# 15

# Maturity assessment of modeling and simulation

In Chapter 1, Introduction, we briefly discussed how credibility is built in modeling and simulation (M&S). The four elements mentioned in that chapter were: quality of the analysts conducting the analysis, quality of the physics modeling, verification and validation activities, and uncertainty quantification and sensitivity analysis. The latter three elements are technical elements that can be assessed for completeness or maturity. Assessment of maturity is important to the staff conducting the modeling and simulation effort, but it is critically important for project managers and decision makers who use computational results as an element in their decision making. It is also important for internal or external review committees who are asked to provide recommendations on the credibility and soundness of computational analyses. This chapter deals with reviewing methods that have been developed for assessing similar activities, and then presents a newly developed technique reported in Oberkampf *et al.* (2007). This chapter is taken in large part from this reference.

## 15.1 Survey of maturity assessment procedures

Over the last decade, a number of researchers have investigated how to measure the maturity and credibility of software and hardware development processes and products. Probably the best-known procedure for measuring the maturity of software product development and business processes is the Capability Maturity Model Integration (CMMI). The CMMI is a successor to the Capability Maturity Model (CMM). Development of the CMM was initiated in 1987 to improve software quality. For an extensive discussion of the framework and methods for the CMMI, see West (2004); Ahern *et al.* (2005); Garcia and Turner (2006); and Chrissis *et al.* (2007). The CMMI, and other approaches discussed here, respond to the need for measuring the maturity (i.e., some sense of quality, capability and completeness) of a process to do one or more of the following:

- improve identification and understanding of the elements of the process;
- determine the elements of the process that may need improvement so that the intended product of the process can be improved;
- determine how time and resources can best be invested in elements of the process to obtain the maximum return on the investment expended;

- better estimate the cost and schedule required to improve elements of the process;
- improve the methods of aggregating maturity information from diverse elements of the process to better summarize the overall maturity of the process;
- improve the methods of communicating to the decision maker the maturity of the process so that better risk-informed decisions can be made;
- measure the progress of improving the process so that managers of the process, stakeholders, and funding sources can determine the value added over time;
- compare elements of the process across competitive organizations so that a collection of best practices can be developed and used;
- measure the maturity of the process in relation to requirements imposed by the customer.

The CMMI was developed by the Software Engineering Institute, a Federally funded research and development center sponsored by the US Department of Defense (DoD) and operated by Carnegie Mellon University. The latest release of the CMMI is CMMI for Development, (CMMI-DEV version 1.2) (Garcia and Turner, 2006; SEI, 2006; Chrissis *et al.*, 2007). The CMMI-DEV is divided into four process areas: engineering, process management, project management, and support (Garcia and Turner, 2006). The engineering process area is further divided into six subareas: product integration, requirements development, requirements management, technical solution, verification, and validation. The meaning of V&V referred to in the CMMI-DEV relates to concepts developed by the Institute of Electrical and Electronics Engineers (IEEE) for software quality engineering (SQE) (IEEE, 1989). As discussed in Chapter 2, Fundamental concepts and terminology, these concepts of V&V are quite different from those used in this book.

A maturity measurement system that has its origins in risk management is the Technology Readiness Levels (TRLs) system pioneered by NASA in the late 1980s (Mankins, 1995). The intent of TRLs is to lower acquisition risks of high technology systems by more precisely and uniformly assessing the maturity of a technology. TRLs are used in both NASA and the DoD. We do not review TRLs in detail in this document, but the interested reader can consult GAO (1999) for more information. TRLs consider nine levels of maturity in the evolution of technological systems. These levels are described by the DoD in DoD (2005) as follows.

- *TRL Level 1: Basic principles observed and reported*. Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of a technology's basic properties.
- *TRL Level 2: Technology concept and/or application formulated*. Invention begins. Once basic principles are observed, practical applications can be invented. The application is speculative and there is no proof or detailed analysis to support the assumption. Examples are still limited to paper studies.
- *TRL Level* 3*: Analytical and experimental critical function and/or characteristic proof of concept*. Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.

- *TRL Level 4: Component and/or breadboard validation in laboratory environment*. Basic technological components are integrated to establish that the pieces will work together. This is relatively low fidelity compared to the final system. Examples include integration of *ad hoc* hardware in a laboratory.
- *TRL Level 5: Component and/or breadboard validation in relevant environment*. Fidelity of breadboard technology, i.e. experimental electrical circuit prototype, increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so that the technology can be tested in a simulated environment. An example is high-fidelity laboratory integration of components.
- *TRL Level 6: System/subsystem model or prototype demonstration in a relevant environment*. A representative model or prototype system, which is well beyond the breadboard tested for TRL 5, is tested in a relevant environment. This represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment.
- *TRL Level 7: System prototype demonstration in an operational environment*. The prototype is near or at the planned operational system. This represents a major step up from TRL 6, requiring the demonstration of an actual system prototype in an operational environment with representatives of the intended user organization(s). Examples include testing the prototype in structured or actual field use.
- *TRL Level 8: Actual system completed and operationally qualified through test and demonstration*. The technology has been proven to work in its final form and under expected operational conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended or pre-production configuration to determine if it meets design specifications and operational suitability.
- *TRL Level 9: Actual system, proven through successful mission operations*. The technology is applied in its production configuration under mission conditions, such as those encountered in operational test and evaluation. In almost all cases, this is the last *bug* fixing aspect of true system development. An example is operation of the system under operational mission conditions.

