SANDIA REPORT

SAND2015-7455
Unlimited Release
Printed September 2015

V&V Framework

Richard G. Hills, David C. Maniaci, Jonathan W. Naughton

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone:	(865) 576-8401
Facsimile:	(865) 576-5728
E-Mail:	reports@adonis.osti.gov
Online ordering:	http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.fedworld.gov
Online order:	http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2015-7455
Unlimited Release
Printed September 2015

V&V Framework

Richard G. Hills
Validation and Uncertainty Quantification Processes Department

David C. Maniaci
Wind Energy Technologies Department

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-0828

Jonathan W. Naughton
University of Wyoming
1000 E. University Avenue
Laramie, WY 82071

Abstract

A Verification and Validation (V&V) framework is presented for the development and execution of coordinated modeling and experimental programs to assess the predictive capability of computational models of complex systems through focused, well structured, and formal processes. The elements of the framework are based on established V&V methodology developed by various organizations including the Department of Energy, National Aeronautics and Space Administration, the American Institute of Aeronautics and Astronautics, and the American Society of Mechanical Engineers. Four main topics are addressed: 1) Program planning based on expert elicitation of the modeling physics requirements, 2) experimental design for model assessment, 3) uncertainty quantification for experimental observations and computational model simulations, and 4) assessment of the model predictive capability. The audience for this document includes program planners, modelers, experimentalist, V&V specialist, and customers of the modeling results.

CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

Due to the rapid development of computational physics based engineering models and the increased reliance on these models for decision support, model development and assessment processes have undergone increased scrutiny. This scrutiny has lead to the establishment of formal Verification and Validation (V&V) processes, such as those developed by NASA, DoE and various AIAA and ASME codes and standard organizations (AIAA, 1998, ASME, 1998, 2006a, 2006b, 2009, Oberkampf et al., 2007, Pilch et al, 2001, Trucano et al., 2002).

The development of computational engineering models has also changed the relationship between modeling and experimental work for engineered system design and evaluation. The use of modeling to support design has reduced the number of design, build, and test cycles required to ensure that complex systems to meet programmatic goals. As a result of increased dependence on modeling, there has been an increased emphasis on experimental testing for the direct support of model development and model assessment.

The primary focus of this document is to provide guidance on the development and execution of tightly integrated modeling/experimental programs based on well-established V&V practices for the purpose of model assessment. The framework emphasizes a highly collaborative subject matter expert driven planning processes as well as data driven validation procedures.

The intended audience of this document includes program planners, modelers, experimentalist, V&V specialist, and customers of model predictions. To facilitate this broad audience, this document is organized into multiple parts.

The introduction provides background on the V&V processes including definitions, and summarizes the purpose and scope of this document.

Part 1 focuses on program planning based on the physics simulation requirements needed to meet the customer's needs. All readers who participate in program planning, modeling, experimental work, and model assessment will find the content of Part 1 useful.

Part 2 focuses on the collaborative approach to experimental design that requires significant contribution from modelers as well as experimentalist. Modelers, experimentalist, and those that assess the models (i.e. V&V specialist) should read Part 2.

Part 3 addresses the estimation of uncertainty in both the experimental data and in the model predictions. This part contains more technical content than the previous parts, and is focused on providing guidance to modelers, experimentalist, and V&V specialist in planning experimental and modeling uncertainty quantification efforts.

Part 4 addresses model assessment, both from a validation (i.e. direct comparison to experimental data) and subjective (structured expert elicitation) point of view. Part 4 contains significant technical content and is of direct interest to those that assess the predictive capability of the model, such as V&V specialist. Program planners and customers may find the technical content less interesting, but will find that the product of the model assessment – the evaluation of model bias and uncertainty in this bias to be of high interest.

Following Part 4 is the concluding chapter and references.

1.0 INTRODUCTION AND BACKGROUND

1.1 What is a validation directed program?

A model validation directed program focuses on the development and execution of combined computational modeling/experimental tasks specifically designed to assess predictive capability of computational or analytical models for specific applications in a focused, well-structured, and formal manner. The applications that are typically targets of these formalized approaches are those that involve multiple physics on multiple scales, for which the predictive capability of the computations models can have significant economic, environment, or safety impact.

1.2 Role of computational modeling in the decision process

The relative importance of computational modeling and experimental work on the design or qualification of a system design varies from application to application. In some cases, computational models provide critical information during the design of a system whereas qualification is based on test data of a prototype of the final design. In other cases, modeling and testing serve complementary roles where the testing is performed under limited conditions due to economic and other constraints, and modeling is utilized to extend the assessment to other untested conditions. For other cases, modeling serves as the primary source of evidence that a system design meets requirements. Often, the system is a one of a kind, and the scale of the system is such that prototypes at the full scale will not be built. As the impact of modeling on the decision process increases, the importance of evaluating model capability using experimental data increases.

As computational models mature, computational resources increase in capability (i.e. High Performance Computation), and full-scale prototype development becomes less practical due to the complexity of the desired engineered systems, the role of experimental work shifts from providing data for system testing to providing data for model validation. As a result, the formalization of the process to maximize effectiveness of experimental work to support model validation becomes a primary driver in program planning and execution.

1.3 What is validation?

ASME V&V10-2006 (ASME, 2006) defines model validation to be “the process of determining the degree to which a model is an accurate representation of the real world, from the perspective of the intended uses of the model.” This statement can be broken down into several concepts:

- Validation is a measure of accuracy in representing the real world as approximated by measurements from validation experiments. As stated in ASME V&V 20-2009 (ASME, 2009), “There can be no validation without experimental data with which to compare the results of the simulation.”

Validation is a necessary component in the process of providing evidence of model suitability. Validation is not a binary statement about whether a model is valid or invalid, but rather a critical component in the overall assessment of the suitability of the computational model for the intended application. Other evidence of model suitability includes the Phenomena Identification and Ranking Table (PIRT) (Oberkampf and Roy, 2010) and the Predictive Capability Maturity Model (PCMM) (Oberkampf, et. al, 2007) discussed in later chapters.

- Validation focuses on an intended application, which limits the conditions for which the model is to be evaluated. Because computational models are usually intended to be predictive, validation may assess model accuracy for conditions that are different than those for the application.

- When validation experiments cannot be performed at the conditions of the intended application, validation should be performed over a hierarchy of experiments designed to test the various features of the computational model that are important to the application. While not providing direct evidence of model validity at the application conditions, the tests over the validation hierarchy provides evidence that the capabilities of the computational models have been assessed.

1.4 Purpose of this document

The purpose of this document is to provide guidance on the processes of validation driven program planning and execution that are based on methodology developed over the years by various organizations such as NASA, DoE and various AIAA and ASME codes and standard organizations (AIAA, 1998, ASME, 1998, 2006a, 2006b, 2009, Oberkampf et al., 2007, Pilch et al, 2001, Trucano et al., 2002) to help ensure that the model assessment process is complete and rigorous. Because the development of a validation process for a particular application relies heavily on Subject Matter Expertise (SME) to design a validation program that is reasonable given the resources (time, personnel, computational and experimental resources, and funding), the present document will emphasize the SME driven planning processes as well as data driven validation procedures.

The development and execution of this process requires well integrated team planning among those responsible for programmatic needs, computational model developers, model users, and experimentalists, and should consider the needs of the eventual customers of the modeling capability and results. The communication and tight coordination between these team members is one of the more significant benefits of this process, greatly increasing the chances of a successful model validation dataset and campaign.

The methodology presented in this document addresses the approach used to engage scientific/engineering subject matter experts to characterize and prioritize the issues associated with model prediction for the intended application. The development of a business plan to accomplish the results of the scientific planning is beyond the scope of this document and not addressed. However, the customers (internal such as program directors or external such as commercial users of the resulting software) of the modeling efforts are included into the planning process as the customer defines the requirements for the models and the anticipated scenarios to which the models will be applied, as well as understands resource limitations of the program.

1.5 The Process for Validation Directed Programs

The validation directed program and experimental planning processes are summarized in Figure 1.1. This figure is based on that presented by Trucano et al. (2002). The content in the upper blue box represents the integrated program planning that defines, justifies, and prioritizes the hierarchy of validation experiments. The lower blue box represents the design, execution, and computational modeling of specific validation experiments that have been identified for the validation hierarchy.

At the completion of the validation program planning (upper box), one should have a definition of the quantities of interest that are to be predicted at the system level (e.g., some measure of performance, model based environmental specifications or impact, or the probability that the a system remains safe), an assessment of the physics that must be adequately modeled to predict these quantities, a high level identification of the types of experiments required to address questions of predictive capability of the computational models, the preferred scale of these experiments (both physical scale and complexity), a prioritization of these experiments, and the associated planning document. One should think of this planning as a living process, with on-going changes expected due to knowledge gained from the execution of the validation experiments, additional model development efforts, and due to program resource reallocation.

Implementation of the steps indicated in Figure 1.1 should occur in the order shown. This figure is based on ongoing computational/experimental programs that were originally designed for scientific discovery rather than for model validation and have generally evolved through a less formal process. As the focus of these programs move from scientific discovery and associated model building, to prediction of performance of complex engineered systems using computational models, the formalization of the validation process helps focus the program goals, prioritize program needs, and adds transparency to the program decision process.

Because many of the items addressed in the sub-boxes of Figure 2.1 rely heavily on expert opinion (all items in the upper blue box), the entire planning and execution process is very team centric. The make-up of the teams can vary, depending on the specific items being addressed in the various boxes. More specialized teams are often appropriate for the items in the lower box, especially if the validation hierarchy requires diverse types of experiments and models.

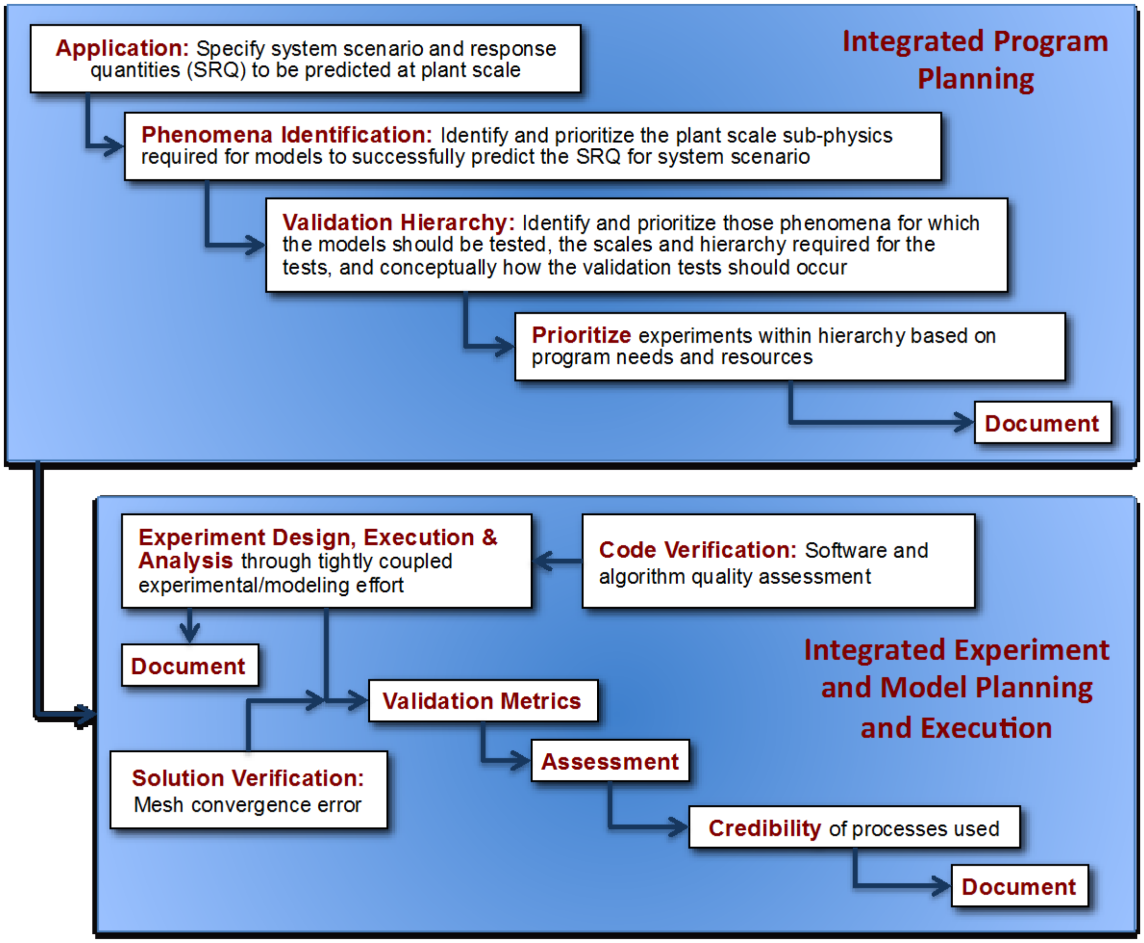


Figure 1.1 Validation directed program planning and implementation

1.6 Validation versus Credibility

Validation requires the comparison between simulation model output and experimental data. Such comparisons provide direct evidence of the ability of a model to simulate the correct physics, for the conditions tested. Engineering computational models are often developed to provide predictions of behavior for scenarios different from those for which validation data are available. As a result, the credibility of the model for the application scenarios requires some expert judgment.

The first step in assessing credibility is to identify what phenomena is important to be adequately captured by the model to meet the goals of its intended use. A well-accepted process to identify and rank the important phenomena is the Phenomena Identification Ranking Table (PIRT, Oberkampf and Roy, 2010). This table is developed using subject matter experts and identifies the important phenomena, classifies the phenomena as high, medium, or low importance; characterizes the current state of the computational model to represent this phenomena, and provides a gap analysis. An extended version of the PIRT will be introduced in the next chapter that provides additional information for program planning.

The assessment of model credibility for the phenomena identified by the PIRT for a specific application is based on sound modeling practices. Formal processes have been developed that break down these practices into six main elements (Oberkampf et. al., 2007). These are

1. Representation or geometric fidelity – are representation errors corrupting the simulation conclusions. For example, does the simplification used to represent bolts in a finite element analysis significantly affect the simulation results?
2. Physics and material model fidelity – how science-based and accurate are the physics and material models? Note that results of science-based models may be more credible that non-science-based models at conditions other than those for which they were tested or calibrated.
3. Code verification, including software quality assurance activities – are software errors or algorithm deficiencies corrupting the simulation results? Are sufficiently formal processes in place to minimize the risk of such errors, such as nightly regression runs to look for unintentional changes in code output due to code development; and code verification test suits to test code predictions against known analytical solutions.
4. Solution verification – are human procedural errors or numerical solution errors corrupting simulation conclusions? – What steps have been taken to ensure that user input errors have been eliminated, what evidence is there that the equation solvers converge, and what steps have been taken to characterize the uncertainty in predictions due to lack of grid convergence (i.e. for finite difference/volume/element algorithms)?
5. Validation – how accurate are the integrated physics and material models. Model validation is an experimental data based assessment of model accuracy, for the conditions of the validation tests, which typically involves coupled physics or other phenomenological effects.
6. Uncertainty quantification and sensitivity analyses – what is the impact of variability and uncertainty on system performance and design margins? The sources of these uncertainties include environmental uncertainties such as those that affect the initial and boundary conditions of the system, model parameter uncertainties such as used in material property relationships or other calibrated behavior, numerical uncertainties due to lack of grid convergence, and model form uncertainties identified through validation tests and through expert judgment.

These six elements are discussed in more detail in a later chapter in Part II of this document. The characterization of the overall risk of using a model for prediction is summarized in Figure 1.2. The left leg of the figure represents the assessment of the important phenomena for the application (PIRT) and the credibility of the computational model based on the six elements considered. This leg represents an assessment based largely on human judgment. The right leg represents the sources of uncertainty that are rolled up to the application prediction. These uncertainties include model parameter, numerical grid convergence, and model form uncertainty uncovered by the validation experiments and other sources. The overall risk of using the model, given model predictions of performance (or safety) margins, their uncertainty, and the credibility assessment can be notionally characterized as shown in Figure 1.3. Note that risk of using model results when the model predicts large design margins relative to the model's estimated uncertainty is less than that if the model predicts small margins relative to the estimated uncertainty. Model results for which the assessment of credibility is higher will likely result in less risk than results for which little credibility has been established based on the six elements discussed above.

The focus of this report is on the validation directed modeling/experimental R&D program planning and implementation and not on assessing risk for the users of the models. However, one should keep in mind that the ultimate goal is to provide the customer with not only predictions, but with information to help identify what the risks are of using the model in the decision making process.

