

Experimental Credibility and its Role in Model Validation and Decision Making

Sarah L. Kieweg
Principal Member of the Technical Staff
Verification & Validation, UQ, and Credibility Processes
Sandia National Laboratories
PO Box 5800, Mail Stop 0828
Albuquerque, NM 87185-0828

Walt R. Witkowski
Manager
Verification & Validation, UQ, and Credibility Processes
Sandia National Laboratories
PO Box 5800, Mail Stop 0828
Albuquerque, NM 87185-0828

ABSTRACT

Experiments are a critical part of the model validation process, and the credibility of the resulting simulations are themselves dependent on the credibility of the experiments. The impact of experimental credibility on model validation occurs at several points through the model validation and uncertainty quantification (MVUQ) process. Many aspects of experiments involved in the development and verification and validation (V&V) of computational simulations will impact the overall simulation credibility. In this document, we define experimental credibility in the context of model validation and decision making. We summarize possible elements for evaluating experimental credibility, sometimes drawing from existing and preliminary frameworks developed for evaluation of computational simulation credibility. The proposed framework is an expert elicitation tool for planning, assessing, and communicating the completeness and correctness of an experiment (“test”) in the context of its intended use—validation. The goals of the assessment are (1) to encourage early communication and planning between the experimentalist, computational analyst, and customer, and (2) the communication of experimental credibility. This assessment tool could also be used to decide between potential existing data sets to be used for validation. The evidence and story of experimental credibility will support the communication of overall simulation credibility.

KEYWORDS: credibility, experimental credibility, data quality, model validation, decision making,

INTRODUCTION

Experiments are a critical part of the model validation process, and the credibility of the resulting simulations are themselves dependent on the credibility of the experiments. A model calibration activity or a validation assessment is only as good as the experimental data. In particular, experimental uncertainty plays a key role in a validation assessment and the determination of model form uncertainty. The American Society of Mechanical Engineers (ASME) V&V guidelines [1, 2], standards [3], and the Coleman and Steele textbook [4] provide excellent resources for how to characterize experimental uncertainty and include it in the validation assessment activities.

The term credibility in the context of model validation and decision-making is not a universally-defined or accepted term. Credibility may encompass the trust a user or decision-maker has in the computational simulation outcomes. It may include an assessment of the correctness or completeness of the MVUQ process, such as the assessment made in a Predictive Capability Maturity Model (PCMM) [5, 6]. A statement of credibility may include both the outcomes as well as the uncertainty in

predictions from a computational simulation. Credibility ideally includes evidence that is communicated in a convincing manner.

The impact of experimental credibility on model validation occurs at several points through the MVUQ process: calibration and uncertainty of model parameters in independent tests (e.g. material properties), measurement and uncertainty of experimental conditions (and thus initial/boundary conditions of a simulation), and measurement and uncertainty of experimental outcomes (used in metrics for comparison with simulation). These last two can apply to both calibration tests (of either unit, subassembly, or full system) and validation tests. In some cases, experiments also play a role in the specification of environmental conditions or the design of the experiments.

In this document, we present attributes of experimental credibility and the scope of experimental credibility in the context of *model validation* and decision making. The focus here is on a class of tests involving large systems or components for validation assessments. Similar frameworks may be proposed in the future for other classes of tests, such as separate effects tests with greater replication or material property tests for calibration. Likewise, tests with a sole use in certification without use in computational simulation are not in the scope of this document and framework. We summarize possible elements for evaluating experimental credibility, drawing from existing and preliminary frameworks developed for evaluation of computational simulation credibility. Finally, we prioritize and categorize the possible credibility elements into a proposed assessment tool.

APPROACH

What is Experimental Credibility and Why is it Important?

For computational simulation credibility, Sandia uses a Predictive Capability Maturity Model (PCMM) [5, 6] to assess the completeness and rigor of the modeling and simulation approaches. An analogous widely accepted model/framework does not exist for validation experiments either at Sandia or in published literature. Previous efforts by others at Sandia and outside Sandia have proposed versions of such assessment frameworks. The previous frameworks differ significantly in possible elements to consider in the assessment and in their context of use. This project seeks to examine previously-proposed and other possible frameworks, evaluate their applicability, and propose an experimental assessment framework with appropriate elements.