The nominal specifications of TRLs as presented above are clearly aimed at assessing the maturity of hardware products, *not* software products. Smith (2004) has examined the difficulties in using TRLs for nondevelopmental software, including commercial-off-the-shelf (COTS) and government-off-the-shelf (GOTS) software and open sources of software technology and products. Clay *et al.* (2007) have also studied the issue of using TRLs to assess the maturity of M&S software. Both studies concluded that significant changes in TRLs are needed before they would be useful for assessing the maturity of software. Stated more directly, TRLs are *not* useful for assessing the maturity of software of any kind. Whether they are useful in assessing the likelihood of the successful completion of a proposed hardware system (in the sense of attaining the planned cost, schedule, and performance), is clearly debatable.

A maturity assessment procedure that deals more directly with M&S processes than the CMMI and the TRLs has been developed by Balci *et al.* (2002) and Balci (2004). He argues that M&S quality can be assessed based on indicators of product, process, and project. By *product* he means either (a) the overall completed M&S application, or

(b) a work product created during the M&S development life cycle such as the conceptual model, M&S requirements specification, M&S design specification, and an executable M&S module. By *process* he refers to the process used to create a work product during the M&S development life cycle, such as conceptual modeling, requirements, engineering, design, implementation, integration, experimentation, and presentation. *Project* refers to the quality indicators of the project plan, the capabilities and experience of the organization conducting the M&S, and the technical quality of the people tasked to develop the M&S application. Some of the attributes of the assessment of quality of product, process, and project are: accuracy, verity, validity, clarity, completeness, acceptability, maintainability, timeliness, reliability, robustness, supportability, understandability, visibility, and maturity.

Harmon and Youngblood (2003, 2005) focus on assessing the maturity of the validation process for simulation models. Their work takes the encompassing view of validation, as is commonly taken by the DoD. As discussed in Section 2.2.3, the *encompassing view* means that the term "validated model" denotes that the following three related issues have been addressed with regard to the accuracy and adequacy of the M&S results.

- The system response quantities (SRQs) of interest produced by the model have been assessed for accuracy with respect to some referent.
- The model's domain of intended use is defined and the model can be applied over this domain.
- The model meets the accuracy requirements for the "representation of the real world" over the domain of its intended use.

It should be noted here that the perspective of validation taken by the AIAA and the ASME is that the referent can *only be experimentally measured data*. The DoD does not take this restrictive perspective. Thus, the DoD permits the referent to be, for example, results from other computer models, as well as expert opinion.

Harmon and Youngblood (2003, 2005) clearly state that validation is a process that generates information about the accuracy and adequacy of the simulation model as its sole product. They argue that the properties of information quality are defined by (a) correctness of the information, (b) completeness of the information, and (c) confidence that the information is correct for the intended use of the model. They view the validation process as using information from five contributing elements: (1) the conceptual model of the simulation, (2) verification results from intermediate development products, (3) the validation referent, (4) the validation criteria, and (5) the simulation results. The technique used by Harmon and Youngblood (2003, 2005) ranks each of these five elements into six levels of maturity. From lowest to highest the six levels of maturity are:

1 we have no idea of the maturity;
2 it works, trust me;
3 it represents the right entities and attributes;
4 it does the right things; its representations are complete enough;
5 for what it does, its representations are accurate enough;
6 I'm confident this simulation is valid.

Pilch *et al.* (2004) proposed a framework for how M&S can contribute to the nuclear weapons program of the US. They suggested that there are four key contributors to M&S: qualified computational practitioners, qualified codes, qualified computational infrastructure, and appropriate levels of formality. As part of qualified codes, Pilch *et al.* described nine elements:

1 request for service,
2 project plan development,
3 technical plan development,
4 technical plan review,
5 application-specific calculation assessment,
6 solution verification,
7 uncertainty quantification,
8 qualification and acceptance,
9 documentation and archiving.

For each of these elements, Pilch *et al.* (2004) described the key issues and the key evidence artifacts that should be produced. They also described four levels of formality that would generally apply over a wide range of M&S situations:

1 formality appropriate for research and development tasks, such as improving the scientific understanding of physical phenomena;
2 formality appropriate for nuclear weapon design support;
3 formality appropriate for nuclear weapon qualification support, i.e., confidence in component performance is supported by simulations;
4 formality appropriate for qualification of nuclear weapon components, i.e., confidence in component performance is heavily based on simulations.

Pilch *et al.* then constructed a table with rows corresponding to the nine elements and the columns corresponding to the four levels of formality. In each element of the table, the characteristics that should be achieved for a given element at a given level of maturity are listed. This table, or matrix, could then be used to assess the maturity of a M&S effort.

NASA recently released a technical standard that specifically deals with M&S as it contributes to decision making (NASA, 2008). The primary goal of this standard is to ensure that the credibility of the results from M&S is properly conveyed to those making critical decisions. Critical decisions are those related to design, development, manufacturing, ground or flight operations that may impact human safety or program/project-defined mission success criteria. The secondary goal is to assess whether the credibility of the results meets the project requirements. This standard is intended to ensure that sufficient details of the M&S process are available to support project requirements and to respond to in-depth queries by the decision maker. The standard applies to M&S used by NASA and its contractors for critical decisions in design, development, manufacturing, ground operations, and flight operations. The standard also applies to the use of legacy as well as COTS, GOTS, and modified-off-the-shelf (MOTS) M&S to support critical decisions. NASA staff developed the standard with an intensive effort over a period of three years.

The following references give a description of its development and examples of the use of the standard in various projects: (Bertch *et al.*, 2008; Blattnig *et al.*, 2008; Green *et al.*, 2008; Steele, 2008).

