two-step time advancement procedures used in geophysical sciences simply becomes a consequence, as a particular case, of the general methodology presented in this work.

Computationally, the most intensive aspect of the methodology presented in this work is the computation of the sensitivities of responses to model parameters, which play a crucial role as weighting functions in all of the expressions for the best-estimated predicted values for parameters, responses and their associated best-estimated reduced uncertainties. For large-scale systems, the most efficient method for computing these sensitivities is the adjoint sensitivity procedure (ASAP), as detailed in [Cac03, Cac81a, Cac81b, Cac05a, Cac05b]. The other computationally intensive aspect in this mathematical formalism is the inversion of the covariance matrix $\mathbf{C}_d(\boldsymbol{\alpha}^0)$ associated with the vector $\mathbf{d} = \mathbf{R}(\boldsymbol{\alpha}^0) - \mathbf{r}_m$, which measures the deviations between the respective computed and experimentally-measured responses. Methods for efficiently inverting of this matrix, as well as for reducing its dimension, (e.g., reduced-order modeling using proper orthogonal decomposition methods) are of substantial interest.

The best-estimate calibrated values for model parameters obtained through the application of the mathematical framework presented in this work can also be used to estimate quantitatively the validation domain of the model under consideration, by computing contours of constant best-estimate uncertainties in the high-dimensional parameter-space. The best-estimate calibrated values can also be used to perform “model extrapolation”, by predicting uncertainties in new environments or conditions of interest. Extrapolation of large-scale models would address both untested parts of the parameter space and higher levels of system complexity in the validation hierarchy.

Ongoing research is currently devoted to the explicit treatment of modeling errors and extending the results by including not only the sensitivities (i.e., first-order information) but also the Hessians (i.e., second-order information) of the responses. Additional work is also ongoing to remove the current restriction to Gaussian distributions. Actually, the de-facto limitation to Gaussian distribution is characteristic of all of the state-of-the-art procedures for data assimilation and model calibration, as evidenced by the scientific literature published thus far. Removing these limitations would contribute significantly to understanding the validation of coupled nonlinear multi-physics models (e.g., of two or more physical phenomena that were not coupled in the initial validation database), particularly the accompanying increase of uncertainty. Developing predictive experimentally validated “best-estimate” numerical models is particularly important for designing new technologies and facilities based on novel processes, while striving to avoid, as much as possible, the costly and lengthy procedures of building representative mock-up experiments, which might confirm -but would not necessarily explain- the predictions of simulation tools.

5. DISCUSSION OF THE WORKFLOW IN PRACTICE

The idealized workflow guidelines we have outlined is both resource intensive and time consuming. This observation should not be surprising. The intent of VUQ analysis is to provide value that is commensurate with the effort put into the process. This value is found through producing a defensible level of confidence in the analysis, which then informs any subsequent decision-making. Any such decision reached via modeling should be made with a quantitative definition of the associated uncertainties.

The level (and expense) of assessment should be appropriate to the risks and consequences of the decision. This document is not about how to make decisions, but it is about how to provide input to the decision making process by encouraging that well-defined uncertainties be associated with modeling.

It is the authors' observation that the quantity and quality of verification, validation, calibration, sensitivity analysis, data assimilation, and uncertainty quantification in scientific simulation has historically been extremely variable. It is undeniable, however, that VVUQ is growing in both recognition and importance. It is our impression that, in practice, scientific simulation studies:

- Often confuse several of these activities with each other;
- Regularly include one of these;
- Sometimes include two of these;
- Virtually never include all of these.

The objective of this manuscript is not for CASL analysts to become experts in all of these (i.e., to do a perfect job in all of them), but to strive to include and become competent in each of them and apply the concepts appropriately, mindful of the following generalizations.

- VVUQ is usually at its best in the lowest level of the hierarchy of Fig. 2 (i.e., for simple, single physics problems).
- VVUQ often breaks down in the middle of the hierarchy, where theory and experiments are sparse.
- Aspects of the VVUQ activity are often absent from the application to the highest level.

5.1 What is Validation?

We will endeavor to define the context for VVUQ in the following sections providing selected excerpts from key papers and discussion to clarify matters on terms that remain difficult to separate from each other conceptually.

In a nutshell, compare simulation and experimental data to assess the adequacy of the simulation model.

The following quotes are from the Trucano et al. “What’s what” paper [Tru06].

Validation’s Purpose

First, we address the purpose of validation, which described well in this quote,

“Validation is the process of quantifying the physical fidelity and credibility of a code for particular predictive applications through the comparison with defined sets of physical benchmarks, consisting of experimental data.”

