SAND2015-20747J

UQ's Role in Modeling and Simulation Planning, Credibility and Assessment Through the Predictive Capability Maturity Model

W. J. Rider, W. R. Witkowski, and Vince Mousseau
Sandia National Laboratories, Albuquerque NM 87185

Summary

The importance of credible, trustworthy numerical simulations is obvious especially when using the results for making high-consequence decisions. Determining the credibility of such numerical predictions is much more difficult and requires a systematic approach to assessing predictive capability, associated uncertainties and overall confidence in the computational simulation process for the intended use of the model. This process begins with an evaluation of the computational modeling of the identified, important physics of the simulation for its intended use. This is commonly done through a Phenomena Identification Ranking Table (PIRT). Then an assessment of the evidence basis supporting the ability to computationally simulate these physics can be performed using various frameworks such as the Predictive Capability Maturity Model (PCMM). Several critical activities follow in the areas of code and solution verification, validation and uncertainty quantification, which will be described in detail in the following sections. The subject matter is introduced for general applications but specifics are given for the failure prediction project.

The first task that must be completed in the verification & validation procedure is to perform a credibility assessment to fully understand the requirements and limitations of the current computational simulation capability for the specific application intended use. The PIRT and PCMM are tools used at SNL to provide a consistent manner to perform such an assessment. Ideally, all stakeholders should be represented and contribute to perform an accurate credibility assessment. PIRTs and PCMMs are both described in brief detail below and the resulting assessments for this project are given.

Phenomena Identification Ranking Table (PIRT)

A PIRT should list all of the important physics phenomena occurring in the physical event being simulated at the specified level of interest [Wilson]. For each individual phenomenon an assessment/ranking must be declared for the "Importance" of this particular phenomenon in the application simulation or, in other words, what is the resulting consequence if the phenomena model is wrong. The rankings are specified as high ("H"), medium ("M") or low ("L"). Then "Adequacy" determinations must be made with respect to how well the phenomena is represented with a mathematical model, how well it is implemented within the simulation code of choice (Sierra/SM for this project), and the level of validation that has been performed for the application space of interest. Once again a ranking is determined for each of these three "Adequacy" areas. For quick visual adequacy analysis, a color-coding scheme is used to identify areas of inadequacy or gaps. Green signifies that the adequacy is acceptable; yellow indicates that the adequacy is marginal and red indicates that the adequacy level is unacceptable and needs to be addressed. The colors are assigned with respect to the “Importance” ranking. If a phenomena’s “Importance” ranking is low, then any level of adequacy is deemed acceptable. If the “Importance” ranking is medium, then acceptable adequacy rankings are medium and high. An adequacy ranking of low produces a yellow square that indicates a marginal adequacy level. When the “Importance” ranking is high, then the only acceptable adequacy level is “high” earning a green square. For a medium adequacy determination a yellow square is used, however, for a low adequacy assessment a red square is used to indicate an unacceptable adequacy level for that particular category.

Table 1: PIRT for simulation of interest

Phenomena	**Consensus**	**Adequacy**		
	Importance	**Math Model**	**Implementation in Code**	**Validation**
Phenomena #1	**H**	**H**	**M**	**L**

For the presented project, the assessment of numerical prediction of important phenomenological aspects of the abnormal fracture problem was evaluated at the start of the project, and presented in **Tabl**. Six different physics phenomena were identified; 5 of them were rated of high importance. Two areas of concern that required attention (shown in red): “Ductile Material Failure” and “Enforcement of Boundary Conditions”. Both of these were of high importance but ranked low on the validation adequacy scale. Since the intents of the project were to better understand failure mechanics modeling and its limitations, validate its applicability for the application being studied and characterize its reliability and usability, the fact that “Ductile material failure” was ranked inadequate was understood and addressed in the scope of the project.

Table 1: Final PIRT for failure predictions for abnormal mechanical environments

Phenomena	**Consensus**	**Adequacy**		
	Importance	**Math Model**	**Implementation in Code**	**Validation**
Large elastic-plastic deformation of metals	**H**	**H**	**M**	**M**