The NASA standard describes a credibility assessment scale for assessing the credibility of M&S results. The scale defines eight factors for assessing credibility (NASA, 2008).

1 *Verification*: Were the models implemented correctly, and what was the numerical error/uncertainty?
2 *Validation*: Did the M&S results compare favorably to the referent data, and how close is the referent to the real-world system?
3 *Input pedigree*: How confident are we of the current input data?
4 *Results uncertainty*: What is the uncertainty in the current M&S results?
5 *Results robustness*: How thoroughly are the sensitivities of the current M&S results known?
6 *Use history*: Have the current M&S been used successfully before?
7 *M&S management*: How well managed were the M&S processes?
8 *People qualifications*: How qualified were the personnel?

These eight factors are grouped into three categories: (a) M&S development (verification and validation), (b) M&S operations (input pedigree, results uncertainty, and results robustness), and (c) supporting evidence (use history, M&S management, and people qualifications). The M&S development and M&S operations categories have two subfactors for evaluation (evidence and technical review), whereas supporting evidence does not have subfactors.

Each of the eight factors given above has five levels of credibility or maturity that are numerically quantified as 0, 1, 2, 3, and 4. There is no uniform description or characterization of the requirements needed to attain a given level across each of the eight factors. That is, each factor has specific descriptors to characterize what is needed to achieve a specific level, except for level 0. Level 0 means either that no evidence exists for that factor, or that the evidence that does exist does not meet even the level 1 criteria. Although no uniform characterization is provided, increasing credibility levels must demonstrate increasing accuracy, formality, and recommended practice.

The final contribution to the literature reviewed comes from the field of information theory. If one agrees with the concept of Harmon and Youngblood (2003, 2005), as we do, that the product of M&S is information, then one must address the fundamental aspects of information quality. Wang and Strong (1996) conducted an extensive survey of information consumers to determine the important attributes of information quality. Stated differently, they went directly to a very wide range of customers that use, act on, and purchase information to determine what were the most important qualities of information. Wang and Strong (1996) analyzed the survey results and then categorized the attributes into four aspects.

- *intrinsic information quality*: believability, accuracy, objectivity, and reputation;
- *contextual information quality*: value added, relevancy, timeliness, completeness, and amount of information;

- *representational information quality*: interpretability, ease of understanding, consistent representation, and concise representation;
- *accessibility information quality*: accessibility and security aspects.

If the user of the information is not adequately satisfied with essentially all of these important attributes, then the user could (a) make minimal use of the information for the decision at hand; (b) completely ignore the information; or (c) misuse the information, either intentionally or unintentionally. These outcomes range from wasting information (and the time and resources expended to create it) to a potentially disastrous result caused by misuse of the information.

## 15.2 Predictive capability maturity model

Building on this previous work, we now describe the *predictive capability maturity model* (PCMM), also referred to as the predictive capability maturity matrix. The present version of the PCMM was first documented in Oberkampf *et al.* (2007) and has been tested in various forms at Sandia National Laboratories since 2005. During this time there was close collaboration with Thomas Zang of NASA and his team in the development of both the Interim NASA Standard (NASA, 2006) and the final NASA Standard (NASA, 2008).

The PCMM was developed to focus more on the computational aspects of M&S as compared to the broader class of models considered in the NASA Standard or the work of Harmon and Youngblood (2003, 2005). There is no single best method for assessing maturity in M&S. One should choose the method that is best suited for the type of M&S activity in question.

### *15.2.1 Structure of the PCMM*

As can be seen in the literature review, a number of similar elements have been identified as contributors to the confidence one should place in the simulation activity itself and in the results of the activity. The PCMM identifies six elements that fundamentally contribute to the credibility of the simulation. These elements are as follows:

- representation and geometric fidelity,
- physics and material model fidelity,
- code verification,
- solution verification,
- model validation,
- uncertainty quantification and sensitivity analysis.

Each of these six elements is defined for minimal overlap or dependency between the elements; i.e., each element attempts to contribute a separate type of information to the simulation activity. The elements listed are some of the most important contributors to simulation credibility, as well as the conceptual issues related to the four aspects of information quality identified by Wang and Strong (1996). When researchers at Sandia attempted to use the approaches discussed in the literature review given above, it was concluded that the primary shortcoming was representational information quality, specifically interpretability. That is, previous work, in our view, lacked a clear and unambiguous specification of what the information meant and how it should be used. We discovered that the primary reason for the shortcoming was that previous work had not adequately segregated some of the underlying conceptual issues, particularly what was being assessed. Was it the quality of the simulation process or the quality of the simulation results that was being assessed? Without improved interpretability, decision makers cannot properly use and act on the information produced in an assessment. These issues will be discussed further in Section 15.2.2.

Of the six elements identified only two, representation and geometric fidelity and physics and material model fidelity, have not been discussed in detail earlier in this book. These will be briefly discussed below because they have been identified as important contributors to simulation maturity. In addition, all the approaches discussed in the literature review agree that some type of graded scale is needed to measure the maturity, or confidence, of each contributing element. This topic will also be discussed in the following.