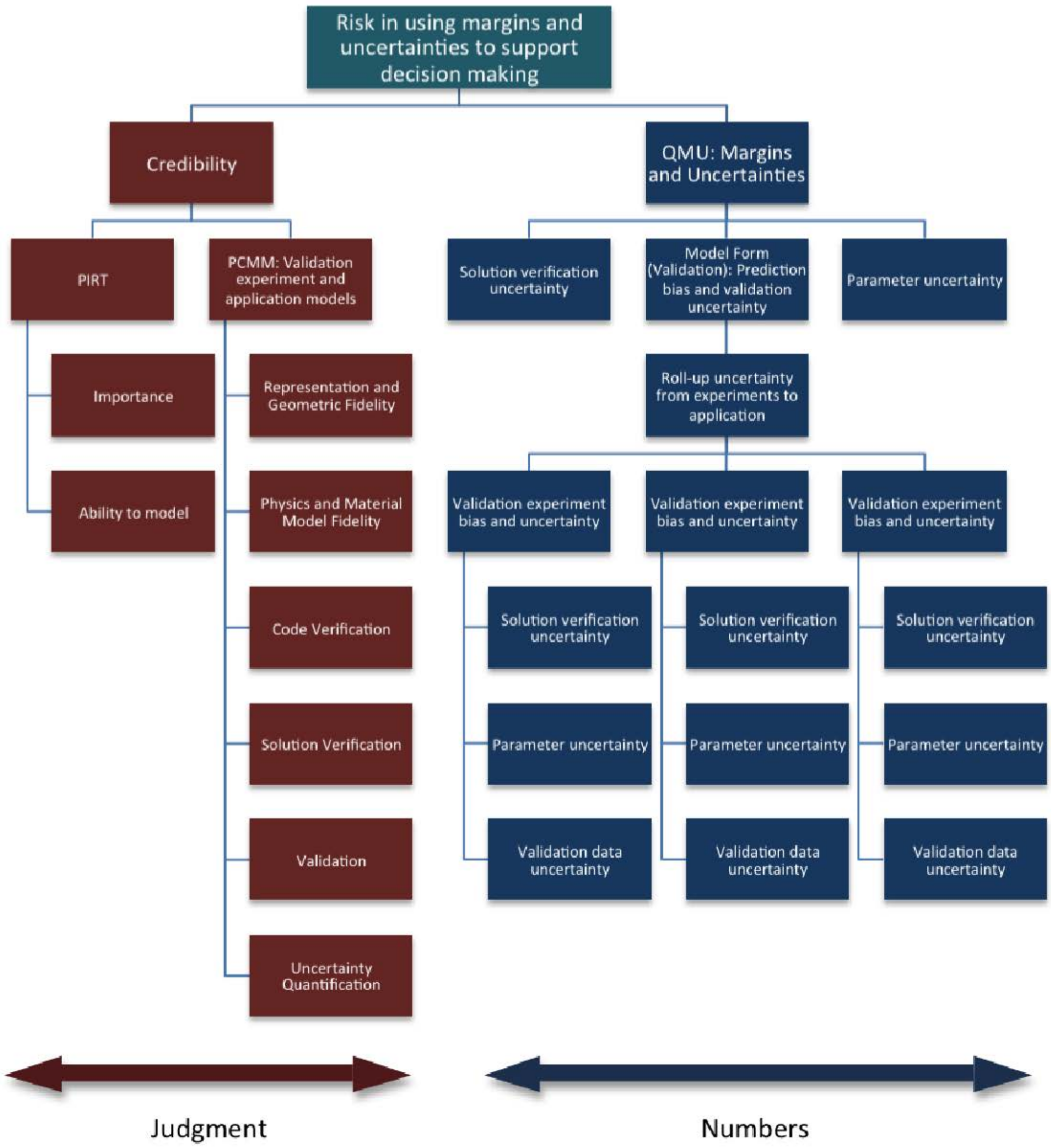


Figure 1.2 Risk of using a model for an application

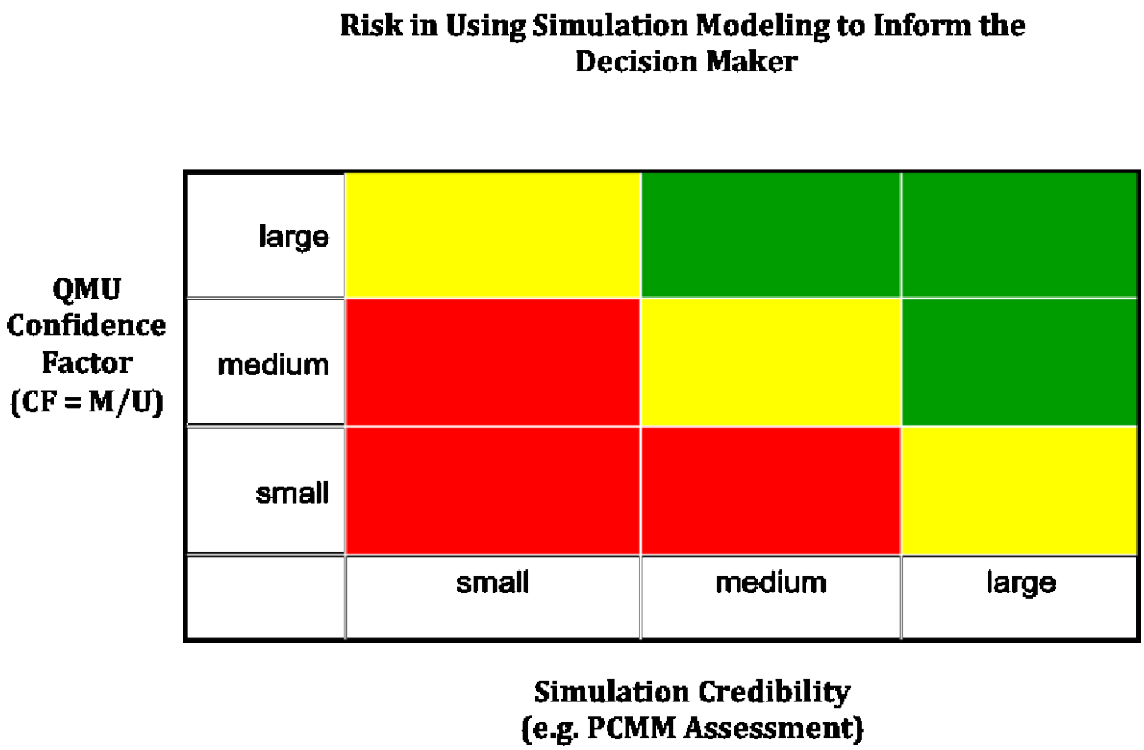
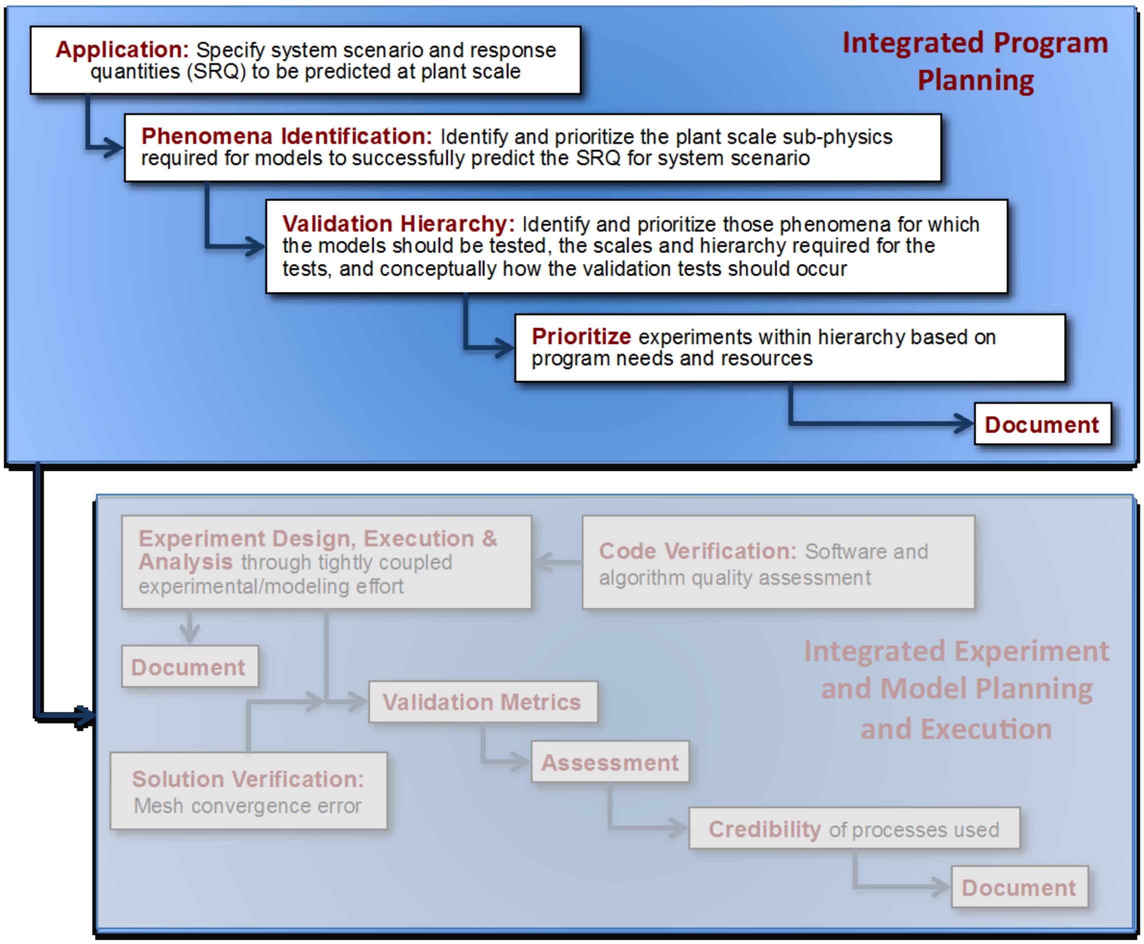


Figure 1.3 **Risk associated with computational simulation**: green – low risk, yellow – intermediate risk, red – high risk. The Predictive Capability Maturity Model (PCMM) is an expert elicitation tool used to assess simulation credibility (discussed in Chapter 11). M and U are the margin, and the uncertainty in the margin, between the predicted performance and performance requirements.

PART 1: INTEGRATED PROGRAM PLANNING



2.0 THE OBJECTIVE

- The modeling objectives specifies in precise terms 1) what the model will be used for, 2) the predicted quantities of interest, and 3) the role of the model in the design decision process for the customer.
- The modeling objectives serve as the basis for modeling and experimental program planning and implementation.
- The development of the objective requires close collaboration between the customers of the model results, experimentalist, model developers and model users.

The first step of integrated program planning is to define the objective or objectives of the computational simulation for the application. All further efforts discussed in the document will be based on the objective or objectives. The analyst requires a clearly stated objective to know what is expected of their models and what quantities will actually be used by a customer for design decision. The experimentalist will provide calibration, characterization, and validation data to the modelers to meet the modeling objective. The objective clarifies to the customers exactly what the modeler will provide which allows the customer to assess how the model will support the customer for the design, planning, and implementation process.

Example Objectives:

1. The computational simulation will be used as a scoping tool to predict the thermodynamic efficiency of various potential engine designs. Prototype engines will be built and tested for the most promising designs to confirm thermodynamic efficiency.
2. The loss of safety due to breach of a specific design of a storage tank, when exposed to a known range of jet fuel based pool fire scenarios, is to be predicted using the computational model. The qualification of the tank will be based on one tank / pool fire test deemed most stressing based on simulation results. The integrity of the tank for other fuel types and wind conditions will be assessed using the computational model.
3. Computation will be used to predict the daily power output of a wind plant, given the inflow conditions, terrain, and plant configuration. Computation will be the primary source of power output estimates prior to construction of the plant.

Note that each example specifies 1) what is to be predicted, 2) the scenario, and 3) the impact of the prediction on the decision process. The first two items are required to define the intended use. The last item specifies the impact that the computational model has on the final design and informs the modeler as to the rigor that must be exercised in developing, validating, and using the computational model.

Also note that none of the statements puts a quantitative specification on the allowed error in the prediction. If a quantitative specification is required, then the modelers and the customers must work together to develop a 'reasonable' specification (see side box).

Often, the meaning of terms in the objective needs further definition. The term 'breach' in the second example objective is nebulous. Does this mean the initiation of breach, or a crack of more than one inch, or a crack of sufficient size to depressurize a container in a defined amount of time? A computational model may not be able to predict breach with high accuracy. The model may be able to satisfactorily predict the initiation of plastic deformation, which can be used as an indicator of breach. In this case, objective 2 could be redefined as follows:

> *The loss of* <u>*assured*</u> *safety due to breach,* <u>*as indicated by the initiation of plastic deformation*</u>*, of a specific design of a storage tank, when exposed to known range of jet fuel based pool fire scenarios, is to be predicted using the computational model. The qualification of the tank will be based on one tank / pool fire test deemed most stressing based on simulation results. The integrity of the tank for other fuel types and wind conditions will be assessed using the computational simulation.*

A cautionary note on specifying model accuracy in an Objective

The ability to specify model accuracy requirements at this early planning stage is very difficult and seldom accomplished. While ballpark estimates of the prediction uncertainty are required to establish if the role of modeling is appropriate for the application, a specific pass-fail uncertainty specification can be counter-productive. Often the customer and modelers do not know the margins of safety that an actual design will have, or the details of the actual scenario. Designs that have large margins of safety can tolerate larger uncertainty in model predictions. Other applications can possess large uncertainties in the input conditions (such as uncertainty in inflow conditions), which greatly effects model prediction uncertainty. Customers are primarily interested in knowing how large the prediction uncertainty is, so that the design can build in enough margin to accommodate this uncertainty.

The phrase 'assured safety' are conditions for which we are confident the system is safe, rather than conditions at which the system transitions from safe to not safe. Note that the redefined objective provides enough information so that the modelers understand what is to be expected of the model in sufficient detail that they can take the next step, that of identifying and ranking the physics required to successfully model the quantity of interest for the scenarios of interest. The process to identify and rank the important physics is the topic of the next chapter.

3.0 PHENOMENA IDENTIFICATION RANKING TABLE

- Provides a structured approach to prioritize physical and other model related phenomena for an intended application1
- Identifies gaps between technical requirements and models, code capabilities, and V&V activities
- Focuses limited resources on prioritized activities that will assess or improve the predictive accuracy

3.1 PIRT: Background

The next step in developing a V&V plan is to identify the physics and non-physics based phenomena that are important to represent in the computational model to meet the Objective defined in the previous chapter. Formalized methodology to identify and rank such phenomena was developed by the nuclear power industry (Shaw, et al, 1988, Wilson and Boyach, 1998) and has been adapted by other organizations such as the DoE nuclear weapons community (Trucano et al., 2002, Pilch et al, 2001), and V&V Code and Standards committee (ASME, 2006) and authors (Oberkampf and Roy, 2010). The basic tool used for this process is the Phenomena Identification Ranking Table (PIRT).

The goal of a PIRT is to ensure both sufficiency[2] and efficiency. Sufficiency is provided through a process of consensus building by expert elicitation for an intended application. Efficiency is provided through prioritization of the phenomena and gap analysis of the simulation and experimental capabilities.

3.2 Who?

The PIRT is developed based largely on subject matter expert (SME) consensus opinion. The PIRT development team should be broad based with the team comprised of modelers, developers, code users, experimentalist, as well as the customers who are familiar with the application as defined by the objective. The inclusion of a Validation and Verification (V&V) specialists is beneficial as they are familiar with many of the processes that have been developed for V&V that are directly relevant to the assessment of model capability. Because the results of the PIRT will be used for program planning, the 'quality' of the team is paramount to the success of a planning effort.

---

[1] Some of the content in this chapter was taken directly from PIRT: How To developed by Amalia Black for internal use at Sandia National Laboratories (SAND2013-6285P). Dr. Black is a co-worker of the first author of the present report and gave us permission to use this content unquoted.
[2] Sufficiency - The goal of model assessment is to assess whether the model is sufficient for the intended application. Note that this does not necessarily require that the assessment of the model be for all phenomena touched by the application (i.e. completeness), but rather for the phenomena that is considered to have a significant impact on the prediction of the QoIs for the intended application.

Expert elicitation by its nature is subjective, but can benefit by utilizing information through a variety of objective methods, such as sensitivity analyses and numerical grid studies using the model, and existing validation results.

3.3 What?

The PIRT is a table that lists the important phenomena in the left column as identified by the team, and continues with a column characterizing importance of the phenomena, and one or more columns addressing the capability of the model to represent these phenomena. A gap analysis is performed with the results indicated by color codes (i.e. a stop light scheme). Additional columns can be added to the PIRT to suit program needs. For the present work, additional columns are added to aid in program planning. These include a description of the issues associated with the identified gaps, proposed responses to mitigate the effect of the gaps, and priority of the responses from a programmatic point of view.

The PIRT is based on information gathered from all relevant sources and should be updated as activities progress. The initial elicitation approach serves to build consensus in the technical community by soliciting and accommodating a broad spectrum of perspectives.

3.4 Scope

Identifying <u>all</u> of the phenomena that are relevant at the application scale for complex applications can be a daunting and even counter-productive task. The team should focus on those phenomena that are important to the Objective that <u>may</u> be inadequately represented by the model. The phenomena considered should be those that are important on the scale of the application. Examples of types of phenomena that may not be well represented by the computational model are listed in Table 3.1. Note that uncertainty quantification can be considered as a phenomenon, if the ability to predict the impact of natural variability on the quantities of interest is important to the application.

3.5 The Expanded PIRT

While many forms of the PIRT exist (Oberkampf and Roy, 2010), a form that is useful for program planning at multiple physicals scales is summarized in Table 3.2. Note that this table lists phenomena that are of high or medium importance to the prediction of the Quantities of Interest (QoIs) for the application for which the models are suspect in their ability to represent, the issues associated with representing the phenomena, and suggested responses to address these issues, including scale of possible tests. The inclusion by scale allows one to define a validation hierarchy for the tests listed in the last column. Not all issues are associated with tests, such as the need to perform a UQ study, a grid convergence study, or improve site characterization for use in the model. Guidelines for the ranking are provided in the information boxes following Table 3.2.

Table 3.1 Examples of phenomena for inclusion in PIRT

Type	Issues	Potential Responses
Physics	Important physics inadequately represented by model	Model development or experimental characterization to better represent the phenomena Model validation to assess the uncertainty associated with the inadequately represented physics
	Not clear if important phenomena is adequately represented by model	Model validation experiments designed to incorporate the effect of the phenomena
	Interactions between important phenomena	Model validation experiments that include the desired interactions
	Ranking of importance of phenomena included in model	Sensitivity analysis to rank importance for the application quantities of interest (QoI)
Model and Geometric Fidelity	Sub-components that affect prediction of application QoI poorly represented (e.g. fasteners represented by tied surfaces, e.g. fully welded)	Sensitivity analysis on subsystem level with higher fidelity model to assess impact of underrepresented components
	Geometric fidelity insufficient to represent behavior (e.g. stress concentrations around fillets)	Sensitivity analysis on subsystem level with higher fidelity model to assess impact of under-resolved geometry
	Grid resolution may be insufficient to capture behavior	Grid studies (solution verification) to characterize uncertainty due to grid resolution
	Fidelity issues due to de-featuring in model due to elimination of sub-components	Sensitivity analysis on impact of de-featuring
Characterization	Inadequate material property characterization	Material property characterization experiments (research existing and/or develop new)
	Inadequate inflow, boundary condition, or site characterization	Refine characterization of inflow, boundary and site conditions to the required fidelity using experimental or other techniques
	Inadequate characterization of model parameter uncertainties	Characterize from experimental data, data provided in literature, or from new experiments
Uncertainty Quantification	Uncertainty in model prediction not adequately characterized due to large run times of model	Approximate methods such as the use of surrogates, or more advanced UQ propagation techniques, to reduce run times

Table 3.2 Expanded phenomenon identification ranking table

Phenomenon	Importance at Application Level	Model Adequacy			Planning Priority	Issue	Response including scale
		Physics	Code	Val			
Phenom. 1	Medium	Low	Medium	Low	Medium	Environment source terms inadequate	Source term development followed by validation test at system scale
Phenom. 2	High	Uncertain	Medium	Low	High	Validation required	Validation test for phenomena at laboratory scale using XXX… test facility
Phenom. 3	Medium	Medium	Medium	Medium	Low		
Phenom. 4	Medium	Medium	Low	Medium	High	Grid not converged	Formalized grid convergence studies for sub-system to estimate uncertainty
Phenom. 5	High	Uncertain	Medium	Low	High	Validation required	Validation test at laboratory scale using a … test apparatus
Phenom. 6	High	Low	NA - Data based model	Low	High	Data to calibrate constitutive models required	Look for suitable data in the literature. If such data does not exist, perform experiments at laboratory scale to develop data to calibrate constitutive equations. Validate based on independent experiments at subsystem scale. These experiments should be …

Guidelines for Importance Ranking
High: First order importance of the phenomena. Model adequacy, code adequacy, and validation adequacy should be at the "High Level".
Medium: Second order importance of the phenomena. Model adequacy, code adequacy, and validation adequacy should be at least the "Medium Level".
Low: Low order importance of phenomena. Not necessary to model this phenomena with high fidelity for this application.
Uncertain: Potentially important. Importance can be explored through sensitivity study, discovery or validation experiments; and the PIRT revised.

Guidelines for Assessing Physics Model Adequacy
High: A mature physics-based model or correlation-based model is used that is believed to adequately represent the phenomenon over the full parameter space of the application
Medium: Significant discovery activities have been completed. At least one candidate model form or correlation form has emerged and is used that is believed to nominally capture the phenomenon.
Low: No significant discovery activities have occurred and model form is still unknown or speculative, or the model is known to provide poor representation of the phenomena.
Response: Inadequacies are addressed through an explicitly stated strategy. This may include further model development, acceptance of the inadequacy, the parallel use of alternate plausible models, the use of stylized bounding models, or other documented strategies.

Guidelines for Assessing Code Adequacy
High: The intended mathematical model is implemented in the code. An adequate regression suite is run routinely, and there are specific problems in the regression suite that test the implementation of the specified model. Verification problems have been run that test the correctness of the numerical implementation. Enabling code features are fully operational. There are no outstanding (reported) bugs or issues that can undermine usage of the model.
Medium: The intended model is implemented in the code. There is an inadequate regression suite or the regression suite does not specifically touch the phenomena of interest. The verification suite does not address the specific numerical implementation for the application. Certain enabling code features are not fully functional. There are no outstanding (reported) bugs or issues that can undermine credibility of the proposed calculations.
Low: The intended model is not implemented in the code. The regression suite or the verification suite inadequate. Certain enabling code features are not functional preventing the calculation from being run. There are out- standing code bugs or issues that must be resolved before model usage.
Response: Inadequacies are addressed through an explicitly stated strategy. This may include acceptance of the inadequacy, workarounds, or other documented strategies.