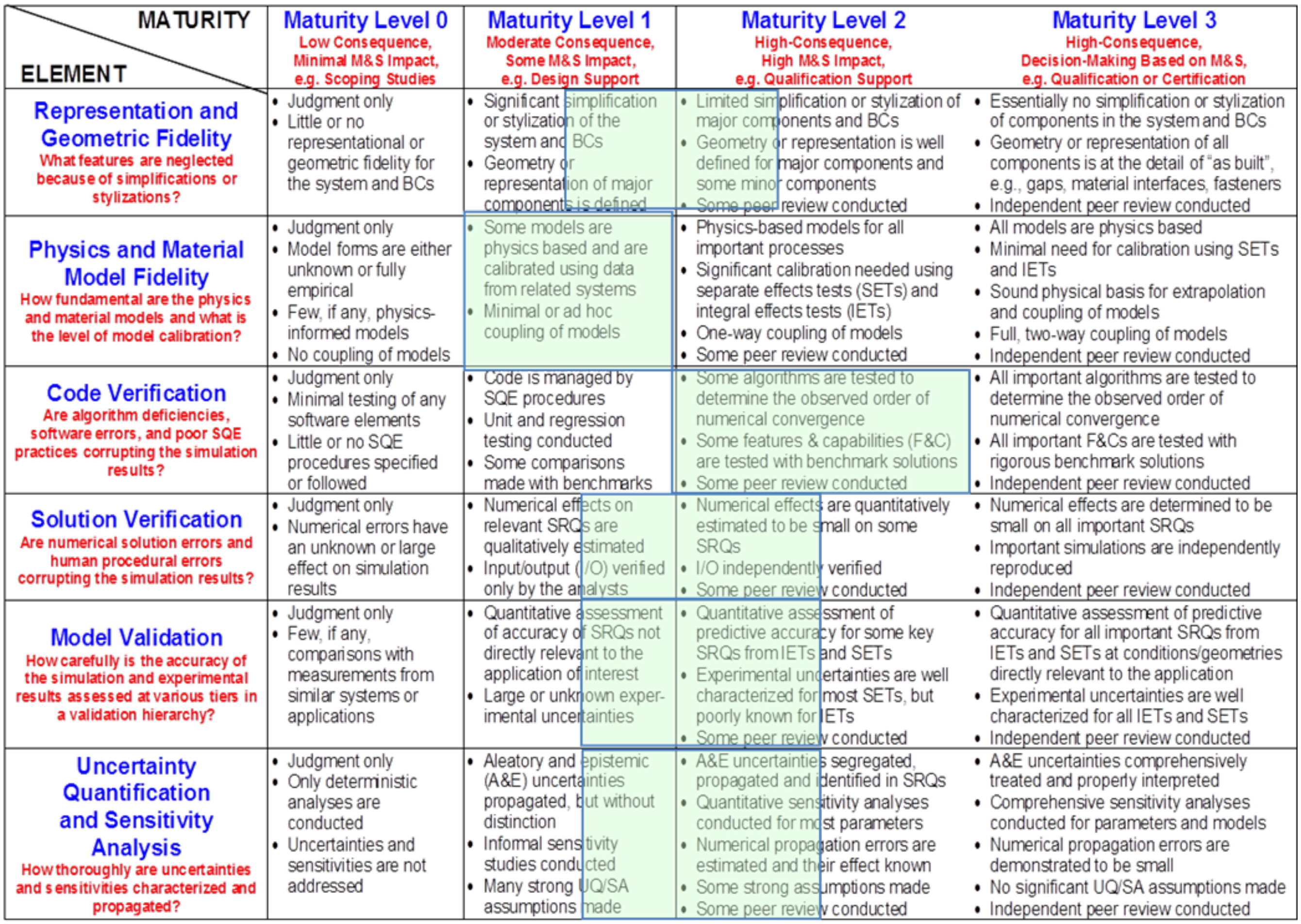
Ductile material failure	**H**	**M**	**M**	**L**
Contact	**H**	**H**	**M**	**M**
Friction between punch and test item	**M**	**M**	**M**	**L**
Enforcement of boundary conditions	**L**	**H**	**H**	**L**
Inertial loads	**H**	**H**	**H**	**M**

The boundary condition assumption gap was also identified and investigated. The model enforces a non-slip boundary condition between the plate and the anchoring fixture. In reality it is possible that the test article could slip in the clamping fixture. Everything was done experimentally to minimize the slippage during impact and penetration. After several experiments were performed, the experimentalists to assess the level of non-compliance of the non-slip boundary condition physically interrogated this situation. They deemed that the plate was not slipping in the clamping fixture. Therefore, the theoretical boundary condition enforcement was deemed acceptable and the "Importance" ranking of "Enforcement of Boundary Conditions" was downgraded to "L" in the final PIRT assessment (shown in blue hatch in **Table 1**) that elevated the "validation adequacy" coloration to green.

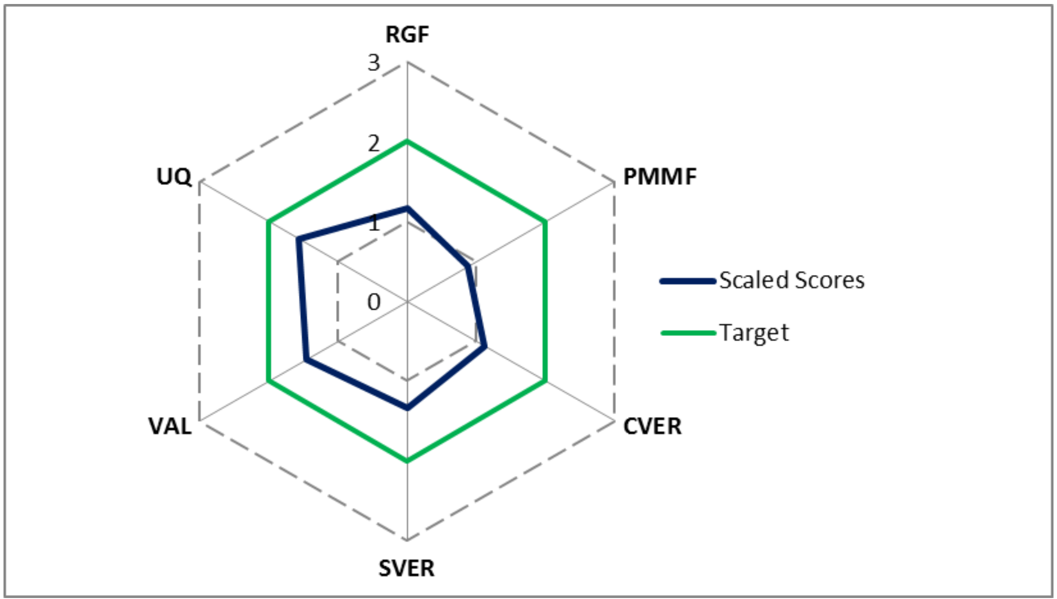
Predictive Capability Maturity Model (PCMM)