Many aspects of experiments (e.g. the planning, UQ, analysis, and documentation) used in the development and V&V of computational simulations will impact the overall simulation credibility. These aspects fall into categories of: experimental planning, experimental UQ, processing of test data, use of data for model calibration/validation, documentation, and archiving. In this document, this collection of experimental aspects will be labeled “experimental credibility for computational simulations.” *Experimental credibility is the assessment of the both the correctness and completeness of these experimental aspects and the process as a whole, and an assessment of the experiment’s use with its stated purpose – in this case, model validation. The assessment focuses on the correctness, completeness, and applicability of these experimental aspects and use.* Because experiments are used throughout the development and MVUQ of computational simulations, the experimental credibility has a strong impact on the overall simulation credibility. In particular, the completeness of the quantification and documentation of experimental uncertainty will impact the quality and use of the validation metric. In addition, an incomplete understanding and documentation of the uncertainty in the experimental conditions will impact the comparison to the simulation.

The overall goal is to provide a framework that is a structured method for assessing the level of correctness, completeness, and applicability of the experimental conditions and measurements that are intended to contribute to model validation. A similar framework (the PCMM) exists for assessing the level of maturity of computational simulation elements [6] [5].

Often, institutions may have existing requirements for testing used to qualify or evaluate a material, component, or system. However, one common misconception is that a test that meets a test plan’s specification by the customer is adequate for MVUQ purposes. This may not always be true, and in fact, without direct involvement of the analyst in the specification and planning of a test, often tests are not designed and specified for validation purposes alone. Thus, it is imperative to create a tool that helps in the planning and assessment of experiments that are intended to contribute to MVUQ processes.

PCMM and other PCMM-like Frameworks

In 2002, Trucano *et al.* published a SAND report [7] which provided many best practices and general concepts for experimental validation of ASCI code applications. Much of these general concepts eventually made it into the PCMM in 2007. In 2007, a SAND report documented the PCMM – the Predictive Capability Maturity Model – for computational simulation. [6] It was based on a maturity model framework concept previously used to evaluate the maturity or quality of a technological process. The report provides a historical summary of these frameworks and the many reasons they are useful. The original PCMM had six elements of computational simulations: (1) representation and geometric fidelity, (2) physics and material model fidelity, (3) code verification, (4) solution verification, (5) model validation, and (6) uncertainty quantification and sensitivity analysis. In 2013, a 4th generation PCMM was proposed [8]. The six original elements were expanded into nine, each with included sub-elements. In particular, the revised PCMM included two elements which explicitly evaluate the experimental data for both constitutive model calibration and overall model validation. Since then, others at Sandia have discussed other possible frameworks focused solely on assessing the experimental data itself, for use in validation.

In 2014, Oberkampf and Smith developed assessment criteria for evaluating CFD validation benchmark experiments. [9] Each element included descriptors for completeness level, in the spirit of the PCMM. While intended for CFD experiments, the elements have some broader applicability. Lance and Smith used that assessment to guide the development and documentation of benchmark experiments. [10, 11] Elele *et al.* developed a risk assessment matrix for tests and facilities based on an assessment for modeling and simulation. [12] Each element was assessed for risk, instead of level of maturity. The elements were divided into categories of capability, accuracy, and usability. Notably, that assessment included elements of intended use and usability.

Summary of Potential Assessment Elements

The following lists of possible attributes, or elements, for assessment of experimental credibility captures most of the elements that appeared in the summarized frameworks in the previous section. It also reflects some of the best practices for validation tests [7] [6]. These elements will be prioritized and categorized in the final proposed assessment tool in the section below. It is possible to group these with different affinities – the proposed element affinity is just one possible framework. In future work, this proposed framework will be demonstrated and evaluated with a set of experimental-computational partners to evaluate its efficacy for improving planning and communication.

