Exploitation of a Validation Hierarchy for Modeling and Simulation

Scott Shaw[1]
MBDA UK Ltd, Lea Field Way, Bolton, BL5 1ED, UK

James M. Luckring[2]
NASA Langley Research Center, Hampton, VA, 23681, USA

William L. Oberkampf[3]
W. L. Oberkampf Consulting, Georgetown, Texas, 78633 USA

Rick E. Graves[4]
U. S. Air Force Research Laboratory, Aerospace Systems Directorate, Wright-Patterson Air Force Base, Ohio 45433-7512 USA

There is an evolving need to increase reliance on physics-based simulation to develop, design, and optimize a wide range of engineering systems. This increased reliance on modeling and simulation has highlighted a growing need to better assess the confidence that modeling and simulation analysts have in their computational results. More specifically, analysts and designers are being asked to reduce physical testing, but still ensure the performance, reliability and safety in their designs. For isolated components of a complex system, where a single discipline may drive product design, this is less of a concern because a single physics process is often straightforward and easy to explain. However, when these isolated components are integrated, and are expected to operate in a multidisciplinary context, new concepts and model assurance standards are required. In this paper, we address this challenge by showing how a model validation hierarchy can be exploited to identify those model validation experiments that will contribute most to increasing confidence and credibility of modeling and simulation predictions. The approach that is adopted contains four main steps. The first step is the construction of a model validation hierarchy that links subsystems, assemblies, and components to a hierarchy of physical experiments that can be used to support model validation. This hierarchy connects in a clear and logical way the perspective of systems engineers focusing on system design, to the perspective of the modeling and simulation analyst who is concerned in simulation accuracy. The structure and content of this hierarchy is then used in a second step to establish which physical phenomena have the greatest impact on overall system performance metrics. A gap analysis technique is then used to prioritize the important computational activities that may damage the accuracy of the simulation results. Unfortunately, a common outcome of such gap analyses is the identification of many important gaps. In the final step of our process, we advocate for the use of a global sensitivity analysis to improve the prioritization of the weaknesses in computational activities.

[1] Platform/Weapon Synthetics SME, Systems Design Airborne Weapon Systems, MBDA UK, Senior Member AIAA.
[2] Senior Research Engineer, Configuration Aerodynamics Branch, AIAA Associate Fellow.
[3] Consultant, Fellow AIAA.
[4] Aerospace Engineer, Design and Analysis Branch, Associate Fellow AIAA.

Abbreviations:

AIAA	American Institute of Aeronautics and Astronautics
ASME	American Society of Mechanical Engineers
DoD	Department of Defense
GSA	Global Sensitivity Analysis
HELP	Hierarchy Element and Listing Procedure
IEEE	Institute of Electrical and Electronics Engineers
M&S	Modeling and Simulation
MSSA	Multidisciplinary Statistical Sensitivity Approach
NAFEMS	National Agency for Finite Element Methods and Standards
NASA	National Aeronautics and Space Administration
NATO	North Atlantic Treaty Organization
PIRT	Phenomena Identification and Ranking Table
QPIRT	Quantitative PIRT
SME	Subject Matter Expert
SRQ	System Response Quantity
STO	Science and Technology Organization
V&V	Verification and Validation

I. Introduction

In the aerospace industry and within government agencies, there is increasing emphasis on predictive numerical simulation to reduce the development time and testing cost associated with new systems. A key element of this initiative, usually referred to as digital engineering, is the use of physics-informed models to identify potential system-level shortcomings in the early phase of the product life cycle. For isolated components, where a single physical discipline may drive product design, this is less of a concern. However, when these isolated components are integrated in a multidisciplinary complex system, innovative methodologies are required to help establish credible model assurance standards. With increased reliance on predictive physics-based simulation, the connectivity and rigor associated with physical experimentation must also mature to keep pace with digital engineering requirements.

To address this concern, industry and government have invested significantly in integration of digital technologies across all areas of the product life cycle to impact the time and cost associated with the design, development, manufacturing, and testing of commercial and defense products. The U.S. Department of Defense (DoD) defines digital engineering as an integrated digital approach that uses authoritative sources of system data and models as a continuum across disciplines to support life cycle activities from concept through disposal [1]. Computational simulation is a common source of digital information that may be leveraged to inform the entire life cycle. In this context, model validation is 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 [2]. For digital engineered systems, new methodologies are needed to provide a system designer and stakeholders quantifiable confidence in the credibility of physics-based models used for the complete system, subsystems, assemblies, and components.

We believe that the foundations upon which any rigorous model confidence assessment are built are code verification, solution verification, model validation, and uncertainty quantification. A number of publications are available on code and solution verification, see for example Refs. [3] and [4]. In this paper, we focus our efforts on model validation, the process of assessing and quantifying model form uncertainty. When assessing the accuracy of a computational model of a complete system, we face three major challenges. Firstly, the challenge of dimensionality due to the large number of potential model input parameters that define the possible operating conditions of the overall system. Unfortunately, the explosion in the number of cases that must be considered if model form uncertainty is to be characterized for all combinations of model inputs is usually intractable and we must instead focus on a limited number of cases that best represent the overall operation space. The second challenge is that of complex, nonlinear coupling between individual physical phenomenon and the potential for resulting emergent behavior that is peculiar to a limited input parameter space. Finally, we must face the fact that there are many sources of uncertainty that make it not only difficult to adequately characterize the initial and boundary conditions required for a representative simulation but may also prevent meaningful assessment of model form uncertainty. (See, for example, the discussion of Roache [3].)

Addressing these challenges raises two fundamental questions that test the viability of digital engineering based on simulation-informed decision making. Firstly, how can we simplify system level validation experiments in a way that allows model form uncertainty to be both measured and related to the model uncertainty of the overall system?

Secondly, recognizing that we are unlikely to be able to perform many of the validation experiments we would like, how should we select from among the large number of possible validation experiments? Stated differently, how should we prioritize those that will make the most important contribution to assessing and quantifying overall system model form uncertainty?

Existing approaches to these questions belong to three main classes: (i) knowledge-based approaches (ii) sensitivity-based approaches and (iii) coverage-based approaches.

A. Expert Knowledge Based Approaches

Subject Matter Expert (SME) elicitation is an important technique for quantifying and synthesizing expert knowledge in situations when uncertainties exist due to insufficient data, ambiguity, or complexity. SME elicitation processes provide powerful frameworks through which to aggregate knowledge and align stakeholders to bring focus on the problem at hand. In modeling and simulation, a particular form of SME elicitation, the Phenomena Identification and Ranking Table (PIRT) [4-7] has been developed by the nuclear power industry where it has been applied to a wide range of nuclear technology issues. The PIRT process begins by constructing a list of phenomena that are expected to influence a particular system response quantity (SRQ) of interest. The process then proceeds by ranking these phenomena according to an agreed set of criteria to identify what is most important relative to the issue in question. The process captures not only the ranking but also the rationale and supporting information obtained to explain the ranking.

Knowledge-based approaches, like PIRT, suffer from three main drawbacks. The first drawback is the potential that cognitive heuristics and bias may distort evidence in the decision-making process. As a consequence, the results may be based upon informed *opinion* rather than an objective point of view. This drawback can be mitigated to some degree by drawing on expert calibration techniques and ensuring that the panel is formed from diverse backgrounds. The second potential drawback is that of completeness. The knowledge of the expert panel is finite and, therefore, the issue in question may go beyond the experience and expertise of the panel members. If this occurs, it may require explicit elaboration of all the elements and combinations of elements related to the issue in question. Finally, the analysis is qualitative in nature in the sense that experts believe certain elements are important or not important, but we have no quantitative basis on which to make finer grain assessments or rankings. Some efforts have been made to overcome this, see for example the work of Yurko and Buongiorno [8] who have developed a Quantitative Phenomena Identification and Ranking Table (QPIRT) approach by using traditional PIRT in combination with a sensitivity analysis.