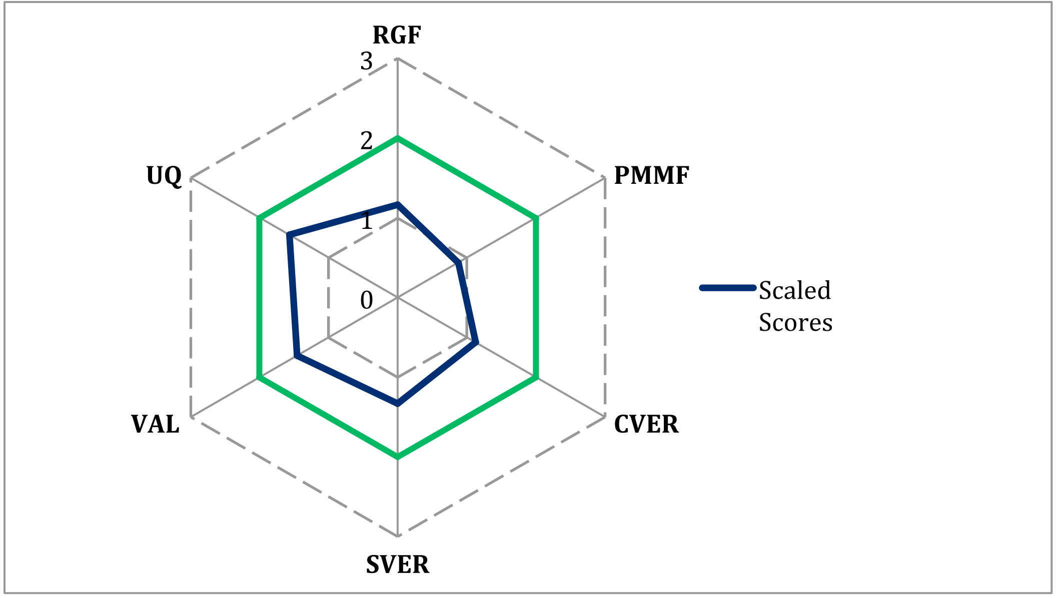
The PCMM was developed at SNL for the DOE ASC program as a means of assessing completeness of modeling and computational simulation activities for a particular application. There have been 3 generations of PCMM with an emphasis evolving from assessment to evidence inventory [Pilch06,Pilch07,Oberkampf07,Oberkampf10]. **Error! Reference source not found.** shows one of the PCMM templates commonly used. There are 6 elements to computational simulation that require investigation with respect to maturity level. Each element has several factors to consider. In the early PCMM versions these were divided into subcategories to be assessed separately but in the newer, current version these are areas to explore, consider and aggregate when determining an overall evaluation of the element. The stakeholders, from analysts to the customers, must determine what the appropriate goals, or required maturity levels, are for each of the elements for the required simulation. The supporting evidence determines the current level of maturity and qualifies the assessment. The difference between the current state and desired maturity level identifies the gaps that need to be addressed. Before the project progresses it must be determined and accepted by all involved whether these gaps are acceptable (and the required maturity level reduced) or what the mitigation plan is to reduce this gap. Often these gaps cannot be closed due to funding, time or technical constraints. Therefore, it must be determined whether a compromise between the stakeholders is possible and viable.

The PCMM assessment performed by the analysis team for the example is given in Table 3 with the specific "grades" highlighted in light green. The programmatic maturity level goal/target was determined to be a level 2 for all elements at the beginning of the project. A Kiviat diagram can be used to provide a visual metric of the state of each of the elements as compared to the specified target level. This simple chart quickly shows which element of providing a credible prediction is lacking. Both elements of solution verification and validation showed as lacking and require further attention.



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





The Predictive Capability Maturity Model (PCMM) was defined to allow the assessment of the state of modeling and simulation (M&S) is a manner that balances all aspects of the M&S workflow. PCMM provides a structured breakdown of the component work within an engineering M&S study. Classically it contains six elements, each requiring specific focus when the PCMM is applied. These six elements are geometry/representation fidelity, model fidelity, code verification, solution verification (numerical error estimation), validation and uncertainty quantification/sensitivity analysis. Each of these elements entails significant complexity and contributes to the overall quality.

Part of the original role of PCMM was to organize and structure the achievement of high credibility M&S for the purposes of decision-making. As PCMM was developed at an engineering laboratory, the focus was necessarily centered upon engineering analysis and thus biased. Over time these biases have become evident upon the application of PCMM to other M&S endeavors. In the process the framework has been modified, extended and its biases lay bare. Each extension has helped the model itself mature and yielded an enhanced understanding regarding its application and form. Among the most important of these extensions is the expansion of the process surrounding PCMM to include pre- and post-conditions and targets and an overall iterative framework for applying the assessment. We have come to view PCMM more as a communication and planning tool than a vehicle for assessment. It is in this broader sense that the PCMM may find its greatest utility in use, the definition of high-credibility M&S for decision making from inception to completion with greater quality. A third way to constructively engage with PCMM is through viewing it as defining the elements of the modeling and simulation workflow. In any workflow the PCMM encapsulates the different issues and topics that must be confronted. Recent work on PCMM has provided a recommended workflow together with the assessment elements.

Need for Frameworks

One of the big problems that the entire V&V enterprise has is the sense of imposition on others. Every simulation worth discussing does "V&V" at some level, and almost without exception they have weaknesses. Doing V&V "right" or "well" is not easy or simple. Usually, the proper conduct of V&V will expose numerous problems with a code, model, and/or simulation. It's kind of like exposing yourself to an annual physical; it's good for you because you learned something about yourself you didn't know, but you might have to face some unpleasant realities. In addition, the activity of V&V is quite broad and something almost always slips between the cracks (or chasms in many cases).

To deal with this breadth, the V&V community has developed some frameworks to hold all the details together. Sometimes these frameworks are approached as prescriptions for all the things you must do. Instead I'll suggest that these frameworks should not be recipes, nor should they be thought of as prescriptions, they are things to be seriously considered. If listed aspects of PCMM are disregarded, it should be justified through rigorous analysis. They are "thou should," not "thou shalt," or even "you might."

Several frameworks exist today and none of them is fit for all purposes, but all of them are instructive on the full range of activities that should be at least considered, if not engaged in.

CSAU – Code Scaling Assessment and Uncertainty [Boyack89,Boyack90] developed by the Nuclear Regulatory Committee to manage the quality of analyses done for power plant accidents. It is principally applied to thermal-fluid (i.e. thermal-hydraulic) phenomena that could potentially threaten the ability of nuclear fuel to contain radioactive products. This process led the way, but has failed in many respects to keep up to date. Nonetheless it includes processes and perspectives that have not been fully replicated in subsequent work. PCMM is attempting to utilize these lessons in improving its completeness.