This offers a concise definition along with the previously described definition of benchmark. Next, we move to a purpose for validation,

“The purpose of validation is to quantify our confidence in the predictive capability of a code for a given application through comparison of calculations with a set of experimental data.”

and

“Validation deals with the question of whether the implemented equations in a code are correct for an intended application of the code.”

We note that the intended application will have an increasing influence on the nature of experiments used in validation as one moves higher in the validation hierarchy. We define how VVUQ and its impact on simulation-based decision making by addressing the impact of evidence and an analogy to the legal system,

“V&V is a process of evidence accumulation similar to the formation of legal cases.”

The workflow should ideally have a logical and linear path of activities with a well-defined sequence,

“In an ideal setting, validation should not be considered until verification analyses have been satisfactorily addressed. In practice, however, for modern simulation codes used to model complex physical phenomena, full resolution of verification and validation questions is essentially impossible. Instead, verification and, more acutely, validation are ongoing processes that should be subject to continual improvements.”

5.2 Metrics

Oberkampf and Trucano [Obe02] present an analysis of useful general characteristics of a validation comparison (called a metric in that paper) and a specific example. Roy and Oberkampf [Roy10] provide a concrete example validation metric use. A different formulation of validation comparisons is found in Zhang and Mahadevan [Zha03] and Mahadevan and Rebba [Mah05].

A validation metric ideally includes the following properties:

1. Has a physically meaningful interpretation, i.e., one that is thought to be relevant to the system response quantity (SRQ) of interest/feature of merit (FOM);
2. Is experimentally achievable;
3. is sufficiently sensitive to discriminate meaningfully different results, but not so sensitive that a “meaningful” result cannot be obtained.

These are characteristics (or shortcomings) of a (community’s) given (favorite) metrics; one needs to know how to weigh them when making a validation evaluation (even if this procedure is heuristic).

They are used to compare experiment with simulation in an application of interest for a purpose of interest. That is, the context and purpose of the comparison should be known at the outset. In the long term, the metrics evolve over time and their definition is iterative in nature based upon feedback from the results found applying the VVUQ processes.

5.3 Role of other assessment techniques in Validation

One obtains a “best estimate model” in validation, which does include uncertainties. (UQ within verification and validation.)

Verification aspect:

- Calculation verification is to be used to quantify numerical uncertainties
- Calculation verification provides a means by which to defensibly estimate the contribution of the numerical algorithm to the uncertainty budget.

5.4 Hierarchical view

VVUQ is necessarily hierarchical in nature just as modeling. Multiscale modeling is an important emphasis in computational science, and by the same token validation is multiscale. For most integrated engineering applications many physical models are joined together. The issue of design of experiment, quality of experimental data and comparison is simplest for a single physical process. As more physical processes are joined together, the entire process becomes more difficult, and uncertainty systematically increases. Ultimately, the specter of calibration becomes a necessity in an increasingly unprincipled manner moving toward the multiphysics end of the spectrum. This alone complicates the VVUQ process substantially.

5.5 Distinguishing Calibration and Data Assimilation from Validation and Uncertainty Quantification

There is some disagreement among experts as to the proper interaction of validation and calibration. Trucano et al. [Tru06] maintain that validation must be undertaken before calibration: “...calibration is logically dependent on the results of validation...”

for the reasons that “Validation provides important information to calibration accounting for model-form uncertainty” and “Validation provides information that is necessary to understand the ultimate limitations of calibration.” Nelson et al. [Nel10], however, see calibration and validation as being different aspects of the same overall “calibration/validation/prediction process”, maintaining that “...calibration is a task that is considered part of validation.”

The process of calibration and validation can be brought together into a single self-consistent framework via data assimilation. The structure of the solution and data is considered as a whole providing an optimal calibration with the simultaneous assessment of the state of validation for the model. This approach is being codified in a separate activity within the CASL program.

To start, given a calibrated model, you need to assess it before applying it to prediction.

The following quotes are from the Trucano et al. “What’s what” paper [Tru06] and are relevant to calibration and validation:

“Calibration and validation are essentially different.”

“Calibration and validation are distinct.”

I think the issue comes down to validation must be done with data NOT used in the putative calibration. The data used for calibration can contain useful information about aleatory uncertainty, but not epistemic uncertainty due to the fundamental nature of each category of uncertainty. The definition of “NOT” in this case could be problematic depending on the difficulty of achieving independent experimental data for a given circumstance. This goes to the ultimate validation providing faith in that the calibration is producing something akin to a good prediction to the system in question. The particular issue that is the most difficult to examine is the quality of the extrapolation of a simulation to situations where experimental data does not exist, yet the situation has application significance. For nuclear reactor operations this includes most severe accident scenarios.