B. Sensitivity Based Analysis

The goal of a sensitivity analysis is to identify how "the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input" [9]. Sensitivity analyses can be broadly classified as local or global. In the case of the former, the primary concern is in understanding the local relationship between model inputs and outputs while global sensitivity is concerned with understanding the variance in model outputs in terms of all contributing uncertainties; all characterized as random variables. The latter is of particular interest in the current work because global sensitivity analysis provides a rigorous mathematical basis with which to address questions such as "Which factors are most important in determining uncertainty in the model output?" and "Which input uncertainty should we tackle to reduce variance of the output the most?" A detailed description of global sensitivity analysis (GSA) is beyond the scope of the present work. The interested reader is referred to the introductory text of Saltelli [9].

A key challenge in global sensitivity analysis is the need to distinguish between aleatory and epistemic uncertainties. Aleatory uncertainty is irreducible and arises due to the random variation of inputs and model parameters that impact model outputs. Epistemic uncertainty is due to a lack of knowledge and can be reduced by learning more. Model form uncertainty is an example of an epistemic uncertainty. This distinction is important because the mathematical techniques needed to handle aleatory and epistemic uncertainties are quite different and pose significant difficulties when both forms of uncertainty must be considered in combination, see for example the discussion of Ref. [10].

Guo and Du [10] have presented a unified uncertainty analysis that represents epistemic uncertainties using evidence theory's basic probability assignment and aleatory uncertainties using probability density functions. Their approach results in plausibility and belief functions that model the combined epistemic and aleatory uncertainty. The difference between belief and plausibility functions is a measure of the epistemic uncertainty that can arise from any of the sources of uncertainty.

In engineering practice, most model parameter uncertainties are of mixed type. That is, model parameters are commonly mixtures of both aleatory and epistemic uncertainty. Mullins and Mahadevan [11] address this problem by assigning a probability density function to represent the aleatory uncertainty in model calibration activities. The calibrated model can then be used in place of the original model to make predictions that capture the mixed nature of the parameter uncertainty.

For multidisciplinary models, the uncertainty propagation challenge is compounded by the need to consider multiple coupled models. Jiang, Chen and German [12] developed a multidisciplinary statistical sensitivity approach (MSSA) to obtain the importance rankings of contributions from various sources of uncertainty. To propagate uncertainties, they apply the concept of model bias correction to multidisciplinary models by introducing additional uncertainties in the linking variables, so called because they link the outputs of one model to the inputs of another. Their technique employs metrics based upon the concept of relative entropy that provide not only information about the variance of the model outputs but also the entire performance probability distribution shape.

Recent developments using imprecise probability theory offer an opportunity to directly characterize epistemic uncertainties in a more flexible mathematical structure. For example, probability bounds analysis allows an epistemic uncertainty to be characterized as a pure interval, without any assumed probability distribution [4]. More capable methods to characterize epistemic uncertainty in imprecise probability theory are also available [13,14]. However, the mathematical procedures and the software packages for conducting GSA using imprecise probabilities are not well developed.

Sensitivity analyses need not be confined to modeling of uncertainties and their propagation. For example, Allaire and his coworkers [15] have developed a measure of system complexity that can be used to compare and rank different candidate designs with respect to SRQs. In addition to ranking of designs, their approach allows key contributors to the complexity to be identified. Such factors may include model adequacy/inadequacy.

Global sensitivity analysis for a complete system, as discussed above, is rarely practiced for two main reasons. First is the difficulty of obtaining a mathematical model of the *complete* system in a form amenable to the propagation of uncertainties. Second, for complex systems, it is well recognized that the model form uncertainty, particularly the coupling of subsystems and subassemblies, is a dominant contributor to the total uncertainty of the SRQs at the top level of the model validation hierarchy. The magnitude of model form uncertainty is commonly underestimated, especially in fields like structural dynamics of complex structures that are built from many substructures and components.

C. Coverage-Based Approaches

Coverage-based approaches for model validation measure and communicate the level of model validation that has been performed in comparison to the amount that could be performed for complete coverage [16]. At their simplest, they measure the degree to which the model input parameters have been assessed with validation experiments as compared to the model's full operational domain [17]. Hällqvist and his coworkers [17,18] have illustrated this concept in the context of aircraft simulation. In their work, they elaborate four general principles for a validation coverage metric: (i) when conducting a validation experiment at an untested model operational point, one should always seek to broaden coverage; (ii) diverse validation experiments yield better coverage than clustered validation settings; (iii) extrapolation results in a degradation of coverage; and (iv) the coverage metric should be objective. To implement these principles, they suggest the use of a sensitivity adapted nearest neighbor metric.

For models whose output is essentially linear with respect to the model, inputs to this approach should work well. However, for many engineering models that exhibit strong nonlinear behavior between model inputs and model outputs, it is unclear that application of their second and third principles is a useful approach. Indeed, we may find that extrapolation is perfectly reasonable in regions exhibiting linear behavior while clustering may make sense where the model exhibits bifurcations or discontinuous behavior.

Olsen and Raunak [16], [19] have attempted to remedy these deficiencies by extending the idea of coverage more generally. In their work, they propose that the elements of the model that are to be validated be identified, and that the coverage metric be adapted to measure the degree to which all the possible model validation experiments that could be done for such elements is covered by the model validation experiments that have been performed. While this addresses the problems with Hällqvist et al.'s principles, it still requires that we find a way to define the span of the model validation experiments that could be performed.

In this paper, we demonstrate how a model validation hierarchy, see for example Ref. [20], can be exploited to identify model validation experiments that are traceable to a specific element in the hierarchy and will contribute the most to reducing program risk. Our identification process contains three main elements. The first element adapts an existing approach known as the Phenomena Identification and Ranking Table (PIRT) described in Refs. [4-7] to make use of the structure and content of the validation hierarchy to understand which physical phenomena have the greatest impact on system performance. We then employ a gap analysis approach from Ref. [7] to prioritize the physical phenomenon, as well as any other computational simulation activities, that contribute to uncertainty in predicted system response quantities. A global sensitivity analysis approach [7] is used to complete the prioritization and identify resultant gaps in predictive capability of the physics-based models.

The paper is constructed as follows. In Section II, we introduce the overarching principles and philosophy of our approach to create a model validation hierarchy. This methodology introduces several definitions related to computational physics-based simulation that we use throughout the paper. In Section III, we then apply the Hierarchy Element and Listing Procedure (HELP) technique, which exploits the content and structure of the validation hierarchy to identify physical phenomena that are likely to be of most importance for the system response quantities of interest. The use of gap and global sensitivity analysis to refine the prioritization further are demonstrated in Sections IV and V before we present conclusions in Section VI.

II. Validation Hierarchy

In this section, we introduce the overarching principles and philosophy of our approach to creating a validation hierarchy. We first present some background terminology and basic concepts of a model validation hierarchy. This includes several definitions that we use throughout this paper. We then introduce a general framework for the construction of the validation hierarchy.

A. Background Terminology and Concepts

It is useful to acknowledge that differing definitions of verification and validation (V&V) have developed over time among several professional communities (e.g., software engineering, systems engineering, and the Department of Defense). The IEEE [21] first established formal V&V definitions, and the software engineering and systems engineering professions follow different variants of the IEEE definitions. Subsequently, the DoD established definitions [22] that are used among computational physics professions such as the AIAA, ASME, NAFEMS, and the nuclear weapons laboratories. Our interests are anchored in computational physics, and, in the following sections, we present the definitions we use in the present work.