PCMM – Predictive Capability Maturity Model developed at Sandia National Laboratories for the stockpile stewardship program in the last 10 years. As such it reflects the goals and objectives of this program and Sandia's particular mission space. It was inspired by the CMMI developed by Carnegie Mellon University to measure software process maturity. PCMM was Sandia's response to calls for greater attention to detail in defining the computational input into quantitative margins and uncertainty (QMU), [Pilch07] the process for nuclear weapons' certification completed annually.

CAS – Credibility Assessment Scale [CAS09] developed by NASA. They created a similar framework to PCMM for simulation quality in the wake of the shuttle accidents and specifically after Columbia where simulation quality played an unfortunate role. In the process that unfolded with that accident, the practices and approach to modeling and simulation was found to be unsatisfactory. The NASA approach has been adopted by the agency, but does not seem to be enforced. This is a clear problem and potentially important lesson. There is a difference between an enforced standard (i.e., CSAU) and one that comes across as well intentioned, but powerless directives. Analysis should be done with substantial rigor when lives are on the line. Ironically, formally demanding this rigor may not be the most productive way to achieve this end.

PMI, Predictive Maturity Index developed at Los Alamos [Hemez10,Unal11]. This framework is substantially more focused upon validation and uncertainty, and it is a bit lax with respect to the code's software and numerical issues. These aspects are necessary to focus upon given advances in the past 25 years since CSAU came into use in the nuclear industry.

MURM, (Model Utilization Risk Management) was developed by the Applied Physics Laboratory [Pace] o provide an explicit scoring system for risks associated with using M&S in decision-making. The MURM's intent is to provide a scale for considering the risk-based maturity of the capability and the risk-based assessment of the decision. This should provide the decision-maker with the information necessary to make an informed choice in utilizing M&S in a process.

Computational simulations are increasingly being used in our modern society to replace some degree of expensive or dangerous experiments and tests or where tests can't be conducted. Computational fluid and solid mechanics are ever more commonplace in modern engineering practice. The challenge of climate change may be another avenue where simulation quality is scrutinized and could benefit from a structured, disciplined approach to quality. Ultimately, these frameworks serve the role of providing greater confidence (faith) in the simulation results and their place in decision-making. Climate modeling is a place where simulation and modeling plays a large role, and the decisions being made are huge.

The question lingers in the mind, "what can these frameworks do for me?" My answer follows:

1. V&V and UQ are both deep fields with numerous deep subfields. Keeping all of this straight is a massive undertaking beyond the capacity of most professional scientists or engineers.
2. Everyone will default to focusing on where they are strong and comfortable, or interested. For some people it is mesh generation, for others it is modeling, and for yet others it is analysis of results. Such deep focus may not lead (or is not likely to lead) to the right sort of quality. Where quality is needed is dependent upon the problem itself and how the problem's solution is used.
3. These are useful outlines for all of the activities that a modeling and simulation project might consider. Project planning can use the frameworks to develop objectives and subtasks, prioritize and review.
4. These are menus of all the sort of things you might do, not all the things you must do.
5. They provide a sequenced set of activities, prepared in a sequenced rational manner with an eye toward what the modeling and simulation is used for. (Intended-use). Again, the framework provides a suggestion and not a straightjacket. Different sequencing can be executed if reasoned analysis calls for it.
6. They help keep your activities in balance. They will help keep you honest.
7. You will understand what is fit for purpose, when you have put too much effort into a single aspect of quality.
8. V&V and UQ are developing quickly and the frameworks provide a "cheat sheet" for all of the different aspects. It assures that the analysis and examination remains current.
9. The framework's flexibility is key, not every application necessarily should focus on every quality aspect, or apply every quality approach in equal measure.
10. Validation itself is incredibly hard in both breadth and depth. It should be engaged in a structured, thoughtful manner with a strong focus on the end application. Validation is easy to do poorly.
11. The computational science community largely ignores verification of code and calculations. Even when it is done, it is usually done poorly, or insufficiently.
12. Error estimation and uncertainty too rarely include the impact of numerical error, and estimate uncertainty primarily through parametric changes in models.
13. Numerical error is usually much larger than acknowledged. Lots of parametric and model calibration is actually accounting for the numerical error, or providing numerical stability rather than physical modeling.
14. Helps identify gaps and associated risks for each simulation
15. Helps you incorporate resource constraints and identify associate risks in forgoing V&V activities