Guidelines for Assessing Validation Adequacy
High: Comprehensive validation evidence to use the model for the intended application. Numerical errors and predictive uncertainties of the model or correlation are quantified over the full parameter space of the application or over the parameter space of the database and the degree of extrapolation to the application is quantified and justifiable. The database used to condition the computational model is relevant to the application.
Medium: Partial validation support for model use in the intended application. Some validation evidence exists, but there are known gaps for phenomena of moderate or high importance. Numerical errors are unknown. Non-statistical comparisons of experiment data such as tabular comparisons or data trace overlays are employed. The degree of extrapolation (if any) may not be quantified. The database may not be fully relevant to the application.
Low: Insufficient validation support for model use. No significant comparisons with experiment data or ad hoc comparison of experiment "pictures" with prediction. The database is not relevant to the application.
Response: Inadequacies are addressed through an explicitly stated strategy. This may include acceptance of the inadequacy, workarounds, or other documented strategies.

Gap Assessment
The gap assessments can be indicated within the PIRT with green, yellow, and red stoplight color coding as shown in Table 3.2. Gaps are defined as shortcoming between the importance level and the current model, code, validation or material adequacy.
Green means that there is no gap, i.e., current adequacy is at the same level as the importance level. For example, a phenomenon with *medium* importance that has *medium* adequacy would be colored green. Yellow means that the adequacy is one step below the importance level, and red means the adequacy is two steps below the importance level. Blue is assigned to phenomena whose importance is currently unknown. The color code also denotes priority by which gaps should be addressed from a scientific perspective; that is, resources should first be focused on red and then yellow, while green requires no new resources.

Guidelines for Issues and Responses
The last two columns of the expanded PIRT provide more information of the issues associated with modeling of the phenomena and the specific responses planned to address the issues. These columns should be completed prior to the planning priority column (see box below). The expanded PIRT addresses the types of experiments that must be performed for characterization and for validation across the scales (or complexity) of the validation hierarchy. A graphical view of this hierarchy is shown in Figure 3.1 for the scales associated with wind plants. The validation hierarchy is discussed further in the next section.

Guidelines for Planning Priority
The gap assessment is based on scientific and engineering subject matter expert opinion and does not consider the resource required to address the issues listed in the PIRT. The priority for planned activities ideally follows the gap assessment results, with the gaps

denoted by red generally receiving the highest priority from a planning/resource perspective. One method to denote planning priority is to specify the anticipate time line of each activity (by quarter, or by year). Some significant gaps may require more resources than are available (time, experimental facility, computational resources) and as a result, be planned for later in the program (i.e. lower planning priority).

The program planning priority will be heavily impacted by the availability of resources. While the subject matter experts can take the first cut at prioritizing the work, the final priority will be very dependent on organizational resources, the needs and resources of the program directors, and the customers. As a result, program decision makers must be included in prioritization process as they will understand resource limitations that will likely have a significant impact on the planning prioritization results.

3.6 Validation Hierarchy

The expanded PIRT is the initial step in identifying the validation hierarchy. Generally, suites of experiments are performed over a validation hierarchy for complex applications. These are often of three types; material characterization experiments, ensemble validation experiments, and accreditation experiments. Ensemble tests can include separate effects tests (designed to test specific physics), integrated effects tests (designed to test interacting physics). Data from material characterization experiments are used to calibrate constitutive models, or to test calibrated models, are generally less expensive to perform, and can produce more and higher quality data (i.e., over multiple material samples). Ensemble validation experiments represent suites of experiments designed to test a computational model's ability to represent various aspects of the physics or subsystems relevant to the application. They generally do not represent the full complexity of the target application of the model. Data and corresponding computational predictions are compared to assess computational model performance. These experiments may or may not provide sufficient data to characterize variability across similar tests. Generally, these experiments are more expensive, producing less data of perhaps lower quality. Accreditation tests can involve sub-system or full system testing with application hardware under conditions more closely representing the design conditions or regulatory requirements of the target application. Such experiments are typically expensive, resulting in very limited data that may have very limited validation quality. Figure 3.1 illustrates one representation of the validation hierarchy. The complexity of the physics represented increases as one moves from the base of the triangle to the top. The layers illustrated range from material and constitutive properties characterization test (i.e. stress-strain curve, temperature dependent thermal conductivity), to separate effects of physics tests (elastic response, thermal radiation), to integrated effect/physics tests (coupled conduction and convection heat transfer), to sub-system tests (typically engineered sub-systems with behavior defined by coupled physics), to full systems at the top of the hierarchy. The experiments in a layer may represent the same physics evaluated under different conditions, or may represent different physics at the same or different conditions expected for the application. Other authors define the layers in the hierarchy differently, but the concept is the same. For example, Oberkampf and Roy

(2010) denote the layers in the hierarchy as 1) unit problem tier at the base, 2) benchmark tier, 3) subsystem tier, and 4) system tier. Other discussions on the validation hierarchy are provided by Pilch et al. (2001), Trucano et al. (2002), ASME and (2006b).

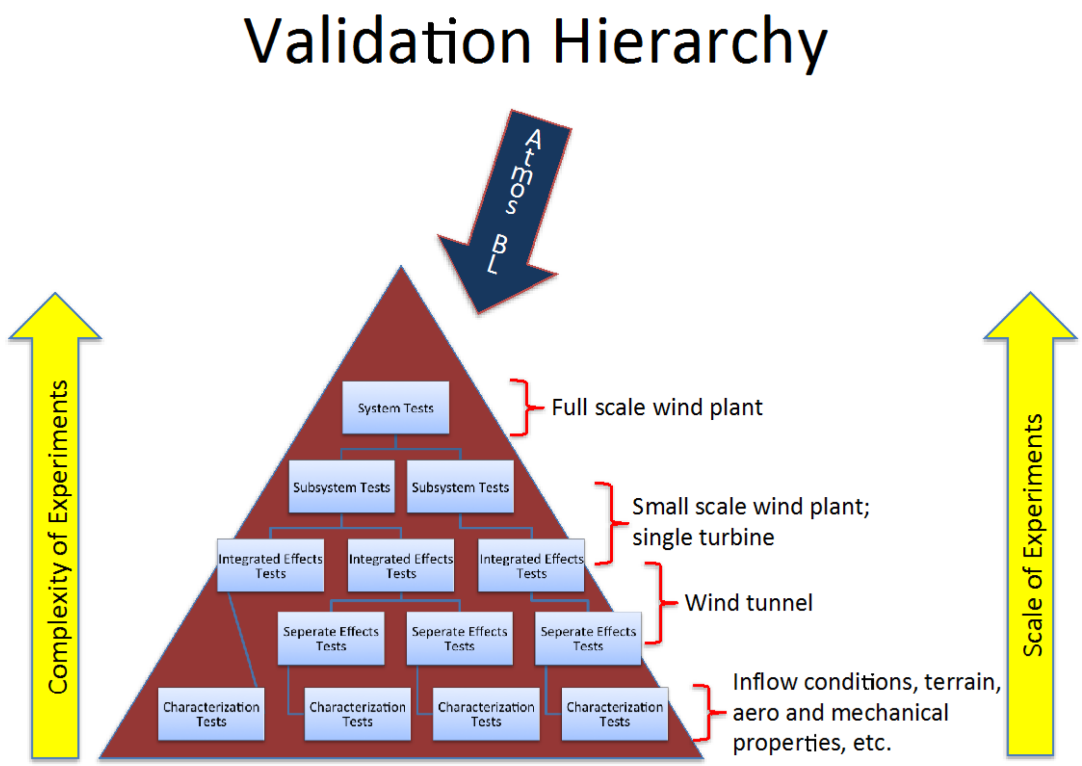


Figure 3.1 The validation hierarchy

Figure 3.2 represents a general relationship between the complexities of the experiments relative to the location in the hierarchy. Note that material characterization experiments generally use geometrically simple material samples and are ideally performed over the range of environmental conditions (for example, the temperature range) expected for the target application. Ensemble validation experiments represent more geometric and physical complexity, but are often not performed over the full range of environmental conditions expected for the target application. For example, ensemble validation experiments may be performed under lab conditions that do not represent the full complexity of conditions expected during the operation of the system (e.g. during a flight). Finally, because fewer accreditation experiments can be performed due to their expense, and because they are performed for a limited number of conditions, they cannot represent the entire design space of the intended application of the computational model. They may be useful as “acceptance” tests for the computational model.

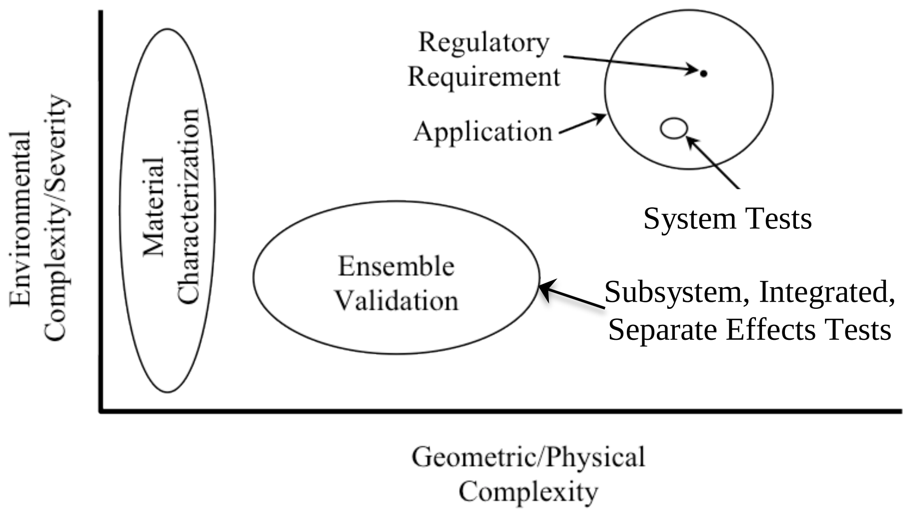


Figure 3.2 Experimental hierarchy complexity (based on Hills et al. 2008)

4.0 HIGH LEVEL PROGRAM PLANNING BASED ON THE PIRT

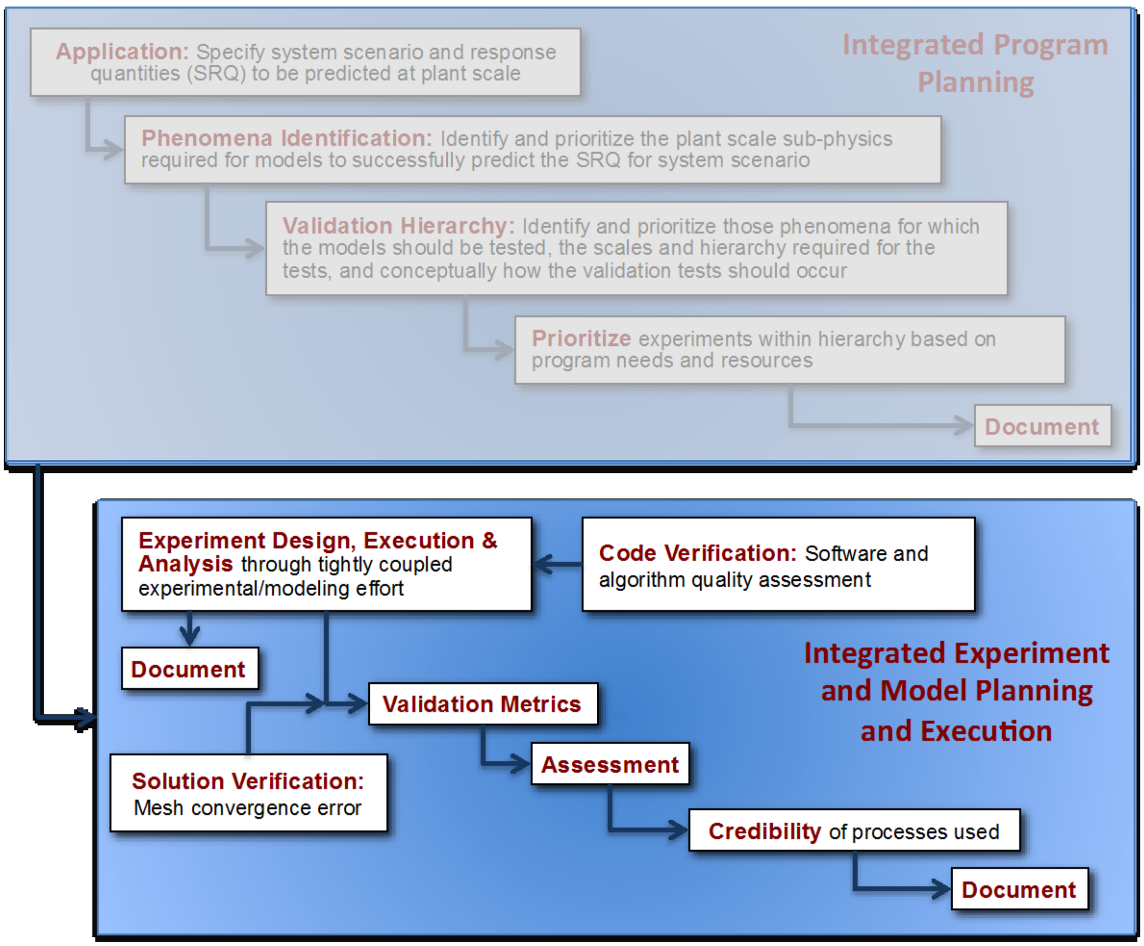
- The planning of the program is based heavily on issues and responses identified in the expanded PIRT and on the resource and other limitations of the program.
- Many of the program limitations are not scientific, such as the lack of sufficient funding, impact of funding cycles, lack of experimental capability, lack of sufficient computational resources, or insufficient model capability to meet the goals in the desired time frame.
- Planning often requires significant compromise and can result in exploring other approaches (such as qualification based on testing) to meet the customers' needs.

With the completion of the 1) Objective and the 2) extended PIRT, one can initiate more detailed planning to address the responses identified in the PIRT. This planning can lead to more specific tasks for 1) model development, 2) exploration of issues that may have an impact on prediction, 3) characterization experiments and the development of characterization methodology for the required inputs for the computational model, and model validation experiments to assess predictability for those issues that are of concern.

Because it is rare that a program has the resources to address all significant items identified in a PIRT, or in some cases to address even some of the high priority issues, compromises must be made during the planning process. The formal processes for such planning is outside the scope of this document and is very dependent on organizational structure and resources, funding sources, the organization's historic approach to planning, and the needs of their customers. Decision making planning teams often include senior scientist/engineers who can provide scientific input on the compromises that result when key issues that have been identified as concerns in the PIRT are either left unaddressed or delayed to later in the program, and can recommend other approaches to meet program goals.

Part II of this document assumes that the decisions have been made as to the types of experiments to be executed (or at least planned). Part II specifically addresses collaborative methodology to develop individual validation experiments to support the objective defined in Chapter 2 of Part I.

PART 2: INTEGRATED EXPERIMENT AND MODELING PLANNING AND EXECUTION



5.0 OBJECTIVE OF THE A MODEL VALIDATION EXERCISES

- The validation exercise objective specifies in precise terms 1) the intent of the experiment, 2) the predicted/measure quantities of interest that will be compared, 3) how the results will be used to support system level prediction, and 4) the basic configuration of the experiment for a validation issue identified in the PIRT.
- The development of the objective requires close collaboration between the customers of the model results, experimentalist, model developers and model users. The customers may be other modelers that use the results as input to their models on the next larger physical or system scale.

5.1 Background

The first step is to define the objective or objectives of the validation exercise. The planning for the modeling and the experimental work, as well as the post experiment validation assessments, should be based on this objective. One of the most important considerations in specifying an objective is that the experimental objective supports the overall system objective defined in Chapter 2, and addresses the issues identified in the PIRT. The three items that a validation objective should address are discussed below.

5.2 Experimental Intent

The experimental intent should be a validation issue that was identified in the PIRT as a program priority. The issue may relate to sub-phenomena (e.g. does the model capture the important influence of temperature dependence in the performance of a system), related to coupled phenomena (does the model adequately represent the decomposition of foam as a function of time when exposed to a high temperature convective environment), related to a component or subsystem performance (does the model adequately predict power output for a single wind turbine), or to a system level response. The validation issue specified in the PIRT may be indirectly addressed by the validation exercise due to the inability to design and perform an experiment that can isolate just that one issue. For example, a validation experiment generally represents several coupled phenomena, one of which may be the one identified as a prime concern in the PIRT.

5.3 Predicted/Measured Quantities of Interest

There are several considerations in choosing the Quantities of Interest (QoI) for the validation experiments.

1. Ideally, the relation between the validation QoI, and the quantity of interest for the full system as defined in the system objective, should be known and quantifiable. For example, if the QoI is output voltage for a component that is the input voltage for the next component, the relation is direct. Unfortunately for complex system models, this relation is often indirect. One may measure temperature and heat flux in a validation experiment whereas the QoI for the system is time of failure of the system due to

response of multiple components to a hot environment. This indirect relationship complicates the interpretation of the results of the validation exercise, but is often unavoidable.

2. The QoI should allow direct comparison between experiment and model results. If the measurement represents an average over some area (e.g. flux), then the model prediction should also represent the same average
3. The QoI may be a post processed quantity based on measurements. The QoI in this case may be power as a function of frequency. The model may directly calculate the power spectrum, or the predicted acceleration time histories may be post processed to give the corresponding spectrum.
4. The uncertainties in the QoI for both the experiment and the model prediction should be estimated. The end goal of a validation exercise is generally to measure discrepancy (bias) between the experimental results and the model prediction and the uncertainty on this bias (ASME, 2009).

5.4 Use of the validation results for the system level

The impact of potential validation results on the use of the model at the system level is not always addressed. As a result, this aspect of the validation is often neglected during the planning stages. Examples of typical model use scenarios based on validation results are listed below:

1. The resulting discrepancy or bias (experimental measurement minus the model prediction) plus or minus its uncertainty for the predicted output of a component will be used to characterize input and its uncertainty for the next component. Expert judgment is required as to when the discrepancy and its uncertainty are sufficiently large to cast doubt on the ability of the model to predict performance of the component over the range of scenarios for which the system is designed but not fully tested.
2. The model prediction, accounting for the resulting discrepancy plus or minus its uncertainty, may still provide conservative but useful results. For example, consider a computational model that under-predicts voltage output for a component. This under-predicted voltage may still exceed the minimum output voltage specifications for the component, and as a result provides a useful conservative estimate.
3. Expert judgment is used to evaluate whether the resulting discrepancy and associated uncertainty represents a risk in using the model for the intended application at the system level. If so, the model will be considered invalid and a mitigation strategy will be developed (e.g., additional model development, experimental characterization of the performance of the component, redesign of the component so that its performance is easier to predict using a model).
4. If the model fails to meet some prediction criteria for a particular application, the model will be calibrated to the validation results, and used with the understanding that the model is only useful for the scenarios covered by the validation experiments.

Clearly, expert judgment plays a key role in the decision to use the computational model for the application. The ideal approach would be to map the model validation discrepancies and their associated uncertainties to the predicted application QoI. This would allow one to better judge the impact of validation discrepancies at the subsystem level on model predictions at the application level. While this topic has been addressed (Hamilton and Hills, 2010a, 2010b, Hills, 2013, Kennedy and O'Hagan, 2001), the robustness of these approaches has not been established for cases that the application QoI is different from those measured in the validation experiments, and as a result, no recommendations as the suitability of these approaches are include here.

5.5 Basic configuration of the experiment

The basic configuration of the experiment is a statement about the configuration of the test article and the test apparatus. For example, the configuration can be a scaled model of an airfoil in a wind tunnel, or a component in an environmentally controlled oven, or a test article tested in an outdoor facility exposed to a well monitored natural environment. More detail on the configuration and data acquisition for the experiment are developed in later planning stages.