1. Validation

For our work, we follow the AIAA [2] and define 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 uses of the model."

In this definition, the 'model' is often a mathematical/computational representation of some physical process (e.g., solid mechanics, fluid dynamics, electrodynamics) and the 'real world' is often a physical experiment or test. Following Roache [23], the intent of validation is to assess whether "we have solved the right equations." With this definition, model validation is an assessment of model accuracy. Model validation is empirical in the sense that it involves the comparison of model realizations with the observed behavior of the system of interest. Tests of model accuracy are also objective; the intent is to quantify the difference between the model realization and physical measurement. As a result, model validation is *not* a judgement on the adequacy of the model for its intended use.

Comprehensive discussions of model validation and related concepts can be found in the texts of Roache [3,23] and Oberkampf and Roy [4].

2. Validation Experimentation

In the context of supporting the development and application of Modeling and Simulation (M&S) capability, a number of authors (e.g., Bradley [24], Marvin [25] and Oberkampf [26]) have identified the need for an additional category of experiment that does not fit within the traditional categorization: the validation experiment. Following Oberkampf and Roy [4]:

Validation experiments are designed, executed, and analyzed for the purpose of quantitatively determining the ability of a mathematical model expressed in computer software to simulate a well characterized physical process. In addition to the traditional experimental emphasis on performing measurements in a controlled repeatable environment, validation experiments are distinguished by the need to characterize initial and boundary conditions to allow more precise comparisons between measurement and simulation. Thus, in the context of model validation measurements, a new priority is that the experiment be well characterized in the sense that all important model input data needed for a simulation are measured.

A recent assessment of validation experimentation was performed with a specialists meeting by the NATO STO [27] and a good example of validation testing can be found with a current campaign to study juncture flows [28]-29].

3. System Environment and Scenario

A particular system environment and scenario must be specified for the construction of a model validation hierarchy. Oberkampf and Roy [4] identify three general environments in which systems can operate: normal, abnormal, and hostile. Normal environmental conditions are those that can reasonably be expected to be encountered by the system of interest over its nominal range of operating conditions. Abnormal environments are defined as any conditions of operation that would be considered adverse, highly unusual, including a range of accident conditions. Finally, hostile environments are defined as any conditions where the system is under any type of hostile action, for example, physical, electronic or software attack, or attempt to disable or defeat the system.

For each environment, one can identify potential scenarios that could be experienced by the system. A scenario is a particular event or sequence of events to which the system is exposed. For the normal environment, identification of scenarios is often straight-forward. Some normal environment scenarios for an aircraft would be takeoff, cruise, and landing. For abnormal and hostile environments, scenarios are much more diverse. Scenarios from all three environments provide different needs for model validation assessments; the environment and scenario establish the context for these assessments.

4. System Response Quantity

A System Response Quantity (SRQ) is a physical quantity that characterizes some aspect of the behavior or response of the system of interest. Example SRQs for the cruise scenario of an aircraft are lift and drag, deflection of the structure, and rate of fuel consumption. SRQs range from aggregate measures related to high-level system performance down to detailed local quantities, for example, the temperature of an electronic component within a guidance computer. System response quantities establish the focus for a model validation assessment. They occur at all levels of a model validation hierarchy.

B. General Framework of Model Validation Hierarchy

In this paper, we employ the concept of a model validation hierarchy introduced by Luckring et al. [20]. A conceptual framework for a validation hierarchy is shown in Fig. 1. The framework contains two distinct perspectives of the system. At the top is a systems engineering perspective. This systems architecture section of the hierarchy acknowledges the fact that the needs for a validation hierarchy are rooted in complete system performance issues. This region would be comprised of multiple tiers organized with a parent/child system decomposition common to systems engineering. This results in a systems architecture where the complexity decomposition, follows the functionality of the system, subsystems, and assemblies.

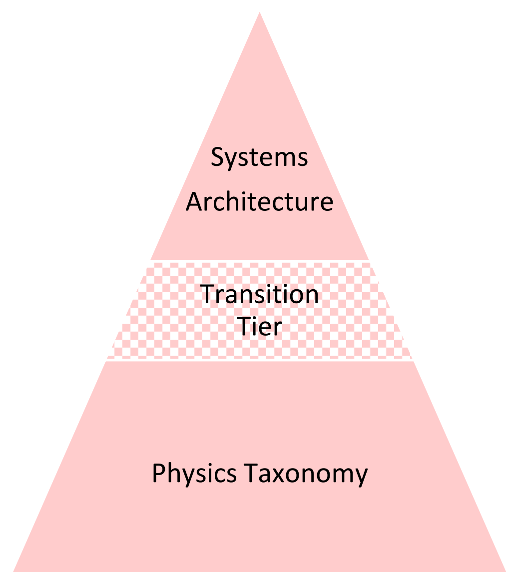

Figure 1: Validation Hierarchy Concept.

The bottom section of the hierarchy is a physics/phenomenological perspective that embraces modeling and simulation. This physics section of the hierarchy acknowledges the fact that system modeling will include highly complex physics that need to be decomposed into simpler problems. The physics taxonomy employs physical phenomena as its organizing principle. This region is comprised of multiple tiers organized with a parent/child physics decomposition. This results in a physics taxonomy where the complexity decomposition follows the decoupling of physical processes. The highest Physics Taxonomy tier would often embrace complex phenomena with complex interactions. The lowest Physics Taxonomy tier would include unit-problem-like topics, i.e., isolated phenomenon without physics interaction effects. Validation experiments, for the purpose of assessing the accuracy of specific simulation models, can be crafted for various tiers within the Physics taxonomy section.

In many ways, systems architecture and physics taxonomy regions represent disparate views within the same validation hierarchy (i.e., systems operations vs. physical processes). In attempting to use previous published work on the construction validation hierarchies, Luckring et al. [20] found that it was necessary to introduce the concept of a transition tier for the purposes of bridging this gap between these two perspectives. The transition tier is neither part of the systems architecture nor part of the physics taxonomy. It transforms the hierarchy from a systems architecture view to a physics taxonomy view. The transition tier is a single tier, and its focus is on modeling the elements in the lowest tier in the systems architecture of the hierarchy.

This approach to the construction of a validation hierarchy is described in detail, with examples, in Ref. [20].

III. Prioritization Based on System Response Quantities

A. Phenomena Identification and Ranking Table Technique

As with many of the methods associated with a rigorous, evidence-based approach to assessing the validity of modeling and simulation capability, the Phenomena Identification and Ranking Table (PIRT) technique was pioneered by the nuclear reactor safety community [5,6]. The technique was developed to provide a systematic approach to the gathering of information from Subject Matter Experts (SMEs) on a specific issue and ranking that information according to its importance to reactor safety in an accident environment. Although originally created with the needs of better understanding safety-critical physical processes, early researchers recognized that the approach could form an important component of any model validation process and it was generalized by Wilson and Boyack [6].

The PIRT technique is a structured expert elicitation process that consists of the following key steps:

(i) Define the issue for which a PIRT is required;
(ii) Define the specific objectives that are to be considered;
(iii) Assemble a team of subject matter experts with a wide diversity of perspectives;
(iv) Define the system, the environment, and the scenario that are to be investigated;
(v) Define the evaluation criterion and the system response quantities used to judge importance;
(vi) Identify, compile, and review the current knowledge base;
(vii) Identify plausible phenomena;
(viii) Develop an importance ranking for the plausible phenomena;
(ix) Assess the current knowledge level regarding each phenomenon;
(x) Document and report the PIRT results.

