SAND2015-0840R

SANDIA REPORT

SAND2015-XXXX
Unlimited Release
Printed February 2015

V&V Framework

Part 1 Release

Richard G. Hills, David C. Maniaci, and Jonathan 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@osti.gov
Online ordering:	http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone:	(800) 553-6847
Facsimile:	(703) 605-6900
E-Mail:	orders@ntis.gov
Online order:	http://www.ntis.gov/search

SAND2015-XXXX
Unlimited Release
Printed February 2015

V&V Framework

Part 1 Release

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

To February 24-25 A2e HFM planning meeting participants:

A2e HFM Objective
Accurately predict, assess and optimize wind plant performance utilizing High Performance Modeling (HPC) tools developed in a community-based, open-source simulation environment to understand and accurately predict the fundamental physics and complex flows of the atmospheric boundary layer, interaction with the wind plant, as well as the response of individual turbines to the complex flows within that plant

One of the focuses of the February meeting is to identify the important physics (phenomena) that should be simulated by our high fidelity models, assess the capability of our models to represent these physics, and to perform a gap analysis of model development, verification, and validation needs associated with the more important physics. The gap analysis provides structured information to help in the planning of an integrated A2e HFM/Experimental program to meet the objective stated above. The identification and ranking of this phenomena and the corresponding gap analysis is one of the more important considerations when utilizing models of complex systems (such as a wind plant).

To help expedite the discussions, we will develop a simple table (called a Phenomenon Identification Ranking Table, PIRT) listing the important phenomenon in the first column, followed by columns with our assessments of the importance of the phenomena relative to the HFM objective (ranked High, Medium, Low), ability of our mathematical models to represent the phenomena, followed by the verification and validation evidence that the computational algorithms solve the mathematical models and represent the behavior of the physics as required to meet the HFM Objective. A simple gap analysis can then be performed based on this information. The PIRT is the most commonly used approach for characterizing model capability and model/validation needs from a physics perspective. This tool was first developed by the nuclear power industry, and has become the standard approach to addressing model capability needs from a planning perspective for complex, multi-physics engineering models.

The attached document is PART 1 of a validation directed planning document that we are developing. Chapter 3 provides a more detailed description of, and guidelines for the use of the PIRT (worth reading before the meeting). We will also provide copies of a consolidated PIRT showing rankings for a limited number of phenomena, and a PIRT listing more phenomena with the rankings left blank. These were developed by others in the A2e program and can be used as a reference for our discussion. We suggest that you consider using this second PIRT as a starting point for your discussions, revising phenomena as appropriate and assigning your own rankings.

The first day of the meeting will have breakout group meetings that will allow the meeting attendees to provide input to the gap analysis by completing the PIRT. The breakout session leads will facilitate this process, and several people who are familiar with the gap analysis process in regards to the PIRT will provide support. The breakout sessions during the morning of the second day will give attendees a chance to provide direct input to the experimental planning process by identifying detailed quantities of interest and the associated measurements that are needed to validate models for the intended applications. The final breakout sessions of the meeting will focus on how different aspects of the models impact simulation software development.

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 manor. 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.