5.6 Example Objectives

Example objectives for the validation experiments are listed below:

1. The voltage output of multiple samples of component A will be measured for a pre-defined range of environmental temperatures and component inputs, and the results compared to model predictions for voltage output. The resulting differences between prediction and observation will be used to establish whether the model can be used to evaluate component response for untested temperatures and to estimate the expected uncertainties in the model predictions.
2. A series of puncture tests will be performed to evaluate the ability of the computational simulation to predict puncture as function of a tapered spike geometry and incoming velocity. The QoIs will be the cross-section area of the puncture crater and the depth of penetration of the spike. If the mean cross-section areas are within 30% of those means observed for each incoming velocity and spike geometry, and show the correct trends, the model will be used to provide insight into crater damage for other spike geometries over the same range of velocities for prototype design purposes.
3. Wind tunnel comparisons between measure and predicted velocity downstream of a test article will be made in several cross-flow planes. The differences between observed and predicted drag on the test article, and outflow velocities at the planes of measurements and the associated uncertainties will be used to access, through expert judgment, whether the model needs further development or whether the model and its uncertainties are suitable for use at the full scale.
4. Wind plant power production will be compared to predicted power production over several time scales, atmospheric boundary layer stability and inflow conditions, and turbine yaw conditions using a specified engineering model. The resulting differences between prediction and test observations of power, coupled with expert judgment,

will be used to assess model adequacy to for use in prototyping plant control methodology.

Note that the basic configuration and intent of the experiment, the predicted/measured quantities of interest, and a brief statement on how the results will be used to support the system level (full scale) analysis are specified in the above statements.

6.0 CONSIDERATIONS IN THE DESIGN AND IMPLEMENTATION OF A MODEL VALIDATION EXERCISE

- The design of successful validation exercises requires close collaboration between the experimentalist and the modelers during the conceptualization, design, execution, and post-processing phase of the experiments.
- The most successful validation exercises are those for which the computational model was used to help design the experiment.
- Uncertainty of the measurements and of the model predictions of the measurements plays a key role in model validation.

The objective defined for the validation experiments discussed in Chapter 5 can now be addressed in the design of the validation exercise. The phrase 'validation exercise' is used to reflect that the validation process involves experimental design including data acquisition design, modeling to insure that the conditions and scope of the experiment are sufficient to adequately test the model, the execution of the experiment and the computational model for the experiment as it occurred, the post-processing of the results including the evaluation of the validation measures and their uncertainties, and the interpretation of the results.

The focus of this chapter is on issues that should be addressed to ensure that the experimental results are suitable for model validation. The process to address these issues requires close and continuing collaboration between the experimentalists and the modelers.

6.1 Quantities of Interest

The experimental objective defines the quantities of interest (QoI) that are tied to the overall system objectives defined in Part I. The focus during the design and execution of a validation experiment should be on accurately measuring the experimental QoIs and assessing the uncertainty in the measured values. However, it is recommended that other quantities be measured during the experiments to obtain additional information on test article response, and if appropriate, the test apparatus response. For example, the validation QoI that will be used to assess the model may be heat flux on a surface. Temperatures can also be monitored at various locations throughout the test article at little added expense relative to the overall expense of the validation exercise. This additional information provides useful insight as to whether the experiment and test article behaved as expected, and often provides valuable information when unexpected results occur with either the experiment or the model predictions.

6.2 Physical Simulation and Computational Simulation

Ideally, a validation experiment provides a good representation of the response of a component or system under conditions that it will see during operation at the system level. Unfortunately, an experiment generally cannot exactly reproduce all of the variables of the anticipated environment; the test apparatus used can only simulate certain aspects of the planned environment; the test article is often a scaled, de-featured, or a prototype version of the article as it will be fielded; and there are often sensors installed on the test article that can affect the test article response. Thus, an experiment is usually, at most, a physical simulation of the article, component, subsystem, or system that will be fielded, and its field environment. Likewise, the use of a computational model to predict a test article response requires significant care and judgment as to how the model is defined, the level of geometric fidelity required to represent the test article response, and the impact of uncertainties associated with the required model inputs. In the end, the definition and execution of the model used to predict the experiment, and the design and execution of the experiment itself, requires significant judgment. One must not lose sight that both the model and the experiment represent simulations of the true behavior of the components, subsystems, or systems, as well as the environments during planned fielded operation.

6.3 Who?

The validation team should be composed of modelers who will support the design of the experiments, modelers whose models will be assessed, experimentalist who will design and execute the experiments, and customers of the results. The inclusion of V&V specialists is useful as these specialists are familiar with the processes described in this document. They generally have the statistical and uncertainty quantification background to help with the assessment of uncertainty in the model error, and can help with the interpretation of the results from a statistical perspective. Note that there may be more specialized representation on the validation team than there was on the PIRT team discussed in Part I, since the design and implementation of a specific validation experiment may require a more specific skill set. The customers may be internal customers who are eventually responsible for delivering a system level simulation product to another customer. The customer may be the systems engineer who will be ultimately be responsible for the design and qualification of a system. Multiple modelers with competing models can be accommodated during the design process to insure that the experiments can be used to assess strengths and weaknesses of various modeling approaches. The important aspect of the validation team is that it includes those that produce the data, those that model the experimental results, and those who do the comparisons (often statistical) and help interpret the results, all with a clear understanding of the customers' needs.

The validation team should remain in close communication throughout the validation exercise. There are often issues that must be addressed that become evident during the experimental design and fabrication process (of the test item and of the test apparatus). The modeler will often have strong opinions on how different approaches to address these issues affect the ability of the model to represent the experiment, and the ability of the experiment to address the issues associated with model validation. For example, based on issues with design or fabrication, an experimentalist may be concerned that a particular input cannot be adequately controlled. The modeler can often provide useful guidance on how well the input really needs to be controlled, or whether monitoring the transient behavior of the actual input conditions is sufficient.

Visits by the modeler to the test facilities often result in additional questions by the modeler on the configuration and conditions of the experiment, as well as provide the modelers with more insight as to some of the modeling assumptions that are made. It is rare for a modeler to leave their models unchanged after visiting the experimental site, as they invariable realize that some of their assumptions about the experiments were incorrect. In many cases, these assumptions can be addressed by modifying the model to better represent the experiment as it is actually configured, resulting in an improved validation exercise.

Because experimental anomalies often arise during the experimental execution, especially for more complex experiments that are further up the validation hierarchy, having a modeler and the V&V Subject Matter Expert (SME) present to see these anomalies and work with the experimentalist to hypothesize the source and impact of these anomalies can often help improve the overall validation process. Sometimes these anomalies can be treated by adjusting the input or configuration of the model, or by post processing the results.

Differences between experimental data and model predictions are expected, even for ‘perfect’ models, due to the presence of uncertainty in both the experiment and in the parameters of the model. However, the observations of trends in the validation differences over time and space suggest that systematic effects are present. Often these trends are due to model form error. These trends can also be due to systematic error in the measurements, such as a thermocouple with a poor thermal contact or high lead losses, or calibration issues associated with the data reduction equations that convert the quantity measured (e.g. electrical resistance) to the quantity desired (strain). These issues are best resolved through collaboration between the modelers and the experimentalist.

6.4 Experimental Configuration and Computational Models

The experiments should be designed to address the model validation issue or issues identified in the PIRT of Part I. The team should understand the basic concept of how the validation results will be used to inform the decision maker as to the suitability of the model for the intended application and the possible additional work that may be required if the validation results indicate that the model is not adequate. The ultimate characterization of what is suitable will rely heavily on interactions between the analyst using the systems model and the customers or their representatives.

For the case of validation, the computational models should simulate the validation experiment as it was realized, requiring sufficient knowledge of the test article, the experimental configuration and its impact on the test (e.g. wind tunnel wall effects), the environmental conditions (e.g. boundary and initial conditions), the impact of sensors on the performance or response of the test article, and the uncertainties such as those associated with the diagnostic error, control of the environment, and test article configuration. In some cases, the model must also model the test apparatus as there may be two-way interactions between the test article and the test apparatus response.

In contrast to discovery experiments where the experimentalist often attempts to carefully control the environment (e.g. maintain a constant temperature over the duration of the experiment), model validation experiments can be less controlled as long as the actual environment is measured during the experiment and the model is capable of incorporating the environment as it occurred. Thus validation experiments may require less control of the environment, but require detailed, accurate measurements on how the environment and other model boundary conditions vary across space and time.

During the design phase, the experimentalist and modelers should agree on the following:

- the basic experimental configuration including the characteristics of the test article and test apparatus as it affects test article response,
- the location and type of instrumentation used to monitor environmental conditions that are required as input to the model,
- the location and type of instrumentation used to monitor the response of the test article and perhaps the test apparatus,
- the sampling rates and spatial resolution of the data required,
- the supplemental experiments required to characterize those properties of the experiment that are required by the model (e.g. thermal properties of the materials in the test article, turbulence characteristics of a wind tunnel), and
- the methodology used evaluate uncertainty in the measurements and the model predictions.

Because the estimation of uncertainty is a key component in a validation exercise, a separate section is provided on uncertainty in a following chapter.

6.5 Experimental Characterization

Experimental characterization is the estimation of those quantities that are used as input for the model of the experiment. Experimental characterization includes

- measured or controlled initial and boundary conditions,
- measured or controlled loading conditions,
- characterization of the material properties, experimental configuration (including geometry), site conditions, or any other quantity that must be used as input for the model. For example, coupon tests are often performed on material samples to evaluate stress-strain curves. Site characterization experiments or analysis is often performed to characterize the topography, soil properties, vegetation, wind characteristics, thermal characteristic, and other characteristics that can be utilized by the computational model, and
- characterization of the uncertainties in the above model input quantities and the measurements so that that the modeler can account for these uncertainties in

evaluating the corresponding uncertainty in the model prediction of the experimental measurements.

Note that the last two bullets represent supplementary experiments that must be planned as part of the validation exercise.

6.6 Failed Validation Exercises

There are several issues that can lead to the failure of a validation exercise.

- Sufficient information about the test article and about the experiment (such as adequate characterization of a boundary condition) as it was executed, was not acquired. As a result, the analyst must make assumptions about certain aspects of the experiment in developing their model, which can call into question the validity of the resulting assessment. Validation experiments are different than discovery experiments or proof-of-concept experiments in that they require a strong focus on the acquisition of the information required so that the computational model can be unambiguously applied to the experiment as it occurred.
- The uncertainty in 1) the environmental conditions, 2) the measurements of response, and 3) the configuration of the test article or test apparatus are not sufficiently characterized. Note that, due to uncertainty, one always expects to see differences between experimental measurement and model prediction, even for perfect models. The question is whether the model simulation is consistent with the experimental results, given the uncertainty in the validation exercise. Answering this question requires estimates of both the experimental and computation uncertainty.
- The experiment, as designed and executed, did not provide data that could be used to assess how well the model represents the phenomena of interest. For example, the conditions of the experiment and the resulting response of the test article may not be sensitive to the phenomena of interest. Or the experiment may be sensitive to multiple phenomena, all of which may not be well represented by the model. In the latter case, one may have compensating model errors suggesting better representation of the phenomena by the model than actually exists. Where significant differences do occur in this last case between experiment and model, establishing which of the competing phenomena is the cause for the difference can be difficult.

Overall, there are many compromises that must be made in designing a validation exercise. Close collaboration between the analyst who will be modeling the experiment and the experimentalist is required during experiment planning, design, and execution to minimize the risk of ambiguous results.

6.7 Modeling Supported Design

The most robust validation exercises are those that are modeled in support of the design of the experiment. Modeling is often used to help define the temporal and spatial extents of the experiments, to help optimize sensor location, and to help optimize the boundary conditions and test article configuration so that the test article response occurs in a fashion that is sensitive to the phenomena of interest.

Modeling during design is also important to the modeler, as it is the modeler's responsibility to work with the experimentalist to obtain the right information so that the model can be applied to the validation experiment. If the model is developed and executed during design, the modeler is required to estimate the information listed in Section 6.5 plus information that is obtained from other sources (e.g. properties for common materials). The modeler thus can develop an inventory of the information required to unambiguously apply the model to the experiment, and can work with the experimentalist to develop such information, when appropriate.

6.8 Validation Metrics and the Role of Uncertainty

The validation team should plan on how the model predictions will be compared to the experimental observations during the design phase. For physics based models, these comparisons are generally based on quantitative validation metrics or measures.

Several types of quantitative validation measures or metrics have been proposed. These measures include mathematical metrics, which requires that the values for the metrics be non-negative and meet other mathematical requirements for the metric. The measure can also be in terms of a probability, i.e., the probability of the observed differences between model predictions of a valid model and the experimental observations, given the modeled and measured uncertainty in the validation exercise. The measure can be a signed quantity, such as a signed difference between model prediction and experimental observation, and the uncertainty in this difference due to experimental and prediction uncertainty.

This last measure often provides the most flexibility for the following reasons:

- Customers of computational model results are generally interested in accuracy as measured by difference between prediction and true value (defined as model bias), rather than some more abstract measure of accuracy, such as the probability of the difference.
- For cases where model errors are significant, customers are often interested in the sign and ranges of possible model errors (i.e. under or over-prediction), as this range can have a significant bearing on the suitability of the computational model to assess whether a design criteria is met in a conservative fashion. These ranges are sometimes characterized in terms of several multiples of standard deviations around the expected model error.
- Measures that assess validity relative to uncertainty in the validation exercise can be misleading, as the larger the uncertainty that exists in a validation exercise, the more likely that an invalid model prediction will appear consistent with the data, within the uncertainty of the exercise. In contrast, estimating model error (e.g. differences or biases) and its uncertainty, leads to ranges of the estimated model error, clearly communicating the impact of uncertainties associated with the validation exercise.

- The estimation of model error and its uncertainty reflects the accuracy of the model predictions relative to the experimental observations, independent of the accuracy requirements of the intended applications. This allows one to characterize the computational model error and uncertainty, and then evaluate acceptance or rejection of the usefulness of the computational model as a separate step as the design evolves and the design margins become more evident.

One of the distinguishing features of validation exercises is the central role of uncertainty in the validation comparisons. This uncertainty is due to the characterized uncertainty in the measurements and the uncertainty in the model predictions. There are often choices as to where to account for the uncertainty. For example, measurement uncertainty is generally considered the uncertainty induced by the instrumentation and data acquisition systems, or due to the uncertainty introduced through the use of data reduction equations (i.e. calibrated models that convert the quantity measured to the QoI such as electrical resistance to temperature, or accelerometer data to shock response spectrum). The International Organization for Standardization on the Guide to the Expression of Uncertainty in Measurements (ISO GUM, 1995), the ASME Code and Standard PTC 19.1 (ASME, 2006a), and Coleman and Steele (2009) provides guidance to estimate measurement uncertainty due to measurement and data reduction equation uncertainty. These standards form the basis for the ASME Verification and Validation standard ASME 20-2009 (ASME, 2009).

The impact of other forms of uncertainty on measurements, such as environmental or boundary condition uncertainty can be considered as either an uncertainty in the measurements or in the model. Unless the experimentalist performs repeated experiments to explicitly evaluate the variability in the measured QoI's due to environmental and other sources of uncertainties, such uncertainties generally require a model to define a relationship between the uncertain conditions affecting the performance of the test item and the uncertainty in its response. Because this model is often the model being tested, the incorporation of these forms of uncertainty is almost always through the computational model. Ultimately, the modeler will be responsible for characterizing the uncertainty in their prediction due to model input uncertainties, such as environmental uncertainties for the anticipated application. For this reason, the following recommendations are made:

Experimental Uncertainty

The estimation of diagnostic and data reduction uncertainty should be the responsibility of the experimentalist. As discussed in ISO GUM (1995), the sources of these estimates can be statistical analysis of repeated measurements or through other sources such as sensor manufacture specifications, analysis of uncertainties in data reductions equations, and expert opinion.

Computational Model Uncertainty

The sources of uncertainty that are typically considered by modelers for validation experiments are 1) model parameter uncertainty which includes parameters associated with initial and boundary conditions, environmental conditions, forcing functions, and constitutive equations, and 2) solution verification uncertainty which is the uncertainty associated with lack of linear or non-linear equation solver convergence and the lack of finite element/volume/difference grid convergence.

Methodology to estimate model uncertainty due to parameter and grid convergence uncertainty is provided in Oberkampf and Roy (2010), in ASME (2009), and in Roache (1998, 2009).

Because the roll of uncertainty is so important to a validation exercise, approaches used to estimate this uncertainty should be identified during the validation exercise planning stage.

Each of these sources of uncertainty are discussed in the following chapters.

6.9 Archiving, Challenge Problems, and Supporting the Broader Modeling Community

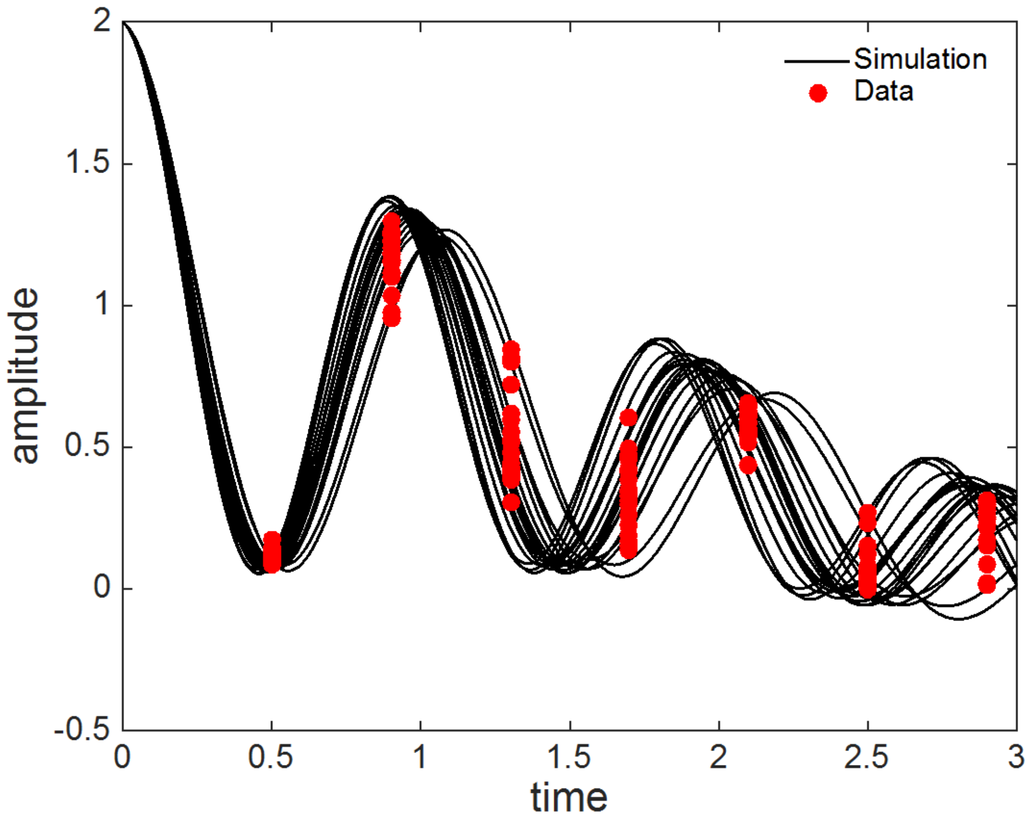
Validation experiments provide data that are very useful to a broader modeling community if the information that was required to test the models is adequately documented and archived. The organization of the validation exercise into a challenge problem provides a well-established approach Babuška, et al. (2008) and Helton and Oberkampf (2004) to insure that the experiment and its data is adequately archived and documented for the broader community.

A challenge problem is a carefully designed validation exercise for which all of the required information is made available to modelers from multiple organizations. The modelers from the different organizations develop and apply their models to predict the experimental measurements to be used for validation. Generally, a workshop is organized for the modelers and experimentalist so that results can be compared, issues identified, and lessons learned. Several approaches are often used in distributing data to the modelers. The first approach is to distribute all data prior to the workshop, including the QoI response data from the experiment that will be used for assessment. This allows the modelers to do the comparisons between experiment and prediction prior to public view. Unfortunately, this also allows the models to be tweaked based on the observed response data prior to presentation at a workshop. A second approach is to perform a blind validation exercise. In this case, all data required to configure and run a computational model for the experiment (geometry, boundary conditions, uncertainties, etc.) are provided prior to the workshop. The modelers then provide their predictions of the QoI to a third party, who compares the results to the experimental response data. The third party reports the results at the workshop and through a report, and releases the remaining data. The reported results can be identified by organization, or can be presented anonymously. If more then one validation experiment is planned, a hybrid approach can be utilized for which the experimental data is released prior to modeling for experiment 1 and held until after independent validation (i.e. blind) for experiment 2.