#### *15.2.1.1 Representation and geometric fidelity*

Representational and geometric modeling fidelity refers to the level of detail included in the spatial and temporal definition of all constituent elements of the system being analyzed. Note that when we refer to a *system*, we mean *any* engineered system, e.g., a subsystem, a component, or a part of a component. In M&S, the representational and geometric definition of a system is commonly specified in a computer-aided design (CAD) software package. The traditional emphasis in CAD packages has been on manufacturing-related dimensional, fabrication, and assembly specifications. As M&S has matured, CAD vendors are now beginning to address issues that are specifically important to computational-analysis needs, e.g., mesh generation and feature definitions that are important to various types of physics modeling. Even though some progress has been made that eases the transition from traditional CAD files to the construction of a computational mesh, a great deal of work still needs to be done. Aside from geometry clean-up and simplification activities, which are directed at making CAD geometries more useful in simulation, there is no general process for verifying that the CAD geometries loaded into calculations are appropriate and consistent with the physics modeling assumptions. A key issue that complicates the mapping of CAD geometries to a geometry ready for construction of a computational mesh is that the mapping is dependent on the particular type of physics to be modeled and the specific assumptions in the modeling. For example, a change in strength of material properties along the surface of a flight vehicle would be important to a structural dynamics analysis, but it may not be important to an aerodynamic or electromagnetic analysis. As a result, the CAD vendors cannot provide a simple or algorithmic method to address the wide variety of feature definitions and nuances required for different types of physics model. The time-consuming task of addressing detailed representation and geometric fidelity issues becomes the responsibility of technically trained staff with different backgrounds, such as CAD package developers, computational scientists, and mesh-generation experts.

#### *15.2.1.2 Physics and material model fidelity*

It is well recognized that improvement in the fidelity of physics modeling has been the dominant theme pursued in most simulations directed toward engineering systems. The range of physics modeling fidelity can vary from empirical models that are based on the fitting of experimental data (empirical models) to what is typically called *first-principles physics*. The three types of model in this range are referred to here as fully empirical models, semi-empirical models, and physics-based models. Physical process models that are *completely* built on statistical fits of experimental data are referred to as *fully empirical models*. These fully empirical models typically have *no* relationship to physics-based principles. Consequently, the fully empirical models rest entirely on the calibration of responses to identified input parameters over a specified range and should not be used (extrapolated) beyond their calibration domain. A *semi-empirical model* is partially based on physical principles and is highly calibrated by experimental data for the system or process of interest. An example of a semi-empirical model that has been heavily used in nuclear reactor safety is the control volume, or lumped parameter, model. Semi-empirical models typically conserve mass, momentum, and energy, but at clearly finite physical scales compared to the system of interest. For example, a 3-D system may be divided into $10\times10\times10$ control volumes in an analysis. In addition, they rely heavily on fitting experimental data as a function of dimensional or nondimensional parameters, such as Reynolds or Nusselt numbers, to calibrate the models. By *physics-based models* we mean models that are heavily reliant on partial differential or integro-differential equations that represent conservation of mass, momentum, and energy at infinitesimal length and time scales relative to the physical scales in the system. Some physicists use the term first-principles, or *ab initio*, physics to mean modeling that starts at the atomic or molecular level. These models, however, are essentially never used in design and analysis of engineering systems.

Another important aspect of physics modeling fidelity is the degree to which various types of physics are coupled in the mathematical model of the system. For fully empirical and semi-empirical models, strong assumptions are made to greatly simplify the physics considered, and little or no coupling of different types of physics is included. For physics-based models, however, the modeling assumptions must include various types of physical phenomena, as well as certain types of physics coupling. As shown in Figure 15.1, two basic approaches are used to couple the physics involved in the physical process:

- one-way causal effect, i.e., one physical phenomenon affects other phenomena, but the other phenomena do not affect the originating phenomenon; and
- two-way interactions, i.e., all physical phenomena affect all other physical phenomena.

In physics-based modeling, each physical phenomenon is represented by a mathematical model with boundary conditions (BCs) and initial conditions (ICs). In one-way coupling (Figure 15.1a), the physics of phenomenon 1 affect phenomena 2 and 3, but there is no direct

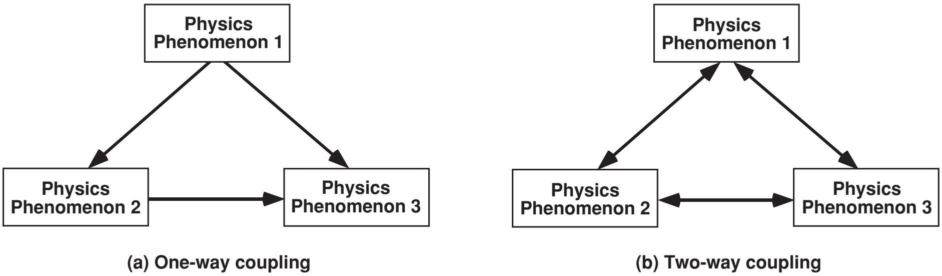


Figure 15.1 Example of two basic types of coupling of physical phenomena (Oberkampf *et al.*, 2007).

feedback from phenomena 2 and 3 to phenomenon 1. Phenomenon 1 affects phenomena 2 and 3 either by way of BCs or through some type of system excitation, e.g., the addition of a source term on the right side of the PDEs for phenomena 2 and 3. In addition, as shown in Figure 15.1a, the BCs or source terms of phenomenon 3 are determined by phenomenon 2. In two-way coupling (Figure 15.1b), all phenomena in the system affect all other phenomena. This two-way interaction can be modeled as strong coupling, where two or more phenomena are modeled within the same mathematical model, or as weak coupling, where the interactions between phenomena occur through BCs between separate sets of mathematical models. If one has a time-dependent mathematical model, this coupling can be either accounted for iteratively within each time step or accounted for in the next time step computed, i.e., time-lagged coupling.