- Planning
- Hardware/material fidelity
- Experimental/environmental fidelity
- Experimental verification: Control code
- Experimental verification: Data analysis code
- Experimental verification: Test facility
- Experimental verification: Evidence test performed correctly
- Experimental validation: Confirmation of measurements
- Intended use: Test conditions
 - Rigor and completeness of characterizing test conditions for use as a validation test.
- Intended use: Test output
 - Applicability and completeness of measured test outputs for purposes of validation assessments.
- Uncertainty Quantification (UQ)
 - Elements of this will be detailed in a sub-assessment tool described in a later section.
 - Applicable to both test conditions and output measurements.
- Peer review of all the above elements
- Documentation

Lessons Learned from the PCMM and PCMM-like Frameworks

The 4th-generation PCMM development project [8] emphasized some key lessons learned that should be applied to a framework for assessing experimental uncertainty and credibility:

- The purpose of the tool should be clear.
- The purpose of the tool is not to “grade” or “score” the work, but rather to provide a tool that facilitates planning and communication. A summary score should not be used – a quantitative score is artificial and misleading.

- The assessment should involve an inclusive team-based SME elicitation process.
- The requirements of the customer should be documented in the evaluation process, and be specific about how the tool and the work will be used to communicate credibility to the customer.

These lessons are applied to the following two proposed frameworks for (1) assessing the correctness, completeness, and applicability of a validation experiment, and (2) assessing experimental uncertainty. The goal is to provide tools that encourage early planning and continual communication. Unlike previous frameworks like the PCMM, this proposed structure does not include any “scoring” activities. Instead, these frameworks include open-ended questions that should encourage discussion of advantages and limitations of each element, and communication of credibility evidence.

PROPOSED FRAMEWORK TO ASSESS VALIDATION EXPERIMENTS

The two primary goals of the proposed assessment tools are to encourage early planning and facilitate communication of credibility of final products. Thus, the format of the assessment has been altered from the format seen in the PCMM in order to promote communication and avoid the graded levels. Rather than use graded maturity levels in these assessment frameworks as was done in the PCMM, descriptors for each element are casted as open-ended questions. Thus, instead of prescriptive levels that dictate what should be done at each level, the frameworks proposed below provide prompts and open-ended questions to facilitate planning and communication.

Rather than simply indicating the maturity level for each element, users are expected to respond with a description addressing the prompts and open-ended questions. The users are asked to consider the strengths and weaknesses of the experiment using each element to guide their assessment. The written responses allow users to differentiate how individual aspects of that element are assessed. The user’s descriptive discussion of evidence for each element can contribute to a written credibility story for the simulation as a whole (i.e. augment the PCMM analysis of the simulation), and aid communication to the customer.

Two frameworks are proposed. First, is an “Assess Validation Experiment” tool (see Table I) for assessing the correctness, completeness, and applicability of a validation experiment. The second is an “Assess Experimental Uncertainty” tool (see Table II). The second tool is used to address the Uncertainty Quantification element of the first tool. The second tool also includes a list of best practices for quantifying experimental uncertainty, most of which were summarized from published pre-test and post-test uncertainty analysis guidelines. [3, 4] For example, best practices include the identification of all sources of uncertainty, e.g. measurement, data analysis, repeatability, and unsteadiness. Guidelines, standards, and textbooks provide methods for characterizing types of uncertainty, quantifying uncertainty, and using experimental uncertainty in the validation assessment. [1-4]

CONCLUSIONS

The proposed frameworks are expert elicitation tools for planning, assessing, and communicating the completeness, correctness, and applicability of an experiment (“test”) and its uncertainty in the context of its intended use. The goals of the assessment are (1) to encourage early communication and planning between the experimentalist, computational analyst, and customer, and (2) the communication of experimental credibility. The evidence and story of experimental credibility will support the communication of overall simulation credibility to the customer. The primary application of this tool is for tests used for validation of simulations. Ideally, such tests are designed specifically for validation, but in practical conditions, often existing tests developed for another purpose will need to be used for validation purposes. Thus, these tools can also be used to evaluate existing data sets for potential use in validation. In future work, this proposed framework will be demonstrated and evaluated with a set of experimental-computational partners to evaluate its efficacy for improving planning and communication.