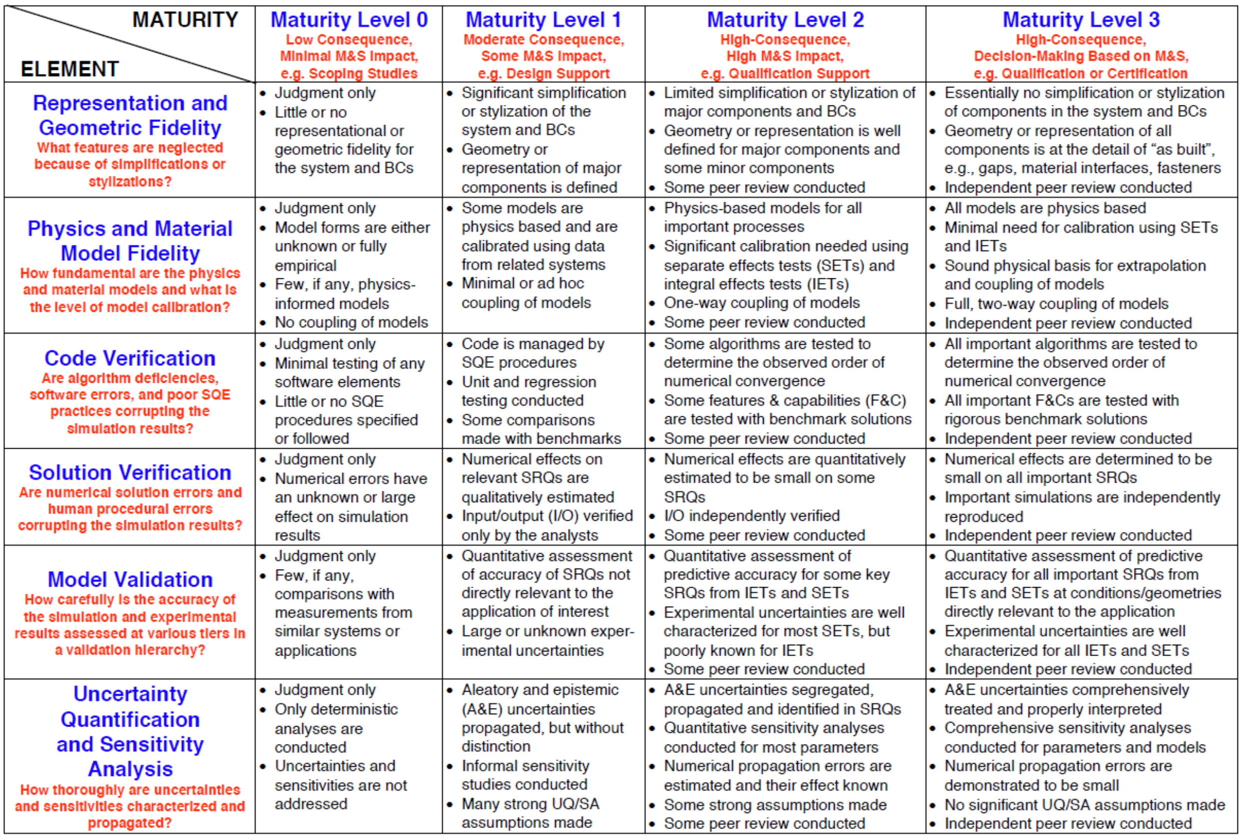
Different Frameworks and Their Vision

The Original PCMM

As originally constructed (and reported in [Oberkampf07]) the PCMM addressed six elements that were identified to be essential for successful application of modeling and simulation. These elements were:

- representation and model fidelity
- Physics and material model fidelity
- Code verification
- Solution verification
- Model validation
- Uncertainty quantification (UQ) and sensitivity analysis (SA)

In the PCMM process a general set of attributes were identified for each of these elements to permit characterization of each element into one of four maturity levels. This resulted in defining a matrix of maturity levels as illustrated in Figure 1.



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

Figure 1: PCMM Classification Guidance (from [Oberkamp07])

As can be seen from the information contained in Figure 1, the PCMM maturity levels constitute a hierarchy that represents increasingly greater levels of sophistication and computational fidelity (and expense and resources). The levels contained within this hierarchy can be summarized as follows.

- Level 0: At this level there is little or no assessment of completeness and accuracy and the capabilities for the element being assessed are highly reliant on personnel experience and judgment.
- Level 1: For this level an informal assessment of completeness and accuracy has been performed using internal peer review groups.
- Level 2: This level applies a formal process to assess completeness and accuracy of the element being evaluated. At this level use is made of external peer reviews for at least some of these assessments.
- Level 3: Finally, at this level a formal assessment of the element has been completed with the assessments predominantly being conducted by external peer reviews.

We note that the assessment of the levels of maturity does not, by itself, indicate the degree to which the M&S capability will be successful at meeting the requirements identified to address a particular application (e.g. a licensing application or a regulatory requirement, or a design specification). To identify the degree to which such a capability will be present for a particular application, one would compare the assessed maturity level for the PCMM elements with an objective level of maturity that is identified as necessary for the application of interest. Any application will provide specific quantities of interest associated with defining a successful outcome for the intended purpose. A hypothetical example of this is provided in Figure 2. In this Figure the coloring of each element indicates the degree to which the assessed maturity of that element meets the identified requirements (with obvious increasing divergences in meeting the requirements as one progresses along the color scale).

Variations on a Theme

PCMM is a framework to organize that entire V&V and UQ landscape into neat little boxes. Of course reality is never so neat and tidy, but structure for such a complex and potentially unbounded activity is good. PCMM has gone through numerous revisions, extensions and rearticulations, and the safest conclusion is that the framework will never be complete.

A close examination of PCMM provides an insight to its intrinsic bias toward a certain class of engineering calculations. This is clearest in the emphasis on geometric fidelity and solution verification, which belie its basis in mesh based calculations. Nonetheless, peeling back a layer, one should not lose sight of the nature of these entries in the framework. The geometry is really a view of the representation of reality in the overall computational simulation, while the solution verification is the estimate of error in the same. Both aspects are essential to determining the quality and assigning confidence to the simulation results. Other aspects of the PCMM translate across fields more readily. Code verification and software quality are necessary elements in providing reliable computer codes for simulation. Modeling and its credibility are universal in the field. Finally validation provides the tangible and measured connection to reality, and uncertainty quantification is key to decision-making.

Other elements have been added due to their importance and focus when applying the framework. Recent work has taken the validation element and broken it up into four separate activities to be examined. Two elements now involve the quality and acquisition of experimental data; followed by the assessment of the simulation with respect to the data. The validation exercise is divided into two sections, one with data and validation applied to the underlying and low-level models in the code, and the second applying to the application-level or high-level modelling and associated data for comparison. In addition we could have added elements to PCMM associated with the requirements arising from the simulation customers, and the impact of the code user/analyst/engineer on the results obtained. This area is sensitive and controversial because the human impact of M&S analysis is a “hot-button” issue with many.

A useful concept in examining complex, difficult problems is the characterization of the problem as “wicked”. A wicked problem has a number of characteristics that make it special and difficult to solve. For example, the problem cannot be fully understood before attempting to solve it. As the problem is tackled, new aspects of the problem are unveiled, and the solution to the problem needs to be rescoped. This is a continual aspect of the problem, and it bedevils those who attempt to apply project plans and complete predictability to the problem. There are even super-wicked problems that are more difficult due to counter-intuitive feedbacks between the problem itself and the solution. Characteristics of super-wicked problems are that those solving the problem are also causing the problem, and future results are irrationally discounted presently among others. V&V with PCMM as an example can probably characterized as being wicked and perhaps even super-wicked. Thus the activity defies full articulation.