In the context of PIRT, the term phenomena can refer to the condition of the system (or its associated subsystems and components), physical phenomena related to the operation of the system, physical or engineering approximations made in modeling the system, or anything else that might change the system response quantities of interest.

The results of a PIRT investigation are usually presented in a tabular form [4], such as that shown in Table 1. The basic anatomy of the table is as follows:

(i) Identification of the specific environment and scenario that is under investigation;
(ii) A column for the plausible phenomenon identified and a column for each of the system response quantities under investigation;
(iii) A row for each of the plausible phenomenon identified;
(iv) An indication of the importance of the phenomena for each system response quantity.

Table 1: Phenomenon Identification and Ranking Table.

(a) PIRT Table

Environment A, Scenario B			
Phenomena	**SRQ1**	**SRQ2**	**SRQ3**
A			
B			
C			
D			

(b) Ranking Criteria

Importance	
High	**the phenomenon has a clear influence on the SRQ**
Moderate	**the phenomenon has some influence on the SRQ**
Low	**the phenomenon has little or no influence on the SRQ**

B. Hierarchy Element Listing and Prioritization Technique

The 10 key steps mentioned earlier for the PIRT process are shown in Table 2. These steps can be arranged in three major phases. An initial *formation* phase in which the issue to be addressed is identified, terms of reference are created, a team is formed, and the objectives of the team are agreed. This is followed by a second *identification* phase in which the available information is reviewed, and all plausible physical phenomena are identified. The final phase in the process is *prioritization* of the plausible physical processes regarding their expected impact on quantities of interest and then documentation of the results of the PIRT. This grouping of the 10 steps is illustrated in Table 2.

At this level of abstraction, we can see that, in principle, the PIRT technique and the process for creating a validation hierarchy detailed by Luckring et al. [20] have two phases in common, formation and identification. That is, the steps discussed in detail in Luckring et al. [20] serve the same purpose, formulation of the system being analyzed and identification of the elements and physical processes occurring in the system of interest. Given this correspondence between the two techniques, our approach is to replace the formation and identification phases of the PIRT with the corresponding activities employed to create the validation hierarchy. Then the prioritization phase is conducted, but the process is applied to the elements identified in the model validation hierarchy. We name this new approach the Hierarchy Element Listing and Prioritization (HELP) technique.

Table 2: Arrangement of PIRT steps in three major phases.

Phase	Steps
Formation	define the issue for which a PIRT is required; define the specific objectives that are to be considered; assemble a team of knowledgeable individuals and subject matter experts; define the system. the environment and the scenario that are to be investigated; define the evaluation criterion, the system response quantities, used to judge importance.
Identification	identify, compile, and review the current knowledge base; identify plausible phenomena.
Prioritization	develop importance ranking for phenomena; assess the current knowledge level regarding each phenomenon; document and report the PIRT results.

The HELP technique consists of 4 basic steps:

(i) Create (or obtain) a model validation hierarchy for the system;
(ii) Develop importance ranking for each of the elements in the validation hierarchy;
(iii) Assess the current knowledge level regarding each element of the validation hierarchy;
(iv) Document and report the HELP results.

The results of a HELP investigation can be presented in a tabular form, such as that shown in Table 3. The table is analogous to the PIRT (Table 1) and contains the following key information:

(i) Identification of the specific environment and scenario that is under investigation;
(i) A column for the hierarchy element identifiers and for each of the system response quantities of interest;
(ii) A row for each of the hierarchy elements;
(iii) An indication of the importance of the hierarchy element for the system response quantities of interest.

Table 3: Hierarchy Element and Listing Prioritization Table.

(a) HELP Table

Environment A, Scenario B			
Hierarchy Element	**SRQ1**	**SRQ2**	**SRQ3**
A			
B			
C			
D			
E			

(b) Ranking Criteria

Importance	
High	**the element has a clear influence on the SRQ**
Moderate	**the element has some influence on the SRQ**
Low	**the element has little or no influence on the SRQ**
Not Assessed	**the element has not been assessed because an antecedent has been judged to be of low importance**

The qualitative nature of the assessment is such that only a small number of relative ranks should be defined. In Table 3b, four relative ranks are defined: high, moderate, low, and not assessed. The correspondence of the first three ranks with those used in the PIRT table is the same, while the fourth ranking, *not assessed*, is new. The need for this additional ranking is discussed below.

The concentration of the HELP processes is on hierarchy elements rather than plausible physical phenomenon. Above the transition tier, the hierarchy elements deal with the system rather than the detailed physical phenomena that are likely to be of immediate interest in the prioritization process. However, below the transition tier then, there is a closer correspondence between model validation hierarchy elements and the physical phenomenon of the PIRT process.

The inherent structure of a hierarchy, in particular the clear parent-child relationships between elements, allows the quick elimination of many of the elements that appear in lower tiers. This is because having judged that a parent element is of low importance, then it must follow that all its child elements are also of low importance. Thus, without the need for detailed assessment, all the descendants of an element may be marked of low importance. To show this in the HELP table, we introduce the additional assessment rank ‘not assessed,’ which should be interpreted as ‘not assessed’ because an antecedent has been judged to be of low importance.

We add two features to the HELP table to reflect the nature of the validation hierarchy, see Table 4. The first feature is to use indentation to indicate the tiered relationships between elements. This is shown in Table 4 where the hierarchy elements A1, A2, and A3 exist on the same tier within the hierarchy, with element A as their shared antecedent. Element A3 also has child elements on the next lower tier of the hierarchy, and this is represented by further indentation. By making use of this visual cue, we can represent the structure of the hierarchy in the table.

Table 4: Adjustments to the Hierarchy Element and Listing Prioritization Table.

Environment A, Scenario B			
Hierarchy Element	**SRQ1**	**SRQ2**	**SRQ3**
- A			
+ A1			
+ A2			
- A3			
+ A3a			
A3b			
+ B			

Importance
High
Moderate
Low
Not Assessed

The second feature is to indicate whether an element does, or does not, have elements below itself. Including all tiers and all associated elements can make visual presentation of the table unmanageable, particularly when many elements are of low importance for the system response quantities. To reduce this visual clutter, we make use of the notation + to represent elements that have descendants in the hierarchy that are not shown (for example A1 in Table 4). The notation – is used to show that the element descendants are shown in the table (A3 in Table 4). If an element has neither a + nor a –, then it is a leaf in the tree and has no descendants within the hierarchy (A3b in Table 4). By making these simple adjustments to the visual presentation of the HELP table, we retain the hierarchical character of the information and allow easy identification of high-priorities without losing sight of low-priority topics.

In summary, the proposed HELP technique shares the fundamental features of PIRT (Formation-Identification-Prioritization) but replaces the first two steps with a model validation hierarchy. The HELP process requires no changes to the validation hierarchy itself. By performing the initial prioritization at higher levels within the hierarchy, one can quickly eliminate the need to consider lower-level elements that are of limited importance for the system response quantities of interest. This allows easy identification of high priorities without losing sight of low priority topics. In most cases, the elements that are of high priority will be at the lower levels of the hierarchy and will be easily associated with individual physical phenomena.

IV. Prioritization Based on Gaps in Modeling and Simulation Activities

A gap analysis involves the assessment of current performance compared to the desired or needed performance. By establishing the difference between the current and desired performance, one can identify developments that will lead to improvement. Gap analysis is the basis of many common business improvement tools. Examples include companies using gap analysis to compare their performance to competitors, to support benchmarking exercises, and to allocate resources to move into new markets.