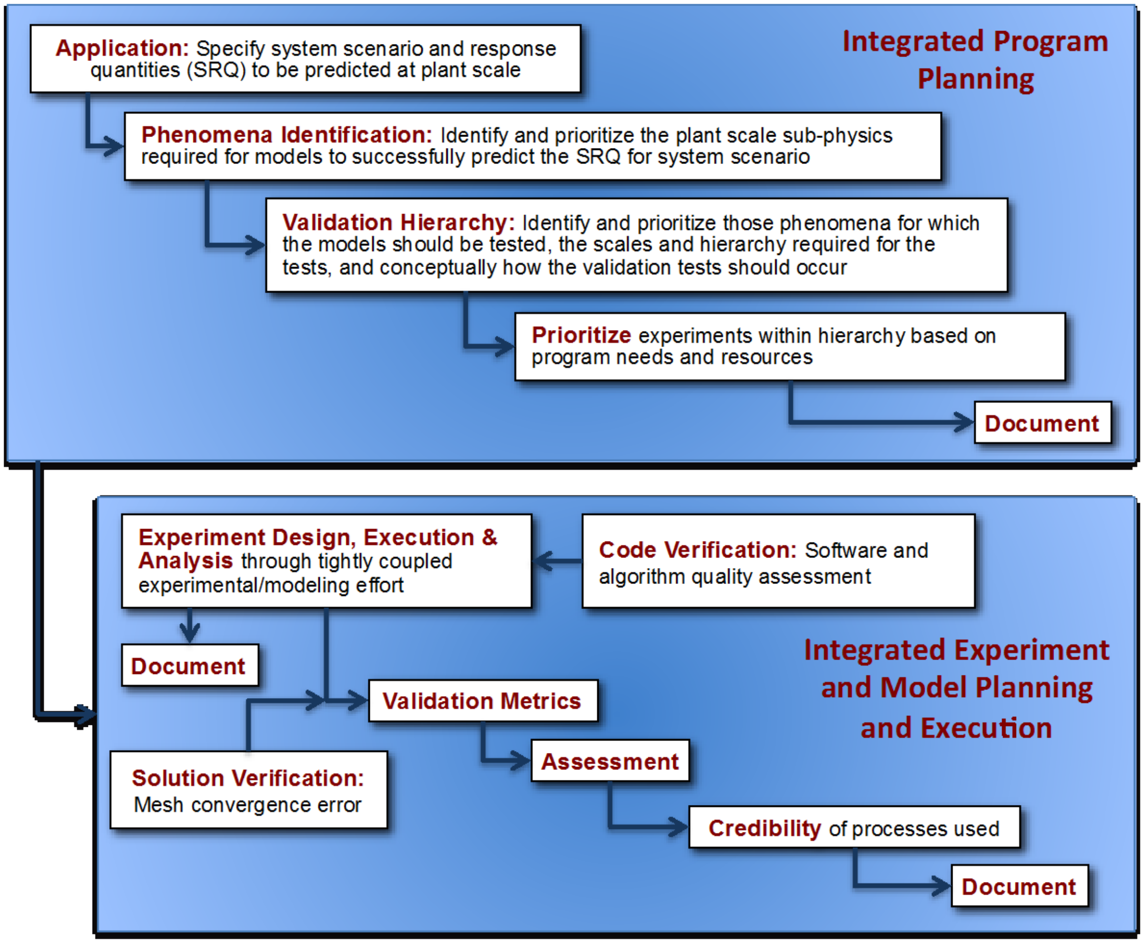
The development and execution of this process requires well integrated team planning among those responsible for programmatic needs, computational model developers, model users, and experimentalists, and should consider the needs of the eventual customers of the modeling capability and results. The communication 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.

Figure 1.1. Validation directed program planning and implementation

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.

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.

1.7 Report Organization and Scope

This document is designed to be use as a guide to the implementation of the process summarized in Figure 2.1. The guide will focus on the 'how' through a step-by-step procedure. References will be provided to the literature that lead to the development of this process, as well as literature that addresses specific technical aspects of the process. The remainder of this document will be divided into two parts. Part 1 addresses the integrated program planning illustrated in the upper blue box of Figure 2.1. Part 2 addresses issues associated with specific validation experiment design, model validation (a quantitative process), model credibility (a judgment process) and issues of uncertainty associated with the lower blue box of Figure 1.1. Part 2 of this document is not included in the present release.