The challenge problem approach to a validation exercise has several advantages.

- The use of the data by external organizations helps insure that the validation experiments are well documented.
- Multiple approaches to modeling are demonstrated and tested.
- The models from multiple organizations tend to be fairly well developed as the results will be presented at a workshop.
- The variability between model approaches (often reflects analyst to analyst variability) becomes evident as the results are shown.
- The information archived and documented from the validation exercise can be used by participants and non-participants well after the workshop, to assess and to drive future development of computational models.

PART 3: UNCERTAINTIES IN MODEL VALIDATION



7.0 OVERVIEW OF UNCERTAINTIES RELEVANT TO MODEL VALIDATION

- Uncertainty in the measurements and in the model predictions plays a key role in model validation.
- There are many approaches to classify uncertainty. The choice should be based on the needs of the application.
- Methodology is well established to estimate measurement uncertainty and to quantify several sources of model prediction uncertainty.

The word "uncertainty" in a measurement or in a model prediction conveys the concept of doubt in the validity of the results to represent the true value of interest. Many approaches are used to classify or characterize uncertainty, and there can be different interpretations of particular approaches. The following is a brief overview of different concepts of uncertainty. More detail in estimating model uncertainty will be provided in a later chapter. We begin with a discussion of aleatory and epistemic uncertainty.

7.1 Aleatory and Epistemic Uncertainty

Aleatory or random uncertainty (also called statistical uncertainty) is uncertainty that is due to a natural randomness in a phenomena or process. Because this randomness is naturally occurring, it is irreducible. Consider multiple samples of a component. The performance of this component will vary from component to component due to the variability introduced due to manufacturing tolerances, and due to the variability in the materials themselves. This variability is inherent in the performance of the components and cannot be reduced without improving the manufacturing process, or changing material selection and component design. This form of uncertainty in performance is represented by the probability of occurrence of the performance outcome, given the full population of components. Familiar statistical methodology is applied to estimate either characteristics of the uncertainty (standard deviation), or to estimate the population probability distribution function (PDF). Note that, while the PDF is fixed for the population of components, this distribution must be estimated if one cannot sample the full population.

Epistemic uncertainty is due to lack of knowledge about the process that can be reduced if more knowledge is available. For example, the height of John sitting at the table is unknown. However, the uncertainty in estimating his height (epistemic) can be reduced by 1) asking John to stand up, and further reduced by 2) measuring John's height. If one is interested in the height of *any* person who *might* enter the room, then there is no specification of which person one is talking about, and their potential height could be any height from the population of heights (aleatory). Thus the question being asked (specific vs. any) can have an impact on the classification as to whether an uncertainty is epistemic or aleatory (Hofer, 1996). One can use sampled heights from the population to estimate a probability density function (PDF) for the population. The uncertainty due to randomness of heights within the population is aleatory, the true PDF for the population is fixed; and the lack of knowledge of this true PDF is epistemic (i.e. we must estimate the true PDF from a finite number of samples). The epistemic uncertainty associated with estimating these PDF's can in principle be reduced with additional random samples from the population.

Some application fields refer to aleatory uncertainty as variability, random uncertainty, or irreducible uncertainty; and epistemic uncertainty as reducible uncertainty, systematic uncertainty, or simply uncertainty.

The approaches used to characterize epistemic uncertainty are varied. Example approaches include Dempster-Shaffer evidence theory, Probability-Box methods, Fuzzy Logic, and Bayesian and Maximum Entropy approaches for which the epistemic uncertainties are conceptualized as probability based plausibilities (Dempster, 1967, Shafer, 1976, Ferson et. al., 2003, Zadeh, 1978, Jaynes, 2003, Gzyl, 1995). As these approaches and associated tools mature, the ability to use these methods for applications will increase.

7.2 When is the distinction between epistemic and aleatory uncertainty useful?

7.2.1 Limited number of samples

One of the more common examples of epistemic uncertainty encountered in statistics is that associated with the estimation of probability distributions based on a finite number of samples from an aleatory population representing a random process. Note that if one had access to the entire population, one could calculate its mean, standard deviation, and as well as the probability or cumulative probability of an occurrence of any outcome without uncertainty. In reality, one generally has a limited number of samples from a population, resulting in uncertainty in the estimation of the mean, standard deviation, or the estimated probability density function characterizing the population. Because the population is defined (i.e. all components manufactured during a specified time period), the population has a unique mean, standard deviation (assuming the second moment exists), and PDF. The uncertainty in the mean, standard deviation, PDF, or any other statistical characterization of the population is epistemic, reducible if more data is available.

7.2.2 Design criteria is a probability

For cases where the design criteria is a probability, such as a requirement that there must be less than 1 in 10,000 chance that the component will fail due to manufacturing defects, then an important uncertainty is the uncertainty in estimating the failure probability from a limited number of samples. The uncertainty (epistemic) in estimating a percentile location in an aleatory distribution is often characterized by a tolerance interval (Hahn and Meeker, 1991). We may wish to know what the 90% confidence interval is (epistemic) that at least 99% of the population of a particular model (aleatory) of an automobile obtains an EPA gas mileage rating of 30 MPG.

Probabilistic risk assessments for high consequence accidents are used for environment assessment (EPA), for nuclear reactors (NRC), and for nuclear weapons (DOE/SNL). The design requirements often require that the Probability of Loss of Assured Safety (PLOAS) does not exceed a specified probability. Thus the QoI is a probability, and any estimate of this QoI will possess uncertainty. A common approach in these types of assessments is to estimate the PLOAS base on random variability in the system (i.e. aleatory), and to estimate the uncertainty in the true value for PLOAS as epistemic uncertainty due to lack of knowledge. In these applications, the epistemic uncertainty is usually also represented by probability distributions. This analysis process is often referred to as a probability (subjective) of frequency (i.e. frequency of failure due to random effects) and is the recommended approach if the design QoI is a probability. This approach requires a separation of these epistemic and aleatory uncertainties when using the model to predict uncertainty (Helton, 2011, Hofer, 1996, Pilch, et. al., 2011).

7.2.3 Classification of reducible uncertainties

Another case for which the distinction between aleatory and epistemic uncertainties are useful is when one wishes to characterize sources of uncertainty by type so that the overall uncertainty is reduced. Aleatory uncertainty is not reducible without changing underlying system so that it possesses less randomness (i.e. more rigorous manufacturing tolerances or a system redesign to be less sensitive to material variability). Epistemic uncertainty can be reduced by taking more samples, investing in instrumentation that has less bias in the measurements, re-calibrating data reduction equations, all of which does not require a redesign in the underlying system being measured. Classifying the more significant sources of uncertainty in terms of aleatory or epistemic can be a step towards deciding what approach to use to reduce the uncertainty in the estimated performance of a design, or the uncertainty in a validation exercise.

7.2.4 Less utility if interested in total uncertainty

For case where the QoI is not a probability but a physical quantity (e.g. power, maximum temperature), the separation of uncertainty into epistemic and aleatory has less utility. Generally, one is interested in a best estimate of the QoI and the aggregated aleatory/epistemic uncertainty in this estimate.

7.3 Type A or Type B

The International Organization for Standardization on the Guide to the Expression of Uncertainty in Measurements (ISO GUM, 1995) and the ASME Code and Standard ASME PTC 19.1 (ASME, 2006a) provides guidance on the estimation of measurement uncertainty. These documents and the concepts behind them form the basis for the ASME Verification and Validation standard ASME V&V 20-2009 (ASME, 2009) and is the approach recommended here. While ISO GUM acknowledges epistemic and aleatory types of uncertainty, the guide focuses on the sources of information used to estimate uncertainty. Specifically, the guide defines Type A and Type B uncertainty based on how the uncertainty is estimated.

Type A corresponds to aleatory uncertainty characterized using repeated measurements. A characterization statistic of this uncertainty is the familiar standard deviation.

Often, one does not have repeated measurements, and other approaches are used to estimate uncertainty. ISO GUM refers to this type of uncertainty as Type B. For example, the estimated uncertainty in measurements may be provided by the manufacturing specifications on the measurement device, from an error analysis on a data reduction (e.g., calibration) equation, or from expert opinion. Type B uncertainty can be either aleatory, epistemic, or both, depending on whether the particular source of uncertainty is random or represents lack of knowledge. In either case, the ISO GUM assumes that the Type B uncertainty can be represented by a probability distribution, and as such, the uncertainty can be characterized by a standard deviation. Note that, in the case of epistemic uncertainty, the probability distribution is a subjective probability, representing degree of belief. Because this uncertainty is not directly estimated from samples from a population, ISO GUM generalizes 'standard deviation' to the phrase 'standard uncertainty'. Standard uncertainty and standard deviation have the same probabilistic meaning, but standard uncertainty is a generalization of standard deviation to cases when a standard deviation may or may not be evaluated directly from repeated measurements or samples. The ASME Code and Standard ASME 20-2009 utilizes this terminology in their document, and this terminology is adopted here.

7.4 Sources of Uncertainties Associated with Validation Exercises

Today's concept of model validation (ASME, 2009, Roach, 2009) emphasizes the need to account for uncertainty in comparing model predictions to experimental observations. This uncertainty contains contributions due to experimental as well as model prediction uncertainty. The uncertainties that are typically considered during a validation exercise are

- Solution verification uncertainty for the application – uncertainty due to lack of finite element or finite difference spatial and temporal grid convergence. Uncertainty due to lack of stochastic convergence of particle-based methods, or lack of convergence of iterative equation solvers are also considered under this category
- Computational model parameter uncertainty – uncertainty in the correct values of the model parameters for the experiment or application
- Experimental data uncertainty

While other forms of uncertainty exist (such as code verification uncertainty[3]), the present document focuses on the uncertainties listed above. The following chapter addresses model prediction uncertainty in more detail.

[3] Includes uncertainty characterization of detected code or algorithmic deficiencies for the intended application

8.0 MODEL PARAMETER UNCERTAINTIES: SCREENING ANALYSIS AND CHARACTERIZATION

- A model parameter screening analysis identifies the model parameters whose uncertainty has significant impact on model prediction uncertainty.
- The screening analysis allows one to down-select that model parameters for which the associated uncertainties should be well characterized
- The resulting uncertainties will be used as input uncertainties for the computational model to estimate the corresponding uncertainties in the model predictions

Model validation compares model predictions to experimental observations, and accounts for uncertainty in these comparison. This requires that the uncertainty in model predictions due to uncertainty in the model parameters be analyzed.

Model parameters (as opposed to algorithmic parameters) are parameters that are utilized in the computational model to represent a particular application. Examples include the parameters used in constitutive equations for the materials, and parameters that appear in boundary or loading conditions (e.g. internal energy sources, external loads) to represent the environment.

Computational models can utilize hundreds of parameters, many of which have some level of uncertainty. The uncertainty in many of these parameters may have little impact on a prediction for a particular application. The characterization of model parameter uncertainty generally requires significant effort, and the propagation of this uncertainty though a model can be computationally expensive. As a result, the identification of a subset of parameters for which the uncertainty has a significant impact on a computational prediction is an important step in performing an uncertainty analysis. This chapter overviews methodology to perform such a screening analysis, as well as briefly discusses methods to characterize the uncertainty of the resulting down-selected parameters.

8.1 Initial Screening

The initial screening to identify this subset of parameters is generally done through the expert judgment of the analyst or a group of analyst, code developers, and experimentalist. The purpose of this initial screening is to identify those parameters that the team has high confidence that the uncertainty does not significantly impact the model predictions. The starting point for this initial screening analysis is to consider those parameters associated with the phenomena identified as important (e.g. the parameters associated with radiant heat transfer) during the development of the PIRT. Those parameters that are identified as important are then analyzed through a more formal sensitivity analysis as discussed in the following sections.

8.2 Quantitative Screening Analysis

The next step is to perform a formalized screening analysis on the identified parameters (Saltelli et al. 2000). The approaches discussed here fall into two broad groups; gradient-based (first order sensitivity analysis) and sampling based (such as experimental design and LHS sampling). Both approaches require nominal values for the parameters and some measure of uncertainty (e.g. standard deviation, interval width) associated with the uncertain parameters. The first order gradient-based approaches can require fewer computational model evaluations, but cannot capture the effect of nonlinear dependence of the QoI on model parameters. Experimental design approaches can capture some of the nonlinear behavior but generally require more computational model evaluations.

8.2.1 Gradient-Based Screening Analysis

Gradient-based techniques utilize estimates of the gradients of the relevant QoI with respect to the uncertain parameters to characterize the impact of model parameter uncertainty on a prediction. Consider the following model

$$S = f(\mathbf{x}, \boldsymbol{\alpha}) \tag{8.1}$$

where $\mathbf{x}$ is the vector of independent variables and $\boldsymbol{\alpha}$ is a vector of the model parameters of interest. Generally, the evaluation of *f* requires the solution of PDEs (or ODEs). A change in *S* due to a change in the i^{th} parameter α_i to first order is

$$\Delta_{\alpha_i} S \approx \frac{\partial f(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_i}\, \Delta\alpha_i, \quad i = 1,2, \dots, n \tag{8.2}$$

where *n* is the number of parameters under consideration. The change in *S* due to the change in the vector $\boldsymbol{\alpha}$ can be approximated to first order by

$$\Delta S \approx \boldsymbol{\nabla}_{\boldsymbol{\alpha}} f(\mathbf{x}, \boldsymbol{\alpha})\, \boldsymbol{\Delta\alpha} \tag{8.3}$$

where is $\boldsymbol{\nabla}_{\boldsymbol{\alpha}}$ is the gradient with respect to the parameter vector $\boldsymbol{\alpha}$. Given an estimate of the gradient of the model *f* to the parameter or vector of parameters, and a characteristic representation of the uncertainty (standard deviation, range, etc.) of the parameters, one estimates the uncertainty in *S* to first order. For example, given uncertainty characterized by the standard deviation, the corresponding change in *y* is

$$\Delta_{\alpha_i} S \approx \frac{\partial f(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_i}\, \sigma_{\alpha_i}, \quad i = 1,2, \dots, n \tag{8.4}$$

The changes relative to each standard deviation can then be used to rank the importance of the parameter uncertainty to the prediction of *S*.

Techniques that can be used to estimate the gradients include the following.

- Gradient estimation using finite differences
    - Requires n+1 (forward or backward finite differences) or 2n+1 (central finite differences) model evaluations at n+1 or 2n+1 sets of the model parameters
    - Represents an approximate gradient over a neighborhood of the nominal values for the parameters, with the neighborhood size reflected by the size of the model parameter differences used to estimate the gradients
- Gradient estimation using sensitivity equations
    - These equations are developed by taking the derivatives of the dependent variables in their governing PDE's, and the boundary and initial conditions, with respect to each model parameter (Beck and Arnold, 1977, Saltelli, et. al., 2008).
    - Solve the resulting sensitivity equations (also PDE's) to evaluate the gradient of f relative to the relative model parameters
    - Useful when the code contains the ability to evaluate and solve the sensitivity equations
- Gradient estimation using sampling
    - Fit a hyperplane to the f that results from n+1 or more Monte Carlo samples (each sample corresponds to a function evaluation) over a neighborhood of $\boldsymbol{\alpha}$ of the size corresponding to the uncertainty in the parameters. For example, Latin Hypercube Sampling (LHS) can be used where the range of each parameter represents an interval of possible parameter values, or normally distributed LHS sampling can be used where the range of the parameter values are characterized by the means and standard deviations or covariance of the parameter uncertainty.
    - Represents an effective gradient over a neighborhood of the nominal values for the parameters
    - With a sufficient number of samples, can provide the most robust first order sensitivity analysis results for a neighborhood

8.2.2 Experimental Design Screening Analysis

Statistical Experimental Design (ED) (Wu and Hamada, 2000; Saltelli et. al., 2000; Cacuci, 2003) techniques have the advantage, relative to the gradient-based techniques summarized previously, that they can capture the impact of nonlinear interactions with and between the various model parameters. Experimental design techniques are constructed to require a minimum number of experimental data points or computational model evaluations to estimate a particular order of sensitivity. For example, the estimation of the full quadratic sensitivity of f relative to n model parameters requires a minimum of $(n+2)(n+1)/2$ model evaluations to estimate the $(n+2)(n+1)/2$ coefficients that appear in the n dimensional quadratic surface. For n=7, the number of model evaluations is 36. Experimental design methodology defines the series of design levels (i.e., the set of point values of the parameter vector) at which the experiment is performed or the computational model is evaluated. These levels for a 3 level design are denoted -1, 0, 1, with -1 denoted the minimum value of the parameter, 0 the mean value, and 1 the maximum value. For the case of parameter ranges characterized by the uncertainty distributions, one can use multiples of standard deviations to represent the minimum, mean, and maximum levels (e.g. -1, 0, 1 can correspond to the mean value of a parameter minus two standard deviations, the mean value, and the mean value plus two standard deviations, respectively). Consider a Box-Behnken (Box and Behnken, 1960) design for a model in 3 parameters. This design is one of the oldest and is still heavily used to estimate a quadratic response surface. The resulting design levels are shown in Table 8.1.

Table 8.1 Box-Behnken design levels for a 3 parameter design

Design Levels		
X_1	X_2	X_3
-1	-1	0
-1	1	0
1	-1	0
1	1	0
-1	0	-1
-1	0	1
1	0	-1
1	0	1
0	-1	-1
0	-1	1
0	1	-1
0	1	1
0	0	0
0	0	0
0	0	0

Note that the design requires the use of the model evaluation at (0, 0, 0) three times. This reuse provides the correct weighting to the center points relative to the other points to insure a better estimate of the quadratic response over the design space. Note also that there are no design levels at the corners of the design (e.g., 1, 1, 1). Other designs that include these points can better represent interaction effects at the corners of the design space. The resulting values for the function evaluations of *f* at the design points are then used to evaluate the corresponding quadratic response surface. This response surface is then used to investigate which terms in the quadratic have a significant impact on the quadratic approximation for *f* over the parameter space. One may find, for example, that a first order term in a particular parameter is not important, but that a term corresponding to a product of this parameter and another is important (i.e., results in significant changes in the QoI over the space). In this case, both parameters should be retained for the model when used in an uncertainty analysis, even though a first order sensitivity analysis indicates that the first parameter uncertainty is not important. Other experimental designs include space-filling designs optimized to fill high dimensional spaces with a given number of model evaluations. For example, Latin Hypercube designs are approximately space filling designs that are well accepted. Specific experimental designs are also available (Saltelli, 2008) to optimize the representation of a non-linear process by a Gaussian process model. Many statistical software packages provide the design points for the various types of designs, as well as provide post diagnostics to characterize and rank order the model's dependence on the parameters.

8.2.3 Sensitivity Analysis using Variance-Based Techniques

The primary advantage of variance-based techniques (Saltelli, et. al, 2000) over those discussed above is that they characterize the sensitivity of a computational model response to the model parameters, without requiring assumed forms of the model (e.g. linear, quadratic) over the sampled neighborhood. In contrast, the gradient-based technique discussed above captures the first order (i.e., linear) sensitivity to the parameters. Box-Behnken experimental design techniques capture quadratic behavior. The variance-based techniques, which utilize samples (i.e., LHS sampling), allow for the characterization of more complex non-linear behavior if the model response can be practically sampled a sufficient number of times. Examples of the application of variance-based sensitivity analysis are provided in McKay (1996), and Saltelli and Tarantola (2002).