During the Advanced Simulation and Computing (ASC) program, researchers at Sandia National Laboratories recognized the power of gap analysis applied to modeling and simulation needed for the nuclear weapons stockpile, see for example Refs. [7,30,31]. Their work emphasized understanding the gap between the capabilities that were needed to adequately represent individual physical phenomena and the existing capability of modeling and simulation to address those needs.

Sandia's gap analysis begins with a completed PIRT. The physical phenomena that are most important for each SRQ of interest are then selected for examination. Next, the capability of the complete software package to represent the physical phenomena are considered against a number of credibility assessment criteria. These criteria commonly include Physics Modeling, Code and Solution Verification, Model Validation, and Uncertainty Quantification and Sensitivity Analysis. Code and solution verification correspond to the potential sources of numerical error, while the others are concerned with model accuracy and predictive capability. The status of the assessment is recorded, usually as adequate, inadequate, or unknown; meaning that information is missing to make an assessment. For a consolidated discussion of the gap analysis procedure, see Ref. [4].

As with the PIRT and HELP techniques, the outcomes of a gap analysis are commonly presented in tabular form, an example gap analysis table is shown in Table 5. The contents of the table are summarized as follows:

(i) Identification of the specific environment, the associated scenario, and the SRQ that is under investigation;
(ii) A column for the plausible phenomenon identified and a column for each of the assessment criteria;
(iii) A row for each of the plausible phenomenon identified;
(iv) An indication of the assessment criteria for each of the phenomena.

In contrast to the PIRT and HELP techniques, where there was only one table for each scenario, a gap analysis is performed independently for each SRQ, and it produces as many tables as there are SRQs of interest for each scenario. The details of the assessment criteria (the columns of Table 5) used in a gap analysis are described below.

Table 5: Gap Analysis Table.

(a) Gap Analysis Table

Environment A, Scenario B, SRQ C					
Phenomena	**Physics Modeling**	**Code Verification**	**Solution Verification**	**Model Validation**	**Uncertainty Quantification and Sensitivity Analysis**
A					
B					
C					
D					

(b) Evaluation Criteria

Status	
Unknown	the measure could not be assessed due to a lack of information
Inadequate	the measure is largely absent from the model. There is weak evidence and/or supporting processes have been poorly executed.
Adequate	the measure is clearly manifest in the model. There is strong evidence and/or supporting processes have been well executed

A. Gap Analysis Process for Modeling and Simulation

The use of an existing validation hierarchy to support gap analysis requires no significant changes to the approach described above. We simply substitute the PIRT with a HELP table and then follow the gap analysis procedure assessing the hierarchy elements. That is, given a completed HELP table with its prioritization indicated, a gap analysis for modeling and simulation activities is conducted on the high importance hierarchy elements.

1. Physics Modeling Criterion

This assessment criterion addresses the suitability of the underlying mathematical model, the data associated with the system of interest, and initial and boundary conditions for the issue at hand. This criterion determines the extent to which the idealization of the problem is adequate for the physical processes being considered. For physics-based models, the appropriateness of the underlying assumptions should be challenged, while for empirical models, we should establish whether the model is being exercised only within its calibrated range. While most of the questions related to this criterion address the mathematical formulation of the problem (the model equations), it should be recognized that in many instances the geometry also undergoes significant idealization, and the impact of such simplifications must be considered. Questions that should be addressed include:

- Does the mathematical model describe the important phenomena over some or all of the parameter space of interest?
- Does the software provide alternative models to describe the phenomenon of interest? Are the material models being used adequate for the question at hand? Are the models physics-based, semiempirical or empirical?
- Is the software flexible enough to support modeling of the geometry to the required fidelity?

2. Code and Solution Verification Criterion

This assessment criterion establishes the extent to which the calculated solution corresponds to the exact solution of the mathematical model equations. We are primarily concerned with two factors. The first, code verification, relates to whether the model equations and numerical algorithms are implemented correctly in software. The second, solution verification, relates to the degree to which the solution of the discretized problem corresponds to the exact solution of the mathematical model, e.g., the partial differential equations.

The objective of code verification is to demonstrate that the implementation in software of the mathematical model is correct. It is the process through which we identify coding mistakes and inconsistencies in the computational algorithms. Code verification is primarily interested in the software development practices used by the developers, as well as the correctness and reliability of the discretization methods employed. Questions that might be asked include:

- Is there sufficient unit and component test coverage?
- What is the observed order of accuracy of the software when numerical solutions are compared with exact solutions of the mathematical model?

Determining the status of code verification may be problematic for users of commercial-off-the-shelf software packages as they may not have access to the detailed information required to make a robust judgement.

Unlike code verification, which should be performed for each option of the software that is exercised for the application at hand, solution verification can, and should, be performed for every solution. In assessing the status of solution verification, we are concerned with the numerical error of the discretized solution relative to the exact solution of the continuous model. Since the exact solution of essentially every practical problem is not known, the process of solution verification is concerned with the estimation of numerical errors due to discretization error, computer round-off error, and iterative solution convergence error.

Iterative methods are generally required for complex nonlinear systems of equations and even for large sparse systems of linear equations. Iterative convergence errors occur due to incomplete iterative convergence of the discrete system. Iterative convergence can be measured using many different metrics, for example changes in residual of individual equations, or residuals of the system response quantities of interest.

Discretization error measures the degree to which the numerical solution of the discretized problem satisfies the continuum integro-differential equations. While there are techniques to estimate this error a priori, they are heavily problem dependent [13], and consequently most successful approaches deal with a posteriori estimation of this source of error. Extrapolation based approaches, for example Richardson extrapolation [4, 23], usually rely on two or more solutions that have been obtained with different discretization step sizes (h-type) or with different order of accuracy (p-type). Such methods usually provide both an estimate of the continuum solution and an associated uncertainty.

3. Validation Criterion

Model validation can be defined as the process through which one demonstrates the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model [2]. Unfortunately, this definition, in particular the phrase ‘from the intended uses of the model,’ is open to interpretation by those who have a limited background in computational modeling and simulation. In clarifying the definition, Oberkampf and

Trucano [33] identified three different aspects of model validation. The first aspect relates to quantifying the accuracy of numerical simulations with respect to a physical experiment for clearly identified SRQs. The second aspect relates to estimating the uncertainty of the predicted SRQs for the application conditions of interest, since most conditions of interest do *not* have experimental data available. The last aspect relates to whether the model accuracy, including the estimated uncertainty, is sufficient to meet the requirements for the decision at hand. Oberkampf and Trucano [33] refer to model validation processes that consider only the first of these aspects as a restricted view of model validation, while processes that account for all three aspects are referred to as the encompassing view of model validation. In this report, we are concerned with the restricted view of model validation and any subsequent use of the term model validation should be read as "restricted view of model validation."

Model accuracy is quantitatively estimated by estimating the differences between the computational results and experimental measurements for each SRQ of interest as a function of the input or control parameters. In assessing model accuracy, we need to address issues around:

(i) The nature of the SRQ and how it has been measured in the experiment;
(ii) The sparsity of experimental data with respect to the model input parameter space;
(iii) The sparsity of computational data with respect to the model input parameter space;
(iv) How uncertainties in both the experimental and computational data are handled;
(v) Whether the measured data are deterministic or probabilistic, i.e., a set of repeated measurements;
(vi) Whether the simulation is probabilistic, i.e., including model input and numerical uncertainty, or deterministic.