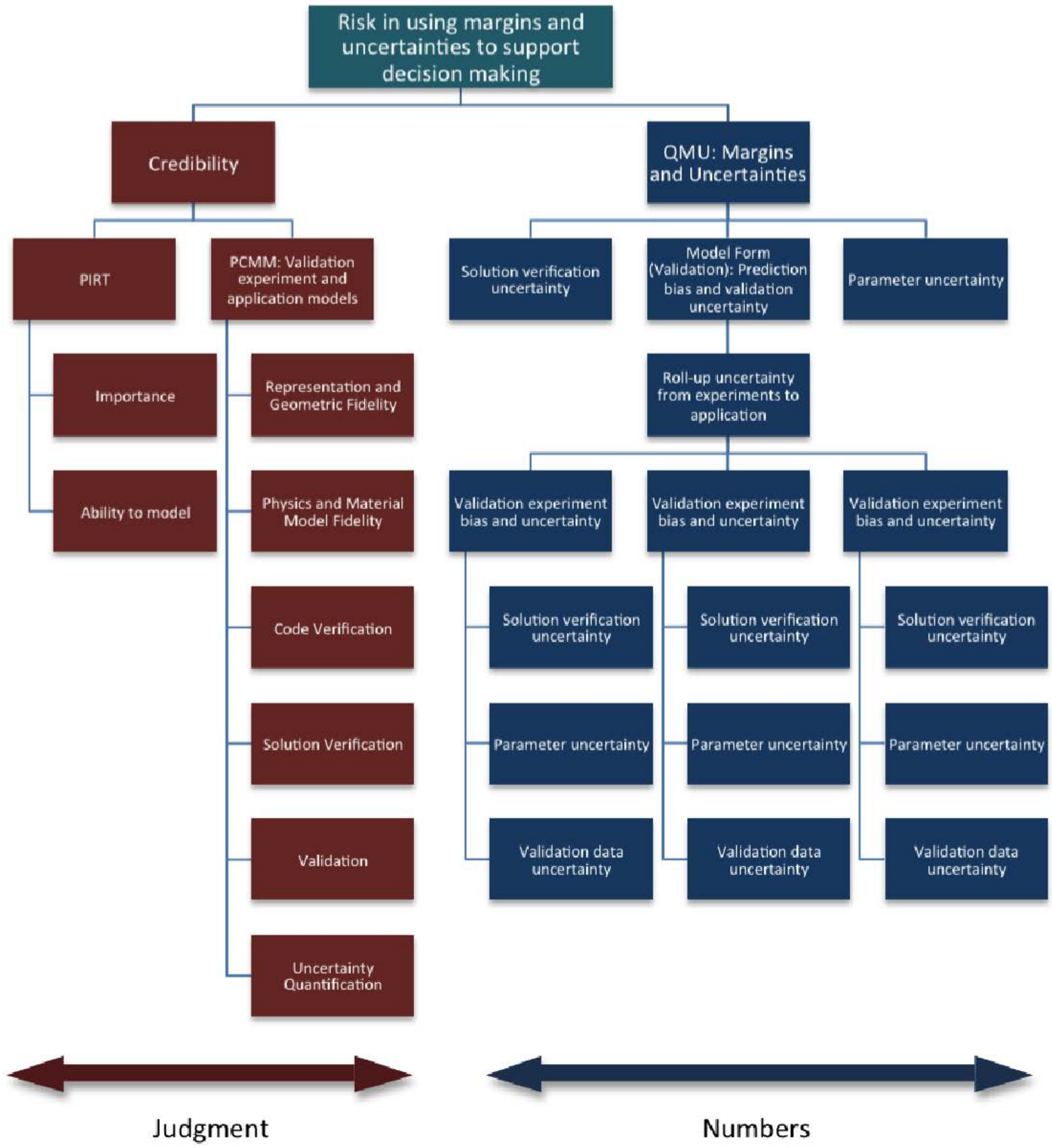


Figure 1.2. Risk