“Validation and calibration in CS&E both depend on results of verification. We also claim that calibration is logically dependent on the results of validation, which is one way of emphasizing that calibration cannot be viewed as an adequate substitute for validation in many CS&E applications.”

“Calibration is conditioned by the probability that validation has been successfully performed.”

“Validation provides important information to calibration accounting for model form uncertainty.”

“Validation provides information that is necessary to understand the ultimate limitations of calibration.”

5.6 What Is Uncertainty Quantification and Its Purpose?

Establish defensible and credible bounds on a prediction. A prediction can be based on simulations, experiments or their combination. The focus is to establish credible bounds on predictability of the model on an intended application in a regime of interest. We have little to add that has not already been given at length by other authors. With this in mind we provide a series of relevant quotes from Trucano et al. [Tru06].

The following quotes are from the Trucano et al. “What’s what” paper [Tru06] regarding the definition and reasoning behind UQ:

“Quantification of Uncertainty is driven by the identification, characterization, and quantification of the uncertainties that appear in the code predictions of “Best Estimate” calculations. The thrust of Best Estimate plus Uncertainty (BE+U) is that prediction is probabilistic precisely because of our inability to complete V&V in some definitive sense and because of uncertainties intrinsic to complex modeling activities.”

The definition of prediction is central to the use,

“Here, a *prediction* is a calculation that predicts a value a set of these values prior to or in lieu of their physical measurement.”

We echo the the terminology and sentiment of Trucano et al., who focus on “confident prediction”:

“In computational science, of course, to some degree confidence is correlated with belief in the quantitative numerical, mathematical and physical accuracy of a calculated prediction. Intuitively, confident prediction then implies having some level of confidence or belief in the quantitative accuracy of the prediction. This further implies a willingness to use the prediction in some meaningful way, for example in a decision process.”

A key word associated with UQ is “predictability”.

UQ is about the model (but it must include experimental data + their uncertainties.) UQ does not (cannot) replace validation or verification; model is accepted as input to the UQ process.

What is the role of sensitivity analysis? Quotes from the Trucano et al. “What’s What” paper [Tru06]:

“Sensitivity analysis underlies the determination of the importance, hence priorities, of code elements that must be subjected to V&V in particular studies.”

“Parameter sensitivity is also important in guiding our studied reaction to model uncertainty. Parsimony, the reduction of the size of the parameter vector, is guided by sensitivity analysis and remains an important model selection principle. Sensitivity analysis is required for understanding the extent to which a model is complicated enough to be credible but not too complicated.”

“First, sensitivity analysis directly contributes to the definition of planned validation activities that culminate in the definition application of validation benchmarks as defined above. This centers on the use of a Phenomenology identification and Ranking Table (PIRT) in defining the key requirements of planned validation tasks.”

“Second, we stress that from the perspective of prediction, calculation of parametric uncertainties of calculations of benchmarks, either local or global, suggests the need to predict these sensitivities off the chosen benchmark sets.”

6. CONCLUSIONS AND RECOMMENDATIONS

In this document, we have described the concepts and flow of activities in VVUQ through a process by which uncertainties in modeling can be evaluated. Our context is broad and applies to the full spectrum of the modeling hierarchy. Moreover, the workflow we have described is general and is intended in the spirit of more of guidelines than as a “cookbook”. In particular, each component activity can be conducted with a fair amount of flexibility to define uncertainties in a manner appropriate for the physical model and the available experimental data. The important aspect to be emphasized is that each element—Verification, Validation, and Uncertainty Quantification—must be parts of an overall assessment of simulation quality.

While the suggested approach to VVUQ is moderately well codified and used in some quarters, practical details of VVUQ vary widely. Most validation cases encountered by modelers require significant application of domain-specific knowledge and experience. Unless the analyst has chosen very simple problems, each particular modeling problem will likely present its own challenges that will require insight, innovation, and determination on the part of the analyst to resolve. Despite these obstacles, VVUQ is a necessary part of the “due diligence” of scientific modeling. The outcome of the VVUQ process is a quantitative assessment of uncertainty, which provides the decision-making authority a degree of confidence to place in the modeling activity’s contribution to the process.