Despite this intrinsic futility we would claim that PCMM is useful. In a real sense the activities within PCMM could be thought of as a menu of activities that one might consider in determining and driving simulation quality. This is as important as the measurement of the quality itself. Being such a complex and potentially unbounded activity means that there is large probability for details to slip from attention. PCMM provides the list of activities for simulation quality improvement and assessment, which can be utilized effectively by code development and application projects to draw upon. A reasonable recommendation is to examine the PCMM for “low hanging fruit” that can easily be incorporated into the simulation process. As we will now describe this process can be further refined by dividing the PCMM into two pieces; one that applies primarily to code development and the foundation of simulation capability, and a second that is more specific to the application of interest. Another practicality is the potentially overwhelming nature of PCMM. It is therefore rational and reasonable to apply less than the complete framework initially.

Thus a useful observation is that one can decompose the whole computational confidence issue into two relatively neat and tidy pieces: that which is more tool (code)-specific, (or foundational) and that which is problem or application-specific. Thus some work provides the general characterization of the computational tools to be used for analysis, while other work is directly applied to the application of that tool. For example, the general software development approach and documentation can be used over and over for different applications, while the specific options from a code used for a given application are narrow and specific to that application. This principle can be applied repeatedly across the span of activities that comprise the quality pedigree for a given code and its intended application. This principle applies for nearly every area of code quality investigation as laid out below.

Foundational PCMM

Software quality is one of the obvious lynchpins of the foundation for a code. The practices applied in the development of the code provide the degree of confidence in the code's correctness and stability. High quality code practices provide important tractability to the overall pedigree of the simulation. These practices will apply to every application of the code. We acknowledge that a subset of the software activities will be more directly relevant to a given activity and may be subject to different requirements.

Code verification is another simulation code quality practice that applies across the entire spectrum of potential applications. Code verification in a nutshell provides the evidence and confidence that the solution algorithm in the code is implemented correctly, and the given mathematical description is actually being solved. This is a distinctive to numerical simulation and applies in a complementary manner to the overall software quality approach. Again, some of the code verification will be more specifically applicable to a certain problem attacked by the code.

The decomposition of validation into two sets allows part of the validation activity to be considered foundational. Many models are common to a large number of simulations and comprise the core of the modeling capability of the code. These models must be validated so that their fidelity can be fully assessed away from the intended application. Furthermore, these models can be compared to special purpose experiments that have relatively small errors compared to many application settings. This separation allows some of the modeling capability to be assessed in a manner that should greatly increase confidence in the code. When these errors are convoluted with the integral-large scale data from the applications, the source of discrepancy can become hidden.

Accompanying the basic low-level validation of the code should be assessment of the concomitant uncertainty and sensitivity of the models associated with the code's simulation foundation. This should include the impact of model form and parameters as well as numerical integration effects (i.e., some solution verification). Again, this activity is undertaken to provide a baseline uncertainty and sensitivity away from the convoluted situation offered by the full application setting.

The capability of the code to model circumstances is provided by the user interface. This aspect of the assessment is relatively small in scope, but quite important. In many respects the flexibility offered by the user interface bounds what a code user can achieve. The importance of this activity has increased dramatically in recent year as user interfaces have become codes unto themselves (i.e., the input is itself executable where for example python, or other advanced scripting languages are used).

The foundational aspect of the customer requirements is applied to whomever is providing the resources for the development of the code. This customer can be distinct from the application use of the code, but not necessarily. Nevertheless, the customer has imposed requirements on the code development, and the assessment should provide a check to whether these have been complied with.

Application Specific PCMM

For any given application there is a computational representation of the problem being solved. This can take many forms including detailed meshes that can be compared with a CAD description. In other cases the representation is simplified as a lumped parameter model and the geometric fidelity is intentionally suppressed.

In every case there is a model of reality that is contained in the computational simulation. This model may be a set of continuously differential equations, integral conservations laws, or algebraic relations. In each case there should be an assessment of the model's capacity to simulate the desired situation in the application. For example, one might have a code that contains the incompressible Navier-Stokes equations, and the degree to which incompressibility applies should be examined. Two-phase flow is replete with complexity where for example one should see whether the description in the code is appropriate for the situation (is slip between the phases important, and do the equations appropriately describe the phenomena?).

For software quality and code verification the application specific assessment is bounded. The appropriate question is how much of the foundational aspect of the code's quality is specific to the application. Given the features of the code that the application depends upon are they tested in the software quality or code verification suites? How deep is this coverage and does it provide confidence in the code pedigree in a manner specific to a given application?

Solution verification is quite often overlooked in the practical use of M&S in engineering there is no excuse for this. The degree that the model representation and detail impacts solutions must be assessed as part of the overall uncertainty estimation. Too often the numerical error is simply calibrated for, or muddled with other modeling errors. The key to this step is the clear separation and articulation of this aspect of error and uncertainty apart from other sources. In many cases the solution verification may involve a simple bounding estimate that gives the idea of the magnitude of the effect on the results.

The integral validation aspect of PCMM comes naturally to the nuclear engineer. The unnatural part of the exercise is properly casting the process with all the other elements of PCMM. PCMM is in essence the deconvolution of many effects that often comprise the validation exercise. The foundational aspects of validation, uncertainty and solution verification attempt to peel away this complexity leaving the core of error to be examined. This is the task with integral validation to first understand how well the uncalibrated code models the circumstance, and then calibrate the solution without undoing any of the foundational work. Thus the calibration can be fully exposed to scrutiny and hopefully underlie the capacity to more directly attack the basis of lack of credibility.

Application uncertainty and sensitivity analysis is then quite clearly defined. Again the separation between foundational model validation with requisite uncertainty gives a basis for attacking the specific application uncertainty with clarity. Both aspects of uncertainty impact results and are included in the assessment, but the goal of clearly identifying the application model-specific uncertainties is obtainable through following the structure outlined here.

User qualification has a large impact on the results, yet is rarely assessed. An ideal case would be to have independent models of the same application. This is often not possible despite numerous studies showing that the user effect is large (larger than model differences in cases).