#### *15.2.1.3 Maturity assessment*

In the review of the literature given above, three methods were described for ranking the maturity of the various simulation elements. The five-point maturity ranking scale of Harmon and Youngblood (2003, 2005) was dominated by the concepts of credibility, objectivity, and sufficiency of accuracy for the intended use. The four-point scale of Pilch *et al.* (2004) was dominated by the level of formality, the degree of risk in the decision based on the simulation effort, the importance of the decision to which the simulation effort contributes, and sufficiency of accuracy for the intended use. The five-point scale of NASA (2008) displays increasing accuracy, formality, and recommended practice, *excluding* the adequacy requirements of the simulation results. NASA clearly separated the ideas of credibility assessment of the simulation results from the requirements for a given application that use the simulation results.

Comparing each of these three maturity-ranking methods, we first note that the methods use scales of different magnitude for ranking maturity. We believe, however, that this difference is not fundamentally important. The key difference in our opinion between the three methods is that only the NASA scale *explicitly excludes* the issue pertaining to adequacy in the maturity assessment; adequacy is addressed *after* the assessment. We believe this was a major step forward in the interpretability of the assessment of a simulation effort because it segregates the assessed maturity of the results from the issue of required maturity (or credibility) of the results, as might be independently specified. We expect that some users of a maturity assessment would prefer to have the maturity scale include, or at least imply, the adequacy of the results for some intended use because it would seem to make their decisions easier. However, we strongly believe that the issues of maturity assessment and maturity assessment requirements should be dealt with independently as much as possible to reduce misunderstandings or misuse of a maturity assessment.

Pilch *et al.* (2004, 2006) discuss a method for assessing the maturity of each simulation element based on the risk tolerance of the decision maker. Stated differently, the maturity scale would be given an ordinal ranking based on the risk assumed by the decision maker who uses results generated by the simulation effort. This approach has some appealing features, but it also introduces additional complexities. We mention three difficulties in using a risk-based scale that have practical impact when constructing a maturity scale.

First, risk assessment is commonly defined as having two components: (a) likelihood of an occurrence and (b) magnitude of the adverse effects of an occurrence. We argue that the estimated likelihood of the occurrence, the identification of possible adverse occurrences, and the estimated magnitude of the adverse consequences are very difficult and costly to determine for complex systems. Consequently, complicated risk assessments commonly involve significant analysis efforts in their own right. Further, combining these complicated risk assessments with the maturity ranking of a simulation effort is difficult to achieve and certainly difficult for anyone to interpret.

Second, the risk tolerance of decision makers or groups of decision makers is a highly variable and a difficult attribute to quantify. The original discussion of Pilch *et al.* correlated the risk-tolerance scale with the increased risk perception of passing from exploratory research to qualification of weapon systems and subsystems. There are certainly other possibilities for quantifying risk aversion.

Third, the risk tolerance of decision makers inherently involves comparison of the apparent or assessed risk with the requirement of acceptable risk from the perspective of the decision makers. As discussed previously, we reject the concept of incorporating requirements into the maturity assessment. As a result, the maturity ranking scale proposed here will not be based on risk or on the risk tolerance of the person or decision maker who uses the information.

Because of these challenges, we take an alternative path. The PCMM levels discussed here are based on two fundamental information attributes discussed by Wang and Strong (1996):

- *intrinsic information quality*: accuracy, correctness, and objectivity;
- *contextual information quality*: completeness, amount of information, and level of detail.

The use of maturity levels is an attempt to objectively track intellectual artifacts or evidence obtained in an assessment of a simulation effort. Any piece of information about the simulation effort can be considered an artifact. As one moves to higher levels of maturity, both the quality and the quantity of intrinsic and contextual information artifacts must increase. We define four levels with the general characteristics of maturity that apply to all six elements as follows.

- *Level 0* – Little or no assessment of the accuracy and/or completeness has been made; little or no evidence of maturity; individual judgment and experience only; convenience and expediency are the primary motivators. This level of maturity is commonly appropriate for low-consequence systems, systems with little reliance on simulation, scoping studies, or conceptual design support.
- *Level 1* – Some informal assessment of the accuracy and/or completeness has been made; generalized characterization; some evidence of maturity. This level of maturity is commonly appropriate for moderate consequence systems, systems with some reliance on simulation, or preliminary design support.
- *Level 2* – Some formal assessment of the accuracy and/or completeness has been made; detailed characterization; significant evidence of maturity; some assessments have been made by an internal peer review group. This level of maturity is commonly appropriate for high-consequence systems, systems with high reliance on simulation, system qualification support, or final design support.
- *Level 3* – Formal assessment of the accuracy and/or completeness has been made; precise and accurate characterization; detailed and complete evidence of maturity; essentially all assessments have been made by independent peer-review groups. This level of maturity is commonly appropriate for high-consequence systems in which decision making is fundamentally based on simulation, e.g., where certification or qualification of a system's performance, safety, and reliability is primarily based on simulation as opposed to being primarily based on information obtained from complete system testing.

We have not mentioned the roles or importance of reproducibility, traceability, and documentation of the artifacts. We have excluded these attributes because they do not directly measure the quality of the information produced. Rather, these attributes fundamentally contribute to (a) *proof* of the existence of the artifacts, and (b) the details of the information that is available for each element. We believe that reproducibility, traceability, and documentation of the artifacts are important in any simulation effort, particularly those that support certification or qualification of the safety and reliability of high-consequence systems. For example, the roles of reproducibility, traceability, and documentation of all artifacts produced by computational analyses in risk assessments for nuclear reactor safety, as well as in performance assessments for the Waste Isolation Pilot Plant (WIPP) and the Yucca Mountain Project, are well recognized and mandated by regulatory policy. Not withstanding this experience, the PCMM will exclude reproducibility, traceability, and documentation of the artifacts. If appropriate for the M&S effort, one could add an element that assessed the level of reproducibility, traceability, and documentation of all of the other six elements in the PCMM.