of Using a Model for Application

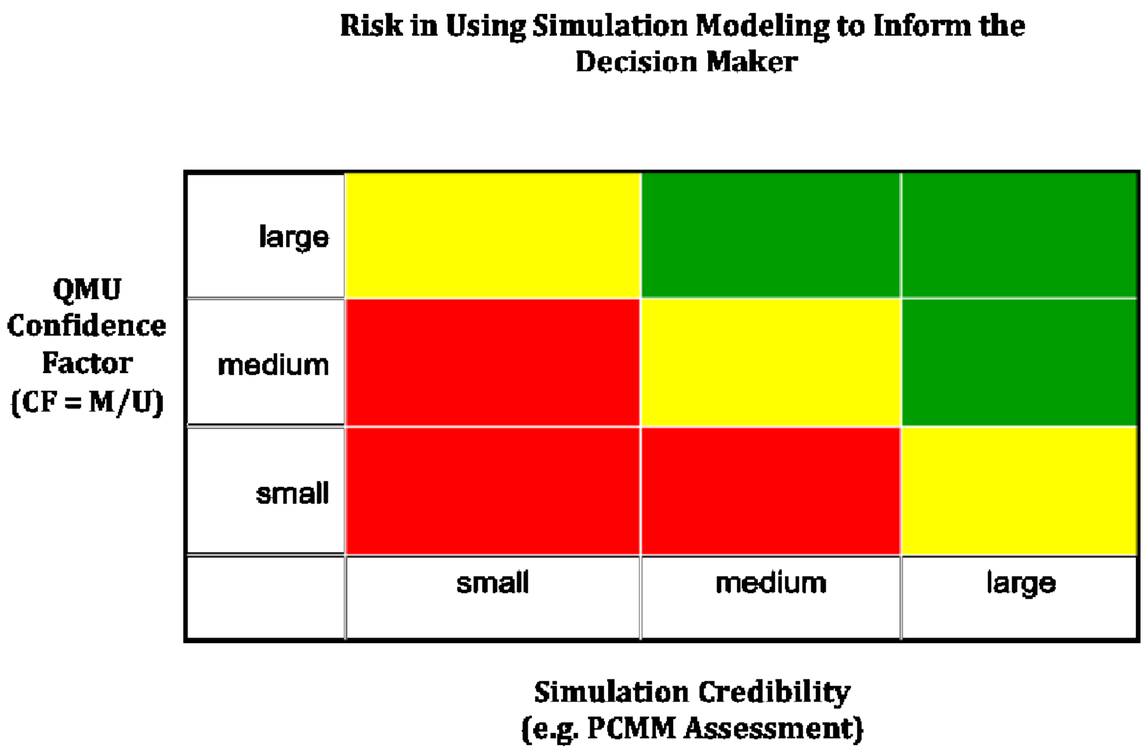