The availability of experimental and computational uncertainties and how they are handled has a significant bearing on the nature and robustness of the comparison. The simplest approach is visualization, e.g., of contour plots. Such comparisons are often entirely qualitative in nature with little or no quantitative information. As such, judgements of accuracy are generally subjective and entirely dependent on the judgement of an SME. The most common approach is direct comparison of the experimental data with measurements extracted from a single computation. While such comparisons are quantitative, they take no account of variability or uncertainty in either the experiment or the computation. At the opposite end of the spectrum are comparisons that account for uncertainty in the experimental measurements and numerical errors and that deal with variability in the experiment using nondeterministic simulations.

Finally, it is important to remember that the emphasis of model validation is not on identifying whether the model is right or wrong, but instead on estimating the accuracy of the model for the conditions of the experiment conducted. After all, a model that can reliably predict the SRQ within the accuracy requirements of the decision to be made can still be useful, even if it is far from perfect.

4. Uncertainty Quantification and Sensitivity Analysis Criterion

Most computations performed with an engineering simulation model are predictions for which no corresponding physical experiment is available for comparison. The estimated accuracy of these predictions must be inferred from the model validation database, as well as all the contributing uncertainties, such as numerical solution error. The final criterion used in the gap analysis process assesses this predictive capability, that is "the use of a computational model to foretell the state of a physical system under conditions for which the computational model has not been validated [2]." This perspective addresses the second aspect of the encompassing view of validation discussed above, i.e., the estimated total uncertainty of the simulation for the application conditions of interest.

There are several ways in which the accuracy of model predictions can be inferred from the comparisons made between simulation and experiment for conditions of the experiments. The first is to assume that the accuracy of the predictions observed for the experiments conducted are comparable to those for the actual system of interest. Even if the experimental conditions are similar to the actual system of interest, there are uncertainties that exist in the system of interest that are not represented in the experiments. Some examples are variability in the system of interest due to manufacturing and assembly of the actual system, a wide range of the operating conditions of the actual system as opposed to those existing in the experiment, and coupled physics that occurs in the actual system as opposed to that occurring in the experiment. As a result, assuming that the accuracy of the simulation is the same as the experimental conditions commonly leads to misinforming decision makers.

The approach of using nondeterministic simulations has been shown over decades to be a more robust approach. There are 4 key steps in a nondeterministic simulation [4]:

(i) Identify and characterize all sources of uncertainty;

(ii) Estimate the numerical error for each SRQ for the application conditions of interest;
(iii) Estimate the total uncertainty for each SRQ for the application conditions of interest;
(iv) Perform a sensitivity analysis to identify the major sources of uncertainty in predictions.

In assessing this criterion, we must therefore judge whether all the contributing uncertainties are properly characterized for decision making, in addition to judging whether the tools and expertise exist to adequately address each of these steps. Further detailed discussion of all 4 criteria can be found in Ref. [4].

B. Some Practical Observations on Performing a Gap Analysis for Modeling and Simulation

In principle, performing a gap analysis is straightforward. One need only compare the modeling and simulation needs of relevant physical phenomenon relative to the credibility criteria of the specific M&S capability discussed above. However, this apparent simplicity masks the practical problems that are faced performing such an assessment.

Perhaps the most challenging difficulty faced in performing a successful review is assembling a team that has the knowledge and experience to assess the capability against the credibility criteria. Conducting a gap analysis for the five credibility criteria requires a great deal of knowledge about the existing capability. This knowledge touches on the fundamental physical modeling, the mathematical details of the discretization scheme and numerical methods employed, the programming of the methods in software, and practical aspects of the intended application. This knowledge is unlikely to reside in a single analyst. Indeed, with the growing complexity of M&S software and increasing reliance on off-the-shelf software, there is a growing separation between the developer of the software and the analyst who uses it. This separation means that it is not only unlikely that an individual can perform the task but also that the complete knowledge base required to perform the exercise may not even reside within an individual organization. For example, if a commercial software package is used for a given element of the hierarchy, then the software company experts may also be needed for the gap analysis.

Assessment of some of the credibility criteria are needed regardless of the SRQ that is specified in the HELP table. For example, no matter what SRQ is being assessed, solution verification capabilities, uncertainty quantification and sensitivity analysis are always required.

Some SRQ specified in the HELP table depend very heavily on specific credibility criteria. For example, the type of physics modeling needed for a particular SRQ may be very different from what is needed for another SRQ. In the example of the following section, moving the focus from aerodynamic loading to aerodynamic heating through an airframe structure would not only change the HELP table fundamentally, but might also require changes to the team membership.

Assessing the code verification status of a particular software package depends very heavily on the modeling options that have been selected. It is important that, where it is possible, the verification status of the individual options that are used for the SRQ of interest are established rather than simply relying on a generic assessment of the overall software suite.

Finally, we note that the common result of a gap analysis is that there are many inadequacies across the five computational activities listed. This can be disconcerting when performing a gap analysis for the first time and may raise questions about whether the software should be employed at all. However, it is important to recognize that a gap analysis makes no judgement as to whether a capability is suitable or adequate for its intended application; it merely helps to understand the strengths and weaknesses of the capability. Provided steps are taken to mitigate the gaps, either by addressing them directly or putting in place appropriate risk management, the capability may be suitable for its intended application even in the presence of many inadequacies.

V. Prioritization Based on Global Sensitivity Analysis

If multiple gaps are discovered to exist, both Inadequate and Unknown gaps, a rational actor will tend to address those that provide the largest net benefit. That is, a procedure should be used to prioritize future work on the gaps that maximize the difference between the benefits of a gap reduction and the associated costs of reducing the size of the gap. Initially, the idea of employing cost-benefit as the basis of a ranking of possible corrective actions seems straight forward, but there are important conceptual details that require further clarification. Such details include defining precisely what we mean by cost and benefit and most importantly, how to measure each of them.

Paez, Paez, and Hasselman [34] have addressed some of these details in the context of the Sandia 2014 V&V challenge problem [35]. They developed an economic analysis of modeling and simulation to help understand which of three alternatives (no testing and no modeling, testing only, and modeling only) was most economically viable. By testing, they mean additional experimental measurements of the system of interest. The measurements from the testing would be used for improved test-based decision making, but these measurements would *not* be used for improvement in M&S. The context of the present cost-benefit analysis is different. Here, we are interested in estimating the cost of closing gaps through improved M&S activities (physical modeling and experiments, code and solution verification, model validation and experiments, and uncertainty quantification and sensitivity analysis) and comparing those costs with the benefits of improved predictive capability of M&S.

When assessing the benefit of an improvement to the M&S capability, the importance of establishing the aleatory and epistemic uncertainties associated with the model cannot be overstated. Aleatory uncertainty, also referred to as variability or randomness, is uncertainty due to the result of a random process. Epistemic uncertainty, also referred to as lack of knowledge, is uncertainty where the only feature known is the bounds on the uncertainty. Aleatory uncertainty is characterized as a random variable and epistemic uncertainty is characterized as an interval-valued uncertainty. Without this characterization, it is impossible to reliably assess the uncertainties associated with individual predictions. Consequently, whilst we can identify that a gap exists, we cannot accurately quantify the size, nor can we assign any likelihood to the reduction of the uncertainty over the magnitude of the gap.

As an example of an epistemic uncertainty gap, consider the uncertainty due to turbulence modeling of a flow field. If multiple turbulence model options are available in the software package used, it may be relatively simple to exercise various models to estimate the cost benefit of the physics modeling gap on the SRQs of interest. However, it is difficult, if not impossible, to know which of these models is more accurate than another regarding modeling uncertainty. Furthermore, no likelihood can be claimed how much better one model is than another, or that *any* of the models is correct.