8.3 Characterization of Parametric Uncertainty

Once the screening analysis is complete, the uncertainties of the down selected model parameters are statistically characterized for use in the Uncertainty Quantification (UQ) analysis. Uncertainty quantification is the propagation of model parameter uncertainty through a model to estimate the associated prediction uncertainty. This propagation can be approximately performed using results from the sensitivity analysis discussed previously, or can be done though sampling techniques. Both techniques are discussed in the following chapter.

Sampling techniques require the specification of the full probability distribution functions (PDF) characterizing parameter uncertainty, whereas the sensitivity analysis approach can be accomplished with knowledge of only the means and covariance matrix for the vector of down selected uncertain parameters. Sampling techniques can be more robust for models that are highly non-linear over the parameter range associated with the uncertainty in the parameters. For carefully controlled validation experiments, the uncertainty in these parameters is often sufficiently small that the linearized sensitivity analysis approach provides useful results. Generally, evaluating the uncertainty in the values for these parameters that are inferred from experimental data are the responsibility of the experimentalist or analyst who estimated these parameters from the data (e.g. material characterization experiments, inflow characterization measurements). The best approach to characterize parameter uncertainty depends on how much data is available, and whether prior knowledge is available that can be characterized through prior probability. A graded approach is suggested for the characterization of uncertainty as summarized below.

Physical data with no prior knowledge

- Base approach: Use standard statistical techniques to estimate parameters in parametric distributions (mean and standard deviation)
- Improvement: Use standard statistical techniques to estimate the parameters, the uncertainty in these parameters, and the goodness of fit of the distribution to the data.

Physical data with prior knowledge

- Use Bayesian estimation to estimate plausibility distributions for these parameters. Use carefully selected priors based on the prior knowledge of the parameters of interest.

No useful physical data

- Base case: Use expert opinion to characterize uncertainty.
- Improvement 1: Use improved expert opinion solicitation methodology (Mayer and Booker, 1991, O'Hagan et al., 2006) and associated uncertainty characterization.

8.4 Random Fields

In the previous discussion, the focus was on random variables rather than random fields. Typical random fields in engineering are those that are dependent on space. For example, a thermal property may vary continuously but randomly with position over a solid. Polynomial Chaos Expansions (PCE) (Wiener, 1938; Xiu, 2010) can be used to characterize such random fields and have the advantage of being more computationally efficient than Monte Carlo sampling techniques for many applications. This approach expands the physical space to a product space that includes stochastic dimensions associated with each term in a truncated PCE expansion. Orthogonal polynomials (the optimum choice of which depends on the type of distribution used for the random variables/fields) are used, with the coefficients as a function of physical space (or time). The PCE approaches have the advantage that they can represent such spatial/temporal variability with a minimal number of basis functions due to their convergence properties (exponential convergence with the appropriate orthogonal polynomials).

9.0 COMPUTATIONAL MODEL UNCERTAINTY

- The characterization of uncertainty in model predictions and measurements is an important step in model validation and is used to characterize the uncertainty in the model error.
- Methodology for estimating the effect of model parameter uncertainty and grid convergence uncertainty is well established and should be utilized.
- The most common approach to estimate the numerical uncertainty due to lack of grid convergence is the Grid Convergence Index approach based on Richardson extrapolation. This approach is discussed here.
- Two of the most common methods used to characterize model parameter uncertainty are a first order sensitivity analysis approach and a sampling approach. These approaches are discussed here.

9.1 Background

The focus of the present chapter is to summarize methodology to evaluate the effect of model parameter uncertainty and numerical uncertainty on model prediction uncertainty. The following nomenclature is used:

S	Output of a model simulation
u_{num}	Standard uncertainty in S due to solution verification (i.e., numerical) uncertainty
u_{input}	Standard uncertainty in S due to model parameter uncertainty
$u_{num+input}$	Standard uncertainty in S due to both model solution verification uncertainty and parameter uncertainty

Here, the uncertainties u_{num} and u_{input} represent standard uncertainty in the desired simulation output due to solution verification uncertainty (lack of spatial and temporal grid, equation solver, and stochastic convergence – also referred to as numerical uncertainty) and standard uncertainty due to model parameter uncertainty (i.e., input parameter uncertainty). Following ISO GUM (1995), PTC 19.2-2005 (ASME, 2006a) and V&V 20-2009 (ASME, 2009), standard uncertainty is the uncertainty in a single measurement or model prediction expressed as a standard deviation. For the case of repeated data, standard uncertainty for a single measurement can be estimated directly from the samples of data using the normal procedures for estimating the standard deviation. Standard uncertainty can also be estimated or defined from sources external to the measurements, such as from calibrated measurement standards, certified reference materials, data obtained from handbooks, or scientific judgment (ISO GUM, 1995).

9.2 Numerical Uncertainty (grid convergence uncertainty)

Solution verification addresses the uncertainties associated with lack of convergence for iterative equation solvers; lack of grid convergence for finite element, finite volume, finite difference grids (including temporal grids); and stochastic convergence in estimating probability density functions or their properties (e.g., moments). For many applications, the choice of the methodology to address these uncertainties represents an open area of research and requires judgment from the analyst.

The focus of the present discussion is on spatial grid convergence. As additional techniques evolve, the inclusion of other forms of solution verification uncertainty (i.e. geometric fidelity uncertainty, grid convergence when utilizing sub-domain or sub-grid models; convergence of models that are heavily influenced by aleatory processes, such as crack initiation in brittle materials, crush of heterogeneous materials, weld failure, etc.) can be considered.

There are several approaches to evaluating the uncertainty associated with lack of grid convergence. An extensive discussion of these approaches is provided in Part III of Oberkampf and Roy (2010). The methodology presented here (Grid Convergence Index) is the most widely used methodology and is based on the concepts underlying Richardson extrapolation (Richardson, 1911). More recently, Rider has developed methodology (Rider and Kamm, 2012, Rider 2013) that is based on the regression of error equations to data obtained from the computational solutions on multiple grids using multiple regression norms and regression functions. This new methodology has been applied to complex applications and promises to provide more robust estimation of solution error.

9.2.1 Grid Convergence Index[4]

The Grid Convergence Index (GCI) represents the most commonly used technique to characterize grid convergence error in the computational fluid dynamics field since the 1990s. The GCI method utilizes solutions for the QoI on multiple grids and provides an estimate of the amount of corresponding discretization error in the finest grid solution [Roache, 1994 and 1998]. Celik et. al. (2008) recommended the method for Journal of Fluids Engineering publications. The method is incorporated in the ASME standard V&V20 – 2009 (ASME, 2009). A least squares version of the method was first presented by Eça and Hoekstra (2002) (see also Pelletier and Roache 2006, Eça, Hoekstra and Roache 2005, Eça, Hoekstra, Hey and Pelletier 2007, Eça and Hoekstra 2009) and provides an improved estimate of grid convergence error based on the application under consideration.

The GCI method can be applied in a post-processing step utilizing solutions for the QoI on multiple grids and can be applied to any grid based partial differential equation solver (finite difference/element/volume). However, if interpolation is required to obtain solutions of the QoI at the same spatial location for different grids, one should use an

[4] Much of this section was taken directly and without quotes from Hills et. al. (2015).

interpolating method that is equal to or higher order than the underlying theoretical order of the elements in the grid.

The GCI method is based on Richardson extrapolation and utilizes solutions for the QoI on multiple grids, all of which should be in the asymptotic region of grid convergence. The characteristic lengths of the grid cell sizes should vary by at least 30% between one grid and the next finest grid. As a result, it is sometimes difficult obtain solutions on 3 or more grids for multi-physics applications. Also, formal order of convergence can degrade in local regions due to highly non-linear effects and due to algorithmic techniques, such as those used for shock capturing. Convergence rates less than the expected formal order can also indicate a software defect (bug), inconsistent geometry between refinements, non-systematic grid refinement, or insufficient convergence of the iterative equation solvers. Convergence rates larger than the expected formal order of the algorithm should be investigated as these generally indicate that the solution is outside the asymptotic region of convergence.

9.2.2 The Basic GCI method

ASME V&V 20-2009 (ASME, 2009) provides a step-by-step process to estimate numerical uncertainty u_{num} based on the GCI method. This process is summarized below.

Evaluate representative cell size. In three dimensions, this representative size can be estimated from

$$h = [\Delta x_{\max}\, \Delta y_{\max}\, \Delta z_{\max}]^{1/3} \tag{9.1}$$

for structured grids, or

$$h = \left[\frac{1}{N}\sum_{i=1}^{N} \Delta V_i\right]^{1/3} \tag{9.2}$$

for unstructured grids. Note that in the second case, h is based on the cube root of the average volume of the cells across the grid (N total cells, with ΔV_i equal to the volume of cell i).

Perform the computations of the QoI, S, on three grids with a refinement ratio $r = h_{course} / h_{fine}$ greater than 1.3. Denote the three grids representative cell sizes by h_1, h_2, h_3 with $h_1<h_2<h_3$, and $r_{21} = h_2/ h_1$, $r_{32} = h_3/ h_2$. The order of convergence, p, is estimated from

$$p = \frac{\left|ln\left|\frac{S_3-S_2}{S_3-S_2}\right|\right| + q(p)}{\ln r_{21}} \tag{9.3}$$

where

$$q(p) = ln\left(\frac{r_{21}^p - \gamma}{r_{32}^p - \gamma}\right) \tag{9.4}$$

$$\gamma = \mathrm{sign}\left(\frac{S_3-S_2}{S_3-S_2}\right) \tag{9.5}$$

The estimated order of convergence can now be used to estimate the extrapolated (corrected) value for the QoI, S.

$$S_{extrap} = \frac{r_{21}^{p} S_1 - S_2}{r_{21}^{p} - 1} \tag{9.6}$$

Note that 9.6 gives the extrapolated or estimated corrected value for S for a converged grid based on Richardson extrapolation. This correction is an approximation because one is ignoring the higher order terms in the Taylor's series expansion that lead to Richardson extrapolation. To address this issue, the GFI method incorporates a safety factor as follows.

$$\mathrm{GCI}_{fine}^{21} = F_s \left| \frac{S_1 - S_2}{S_\mathrm{n}} \right| \frac{1}{r_{21}^{p} - 1} \tag{9.7}$$

where the normalization quantity S_n typically is take to be $S_\mathrm{n} = S_1$. GCI is a characterization of normalized grid convergence error. The normalization value S_n can be replaced with values other than the local values, especially if S_1= 0 (see ASME, 2009). F_s is a safety factor who's value is based on empirical studies of over 500 CFD cases. A safety factor F_s = 1.25 provides a GCI_{fine}^{21} that bounds 95% of the actual normalized errors (when compared to solutions on fully converged grids). For unstructured grids the more conservative value F_s = 3 is recommended (ASME, 2009).

Here we are interested in the standard uncertainty, which does not correspond to the 95% confidence level. To obtain the standard uncertainty from (9.10), V&V 20-2009 (ASME, 2009) proposed that

$$u_{num} = \frac{|S_n|}{1.15} \mathrm{GCI}_{fine}^{21} \tag{9.8}$$

Note that u_{num} is non-normalized standard error. This represents a conservative estimate of the standard uncertainty due to uncertainty in the actual distribution of the normalized errors that lead to the GCI. Note that if these errors were Gaussian distributed, 1.15 would be replaced with 2. Given the estimate for the standard uncertainty due to lack of grid convergence, we can now move on to the model prediction uncertainty due to parameter uncertainty.

9.3 Model Parameter Uncertainty

Section 8.3 summarized the methods used to characterize model parameter uncertainty. In the following sections, the propagation of these uncertainties through the model to the predicted QoI, S, (including the validation QoI) to obtain the corresponding prediction uncertainty u_{input} is discussed. Two approaches will be summarized below. The first is based on the sensitivity analysis performed in the previous chapter. The second is based on sampling.

The sensitivity analysis approach discussed here is a linear analysis and is suitable when the model for the QoI is approximately linear over the uncertainty range of the model parameters. Because validation experiments are often designed to minimize uncertainty in these parameters, non-linear models can often be well approximated as linear over these limited ranges. A second issue associated with the first order sensitivity approach is that it can only provide estimates of the covariance matrix for a vector of QoIs and not the full PDFs for the QoIs.

In contrast, the sampling approach allows one to characterize the non-linear response of the QoI to the uncertain model parameters. Parameter value sets are sampled from the parameter PDFs, these value sets used to calculated the corresponding QoIs. As the number of sample sets increases, the ability to estimate the PDFs of the resulting QoIs increases. However, one must be able to use a very large number of samples to resolve output distributions for the QoI. The more important the tail of these distributions to the application, the more samples must be obtained. The use of model surrogates or response surfaces will be discussed as a technique that will allow one to reduce the number of required model evaluations in a later section.

9.3.1 Gradient-Based Analysis for u_{input}

Gradient-based analysis can be used to estimate a covariance matrix characterizing the uncertainty for multivariate simulation output given the covariance matrix of the uncertainty for the model input parameters. While the covariance alone cannot define a distribution, the covariance can be used to characterize the widths of distributions at various prediction points. Gradients can be estimated using finite differences using the code as a black box (see previous discussion on gradient based parameter screening analysis and Eqs. 8.1 – 8.4), or one can develop and solve systems of partial differential equations governing the sensitivities of the desired system response variables relative to the uncertain parameters, or one can develop and solve adjoint governing equations or adjoint matrix equations associated with the underlying finite difference or finite element method. The latter two methods are not black box in the sense that they are embedded in the code.

The two primary disadvantages of gradient-based analysis are

- The analysis represents the behavior of the simulation model about some nominal location (usually the mean) as linear in the uncertain parameters (i.e. a first order sensitivity analysis).
- The evaluation of the covariance of the simulation model outputs assumes that such a covariance exists. There are heavy-tailed distributions, often associated with rare events, for which these second moments of the distributions do not exist.
- The analysis provides only a mean and covariance of the simulation output. However, if the uncertainty in the simulation output is normally distributed, the mean and covariance fully characterizes the uncertainty distribution in the model output due to model parameter uncertainty.

One can expect approximately normally distributed computational simulation results if

- The model is approximately linear in those uncertainty parameters that dominate the uncertainty over the range of the uncertainty, and

- The uncertainties in these dominant model input parameters are normally distributed.

Even if the above conditions are not true, the covariance is often fairly well approximated by the sensitivity analysis. In contrast, the tails of the output distributions for the QoIs are much more sensitive to differences in the model parameter PDF's and to the nonlinearity of the simulation model. As a result, the applicability of the sensitivity method for estimating the PDF tail behavior is limited to linear models with normally distributed model input parameters.

Once the gradients with respect to the important uncertainty parameters are estimated, the change in a vector of model simulation outputs due to a change in a vector of uncertain input parameters can be approximated to first order as (see Eq. (8.3)).

$$\Delta \boldsymbol{S} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \mathbf{f}(\mathbf{x}, \boldsymbol{\alpha})\ \boldsymbol{\Delta\alpha} \tag{9.9}$$

The model arguments $(\mathbf{x}, \boldsymbol{\alpha})$ represent those arguments whose values are know with certainty (**x –** typically are independent variables), and those known with uncertainty ($\boldsymbol{\alpha}$). Note that if an independent variable has uncertainty (for example the time of a physical measurement), then this variable should be included in $\boldsymbol{\alpha}$ rather than in **x**. **S** represents a vector of model predictions. The i^{th} row of $\boldsymbol{\nabla}_{\boldsymbol{\alpha}} \mathbf{f}(\mathbf{x}, \boldsymbol{\alpha})$ corresponds to the sensitivity of S_i to the vector of parameters $\boldsymbol{\alpha}$. Post multiplying by the transpose of $\Delta \boldsymbol{S}$ gives

$$\Delta \mathbf{S} \Delta \mathbf{S}^{\mathrm{T}} = \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \mathbf{f}\ \boldsymbol{\Delta\alpha\Delta\alpha}^{\mathrm{T}} \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \mathbf{f}^{\mathrm{T}} \tag{9.10}$$

The expected value of Eq. (9.10) is the covariance of **S** and gives the desired result for uncertainty.

$$\operatorname{cov}(\mathbf{S}) = \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \mathbf{f}\ \operatorname{cov}(\boldsymbol{\alpha}) \boldsymbol{\nabla}_{\boldsymbol{\alpha}} \mathbf{f}^{\mathrm{T}} \tag{9.11}$$

The diagonal elements in the cov(**S**) contain the squares of the first order approximations to u_{input} for the simulation output due to the parameter uncertainty.

$$u_{input_i} = \sqrt{[\operatorname{cov}(\mathbf{S})]_{i,i}} \tag{9.12}$$

9.3.2 Sample-Based Analysis for u_{input}

When the model input parameters are not normally distributed, or when the model response is nonlinear about the nominal values for the parameters over the range of the uncertainty in the parameters, gradient based methods will not provide accurate estimates of the prediction uncertainty in the tails of the PDF. Sampling based methods represent a 'gold standard' for application to problems possessing one or both of the above features. Unfortunately, sampling methods require many more function evaluations. While simple Monte Carlo Sampling can be used, stratified sampling or importance sampling, such as Latin Hypercube Sampling (LHS, Fishman, 1996), is often used to reduce the number of model evaluations. LHS has the advantage that it can provide estimates of low order statistics (mean, standard deviation) with fewer samples, but provides less advantage when used to estimate tail statistics (i.e. the location of 0.999 cumulative probability).

Most statistical packages provide LHS samples for the cases for which all of the parameters are either uniformly distributed or normally distributed.

For the case of combined aleatory/epistemic uncertainties, the parameter uncertainties can be propagated through the model as follows.

- Sample a realization of the parameters from the PDFs that have been chosen to represent their uncertainty.
- Evaluate the simulation output, *S*, at the corresponding parameter values.
- Repeat the previous steps until a sufficient number of samples have been obtained to characterize the PDF for the uncertainty in S.

The covariance matrix for *S* can be estimated from the sampled simulation results as follows:

$$\mathrm{cov}(\mathbf{S}) = \mathrm{cov}(S_i, S_k) = \frac{1}{n_r - 1} \sum_j^{n_r} (S_{i,j} - S_{i,\mathrm{mean}}) \left(S_{k,j} - S_{k,\mathrm{mean}}\right); \;\; i, k = 1 \ldots m \qquad (9.13)$$

where m is the number of elements or predicted measurements in the vector **S** and n_r is the number of sample sets used for the evaluation. The resulting u_{input} are given by the square-root of the diagonal elements of cov(**S**).

$$u_{input_i} = \sqrt{[\mathrm{cov}(\mathbf{S})]_{i,i}} \qquad (9.14)$$

9.3.3 Response Surfaces

The number of samples required to resolve characteristics of PDFs depends on the statistical quantities of interest for the prediction uncertainty. For example, means can be estimated with fewer samples, variances/covariance requires more samples, and the estimation of the cumulative distributions in the tails of distributions requires a very large number of samples. Because model simulations of are often very computationally intensive, the number of simulations that can be performed is limited. As a result, response surfaces are often used as surrogates for the computational models for sampling based uncertainty quantification (UQ). Response surfaces, when used for UQ, are fits of simplified model surrogates to the sampled computational model outputs as a function of the uncertain model parameters. The ability to estimate an accurate response surface depends on the number of points sampled and the choice of the response surface. Ideally, an analyst decides the number of computational simulations achievable, choose a sampling scheme such as LHS or through a design of experiment, executes a computational simulation at each of these sampled points, and estimates the coefficients in the mathematical expressions for the response surface given the computational results. The response surface is then used as a surrogate for the model over this space for UQ sampling. The goal of using a response surface is that, by replacing the computational simulation model with a response surface, one can generate many more samples from the surrogate to more accurately quantify the uncertainties associated with u_{input}. Of course, the accuracy of the constructed response surface in representing the behavior of the original computational model can be an issue.

The most appropriate response surface techniques are those that not only estimate a response surface, but also characterize the uncertainty associated with the residuals between the computational model values and the response surface values. These approaches allow this uncertainty to be aggregated into the evaluation of u_{input}. Response surfaces based on Gaussian Process (GP) models are examples of such approaches (see Higdon et. al, 2008, Williams, 2002).