ACKNOWLEDGEMENTS

The authors thank Martin Pilch, PhD, and Angel Urbina, PhD, for valuable conversations and feedback. This project was funded by the Advance Simulation and Computing (ASC) Program of the Department of Energy’s National Nuclear Security Administration and the Weapons Systems Engineering Assessment Technology (WSEAT) Program at Sandia National Laboratories.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525. SAND2017-#####

REFERENCES

1. ASME V&V 10-2006, *Guide for Verification and Validation in Computational Solid Mechanics*. 2006, New York, NY: ASME.
2. ASME V&V 20-2009, *Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer*. 2009, New York, NY: ASME.
3. ASME Performance Test Codes Supervisory Committee, *ASME PTC 19.1-2013: Test Uncertainty 2013*, New York, NY: ASME.
4. Coleman, H.W. and W.G. Steele, *Experimentation, Validation, and Uncertainty Analysis for Engineers*. 3rd ed. 2009, New York: Wiley.
5. *What is Predictive Capability Maturity Model (PCMM) SAND2016-7399TR (UUR)*. 2016: Sandia National Laboratories.
6. Oberkampf, W.L., M. Pilch, and T.G. Trucano, *Predictive Capability Maturity Model for Computational Modeling and Simulations. SAND2007-5948 (UUR)*. 2007: Sandia National Laboratories.
7. Trucano, T.G., M. Pilch, and W.L. Oberkampf, *General Concepts for Experimental Validation of ASCI Code Applications. SAND2002-0341 (UUR)*. 2002: Sandia National Laboratories.
8. Hills, R.G., et al., *Development of a Fourth Generation Predictive Capability Maturity Model. SAND2013-8051 (UUR)*. 2013: Sandia National Laboratories.
9. Oberkampf, W.L. and B.L. Smith. *Assessment Criteria for Computational Fluid Dynamics Validation Benchmark Experiments*. in *52nd Aerospace Sciences Meeting*. 2014. National Harbor, MD.
10. Lance, B.W., J.R. Harris, and B.L. Smith, *Experimental Validation Benchmark Data for Computational Fluid Dynamics of Mixed Convection on a Vertical Flat Plat*. *ASME Journal of Verification, Validation and Uncertainty Propagation*, 2016. **1**(September 2016): p. 021005.
11. Lance, B.W. and B.L. Smith, *Experimental Validation Benchmark Data for Computational Fluid Dynamics of Transient Convection From Forced to Natural With Flow Reversal on a Vertical Flat Plat*. *ASME Journal of Verification, Validation and Uncertainty Propagation*, 2016. **1**(September 2016): p. 031005.
12. Elele, J., et al., *Applying Modeling and Simulation Verification, validation and Accreditation (VV&A) Techniques to Test and Laboratory Facilities*, in *ASME V&V Conference*. 2012.

Table I: “Assess Validation Experiment” Tool. (CompSim = Computational Simulation”

Assess Validation Experiment	Prompts to Consider	How did these elements impact the strength and weakness of this test for the purpose of CompSim validation?
Element	Prompts to Consider	Assessment Commentary
Planning	<ul style="list-style-type: none"> • Was/is the purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? • Was the test originally intended for validation purposes? • How much communication was there between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Did this create any strengths or weaknesses to the outcomes? • Was CompSim involved in the planning of this experiment, and in what way? 	
Sample / Geometric / Material Fidelity	<ul style="list-style-type: none"> • Was the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and validation assessment? • Do you know the pedigree? • Was there any pre-processing of the sample/material that could impact applicability? 	
Experimental / Environmental Fidelity	<ul style="list-style-type: none"> • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and validation assessment? • What could have been changed to improve the applicability? 	
Experimental Verification	<ul style="list-style-type: none"> • Was the code that controls the testing apparatus verified? • Was the code that post-processes the raw data verified? • Are the test facility and equipment documented well and calibrated? • How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test performed correctly? 	
Intended Use / Validation	<ul style="list-style-type: none"> • Were the test conditions characterized well enough for the intended validation assessment? Where any conditions missing, not well-characterized, or suspicious? • Were the output measurements characterized well enough for the intended use as a validation test? Were enough quantities of interests measured, and were the right ones measured? • Where the validation metrics and criteria specified before the testing, or after? 	
Uncertainty Quantification	<ul style="list-style-type: none"> • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for making the validation assessment? • To assess the uncertainty quantification, use the elements of the "Assess Experimental Uncertainty" framework. 	
Peer Review and Documentation	<ul style="list-style-type: none"> • Which of the above elements of the test were reviewed by subject matter experts? Which elements were not, and of those, which may need further review and why? • Are the above elements of the test all documented? Does the documentation serve the need for making the validation assessment, and helping write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? 	