When gaps are reduced at the lower levels of the model validation hierarchy, the impact on the next higher level of the hierarchy is difficult to assess. It is increasingly difficult to assess the impact of the gap reduction at even higher levels of the hierarchy. For example, how should one relate improvements in an aerodynamics model that is coupled to a structural dynamics model at a higher level in the hierarchy? It should be clear that these conceptual issues regarding the combined characterization of aleatory and epistemic uncertainties, as well as their propagation upward through the validation hierarchy, is a difficult research topic. In the near term, however, the most promising approach is to characterize all epistemic uncertainties as uniform random variables over estimated possible ranges and use traditional global sensitivity analysis techniques.

A. Global Sensitivity Analysis

Predicting the performance of a complete multidisciplinary system inevitably requires the aggregation of the various sources of uncertainty associated with individual elements of the system model. To assess the overall uncertainty in an SRQ, we need not only a model of the overall system but also the means to propagate uncertainties originating from the outputs of individual submodels through the complete model.

Global sensitivity analysis is “the study of how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input [9].” Global sensitivity analysis should be contrasted with local sensitivity analysis, where the primary concern is in understanding the local relationship between model inputs and outputs. In other words, global sensitivity is concerned with understanding the variance in model outputs in terms of all contributing uncertainties; all characterized as random variables. Global sensitivity analysis provides a rigorous mathematical basis with which to address questions such as “Which factors are most important in determining uncertainty in the model output?” and “Which input uncertainty should we tackle to reduce variance of the output the most?” A detailed description of global sensitivity analysis (GSA) is beyond the scope of the present work. The interested reader is referred to the introductory text of Saltelli [9] and applications such as that of Jiang [12] and Xiao [36]. Here we only highlight the main steps:

(i) Formulate the goals of the analysis. In the present context, the goal of the analysis is to establish the potential benefit of closing a gap for specific SRQs for which the gap analysis is performed.

(ii) Identify all relevant sources of uncertainty. For the purposes of a sensitivity analysis, we are primarily concerned with identifying which sources can be considered as model input uncertainties, model form uncertainties, numerical solution uncertainties, and those that are considered as deterministic, i.e., fixed for a given analysis.

(iii) Characterize model input uncertainties. Model input uncertainties generally belong to two groups, system input data and system surroundings input data. System input data include aspects such as the geometry, initial conditions, and physical modeling parameters, such as thermal conductivity and Young’s modulus. Surroundings input data include boundary conditions and external excitation of the system.

(iv) Characterize numerical solution uncertainties. This aspect relates to the solution verification issues, such as mesh resolution error and iterative error. Each of the sources of uncertainty is a numerical error, but we choose to characterize them as random variables because their exact values are seldom known in practice. Note that the GSA procedure discussed here ignores the impact of code verification errors, such as software bugs and errors in numerical solution procedures that cause a numerical solution to converge to an incorrect solution.

(v) Characterize model form uncertainties. This aspect relates to the model validation status of the M&S capability and is caused by uncertainties in the underlying model assumptions and approximations rather than model inputs. Model form uncertainty is difficult to treat mathematically because it is not a random behavior, instead exhibiting similar characteristics to those of systematic bias uncertainties. However, this behavior must be characterized as a random variable in order to use traditional GSA techniques. As a result, model form uncertainty is characterized as a uniform probability density function over a range of the possible estimated impact of the model form uncertainty.

(vi) Estimate the total uncertainty in the SRQ. If propagation through only one mathematical model is considered, then this step includes all the uncertainties identified above. If propagation through multiple models is considered, such as propagation through higher levels of the model validation hierarchy, then the sources of uncertainty include not only multiple model form uncertainties, but also the uncertainties associated with coupling of those models, such as coupled physics.

The GSA yields a prioritized list of the most important uncertainty contributors for each of the SRQs of interest. For an SRQ at a given level in the validation hierarchy, the prioritized uncertainties could come from any source at the present level, in addition to any source at any level below the present level. To answer the question “What is the potential impact of closing a gap on our ability to predict SRQs?” with GSA, we see that we need only change the uncertainties associated with the individual gaps in question in anticipation of our improvements. Then, evaluating the GSA with the original uncertainties and the anticipated reduction in uncertainties allows us to produce a measure of the expected improvement with respect to the SRQ of interest.

Global sensitivity analysis for a complete system, as discussed above, is rarely practiced for two main reasons. First is the difficulty of obtaining a complete model of the system in a form amenable to the propagation of uncertainties. This is especially true early in conceptual design of a new system. One approach to reduce this difficulty is to employ simple full-system models, such as lumped-mass models or control volume models. Second, for complex systems, it is well recognized that the model form uncertainty, particularly the coupling of subsystems and subassemblies, is a dominant contributor to the total uncertainty of the SRQs at the top level of the model validation hierarchy. The magnitude of model form uncertainty is commonly underestimated, especially in fields like structural dynamics of complex structures that are built from many substructures and components.

Recent developments using imprecise probability theory offer an opportunity to directly characterize epistemic uncertainties in a more flexible mathematical structure. For example, probability bounds analysis allows an epistemic uncertainty to be characterized as a pure interval, without any assumed probability distribution [4]. More capable methods to characterize epistemic uncertainty in imprecise probability theory are also available [37], [38]. However, the mathematical procedures and the software packages for conducting GSA using imprecise probabilities are not well developed.

VI. Concluding Remarks

In this paper, we have presented an approach to prioritizing validation experiments that provides clear traceability between the needs of systems engineers and the concerns of modeling and simulation practitioners.

The foundation of our approach is the construction of a model validation hierarchy. Our approach to constructing validation hierarchies provides a systematic means for the organization and deconstruction of complex physical systems and physical phenomenon. The validation hierarchy is constructed for a particular operational environment and scenario of the system of interest, for example, an aircraft or spacecraft in a normal operating environment with all subsystems functioning properly.

To achieve an effective prioritization of physical phenomena, we describe a new approach, which is referred to as the Hierarchy Element Listing and Prioritization (HELP) technique. Our approach draws heavily on the ideas and philosophy that underpin the existing Phenomena Identification and Ranking Table (PIRT) technique. Using a validation hierarchy of the system of interest, the HELP technique consists of three main steps. The first step is to create an importance ranking for each of the elements in the validation hierarchy by considering the degree to which the physical phenomenon associated with the element influences specific response quantities of interest. This importance ranking is qualitative in nature and so only a small number of relative ranks are defined. Due to the inherent structure of a validation hierarchy, some elements have antecedents that are found to be of little or no influence and can be ignored in the prioritization process. In the second step, the current knowledge level regarding each element of the validation hierarchy is assessed. In the third step, the results of the assessment are documented. The results of the prioritization can be presented visually in a simple table, with entries in the table colored according to their relative importance. While the HELP process has many similarities with PIRT, we note that the key difference lies in the concentration on hierarchy elements rather than plausible physical phenomenon.

In a second step, the outputs of the HELP prioritization process are used to perform a modeling and simulation gap analysis. The gap analysis technique addresses the question “Where does the existing capability for five attributes of simulation confidence stand relative to the phenomenon and SRQs that have been identified as important?” The gap analysis is structured around the following attributes of simulation credibility; physics modeling, code verification, solution verification, model validation, and uncertainty quantification and sensitivity analysis.

A challenge often faced when performing gap analysis is that many gaps are identified and so in order to complete the prioritization, we propose a final prioritization step in which a global sensitivity analysis is performed to estimate the likely improvement of closing gaps. This additional step allows those validation experiments that will contribute most to increasing confidence and credibility in modeling and simulation results to be identified.

We believe that the combination of a validation hierarchy, the HELP technique, gap analysis, and sensitivity analysis can serve decision makers in relation to many aspects of modeling and simulation capability. The approach is constructive, rigorous and traceable, and provides a firm foundation for simulation-informed decision making, beginning with a system perspective and progressively moving to a computational modeling and simulation perspective.