Because an analyst is generally interested in using the response surface to sample only over the range of uncertainty of the model parameters for a specific application, or over the uncertainty range of the model parameters for the computational model of a validation experiment, the ability to represent model behavior over this range can be better than the ability to represent model behavior over the total parameter space. In addition, one is often interested modeling measurements at specific measurement locations and times. One can develop separate response surfaces for each location and time, which reduces the complexity and increases the accuracy of the resulting surrogates.

Various methods for response surface estimation have been developed; including adaptive fits using locally and globally supported functions (Multivariate Adaptive Regression Splines (MARS), Friedman, 1991, DAKOTA, Adams et. al., 2010).

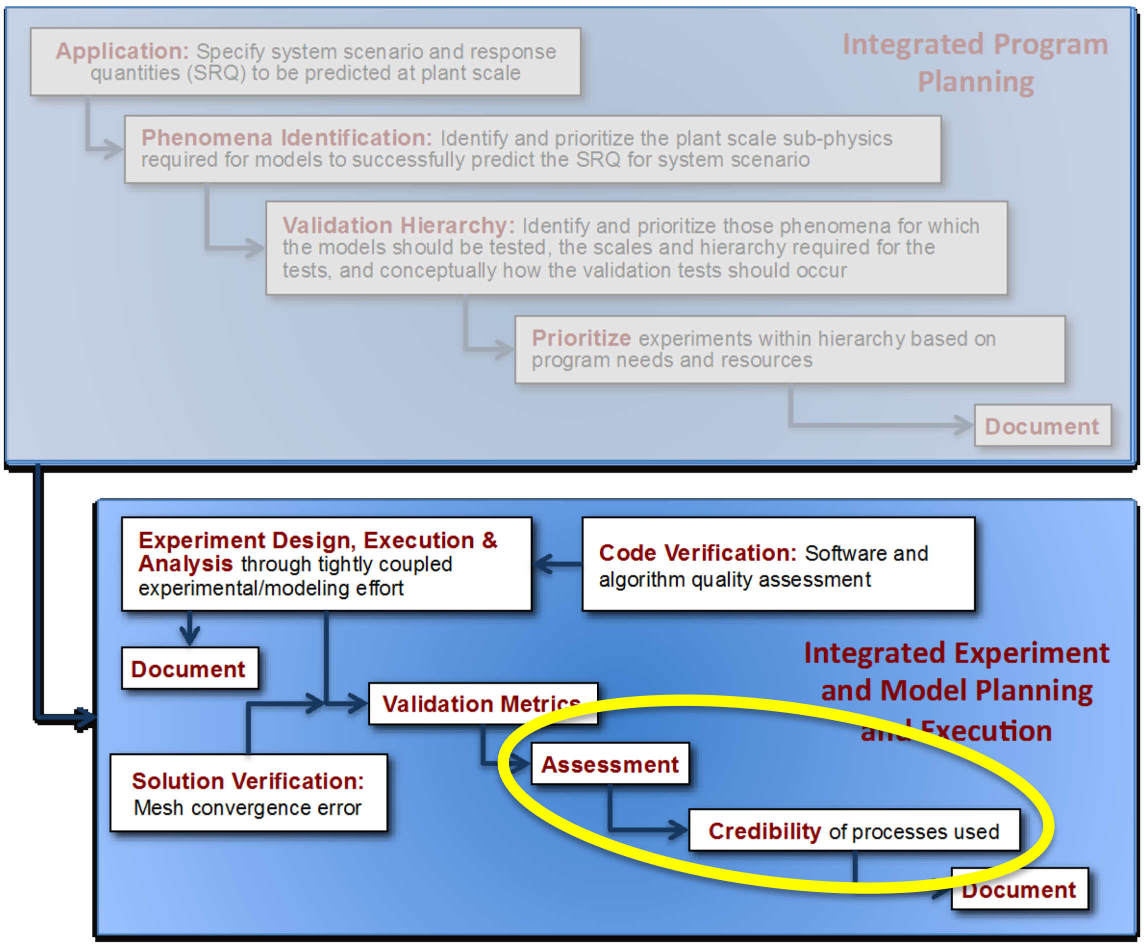
9.4 Combined Uncertainty $u_{num+input}$

The methodology to combine u_{num} from Section 9.2 and the u_{input} from Section 9.3 to obtain a total uncertainty is presented in V&V 20-2009 (ASME, 2009). Note that u_{input} may contain both aleatory and epistemic uncertainty as some of the model parameters may represent random processes and some of the model parameters may possess single values whose values are uncertain due to lack of knowledge. u_{num} is typically epistemic as there is a fixed grid convergence error associated with a predicted QoI, *S*. However, u_{num} may also contain aleatory uncertainty if the grid convergence error is dependent on aleatory model input parameters. Assuming independence between the source of errors in affecting grid convergence and the model parameter uncertainty, and treating epistemic uncertainty as probabilistic (i.e. as is treated V&V 20-2009 (ASME 2009) and the ISO GUM (1995), the total standard uncertainty in the model prediction is

$$u_{input+num} = \sqrt{u_{input}^2 + u_{num}^2} \qquad (9.15)$$

Given a model prediction for the QoI, *S*, the estimated uncertainty for this prediction given by Eq. (9.15), and a corresponding measurement *D* and its uncertainty u_D; one can evaluate a validation metric. This is the topic of the next chapter.

PART 4: MODEL VALIDATION ASSESSMENT



10.0 COMPUTATIONAL MODEL VALIDATION

- Evaluate the measure of agreement between the experimental results and the associated model prediction of the experimental results
- Evaluate the uncertainty in the resulting value for this measure due to measurement, model parameter, and numerical uncertainty.
- Interpret and communicate the results

AIAA (1998) and ASME (2006b) define model validation as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended use of the model." Model validation is based on the direct comparison between experimental data to computational predictions of the data, relative to the uncertainty in the validation exercise (ASME, 2006b). The uncertainty in the validation exercise is based on the uncertainty in the experimental data and the uncertainty in the model predictions.

The focus of the present chapter is on the comparison of experimental data to computational simulation considering data uncertainty and the prediction uncertainties discussed in Chapter 9.

10.1 Estimating Model Error from Validation Data

The ISO GUM (ISO, 1995) and the ASME standard PTC 19.1 (ASME 2006a) define methodology to characterize experimental uncertainty, including the evaluation of uncertainty due to data reduction models. The ISO GUM makes the distinction between standard deviation and standard uncertainty. Standard uncertainty is a generalization of standard deviation to cases where an estimate of its value may be from repeated measurements (i.e. standard deviation); or from sources external to the measurements, such as calibrated measurement standards, certified reference materials, data obtained from handbooks, or scientific judgment (ISO GUM, 1995). This standard considers the case for which epistemic uncertainty can be characterized by a probability distribution. As a result, this standard evaluates the standard uncertainty for the sum of two independent uncertain variables (epistemic and/or aleatory) as equal to the square root of the sums of squares of the standard uncertainties of each variable.

V&V 20-2009 (ASME, 2009) extends PTC 19.1 to computational model validation and applies the same concepts to the uncertainty in simulation output, due to simulation model parameter uncertainty as discussed in Chapter 8, u_{input}; numerical grid convergence uncertainty as discussed in Chapter 8, u_{num}; and data uncertainty, u_D. Data uncertainty corresponds to experimental data acquisition uncertainty associated with diagnostic error. If the errors associated with these uncertainties are independent, then the standard uncertainty in a validation exercise is[5]

[5] V&V 20-2009 (ASME, 2009) present's sensitivity based and sampling based methodology to address correlation between sources of error in the experimental data and the model parameters or input conditions.

$$u_{val} = \sqrt{u_D^2 + u_{input}^2 + u_{num}^2} \quad (10.1)$$

ASME (2009) does not specify rigorous criteria to declare a computational model as valid but rather estimates computational model error, δ_{model}, by the observed difference, *E*, between computational simulation output and physical data. The uncertainty in the estimate of δ_{model} is characterized through u_{val}. For example, see Figure 10.1 for a characterization of the uncertainty in δ_{model} through the interval

$$[E - u_{val}, E + u_{val}] \quad (10.2)$$

where

$$E = S - D \quad (10.3)$$

The analyst/customer/decision maker team must use judgment as to whether the estimated computational model error, *E*, and the estimate for uncertainty in this model error, u_{val}, are significant relative to the intended application of the model.

V&V 20-2009 (ASME, 2009) also presents a sampling approach for evaluating the model prediction bias and the uncertainty in the bias. This process is summarized below:

- Develop PDF's representing the data uncertainty, the grid convergence uncertainty (one typically assumes a normal distribution with zero mean and standard deviation equal to u_{input} using the GCI method discussed previously), and the uncertainty in *S* due to parameter uncertainty.
- Sample a data value, *D*, from the data distribution; a grid convergence error, *G*, from the grid convergence distribution; and a parameter set from the parameter distributions.
- Evaluate *S* using the sampled parameter set, and evaluate the resulting sampled error *E* as follows:
  $$E = S + G - D \quad (10.4)$$
- Repeat the sampling and the evaluation of Eq. (10.4) multiple times to obtain a population of possible *E*.
- Utilize this population of E to estimate the mean model error (i.e. mean *E*) and other statistical characteristics of the population, such as standard uncertainty (i.e. calculated as a standard deviation), quartiles, etc.

10.2 Interpretation

The estimate of the bias error, *E*, and the standard uncertainty in this error, u_{val}, provides a basis for interpreting model validity. Note that several interpretations can be taken depending on model needs.

10.2.1 Scientific Validation

In science, one is often interested in asking whether the model predictions are consistent with the experimental observations, given the uncertainty present in the validation exercise. Or in statistical terms, we can ask what the statistical significance is of a difference. If one has a sufficient number of samples of the differences, then one can evaluate what percentages of the sampled differences are larger in magnitude than the observed difference. This percentage is the significance of the results. For example, if the sampled differences are normally distributed, and the observed difference is $E = 2^* u_{val}$, then the two-sided significance of this difference is 5%. The interpretation of this significance is that one would expect that 5% of the differences would be larger in magnitude for this level of uncertainty, given that the model is valid. Note that this statement does not say that this model is valid, only that if it is, then only 5% of the differences would be larger.

An issue that is often raised with this approach is that the more uncertainty there is in the validation exercise, the more likely that an observed difference is within some uncertainty range of the validation exercise. Once should keep in mind that this approach evaluates whether the evidence is sufficient to consider the model invalid, rather than an evaluation of whether the model is valid. It may simply be that the uncertainty in the validation exercise is such that one cannot resolve model validity relative to the accuracy required.

10.2.2 Engineering Validation

Another approach is to simply state the range model prediction errors, based on the validation exercise. This range may be based on $\pm u_{val}$ around E, or $\pm 2\ u_{val}$ around E. If one can approximate the distribution for the possible E given samples of these differences, then one could also utilize a probability level to define this range.

The advantage of this approach is that the modeler and the customer have an assessment of the range of model error, and can decide after the validation exercise whether that range is acceptable for the application. In engineering, most models are approximate but still useful, especially if one designs sufficient safety factor into the system, so that the model errors are small compared to the safety factor. We strongly recommend the engineering approach to quantify possible ranges of model error.

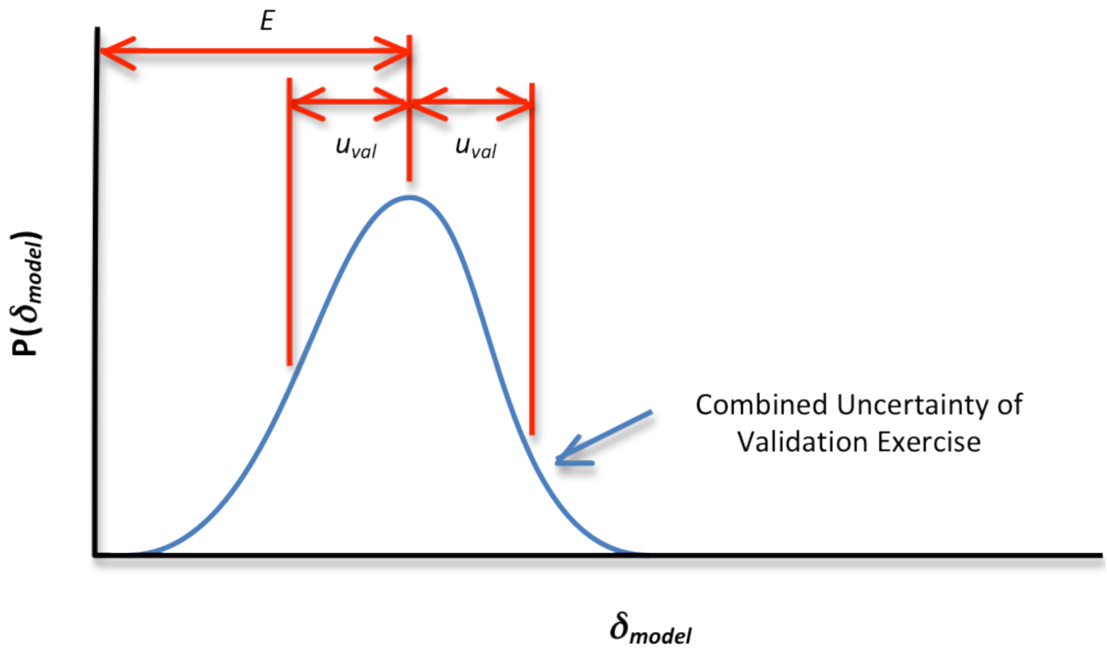


Figure 10.1 Estimate of model error δ_{model}

11.0 PREDICTIVE CAPABILITY MATURITY MODEL

- Computational models are often used for application conditions different from the validation conditions.
- When the application conditions are different, the usability of the model for these conditions requires subject matter expert judgment.
- The Predictive Capability Maturity Model, along with the PIRT, provides a formalized mechanism to elicit and document this judgment.

The previous three chapters overviewed methodology to estimate model error, and uncertainty in the model error, given observed differences between experimental measurements and model predictions, measurement uncertainty, and estimated model prediction uncertainty. These results are critical components in the evidence that the model is suitable for the intended application. Other evidence is more subjective and requires engineering judgment. The focus of the PIRT is to identify, rank, and perform a gap analysis on the physics that is important for an application and is a key element in a model credibility evidence package. While the PIRT does address code quality, verification, and validation at a high level, additional methodology has been developed to provide a more thorough evaluation of computational issues associated with the model. A subject matter expert elicitation tool for this additional evidence is the Predictive Capability Maturity Model (PCMM) (Oberkampf et. al., 2007).

11.1 What is a PCMM?

The development of a PIRT as discussed in Part I is a precondition of the development of the PCMM as the physics, validation hierarchy, and material models addressed in the PCMM are those identified in the PIRT. The PCMM is a deeper dive into computational issues associated with the modeling of this physics. More specifically, the PCMM is an elicitation tool that formalizes the methodology to assess the maturity and completeness of the evidence that a computational model is appropriate for the intended application. The PCMM is based on six primary elements or practices.

- Representative and Geometric Fidelity
- Physics and Material Model Fidelity
- Code Verification
- Solution Verification
- Model Validation
- Uncertainty Quantification and Sensitivity Analysis

Each of these elements is assigned a maturity or completeness level, generally from 0 to 3 based on subjective criteria. The desired level of maturity/completeness depends on the application. A computational simulation that is used for scoping studies will generally require less maturity than one that is relied on heavily for the qualification of a safety critical system. Table 11.2 (Oberkampf et. al., 2007) describes criteria that can be used to assign these levels, by element. Table 11.1 defines the acronyms used in the table. It is acceptable and often necessary to assign fractional levels. The actual levels assigned to each of these six elements are largely subjective and is typically assigned using a team of subject matter experts.

Brief descriptions of the six elements are provided below. More complete discussions can also be found in Oberkampf et. al. (2007).

Representation and geometric fidelity involves the level of detail used to characterize the geometric aspects and features of the system being analyzed. In most cases and disciplines, it is either impossible, unnecessary, or both to model every minute aspect of the geometry.

Physics and material model fidelity involves the (1) degree to which models are physics-based, (2) degree to which the models are calibrated, (3) degree to which the models are being extrapolated from their validation and calibration domains to the conditions of the application of interest, and (4) quality and degree of coupling of multi-physics effects that exist in the application of interest.

Code verification involves the (1) correctness and fidelity of the numerical algorithms used in the code relative to the mathematical model (e.g. the partial differential equations and the constitutive models), (2) correctness of the source code, and (3) configuration management, control, and testing of software through software quality engineering (SQE) practices.

Solution verification involves the (1) assessment of numerical solution errors, such as grid convergence errors, in the computed results and (2) assessment of confidence in the computational results as the results may be affected by human errors (i.e. correct input files).

Model Validation involves the (1) thoroughness and rigor of the accuracy assessment of the computational results relative to the experimental measurements, (2) completeness and rigor of the characterization of the experimental conditions and measurements, and (3) relevancy of the experimental conditions, physical hardware, and measurements in the validation experiments and the validation hierarchy compared to the application of interest.

Uncertainty quantification and sensitivity analysis involves the (1) thoroughness and soundness of the uncertainty quantification effort, including the identification and characterization of all plausible sources of uncertainty, (2) accuracy and correctness of propagating uncertainties through a computational model and interpreting uncertainties in the system response quantities of interest, and (3) thoroughness and precision of a sensitivity analysis to determine the most important contributors to uncertainty in system responses.

The version of the PCMM developed by Oberkampf et. al. (2007) and presented in Table 11.2 is the first generation PCMM. Second, third, and forth generations PCMMs have been developed, with each new generation adding more detailed specifications and descriptions of the issues that should be addressed with each of the 6 elements. In the case of the fourth generation, additional elements have been added, such as customer specifications, quality of experimental data, and relevance of the validation hierarchy to the issues identified in the PIRT. See Hills et al. (2013) for a summary of the first three generations as well as the process used to develop the fourth generation PCMM. The fourth generation PCMM should not be considered superior to any of the other generations, and the choice of which PCMM to use should be left to the project team. The most important aspect of the application of any of the PCMMs is that it be developed in a team environment (i.e. a single subject matter expert should not complete a PCMM) and that the 6 elements listed above be addressed in sufficient detail for the application at hand.

11.2 Why a PCMM?

A completed PCMM can serve several roles. First, it can be used as a planning tool, where the planning is more computationally specific than the higher level planning associated with the extended PIRT. For example, one may decide that solution verification procedures are not sufficiently complete (or rigorous) for the desired application of the model. While this issue may have been identified in the PIRT, the PCMM characterizes these issues in more detail.

The second role is that a completed PCMM can serve as transparent communication and documentation tool. The table structure of the PCMM allows quick communication of issues associated with the credibility (i.e. maturity) of the computational model based on the evidence supporting its credibility. The complete validation evidence package should include, at the minimum, a PIRT, the documentation of the validation exercises (i.e., experiments, models, data) and results over the validation hierarchy, and the PCMM.

11.3 Recommended Elicitation Process for the PCMM

In contrast to the development of a PIRT, we have found that the approaches used to develop a PCMM tends to be more varied (even within an organization or a department) leading to uneven assessments. To address this issue, a formalized PCMM expert elicitation process was developed as part of the fourth generation PCMM (Hills, et. al., 2013). This process is applicable to any of the generational PCMMs and is summarized below:

11.3.1 Overall Goal of the PCMM

To increase communication within and outside the product delivery team as to the maturity of a CompSim to support actionable decisions associated with the customer's needs.

11.3.2 Participants

The team should include at least one customer of the model results (e.g., a designer), one or more analysts (i.e. computational specialist who applies the code to applications), one or more experimentalists, one or more code developers, and a V&V/PCMM specialist. The number of analyst, experimentalist, and developers should be sufficient to provide subject matter expertise to address the major features of the computational simulation that are relevant to the customer's application. Roles of the team members include

PCMM team lead: Responsible for selecting team members, for facilitating the overall process and the meetings, communicating impact, and delivering final product.