Finally the entity paying for the application work has requirements. These requirements should be assessed for how well the simulation has met these.

PCMM Example: QASPR

Here we give an example of what an actual capability assessment looks like in a bit of detail. In this case we will be showing the outcomes and process outline for the QASPR project [SNL] and associated codes, Xyce [FHCL] and Charon [TKHDB, KTHRSR]. The assessment was conducted at Sandia National Laboratories, Albuquerque by Laboratory staff. QASPR was begun an admitadly ambitious effort to replace the experimental testing of integrated circuits and associated semi-conductor material in a high radiation environment. Sandia sought to replace the expensive and risky SPUR reactor with an extensively validated and verified computational capability. The specific assessment was for the maturity of the III-V Npn model predictability within a defined threat environment for the Xyce and

Charon codes. In particular, the device model for the MESA developed using Npn InGap/GaAs.

The first thing to do is pull together a team to conduct the assessment starting with a lead stakeholder. In this case the stakeholder is acting as the customer for the work done by the code. In other cases this may or should be another person altogether. The assessment can take input from any member of the team although each member should not necessarily contribute to each part of the PCMM. Secondly we have a (trained) PCMM assessor who also acts as a moderator for the activity. Next we have a PCMM subject matter expert (SME) whose expertise includes key aspects of V&V. The team is then rounded out by SME's in each major roll for the development and use of the code(s). In this case we have SMEs for UQ, V&V analysis, experiments, applied circuit analysis, calibration and code development for both Xyce and Charon (different people).

The first step done was to meet with the entire team to brief them on the process and gain buy-in. In addition the team receives "training" on the PCMM and the assessment process. In the case of QASPR the project had already conducted an extensive PIRT. Rather than repeat this work, the first step was to review and vet the existing PIRT for the assessment that follows. The team then held a set of meetings to conduct the assessment. It is important that the team be well-represented throughout. After a set of meetings the team prepared some focused feedback on the state of the QASPR project's codes as well as the PCMM process itself.

Part of this effort used an Excel spreadsheet as the vehicle for the assessment, and much of the feedback keyed on this. There was feeling that clarification of the language could be done particularly within the UQ section. Some elements seemed to overlap one another. Overall the Excel Tool was judged to be very beneficial and helps capture a great deal of detail, but can be overwhelming at times during the assessment. We might consider creating a reduced version for real time assessment. Overall it is a great organization tool for the assessor. This sort of feedback should always be included to continually improve the process.

A real positive is that the PCMM assessment generated discussion across different teams consisting of analysts, developers, and experimentalists. As an example of the discussion: For qualification, solution verification will be critical (e.g., input/output file verification) and will need to be better formalized – analysts are now aware of this and QASPR will need to better formalize this workflow. We need another assessment iteration to determine the value of some members on the larger team for the purposes of assessment. Experimentalists seemed to be the odd man out, but may be due to QASPR development of Physical Simulation PCMM. The assessment is also captured in some graphical output, and overall spreadsheet content.

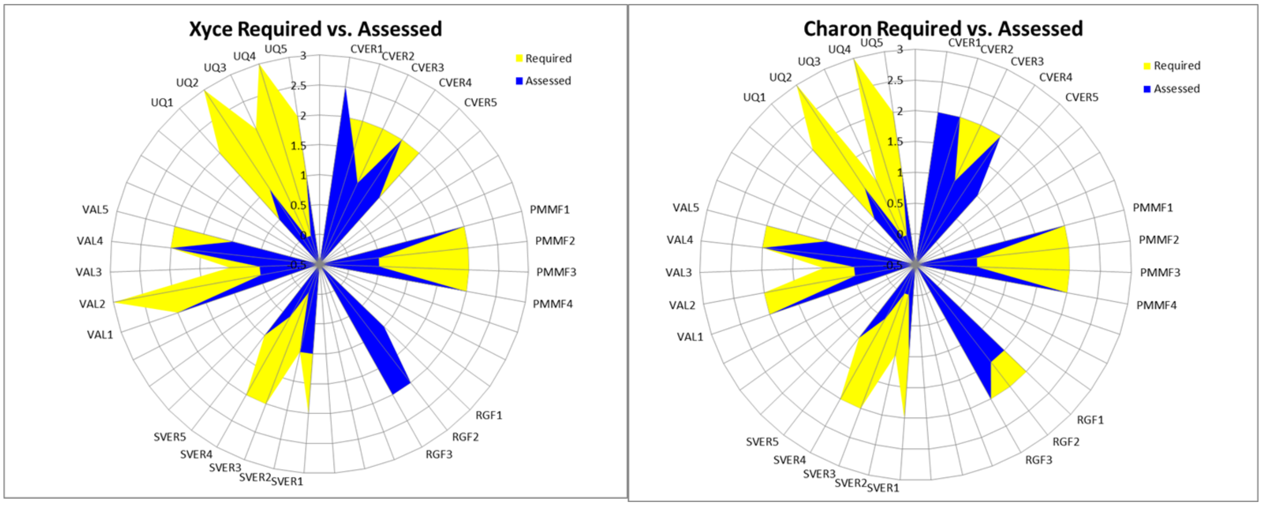


Figure Q1: QASPR: PCMM Kiviat Plots showing the assessed status of Xyce and Charon.

At the end of the assessment we crafted a path forward for further assessments. First, review PCMM elements with PCMM SME prior to assessment to insure all the descriptors are well understood. This was done in real time during the assessment which caused delays. The Excel Tool will be used for future assessments. Evidence may be moved and maintained on a SharePoint site. Will do another iteration of the assessment with the larger group. Continue to schedule review to minimize the time required by participants based on roles. The next assessment will focus on PnP devices and potentially circuits and an update of the Npn capability should also be done.

Outlook

I think it is time for us to shoulder some of the blame and rethink our approach to engaging other scientists and engineers on the topic of modeling and simulation (M&S) quality.