Table II: “Assess Experimental Uncertainty” Tool. This framework includes a list of best practices for quantifying experimental uncertainty, most of which were summarized from published pre-test and post-test uncertainty analysis guidelines. [3, 4]

Assess Experimental Uncertainty	Read these prompts, discuss with team, and write a response assessment for each element. Use this when assessing the Uncertainty Quantification element of the Assess Validation Experiment tool.	Assess the pros and cons of this experiment in terms of its quantified uncertainty
Element	Prompts to Consider	Best Practices
Pre-test planning:	<ul style="list-style-type: none"> Was there pre-test planning between experimentalist and end-user? Was there discussion on use of data and documentation needs? 	<ul style="list-style-type: none"> Discussion initiated pre-test Decide who will do which parts of data analysis and UQ Agree upon level of documentation on data pedigree and UQ Clearly define end use of experiment
Pre-test: Define measurand(s) needed to obtain QOI(s)	<ul style="list-style-type: none"> Are the Quantities of Interest (QOIs) defined and specified how will be measured and/or quantified? How do measurands relate to QOIs? Require post-processing? 	<ul style="list-style-type: none"> Discussion/activity initiated pre-test Plan to measure range of local and globally integrated quantities Specify and document functional relationship between measurand(s) and final QOI(s), and how data processed and/or reduced. Document other unmeasured quantities used to calculate QOI.
Pre-test: Measurement process and management of uncertainties	<ul style="list-style-type: none"> Is the measurement and calibration process well described? Where expected uncertainties considered in experimental design? 	<ul style="list-style-type: none"> Define test objectives Map measurement parameters and nominal level to what calibrations and instruments will determine each. Identify correlated errors (e.g. measurements that come from same calibration/instrument) Specify required uncertainty for each measurand so that final result has required uncertainty
Pre-test/Post-test: Expected and Estimated Uncertainties	<ul style="list-style-type: none"> Is there an uncertainty inventory for all conditions and measurements? What is missing or a limitation for use of test (e.g. UQ and validation)? 	<ul style="list-style-type: none"> Should be done both pre-test (expected) and post-test For each measurand in test, complete spreadsheet of (expected) uncertainties Consider all possible sources of uncertainty Consider documentation, calibration histories, previous tests with similar instruments, previous uncertainty analyses, expert judgement
Pre-test/Post-test: Uncertainty Propagation and Sensitivity Analysis	<ul style="list-style-type: none"> What uncertainty sources are small compared to others? Which uncertainties are not well characterized and can something be done to improve that? What could be done now or in future to reduce predicted or measured uncertainties? 	<ul style="list-style-type: none"> Propagate estimated/measured measurement uncertainties into the expected/measured range of results for the QOI(s). Pre-test analysis: just based on precision limits; Post-test analysis: both precision and bias. Identify which measurand(s) have greatest impact on uncertainty of result Identify if there is a better measurement technique to use. Communicate between experimentalist and analyst on whether expected result uncertainty will be adequate for intended us. If have multiple tests, repeat calculation of results and then find uncertainty of the result directly, and compare to propagated uncertainties from each measurement; extract info about zeroth and first order replication level analysis (e.g. infer sample-to-sample variability with multiple tests).