V&V/PCMM SME: Act as a resource on the use of and interpretation of the items in the PCMM.

Other team members: Responsible for providing individual scores and participation in the deliberations for the group evaluations scores.

11.3.3 Process

The following SME elicitation process was designed to help insure that both individual opinions and group consensus are characterized by the resulting PCMM document.

The availability of a PIRT for the application is a pre-condition for PCMM evaluations. The PIRT addresses the physical phenomena, which are relevant to a PCMM evaluation.

Step 1: The team meets to discuss the PCMM, the elicitation process, and the expectations and use of the resulting PCMM product. Copies of the PIRT should be provided to the team at this or prior to this meeting.

Step 2: After the meeting, the team members individually develop an initial evaluation of those features in the PCMM for which they feel comfortable addressing.

Step 3: The team meets to discuss these individual assessments, to share knowledge that affects these assessments, to reach a team consensus, and to document the consensus PCMM maturity scores.

Step 4: After the meeting, the individual team members reflect on the deliberations and update their own scores if appropriate. Note that individual scores do not have to reflect the team scores. These individual scores are used to document diversity of opinion after the deliberation process is completed.

Step 5: A final meeting of the team is held to discuss the actual or potential impact of the evaluation. The team lead is responsible for providing a summary of the impact. Impacts can be as specific as planned or recommended programmatic adjustments, or softer impacts such increased understanding of the ability (or lack thereof) of the computational simulation to provide the customers with actionable results.

11.3.4 Product

The result of this process is a completed PCMM table by individuals and by team consensus.

11.4 PCMM for Program Planning

The PIRT is used for program planning, whereas the PCMM is generally used as a reporting tool. However, the PCMM can be used as a computational planning tool if utilized early in the model development phases. The advantage of the PCMM relative to the PIRT is the PCMM calls out specific issues, such as geometric fidelity, mesh convergence, or uncertainty quantification, that might have been overlooked when developing the PIRT, allowing one to address these issues before the final evaluation of the PCMM.

Table 11.1 Acronyms used in Table 11.2

M&S	Modeling and Simulation – equivalent to computational simulation
IET	Integrated Effects Test – validation experimental tests that contain multiple (integrated) physics.
SET	Separate Effects Tests – validation experimental tests that are design to isolate single physics types (such as heat conduction, or stress-strain curves from coupon tests)
SQE	Software Quality Engineering
SRQ	System Response Quantity – here we use RQI (Response Quantity of Interest) as the validation quantity of interest for component or subsystems may be different from those of interest at the system level.

Table 11.2 Predictive capability maturity model

MATURITY / ELEMENT	**Maturity Level 0** **Low Consequence, Minimal M&S Impact, e.g. Scoping Studies**	**Maturity Level 1** **Moderate Consequence, Some M&S Impact, e.g. Design Support**	**Maturity Level 2** **High-Consequence, High M&S Impact, e.g. Qualification Support**	**Maturity Level 3** **High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification**
Representation and Geometric Fidelity **What features are neglected because of simplifications or stylizations?**	• Judgment only • Little or no representational or geometric fidelity for the system and BCs	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components is defined	• Limited simplification or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components • Some peer review conducted	• Essentially no simplification or stylization of components in the system and BCs • Geometry or representation of all components is at the detail of “as built”, e.g., gaps, material interfaces, fasteners • Independent peer review conducted
Physics and Material Model Fidelity **How fundamental are the physics and material models and what is the level of model calibration?**	• Judgment only • Model forms are either unknown or fully empirical • Few, if any, physics-informed models • No coupling of models	• Some models are physics based and are calibrated using data from related systems • Minimal or ad hoc coupling of models	• Physics-based models for all important processes • Significant calibration needed using separate effects tests (SETs) and integral effects tests (IETs) • One-way coupling of models • Some peer review conducted	• All models are physics based • Minimal need for calibration using SETs and IETs • Sound physical basis for extrapolation and coupling of models • Full, two-way coupling of models • Independent peer review conducted
Code Verification **Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results?**	• Judgment only • Minimal testing of any software elements • Little or no SQE procedures specified or followed	• Code is managed by SQE procedures • Unit and regression testing conducted • Some comparisons made with benchmarks	• Some algorithms are tested to determine the observed order of numerical convergence • Some features & capabilities (F&C) are tested with benchmark solutions • Some peer review conducted	• All important algorithms are tested to determine the observed order of numerical convergence • All important F&Cs are tested with rigorous benchmark solutions • Independent peer review conducted
Solution Verification **Are numerical solution errors and human procedural errors corrupting the simulation results?**	• Judgment only • Numerical errors have an unknown or large effect on simulation results	• Numerical effects on relevant SRQs are qualitatively estimated • Input/output (I/O) verified only by the analysts	• Numerical effects are quantitatively estimated to be small on some SRQs • I/O independently verified • Some peer review conducted	• Numerical effects are determined to be small on all important SRQs • Important simulations are independently reproduced • Independent peer review conducted
Model Validation **How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?**	• Judgment only • Few, if any, comparisons with measurements from similar systems or applications	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown experimental uncertainties	• Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs • Some peer review conducted	• Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs • Independent peer review conducted
Uncertainty Quantification and Sensitivity Analysis **How thoroughly are uncertainties and sensitivities characterized and propagated?**	• Judgment only • Only deterministic analyses are conducted • Uncertainties and sensitivities are not addressed	• Aleatory and epistemic (A&E) uncertainties propagated, but without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• A&E uncertainties segregated, propagated and identified in SRQs • Quantitative sensitivity analyses conducted for most parameters • Numerical propagation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• A&E uncertainties comprehensively treated and properly interpreted • Comprehensive sensitivity analyses conducted for parameters and models • Numerical propagation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted

12.0 REPORT SUMMARY

This document summarizes recommended best practices associated with a model validation directed experimental/modeling program. These practices utilize tools that have been developed for the modeling of complex engineered systems, such as hydrodynamic modeling for nuclear power plants, safety analysis of nuclear weapons, and aerospace design (commercial and NASA); as well as guides, and codes and standards that have been developed by various international organizations. The recommended practices consider two aspects of a validation directed experimental/model program; 1) program planning and 2) model validation experiments.

Part 1 of this document focused on the utilization of a Phenomena Identification Ranking Table (PIRT) for program planning. The PIRT was originally developed for the nuclear power plant industry, and is presently widely used across many industries when computational multi-physics modeling of engineered systems is central. The development of a PIRT by a team of subject matter experts provides a structured, transparent, and collaborative approach to plan a joint computational/experimental program. The team identifies the important phenomena that should be captured by the model for an intended application; ranks the phenomena as high, medium, or low importance; and assesses current ability to use computational modeling to represent the phenomena. The results are then used to perform a gap analysis, identifying the phenomena for which the importance is high or medium, and the ability to represent the phenomena by the model is thought to be low or medium. This gap analysis prioritizes the phenomena that should be addressed by a model development/validation program.

Part 2 focused on the design of validation quality experiments to address the issues and experiments identified by the PIRT in Part 1, as well as other issues associated with validation and model credibility. The design and execution of validation quality experiments requires tight integration between the experimentalist and the modelers to insure that the experimental results can be unambiguously modeled. The safest way to insure this is to model the experiment during the design phase. This not only insures that the quantities (initial and boundary conditions, material properties, configuration, etc.) required to model the experiment have been identified, but also allows the model to be used to optimize the experimental design from a validation perspective.

Uncertainty plays a key role in validation, and the quantification of uncertainty should receive significant attention. Part 3 address formalized methodology to characterized uncertainty in experimental measurements, in model predictions, and in validation assessments of model prediction error has been developed by various international organizations and documented in guides or codes and standards. These approaches are summarized here, and should be used if possible.

Part 4 addresses model assessment both through model validation and through expert elicitation. Model validation quantifies agreement between model prediction and experimental observation for the conditions of the experiments. However, models are often used to predict system response for conditions other than those of the validation exercise. As a result, judgment must be used as to the relevance of the model verification and validation evidence bases for the application. To formalize and communicate the completeness of this evidence, the Predictive Capability Maturity Model (PCMM) was developed for computational simulation for the nuclear weapons industry. The PCMM is currently being modified and adapted by other industries as the PCMM serves as a comprehensive expert elicitation tool, which asks questions that are relevant to the use of a computational model for high consequence applications. This tool summarizes computational model maturity/completeness based on 6 main elements; representational and geometric fidelity, physics and material fidelity, code verification, solution verification, model validation, and uncertainty quantification and sensitivity analysis.

Overall, the decision to use a computational model to support the design and qualification of a complex engineered system requires the integration of technical data (experimental and computational), significant engineering and programmatic expert judgment, and compromises based on technical and resource limitations. The processes discussed here provide formalism to the design and execution of a computation model development/validation program that is used to develop the evidence basis that a computational model is suitable for the intended application.

13.0 REFERENCES

Adams, B.M., W.J. Bohnhoff, K.R. Dalbey, J.P. Eddy, M.S. Eldred, D.M. Gay, K. Haskell, P.D. Hough, and L.P. Swiler (2010), DAKOTA: A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis. Version 5.0 Users Manual," Sandia Technical Report SAND2010-2183, December 2009; updated "Version 5.1 Users Manual," December.

AIAA (1998), Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, AIAA-G-077-1998, American Institute of Aeronautics and Astronautics, Reston, VA, p. 3.

ASME (2006a), ASME PTC 19.2-2005, Test Uncertainty, The American Society of Mechanical Engineers, New York, NY.

ASME (2006b), Guide for Verification and Validation in Computational Solid Mechanics, ASME V&V 10-2006, The American Society of Mechanical Engineers, New York, NY.

ASME (2009), Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer, ASME V&V 20-2009, The American Society of Mechanical Engineers, New York, NY.

Babuška, I., K. Dowding, and T. Paez (2008), Special Issue: Validation Challenge Workshop, Computer Methods in Applied Mechanics and Engineering, V 197, Issues 29-32, pp. 2373-2665.

Beck, J.M, and K.J. Arnold (1977), Parameter Estimation in Engineering and Science, John Wiley & Sons, New York, p. 411.

Box G. E. P., D. W. Behnken (1960), Some New 3 Level Designs for the Study of Quantitative Variables, Technometrics 2, pp. 455-475.

Cacuci, D. G., (2003), Sensitivity and Uncertainty Analysis: Theory, Chapman &Hall/CRC, New York, NY.

Celik, I.B., U. Ghia, P.J. Roache, et al. (2008), Procedure for estimation and reporting of uncertainty due to discretization in CFD applications, Journal of Fluids Engineering-Transactions of the ASME, Vol. 130, No. 7.

Coleman, H.W. and W.G. Steele (2009), Experimentation, Validation, and Uncertainty Analysis for Engineers, 3rd edition, John Wiley & Sons, New York, 2009.

Dempster, A. P. (1967), Upper and lower probabilities induced by a multivalued mapping, The Annals of Mathematical Statistics V. 38, No. 2: pp. 325–339.

Eca, L., and M. Hoekstra (2002), An Evaluation of Verification Procedures for CFD Applications, In: 24th Symposium on Naval Hydrodynamics.

Eca, L., and M. Hoekstra (2009), Evaluation of numerical error estimation based on grid refinement studies with the method of the manufactured solutions, Computers and Fluids, V. 38, No. 8, pp. 1580–1591.

Eca, L., M. Hoekstra, and P.J. Roache (2005), Verification of calculations: an overview of the Lisbon workshop, AIAA paper, pp. 4728.

Eca, L., M. Hoekstra, A. Hay, and D. Pelletier (2007), Verification of RANS solvers with manufactured solutions. Engineering with computers, V. 23, No. 4, pp. 253–270.

Ferson, S., V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz (2003), Constructing Probability Boxes and Dempster-Shafer Structures, SAND2002-4014, Sandia National Laboratories, Albuquerque, NM.

Fishman, G. S. (1996), Monte Carlo: Concepts, Algorithms, and Applications, Springer, New York, NY.

Friedman, J.H. (1991), Multivariate Adaptive Regression Splines, Annals of Statistics, V 19, No. 1, pp. 1-67.

Gzyl, H. (1995), The Method of Maximum Entropy, Series on Advances in Mathematics for Applied Sciences, World Scientific, New Jersey, V 29.

Hahn, G. J. and W. Q. Meeker (1991), Statistical Intervals: A Guide for Practitioners, John Wiley & Sons, New York, p. 34.

Hamilton, J. R. and R. G. Hills (2010a), Relation of Validation Experiments to Applications, Numerical Heat Transfer, Part B: Fundamentals, V 57, No. 5, pp. 307-332.

Hamilton, J. R. and R. G. Hills (2010b), 'Relation of Validation Experiments to Applications: A Nonlinear Approach, Numerical Heat Transfer, Part B: Fundamentals, V. 57: 6, pp. 373-395.

Helton, J. C. and W. L. Oberkampf (2004), Special Issue: Alternative Representations of Epistemic Uncertainty, Reliability Engineering & System Safety, V 85, pp. 1-369.

Helton, J. C. (2011), Quantification of margins and uncertainties: Conceptual and computational basis, Reliability Engineering and System Safety, V 96, 976-1013.

Higdon, D., C. Nakhlen, J. Gattiker, and L. Ginzburg (2008), A Bayesian Calibration Approach to the Thermal Problem, Computer Methods in Applied Mechanics and Engineering, V. 197: 29-32, pp. 2457-2466.

Hills, R.G. (2013), Roll-up of Validation Results to a Target Application, Sandia Report SAND 2013-7424, Albuquerque, New Mexico.

Hills, R.G., W.R. Witkowski, W.J. Rider, T.G. Trucano, A. Urbina, (2013), Development of a Fourth Generation Predictive Capability Maturity Model, Sandia Report SAND3013-8051, Sandia National Laboratories, Albuquerque, New Mexico.

Hills, R. G., A. Black, K. Cartwright, D. Turner, A. B. Nauble (2015), Framework for Predictive Capability Assessment, Sandia draft report, Albuquerque, NM.

Hofer, E. (1996), When to separate uncertainties and when not to separate, Reliability Engineering and Safety Systems, Elsevier Science Limited, Volume 54, 113-118.

ISO GUM (1995), Guide to the Expression of Uncertainty in Measurement (corrected and reprinted, 1995), International Organization for Standardization, Geneva, Switzerland.

Jaynes, E.T., (2003), Probability Theory: The Logic of Science, Cambridge University Press, Cambridge.

Kennedy, M.C. and A. O'Hagan (2001), Bayesian Calibration of Computer Models, Journal of the Royal Statistical Society: Series B, Vol. 63, Issue 3, pp. 425-464.

Mayer, M. A., and J. M. Booker (1991), Eliciting and Analyzing Expert Judgment; A Practical Guide, ASA-SIAM Series on Statistics and Applied Probability, Society of Industrial and Applied Mathematics and American Statistical Association, London.

McKay (1996), "Variance-Based Methods for Assessing Uncertainty Importance in NUREG-1150 Analyses," LA-UR-96-2695.

Oberkampf, W. L., Pilch, M. and T. G. Trucano (2007), Predictive Capability Maturity Model for Computation Modeling and Simulation, Sandia National Laboratories Report SAND2007-5948, October.

Oberkampf, W. L. and C. J. Roy (2010), Verification and Validation in Scientific Computing, Cambridge.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow (2006), Uncertainty Judgements: Eliciting Experts' Probabilities, John Wiley & Sons, Ltd., England.

Pelletier, D., and P.J. Roache (2006), Verification and validation of computational heat transfer, Handbook of Numerical Heat Transfer, pp. 417–442.

Pilch, M., T. Trucano, J. Moya, G. Froehlich, A. Hodges, and D. Peercy (2001), Guidelines for Sandia ASCI Verification and Validation Plans: Content and Format: Version 2.0. Tech. Report SAND2000-3101, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, January 2001.

Pilch, M., T. G. Trucano, J. C. Helton (2011), Ideas underlying the quantification of margins and uncertainties, Reliability Engineering and System Safety, V 96, pp. 965-975.

Richardson, L. F. (1911), The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam, Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character, V. 210, pp. 307–357.

Rider, W.J. and J.R. Kamm, (2012), Advanced Solution Verification of CFD Solutions for LES of Relevance to GTRF Estimates, SAND2012-7199P, Sandia National Laboratories, Albuquerque, NM, August.

Rider, W.J. (2013), Preliminary Solution Verification of Denovo: Focus on Spatial-Angular Convergence, SAND2013-1421P, Sandia National Laboratories, Albuquerque, NM, February.

Roache, P.J. (1994), Perspective: a method for uniform reporting of grid refinement studies, Transactions-American Society of Mechanical Engineers Journal of Fluids Engineering, V. 116, pp. 405–405.

Roache, P.J. (1998), Verification and validation in computational science and engineering, Computing in Science Engineering, pp. 464.

Roache, P. J. (2009), Fundamentals of Verification and Validation, Hermosa publishers, Socorro, NM.

Saltelli, A., K. Chan, E.M. Scott (2000), Sensitivity Analysis, JohnWiley & Sons, Ltd., New York, NY.

Saltelli, A., K. Chan, E.M. Scott (2008), Sensitivity Analysis, JohnWiley & Sons, Ltd., New York, NY.

Saltelli and Tarantola (2002), "On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal," J. Amer. Statist. Ass., Vol. 97, N. 479, pp. 702-709.

Shafer, G. (1976), A Mathematical Theory of Evidence, Princeton University Press, ISBN 0-608-02508-9.

Shaw, R.A., T.K. Larson, and R.K. Dimenna (1988), Development of a Phenomena Identification and Ranking Table (PIRT) for Thermal-Hydraulic Phenomena during a PWR LBLOCA, Idaho Falls, ID, EG&G.

Trucano, T. G, M. Pilch, W. L. Oberkampf (2002), General Concepts for Experimental Validation of ASCI Code Applications, Sandia Report: SAND2002-0341, Sandia National Laboratories, Albuquerque, NM

Wiener N. (1938), The Homogeneous Chaos, American Journal of Mathematics, Vol. 60, No. 4. pp. 897–936.

Williams, C. (2002), “Gaussian Processes” chapter in The Handbook of Brain Theory and Neural Networks, M. Arbib, ed. Cambridge, MA: MIT Press.

Wilson, G.E., and B.E. Boyack (1998), The Role of the PIRT in Experiments, Code Development and Code Applications associated with Reactor Safety Assessment, Nuclear Engineering and Design, Vol. 186, 1998, pp. 23-37.

Wu, C.F.J. and M. Hamada (2000), Experiments: Planning, Analysis, and Parameter Design Optimization, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York.

Xiu, D. (2010), Numerical Methods for Stochastic Computations: A Spectral Method Approach Princeton University Press.

Zadeh, L.A. (1978), Fuzzy Sets as a Basis for a Theory of Possibility, Fuzzy Sets and Systems, Vol. 1, pp. 3-28.

DISTRIBUTION

External (print and electronic)

6 Department of Energy
Attn: S. Ananthan (1)
J. W. Cline (1)
M. Derby (1)
N. Johnson (1)
M. Robinson (1)
J. Zayas (1)
1000 Independence Avenue S.W.
Washington DC 20585
Room 5H072

External (electronic)

P. Veers	National Renewable Energy Laboratory
P. Moriarty	National Renewable Energy Laboratory
S. J. Schreck	National Renewable Energy Laboratory
W. Shaw	Pacific Northwest National Laboratory
F. Sotiropoulos	University of Minnesota
J. Naughton	University of Wyoming

SNL Internal (electronic)

MS0110	D. E. Womble	1220
MS0825	M. F. Barone	1515
MS0828	R. G. Hills	1544
MS0828	W. R. Witkowski	1544
MS0836	S. P. Domino	1541
MS0845	S. A. Hutchinson	1540
MS1124	D. G. Griffith	6121
MS1124	D. C. Maniaci	6121
MS1124	D. G. Minster	6121
MS1124	B. R. Resor	6121
MS1124	D. L. Laird	6122
MS0899	Technical Library	9536

Sandia National Laboratories