### *15.2.2 Purpose and uses of the PCMM*

For each of the six elements that are used in the PCMM, a maturity level is evaluated. The PCMM differs from NASA (2008) in that the NASA approach assesses the maturity level of the *results* from M&S. It is our view that assessing the maturity or credibility of

Table 15.1 *Table format for PCMM assessment (Oberkampf* et al.*, 2007).*

| Element \ Maturity | Maturity level 0 | Maturity level 1 | Maturity level 2 | Maturity level 3 |
|---|---|---|---|---|
| Representation and geometric fidelity | | | | |
| Physics and material model fidelity | | | | |
| Code verification | | | | |
| Solution verification | | | | |
| Model validation | | | | |
| Uncertainty quantification and sensitivity analysis | | | | |

results is actually a step removed, and more difficult, than assessing the maturity of the processes that occur in each of the elements. We argue that the *representational information quality*, specifically the interpretability and ease of understanding, is significantly improved if one focuses on the maturity of the contributing elements, and not on the broader issue of maturity of results.

The contributing elements and the maturity of each element can be thought of as relatively independent measures, or attributes, of predictive capability. Accordingly, the PCMM can be summarized in a table or matrix format, where the elements form the rows of the table and the maturity levels (0 through 3) form the columns, as shown in Table 15.1.

A PCMM assessment consists of evaluating the maturity level of six individual elements and scoring the maturity level of each element. For each level of maturity, there is a set of predefined descriptors that are used to assess the maturity level of a particular element. If an element is characterized by an assessor (an individual who performs the actual assessment of the maturity level for an element) as attaining the entire set of descriptors at a given level of maturity, the element can be considered to be fully assessed at that level of maturity. An element that is fully assessed at a particular level of maturity will generally be assigned, by the assessor, a score that is equivalent to the maturity level. Thus, for example, if an element were assessed so that it fully met all of the predefined descriptors at maturity level 1, the element would have a score of 1. In preliminary testing of the PCMM table over a range of engineering applications, we have commonly found that some, but not all, of the descriptors at a given level have been achieved. For example, at maturity level 2, only half the descriptors for a given attribute may have been attained. For this common situation, it has proven useful to give a fractional score for the maturity level instead of strictly assigning an integer score at the lower level of maturity. As a result, noninteger maturity scores expressed in tenths, such as 1.5 for partially achieving level 2, should be considered in assessing the maturity level of each element. Upon completion of the assessment, the table would have six individual scores, one score per element.

Before presenting additional details about the table constructed for the PCMM, it is appropriate to discuss more clearly the purpose of this table, certain characteristics of the

Table 15.2 *Example of a PCMM table after maturity assessment (Oberkampf* et al.*, 2007).*

| Element \ Maturity | Maturity level 0 | Maturity level 1 | Maturity level 2 | Maturity level 3 | *Element score* |
|---|---|---|---|---|---|
| Representation and geometric fidelity | | Assessed | | | 1 |
| Physics and material model fidelity | | | Assessed | | 2 |
| Code verification | | Assessed | | | 1 |
| Solution verification | Assessed | | | | 0 |
| Model validation | | Assessed | | | 1 |
| Uncertainty quantification and sensitivity analysis | Assessed | | | | 0 |

table, and how the results (i.e., scores) from completing this table can be used. Simply stated, the purpose of the table is to assess the level of maturity, at a given moment in time, of the key elements in a simulation effort that are directed at an application of interest. The assessment should be conducted, in principle, with little or no regard to any programmatic (or project) requirement for the maturity of the simulation effort. Objectivity, a key ingredient of intrinsic information quality, increases to the degree that maturity assessment is separated from maturity requirements specified by a project.

Table 15.2 gives an example of a PCMM table after a maturity assessment of a simulation effort has been completed.

For purposes of explanation, consider that all elements in Table 15.2 were assessed with the scores shown to the right of the table. The designator *Assessed* would then be placed in the appropriate row and column in the table. Once the assessment has been completed, the set of scores for each element can be compiled. In Table 15.2, the set of scores completed by the assessor(s) for the six elements is [1, 2, 1, 0, 1, 0].

We believe the type of summary information shown in Table 15.2 will prove very useful and informative in many programmatic and review committee situations. Following are some of the experiences derived in preliminary use of the PCMM table.

- In attempting to conduct a PCMM assessment, we found that the assessors are generally not familiar with many of the concepts in the table. By learning about these concepts, the assessors will greatly broaden and deepen their knowledge of many of the elements that contribute to confidence in simulation.
- Conducting a PCMM assessment and sharing it with interested parties, decision makers, and stakeholders engenders discussions that would *not* have occurred without such an assessment. This kind of communication is one of the most significant consequences of a simulation maturity assessment in general. An example of a beneficial interaction would be to initiate a conversation with a stakeholder who may not be familiar with any of the contributing elements to simulation

and help to educate that stakeholder about the importance of these elements and the results of the assessment.

- PCMM assessments made over a period of time can be used to track the progress of simulation efforts. This is useful for managers, stakeholders (decision makers using the results of the simulation effort), and funding sources to determine progress or value added over time.