- [AIA98] AIAA, *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, American Institute of Aeronautics and Astronautics, Reston, VA, AIAA-G-077-1998 (1998).
- [ASM06] ASME, *V&V 10 - 2006 Guide for Verification and Validation in Computational Solid Mechanics*, American Society of Mechanical Engineers, New York, NY, (2006).
- [Bro06] Brock, J. S., Kamm, J. R., Rider, W. J., Brandon, S. T., Woodward, C., Knupp, P., and Trucano, T. G., *Verification Test Suite for Physics Simulation Codes*, Los Alamos National Laboratory report LA-UR-06-8421 (2006).
- [Boy90] Boyack, B. E., "Quantifying Reactor Safety Margins Part 1: An Overview of the Code Scaling, Applicability, and Uncertainty Evaluation Methodology," *Nucl. Engrng Design*, **119**, pp. 1–15 (1990).
- [Cac81] D.G. Cacuci, "Sensitivity Theory for Nonlinear Systems: I. Nonlinear Functional Analysis Approach", *J. Math. Phys.*, **22**, 2794-2802 (1981).
- [Cac81b] D.G. Cacuci, "Sensitivity Theory for Nonlinear Systems: II. Extensions to Additional Classes of Responses", *J. Math. Phys.*, **22**, 2803-2812 (1981).
- [Cac03] D.G. Cacuci, *Sensitivity and Uncertainty Analysis: Theory*, Volume 1, Chapman & Hall/CRC, Boca Raton (2003).
- [Cac05a] D.G. Cacuci and M. Ionescu-Bujor, "Deterministic Local Sensitivity Analysis of Augmented Systems. I: Theory", *Nucl. Sci. Eng.*, **151**, 55-66, 2005.
- [Cac05b] D.G. Cacuci, M. Ionescu-Bujor, and M.I. Navon, *Sensitivity and Uncertainty Analysis: Applications to Large Scale Systems*, Volume 2, Chapman & Hall/CRC, Boca Raton (2005).
- [Cac10] D. G. Cacuci, and M. Ionescu-Bujor, "Model Calibration and Best-Estimate Prediction Through Experimental Data Assimilation: I. Mathematical Framework", *Nucl. Sci. Eng.*, **165**, 18-44, 2010.
- [Han01] Hanson, K. M., Hemez, F. M., "A Framework for Assessing Confidence in Computational Predictions," *Exp. Tech.*, **25**, pp. 50–55 (2001).
- [Hel09] Helton, J., *Conceptual and Computational Basis for the Quantification of Margins and Uncertainty*, Sandia National Laboratories report SAND2009-3055 (2009).
- [Hem05] Hemez, F. M., *Non-Linear Error Ansatz Models for Solution Verification in Computational Physics*, Los Alamos National Laboratory report LA-UR-05-8228 (2005).
- [Kal03] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge (2003).
- [Kam08] Kamm, J. R., Brock, J. S., Brandon, S. T., Cotrell, D. L., Johnson, B., Knupp, P., Rider, W. J., Trucano, T. G., and Weirs, V. G., *Enhanced Verification Test Suite for Physics Simulation Codes*, Los Alamos National Laboratory report LA-14379, Lawrence Livermore National Laboratory report LLNL-TR-411291, Sandia National Laboratories report SAND2008-7813 (2008).
- [Knu03] Knupp, P., and Salari, K., *Verification of Computer Codes in Computational Science and Engineering*, Chapman & Hall/CRC, Boca Raton, FL (2003).
- [Knu07] Knupp, P., Ober, C., and Bond, R., "Measuring Progress in Order-Verification within Software Development Projects," *Engrng. Comp.*, **23**, pp. 283–294 (2007).