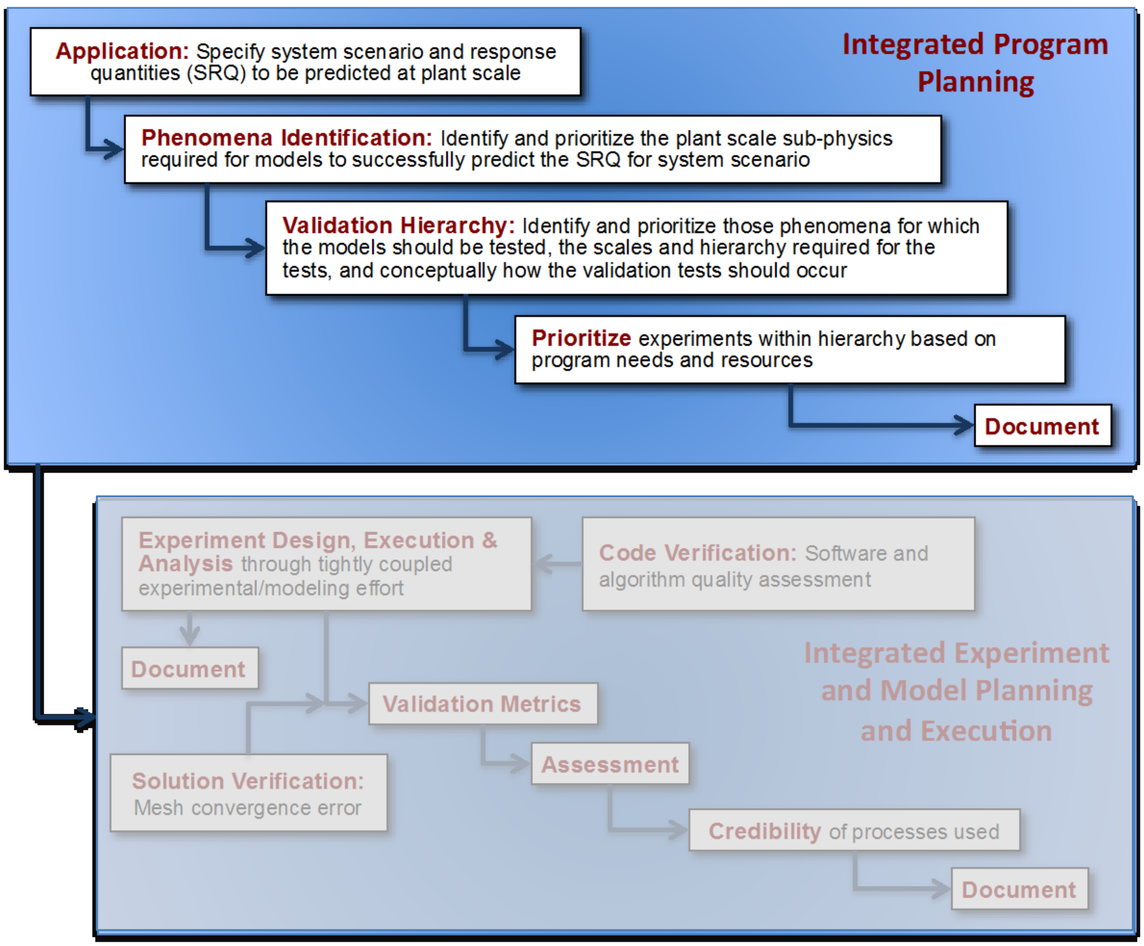