A key practical issue in completion of the PCMM table is, *who* should provide the assessments in the table? We strongly believe that an individual, or a team, that has detailed knowledge of an element should complete that element of the table. These individuals should be very familiar with the elements of the simulation effort and the application of interest. Sometimes, depending on the magnitude of the simulation effort, a simulation project manager is sufficiently familiar with all elements of the table and can complete it in its entirety. The assessed levels in the completed table should represent the actual status of the simulation effort, not some anticipated or expected future status. In other words, the table should measure the maturity of the actual status at a given moment in time, not something that is *nearly* attained or a status that would *look good* in a program review or in a marketing activity.

With the PCMM, we are primarily interested in providing simulation maturity assessment information for an application of interest to program managers, relevant stakeholders, and decision makers. Some applications of interest that commonly involve simulation efforts are (a) design or optimization of new systems; (b) modification or optimization of existing systems; and (c) assessment of the performance, safety, or reliability of existing or proposed systems. In addition, the specification of a system includes the specification of the surroundings of the system and the environment in which the system must operate, e.g., normal, abnormal, or hostile environments. With the system, surroundings, and environment specified, one can then begin to identify particular aspects of each of the six elements that are important to the simulation effort.

An important aspect should be mentioned again concerning the interpretation of scores from a PCMM assessment. Although this aspect was discussed previously, it needs to be stressed and clarified further because it can cause great confusion, as we have seen in testing. We have observed that users of the PCMM commonly interpret an increase in maturity assessment over time to mean that the accuracy of the predictions has improved. This is *not necessarily* true. Stated differently, many people want to interpret the PCMM scores as a predictive accuracy assessment or, similarly, as a measure of the accuracy of the simulation results. As stressed earlier, the PCMM assesses the maturity of the simulation process elements, *not necessarily* the accuracy of the simulation results. The accuracy of the simulation results would commonly increase as the PCMM scores improve, but there is *not* a one-to-one correspondence.

To clarify why this is true, consider an example based on Table 15.2. As explained previously, the maturity level scores shown in Table 15.2 are written as the sequence [1, 2, 1, 0, 1, 0] for the six elements. Suppose that the element of uncertainty quantification and sensitivity analysis was improved from a 0 assessment (the last value in the maturity assessment above) to the condition where multiple simulations were obtained and resulted in capturing some of the uncertainties present in the system being analyzed. For example, suppose the uncertainty quantification analysis began to show a large effect due to variability in the strength of welded joints in a component. With this improved uncertainty quantification, suppose the maturity assessment of the PCMM then became [1, 2, 1, 0, 1, 1], i.e., the last value in the sequence changed from 0 to 1. The decision maker would then have more complete information about the system's uncertainty quantification. The decision maker would then have an estimate of the uncertainty of the SRQs of interest as a function of the variability in weld strength, whereas previously the decision maker may have had no idea of the uncertainty. While the accuracy of the predictions in these hypothetical cases has not changed at all, the decision maker would now be able to recognize some of the contributing uncertainties to the predicted performance of the system.

### 15.2.3 Characteristics of PCMM elements

A brief description of each element of the PCMM table is given in Table 15.3. This table can be used as summary statements of the basic descriptors of each element. Note that the requirements of the descriptors at each maturity level accumulate as one moves to higher maturity levels within an element. For example, to attain a given maturity level for a given element, the descriptors within the specific element of the table must be satisfied, in addition to all descriptors at the lower levels in that row.

A detailed discussion follows for each element of the table.

#### 15.2.3.1 Representation and geometric fidelity

This element is directed primarily toward the level of physical or informational characterization of the system being analyzed or the specification of the geometrical features of the system. For fully empirical and semi-empirical models, there may be little geometric fidelity, e.g., lumped-mass representations or representations that simply deal with the functionality of system components. For physics-based models that solve PDEs or integral equations, significant geometric fidelity may be specified that is then used to prescribe the ICs and BCs for such equations. For other mathematical models, such as electrical circuit or agent-based models, other concepts of representation fidelity are needed. For example, in the case of electrical circuit models, characterization deals with the fidelity of the electrical circuit diagram and level of characterization of the electrical components in the system. For agent-based models, for example, modeling the movement and interaction of robots, representation fidelity might be the geography over which agents move. Geometric fidelity may increase proportionately as physical modeling fidelity increases because of the additional information that is required for the modeling. Thus, the lowest level of maturity assesses geometric fidelity based on convenience, simplicity, and the judgment of the computational practitioner. The higher levels of geometric maturity provide increasingly detailed information that may be more representative of the *as built* geometry; accordingly, levels of stylization of the system and environment decrease. For example, the geometry, material and surface characteristics, and mechanical assembly of the system are typically

Table 15.3 *General descriptions for elements of the PCMM table (Oberkampf* et al.*, 2007).*