- [Lew06] J. M. Lewis, S. Lakshmivarahan, and S.K. Dhall, *Dynamic Data Assimilation: A Least Square Approach*, Cambridge University Press, Cambridge (2006).
- [Mah05] Mahadevan S, Rebba R., “Validation of reliability computational models using Bayesian networks”, *Reliab. Engrng. Syst. Safety*, **87**, pp. 223–232 (2005).
- [Nel10] Nelson, R., Unal, C., Stewart, J., Williams, B., *Using Error and Uncertainty Quantification to Verify and Validation Modeling and Simulation*, Los Alamos National Laboratories report LA-UR-10-06125 (2010).
- [Obe01] Oberkampf, W. L., “What are Validation Experiments?” *Exp. Tech.*, **25**, pp. 35–40 (2001).
- [Obe02] Oberkampf, W. L., and Trucano, T. G., “Verification and Validation in Computational Fluid Dynamics,” *Prog. Aerospace Sci.*, **38**, pp. 209–272 (2002).
- [Obe04] Oberkampf, W. L., Trucano, T. G., Hirsch C., “Verification, validation, and predictive capability in computational engineering and physics,” *Appl. Mech. Rev.*, **57**, pp. 345–384 (2004).
- [Obe07a] Oberkampf, W. L., and Trucano, T. G., “Verification and Validation Benchmarks,” *Nucl. Design Engrng*, **23**, pp. 716–743 (2007); also available as Sandia National Laboratories report SAND2007-0853 (2007).
- [Obe07b] Oberkampf, W. L., Pilch, M., Trucano, T. G., *Predictive Capability Maturity Model for Computational Modeling and Simulation*, Sandia National Laboratories report SAND 2007-5948 (2007).
- [Obe10] Oberkampf, W. L., and Roy, C. J., *Verification and Validation in Scientific Computing*, Cambridge University Press, New York (2010).
- [Pil06] Pilch, M., Trucano, T. G., and Helton, J. C., *Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper*, Sandia National Laboratories report SAND2006-5001 (2006).
- [Rid10] Rider, W.J., Kamm, J.R., Weirs, V.G., *Code Verification Workflow in CASL*, Sandia National Laboratories report SAND2010-7060P (2010).
- [Roa02] Roache, P., “Code Verification by the Method of Manufactured Solutions,” *J. Fluids Engrng*, **124**, pp. 4–10 (2002).
- [Roa04] Roache, P., “Building PDE Codes to be Verifiable and Validatable,” *Comput. Sci. Engrng.*, **6**, pp. 30–38 (2004).
- [Roa09] Roache, P., *Fundamentals of Verification and Validation*, Hermosa Publishers, Albuquerque (2009).
- [Roy05] Roy, C. J., “Review of Code and Solution Verification Procedures for Computational Simulation,” *J. Comput. Phys.*, **205**, pp. 131–156 (2005).
- [Roy10] Roy, C. J., Oberkampf, W. L., “A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing,” submitted to *Comp. Meth. Appl. Mech. Engrng.* (2010).
- [Sal00] Salari, K., and Knupp, P., *Code Verification by the Method of Manufactured Solutions*, Sandia National Laboratories report SAND2000-14444 (2000).
- [Sar98] Sargent, R. G., “Verification and Validation of Simulation Models,” in *Proceedings of the 1998 Winter Simulation Conference*, Medeiros, D.J., Watson, E.F., Carson, J.S., Manivannan, M.S., eds., pp. 121–130 (1998).
- [Sar01] Sargent, R. G., “Some Approaches and Pardigms for Verifying and Validation Simulation Models,” in *Proceedings of the 2001 Winter Simulation Conference*,

- Peters, B. A., Smith, J. S., Medeiros, D.J., Rohrer, M. W., eds., pp. 106–114 (2001).
- [Sch06] Schwer, L. E. “An Overview of the PTC 60 / V&V 10 Guide for Verification and Validation in Computational Solid Mechanics,” Reprint by ASME, available at: <http://cstools.asme.org/csconnect/pdf/CommitteeFiles/24816.pdf> (2006).
- [Sor07] Sornette, D., Davis, A. B., Ide, K., Vixie, K. R., Pisarenko, V., and Kamm, J. R., “Algorithm for model validation: Theory and applications,” *Proc. Nat. Acad. Sci. USA* **104**, pp. 6562–6567 (2007).
- [Ste01] Stern, F., Wilson, R. V., Coleman, H. W., Paterson, E. G., “Comprehensive Approach to Verification and Validation of CFD Simulations—Part 1: Methodology and Procedures,” *J. Fluids Engrng*, **123**, pp. 793–802 (2001).
- [Ste05a] Stern, F., Wilson, R. V., Shao, J., “Quantitative V&V Of CFD Simulations And Certification Of CFD Codes,” *Int. J. Num. Meth. Fluids*, **50**, pp. 1335–1355 (2005).
- [Ste05b] Stewart, J. W., *Measures of Progress in Verification*, Sandia National Laboratories report SAND2005-4021P (2005).
- [Tru02] Trucano, T. G., Pilch, M., Oberkampf, W. L. *General concepts for experimental validation of ASCI code applications*, Sandia National Laboratories report SAND2002-0341 (2002).
- [Tru03] Trucano, T. G., Pilch, M., Oberkampf, W. L., *On the Role of Code Comparisons in Verification and Validation*, Sandia National Laboratories report SAND2003-2752 (2003).
- [Tru06] Trucano, T. G., Swiler, L. P., Igusa, T., Oberkampf, W. L., and Pilch, M., “Calibration, validation, and sensitivity analysis: What’s what,” *Reliab. Engrng. Syst. Safety*, **92**, pp. 1331–1357 (2006).
- [Zha03] Zhang R, Mahadevan S., “Bayesian methodology for reliability model acceptance,” *Reliab. Engrng. Syst. Safety*, **80**, pp. 95–103 (2003).