V&V should be an easy sell to the scientific and engineering establishment. It hasn't been, it has been resisted at every step. V&V is basically a rearticulation of the scientific method we all learn, use and ultimately love and cherish. Instead, we find a great deal of animosity toward V&V, and outright resistance to including it as part of the M&S product. To some extent it has been successful in growing as a discipline and focus, but too many barriers still exist. Through hard learned lessons I have come to the conclusion that a large part of the reason is the V&V community's approach. For example, one of the worst ideas the V&V community has ever had is "independent V&V". In this model V&V comes in independently and renders a judgment on the quality of M&S. It ends up being completely adversarial with the

M&S community, and a recipe for disaster. We end up less engaged and hated by those we judge. No lasting V&V legacy is created through the effort. The M&S professionals treat V&V like a disease and spend a lot of time trying to simply ignore or defeat it. This time could be better spent improving the true quality, which ought to be everyone's actual objective. Archetypical examples of this approach in action are federal regulators (NRC, the Defense Board…). This idea needs to be modified into something collaborative where the M&S professions end up owning the quality of their work, and V&V engages as a resource to improve quality.

The fact is that everyone doing M&S wants to do the best job they can, but to some degree don't know how to do everything. In a lot of cases they haven't even considered some of the issues we can help with. V&V expertise can provide knowledge and capability to improve quality if they are welcome and trusted. One of the main jobs of V&V should be build trust so that they might provide their knowledge to important work. In sense, the V&V community should be quality "coaches" for M&S. Another way the V&V community can help is to provide appropriately leveled tools for managing quality. PCMM can be such a tool if its flexibility is increased. Most acutely, PCMM needs a simpler version. Most modeling and simulation professionals will do a very good job with some aspects of quality. Other areas of quality fall outside their expertise or interest. In a very real sense, PCMM is a catalog of quality measures that could be taken. Following the framework helps M&S professionals keep all the aspects of quality in mind and within reach. The V&V community can then provide the necessary expertise to carry out a deeper quality approach.

If V&V allows itself to get into the role of judge and jury on quality, progress will be poor. V&V's job is to ask appropriate questions about quality as partners with M&S professionals interested in improving the quality of their work. By taking this approach we can produce an M&S future where quality continuously improves.

References

[Boyack90] Boyack, B. E., and G. S. Lellouche. "Quantifying reactor safety margins part 1: an overview of the code scaling, applicability, and uncertainty evaluation methodology." *Nuclear Engineering and Design* 119.1 (1990): 1-15.

[Boyack89] Boyack, B., et al. *Quantifying reactor safety margins: Application of code scaling, applicability, and uncertainty evaluation methodology to a large-break, loss-of-coolant accident*. Nuclear Regulatory Commission, Washington, DC (USA). Div. of Systems Research, 1989.

[CAS09] Babula, Maria, William J. Bertch, Lawrence L. Green, Joseph P. Hale, Gary E. Mosier, Martin J. Steele, and Jody Woods. "NASA Standard for Models and Simulations: Credibility Assessment Scale." In *AIAA Proc. 47th AIAA Aerospace Sciences Meeting, AIAA-2009-1011*. 2009

[FHCL] D. Fixel, G. Hennigan, J. Castro, P. Lin, “Charon parallel semiconductor device simulator,” SAND2010-3905, Sandia National Laboratories, Albuquerque, NM.

[Hemez10] Hemez, François, H. Sezer Atamturktur, and Cetin Unal. "Defining predictive maturity for validated numerical simulations." *Computers & structures* 88.7 (2010): 497-505.

[Hills13] Hills, Richard Guy, Walter R. Witkowski, William J. Rider, Timothy G. Trucano, and Angel Urbina. "Development of a Fourth Generation Predictive Capability Maturity Model." *Sandia National Laboratories, Albuquerque, NM2013* (2013).

[KTHRSR] Keiter, Eric R. and Thornquist, Heidi K. and Hoekstra, Robert J. and Russo, Thomas V. and Schiek, Richard L. and Rankin, Eric L, “Parallel Transistor-Level Circuit Simulation,” Simulation and Verification of Electronic and Biological Systems. pp. 1-21, 2011 Springer Netherlands http://dx.doi.org/10.1007/978-94-007-0149-6_1.

[Oberkampf10] Oberkampf, William L., and Christopher J. Roy. *Verification and validation in scientific computing*. Cambridge University Press, 2010.

[Oberkampf07] Oberkampf, William L., Martin Pilch, and Timothy G. Trucano. "Predictive capability maturity model for computational modeling and simulation." *Development* (2007): 2.

[Pace] Pace, Dale K. "Comprehensive consideration of uncertainty in simulation use." *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 10, no. 4 (2013): 367-380.

[Pilch06] Pilch, Martin, Timothy G. Trucano, and Jon C. Helton. "Ideas underlying quantification of margins and uncertainties (QMU): A white paper." *Unlimited Release SAND2006-5001, Sandia National Laboratory, Albuquerque, New Mexico* 87185 (2006).

[Pilch07] Pilch, Martin M. *Predictive capability maturity model (PCMM)*. No. SAND2007-0102C. Sandia National Laboratories, 2007.

[Roache] Roache, Patrick J. *Verification and validation in computational science and engineering*. Hermosa, 1998.

[SNL] S. M. Holmes, “Tough Enough,” Sandia Research, pp. 11-15, Volume 2, Number 2, July 2014.

[TKHDB] Thornquist, Heidi K. and Keiter, Eric R. and Hoekstra, Robert J. and Day, David M. and Boman, Erik G, “A parallel preconditioning strategy for efficient transistor-level circuit simulation,” ICCAD '09: Proceedings of the 2009 International Conference on Computer-Aided Design, pp. 410-417, 2009, San Jose,

California, ACM, New York, NY, USA.
http://doi.acm.org/10.1145/1687399.1687477

[Wilson] Wilson, Gary E., and Brent E. Boyack. "The role of the PIRT process in experiments, code development and code applications associated with reactor safety analysis." *Nuclear Engineering and Design* 186.1 (1998): 23-37.

[Unal11] Unal, Cetin, et al. "Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors." *Nuclear Engineering and Design* 241.5 (2011): 1813-1833.