Figure 1.3: Risk Associated with Computational Simulation, green – low risk, yellow – intermediate risk, red – high risk. The Predictive Capability Maturity Model (PCMM) is an expert elicitation tool used to assess simulation credibility (discussed in Chapter 11). M and U are the margin, and the uncertainty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

2.1 The Modeling Objective

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:

1. *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*

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.

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.

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 application[1]

- 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.

---

[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.

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.

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

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-	Sensitivity analysis on impact of de-featuring

	components	
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

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 3.2 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).

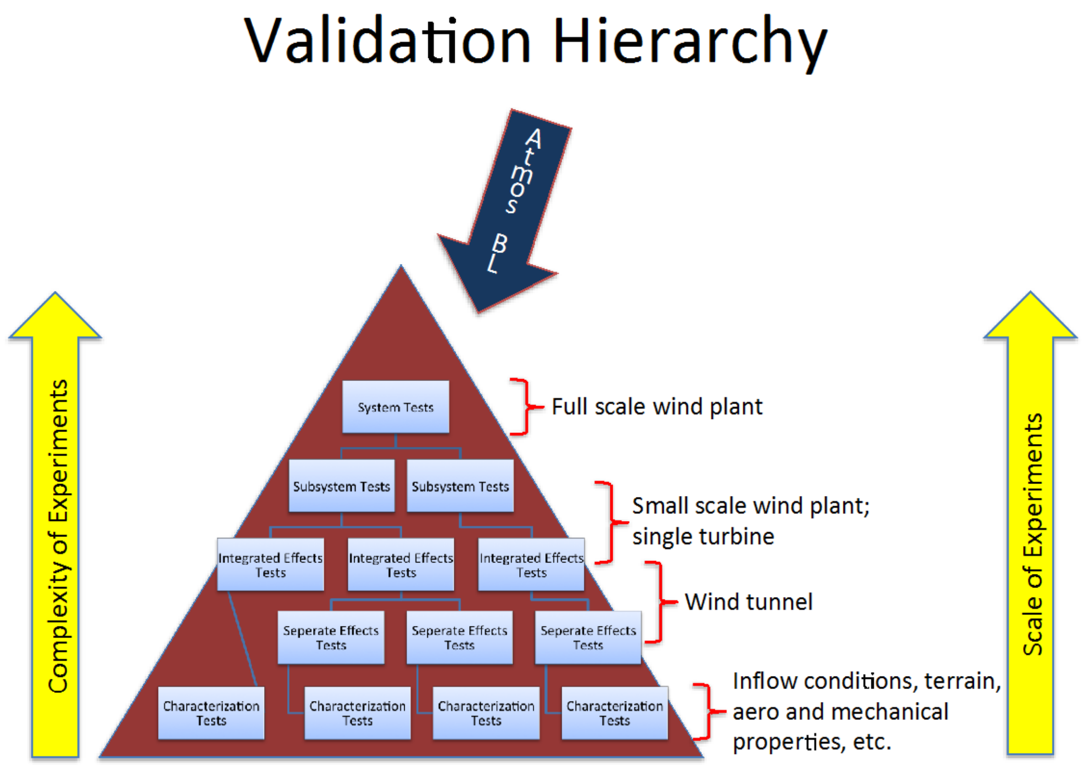


Figure 3.1: The validation hierarchy

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.

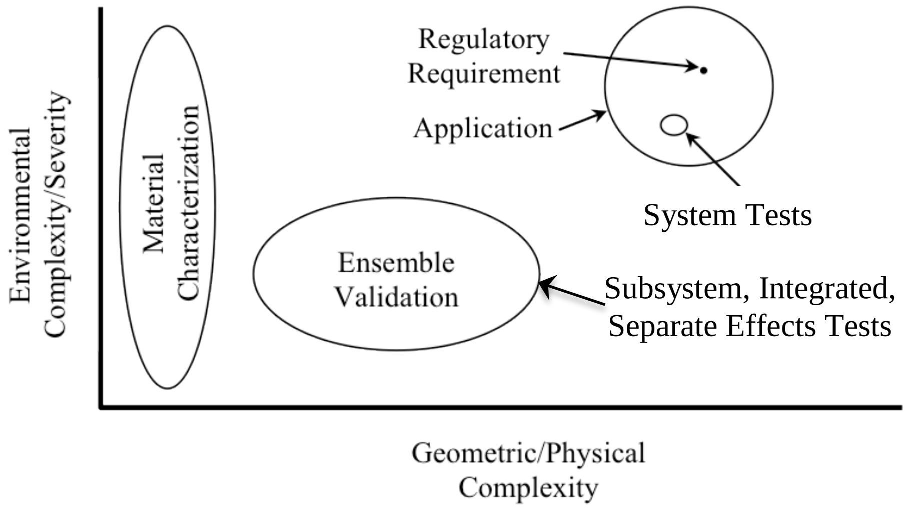


Figure 3.2: Experimental Hierarchy Complexity (based on Hills et al. 2008)

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.

5.0 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 focuses 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 the used to perform a gap analysis, identifying the phenomena for which the importance is high or medium, but 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 focuses 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. Part 2 of this document is not included in the present release, although a summary of it is included here. 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. 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.

A model validation exercise quantifies agreement between model prediction and experimental observation for the conditions of the experiments. 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. A brief overview of this tool is provided in the previous chapter (not included in the present release).

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 judgment, and technical and resource limited compromises. The processes discussed here provide some 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 to the intended application.

6.0 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.

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.