Acknowledgments

The work described in this paper was performed within the framework of the NATO STO AVT-297 activity Development of a Framework for Validation of Computational Tools for Analysis of Air and Sea Vehicles. The authors would like to acknowledge the support, encouragement and detailed review of the work provided by Dr. John A. Benek, Air Force Research Laboratory.

References

[1] Office of the Deputy Assistant Secretary of Defense Systems Engineering, "Department of Defense, Digital Engineering Strategy," June 2018. [Online]. Available: URL: http://ac.cto.mil/digital_engineering/. [Accessed 24 April 2022].
[2] AIAA, "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations," 1998.
[3] P. J. Roache, "Validation in Fluid Dynamics and Related Fields," in Computer Simulation Validation , Springer, 2019, pp. 661-83.
[4] W. L. Oberkampf and C. J. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, 2010.
[5] B. Boyack, I. Catton, R. Duffey, P. Griffith, K. Katsma, G. Lellouche, S. Levy, U. Rohatgi, G. Wilson, W. Wulff and N. Zuber, "Quantifying Reactor Safety Margins, Part 1: An overview of the code scaling, applicability, and uncertainty evaluation methodology," Nuclear Engineering and Design, vol. 119, pp. 1-15, 1990.
[6] G. Wilson and B. Boyack, "The role of the PIRT in experiment, code development and code application associated with reactor safety assessment’," Nuclear Engineering and Design, no. 186, pp. 23-37, 1998.
[7] T. Trucano, M. Pilch and W. Oberkampf, "General concepts for experimental validation of the ASCI Applications," SAND2002-0341, Sandia National Laboratories, 2002.
[8] J. Yourko and . J. Buongiorno, "Quantitative Phenomena Identification and Ranking Table (QPIRT) for Reactor Safety Analysis," Transactions of the American Nuclear Society, vol. 104, 2014.
[9] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, Global Sensitivity Analysis: A Primer, John Wiley & Sons, 2008.
[10] J. Guo and X. Du, "Sensitivity Analysis with Mixture of Epistemic and Aleatory Uncertainties," AIAA J., vol. 45, no. 9, 2007.

[11] J. Mullins and S. Mahadevan, "Bayesian Uncertainty Integration for Model Calibration, Validation and Prediction," Journal of Verification, Validation and Uncertainty Quantification, vol. 1, 2016.
[12] Z. Jiang, W. Chen and B. German , "Multidisciplinary Statistical Sensitivity Analysis Considering Both Aleatory and Epistemic Uncertainties," AIAA Journal, vol. 54, no. 4, pp. 1326-1338, 2016.
[13] T. Augustin, F. Coolen, G. de Cooman and M. Troffaes, Introduction to Imprecise Probabilities, Chichester, UK: Wiley & Sons, 2014.
[14] A. Bernardini and F. Tonon, Bounding Uncertainty in Civil Engineering, Berlin: Springer-Verlag, 2010.
[15] D. Allaire, H. Qinxian, J. Deyst and K. Wilcox, "An Information-Theoretic Metric of System Complexity with Application to Engineering System Design," J. of Mechanical Design, vol. 134, 2012.
[16] M. Olsen and M. Raunak, "A Framework for Simulation Validation Coverage," in Proceedings of the 2013 Winter Simulation Conference, 2013.
[17] R. Hällqvist, M. Eek, I. Lind and H. Gavel, "Validation Techniques Applied on the Saab Gripen Fighter Environmental Control System Model," in Proceedings of the 56th SIMS, Linkoeping, 2015.
[18] R. Hällqvist, R. Braun, M. Eek and P. Krus, "Optimal Selection of Model Validation Experiments: Guided by Coverage," Journal of Verification Validation and Uncertainty, vol. 6, no. 3, 2021.
[19] M. Olsen and M. S. Raunak, "Quantitative Measurements of Model Credibility," in Model Engineering for Simulation, 2019, pp. 163-187.
[20] J. Luckring, S. Shaw, W. Oberkampf and R. Graves, "Development of a Validation Hierarchy for Modeling and Simulation," STO-TR-AVT-297, NATO, 2022.
[21] IEEE, "Guide for Software Verification and Validation Plans," 1994.
[22] Department of Defense (DoD), "VV&A Recommended Practices Guide," 2011. [Online]. Available: https://vva.msco.mil. [Accessed 4 12 2020].
[23] P. Roache, Fundamentals of Verification and Validation, Socorro, New Mexico: Hermoza Publishers, 2009.
[24] R. G. Bradely, "CFD Validation Philosophy," in AGARD CP-437, 1988.
[25] J. G. Marvin, "Accuracy Requirements and Benchmark Experiments for CFD Validation," in AGARD CP-437, 1988.
[26] W. L. Oberkampf, "What are Validation Experiments," Experimental Techniques, vol. 25, no. 3, pp. 35-40, May/June 2001.
[27] NATO STO, "Progress and Challenges in Validation Testing for Computational Fluid Dynamics," STO-MP-AVT-246, 2016.
[28] C. L. Rumsey, D. H. Neuhart and M. A. Kegerise, "The NASA Juncture Flow Experiment: Goals, Progress and Preliminary Testing," in *AIAA Paper 2016-1557*, 2016.
[29] M. A. Kegerise, D. H. Neuhart, J. A. Hannon and C. L. Rumsey, "An Experimental Investigation of a Wing-Fuselage Junction Model in the NASA Langley 14- by 22-Foot Subsonic Wind Tunnel," in *AIAA Paper 2019-0077*, 2019.
[30] K. Geelhood and W. Luscher, "Degradation and failure phenomena of accident tolerant fuel concepts: Chromium Coated Zirconium Ally Cladding," PNNL-28437, Pacific Northwest National Laboratory, Richland, 2019.
[31] S. Tieszen, T. Chu, D. Dobranich, V. Romero, T. Trucano, J. Nakos, W. Moffat, T. Hednrickson, S. Sobolik, S. Kempka and M. Pilch, "Integrated modeling and simulation validation plan for W76-1 Abnormal thermal environment qualification – Version 1.0 (OUO)," SAND2002-1740 (OUO), Sandia National Laboratories., 2002.
[32] B. Boughton, V. Romero, S. Tieszen and S. Sobolik, "Integrated modeling and simulation validation plan for W80-3 Abnormal thermal environment qualification – Version 1.0 (OUO)," SAND2003-4152 (OUO), Sandia National Laboratories., 2003.
[33] W. L. Oberkampf and T. G. Trucano, "Verification and Validation in Computational Fluid Dynamics," *Progress in Aerospace Sciences*, vol. 38, 2002.
[34] P. Paez, T. Paez and T. Hasselman, "Economic Analysis of Model Validation for a Challenge Problem," Journal of Verification, Validation and Uncertainty Quantification, vol. 1, no. 1, 2016.
[35] K. Hu and G. Orient, "The 2014 Sandia Verification and Validation Challenge: Problem Statement," Journal of Verification Validation Uncertainty Qualification, vol. 1, no. 1, 2016.
[36] H. Xiao, J.-L. Wu, J. Wang, R. Sun and C. Roy, "Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach," *Journal of Computational Physics*, vol. 324, pp. 115-136, 2016.
[37] A. Bernardini and F. Tonon, Bounding Uncertainty in Civil Engineering, Berlin: Springer-Verlag, 2010.
[38] T. Augustin, F. Coolen, G. de Cooman and M. Troffaes, Introduction to Imprecise Probabilities, Chichester, UK: Wiley & Sons, 2014.