| Maturity / Element | **Maturity level 0** Low consequence, minimal simulation impact, e.g. scoping studies | **Maturity level 1** Moderate consequence, some simulation impact, e.g. design support | **Maturity level 2** High consequence, high simulation impact, e.g. qualification support | **Maturity level 3** High consequence, decision-making based on simulation, e.g. qualification or certification |
|---|---|---|---|---|
| **Representation and geometric fidelity** What features are neglected because of simplifications or stylizations? | • Judgment only<br>• Little or no representational or geometric fidelity for the system and BCs | • Significant simplification or stylization of the system and BCs<br>• Geometry or representation of major components is defined | • Limited simplification or stylization of major components and BCs<br>• Geometry or representation is well defined for major components and some minor components<br>• Some peer reviews conducted | • Essentially no simplification or stylization of components in the system and BCs<br>• Geometry or representation of all components is at the detail of *as built*, e.g., gaps, material interfaces, fasteners<br>• 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<br>• Model forms are either unknown or fully empirical<br>• Few, if any, physics-informed models<br>• No coupling of models | • Some models are physics based and are calibrated using data from related systems<br>• Minimal or ad hoc coupling of models | • Physics-based models for all important processes<br>• Significant calibration needed using separate effects tests (SETs) and integral effects tests (IETs)<br>• One-way coupling of models<br>• Some peer reviews conducted | • All models are physics based<br>• Minimal need for calibration using SETs and IETs<br>• Sound physical basis for extrapolation and coupling of models<br>• Full, two-way coupling of models<br>• Independent peer review conducted |
| **Code verification** Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results? | • Judgment only<br>• Minimal testing of any software elements<br>• Little or no SQE procedures specified or followed | • Code is managed by SQE procedures<br>• Unit and regression testing conducted<br>• Some comparisons made with benchmarks | • Some algorithms are tested to determine the observed order of numerical convergence<br>• Some features & capabilities (F&C) are tested with benchmark solutions<br>• Some peer reviews conducted | • All important algorithms are tested to determine the observed order of numerical convergence<br>• All important F&Cs are tested with rigorous benchmark solutions<br>• Independent peer review conducted |


https://doi.org/10.1017/CBO9780511760396.021

| | | | | |
|---|---|---|---|---|
| **Solution verification** Are numerical solution errors and human procedural errors corrupting the simulation results? | • Judgment only<br>• Numerical errors have an unknown or large effect on simulation results | • Numerical effects on relevant SRQs are qualitatively estimated<br>• Input/output (I/O) verified only by the analysts | • Numerical effects are quantitatively estimated to be small on some SRQs<br>• I/O independently verified<br>• Some peer reviews conducted | • Numerical effects are determined to be small on all important SRQs<br>• Important simulations are independently reproduced<br>• 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<br>• 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<br>• Large or unknown experimental uncertainties | • Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs<br>• Experimental uncertainties are well characterized for most SETs, but poorly known for IETs<br>• Some peer reviews conducted | • Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application<br>• Experimental uncertainties are well characterized for all IETs and SETs<br>• Independent peer review conducted |
| **Uncertainty quantification and sensitivity analysis** How thoroughly are uncertainties and sensitivities characterized and propagated? | • Judgment only<br>• Only deterministic analyses are conducted<br>• Uncertainties and sensitivities are not addressed | • Aleatory and epistemic (A&E) uncertainties propagated, but without distinction<br>• Informal sensitivity studies conducted<br>• Many strong UQ/SA assumptions made | • A&E uncertainties segregated, propagated and identified in SRQs<br>• Quantitative sensitivity analyses conducted for most parameters<br>• Numerical propagation errors are estimated and their effect known<br>• Some strong assumptions made<br>• Some peer reviews conducted | • A&E uncertainties comprehensively treated and properly interpreted<br>• Comprehensive sensitivity analyses conducted for parameters and models<br>• Numerical propagation errors are demonstrated to be small<br>• No significant UQ/SA assumptions made<br>• Independent peer review conducted |

specified in a CAD file. For systems that may be in a dangerous state of excessive wear, a damaged condition, or crack propagation due to cyclic loading, the specification of the geometry and surface properties can become quite complex and uncertain.

General descriptions of the levels of physical representation and geometric fidelity follow.

- *Level 0*: Simplicity, convenience, and functional operation of the system dominate the fidelity of the representation and the geometry for the system being analyzed. There is heavy reliance on judgment and experience, with little or no expectation or quantification of representation and geometric fidelity.
- *Level 1*: Quantitative specifications are applied to describe the geometry of the major components of the system being analyzed. Much of the real system remains stylized or ignored, e.g., gaps between components, changes in materials, and surface finish.
- *Level 2*: Quantitative specifications are applied to replicate the geometric fidelity of most of the components of the real system. Little of the real system remains stylized or ignored. For example, important imperfections due to system assembly or defects due to wear or damage in the system are included. A level of peer review, such as informal or internal reviews, of the model representation and geometric fidelity has been conducted.
- *Level 3*: The geometric representation in the model is *as built* or *as existing*, meaning that no aspect of the geometry of the modeled real system is missing, down to scales that are determined to be relevant to the level of physical modeling chosen. An example is a complete CAD model for the real system as assembled and meshed for the discrete model with no significant approximations or simplifications included. Independent peer review of the model representation and geometric fidelity has been conducted, e.g., formal review by the simulation effort customer or by reviewers external to the organization conducting the simulation.

#### *15.2.3.2 Physics and material model fidelity*

This attribute primarily addresses the following:

- the degree that models are physics based, i.e., are they fully empirical, semi-empirical, or physics based;
- the degree to which the models are calibrated;
- the physics fidelity basis with which the models are being extrapolated from their validation and calibration database to the conditions of the application of interest;
- the quality and degree of coupling the multi-physics effects that exist in the application of interest.

Generally, as the physical fidelity of the model increases, and the needed input data is provided to the model for a simulation, the model is increasingly able to provide physics-based explanatory power for the particular physical phenomenon of interest. Within the broad class of physics-based models, there are important distinctions in the degree to which the model is calibrated. For example, does the model require recalibration (updating) even if there are relatively small changes in the system design or small changes in the BCs or ICs? Alternately, does the model require calibration only at lower levels in the validation hierarchy, i.e., separate effects tests (SETs), in order to yield accurate predictions? Or, does the model also require calibration or recalibration at higher levels of the validation hierarchy, i.e., integral effects tests (IETs), to attain accurate predictions? For two models yielding *the same* level of agreement with experimentally measured responses at the IET level, one model calibrated only at the SET level and one model calibrated at the IET level, the model that requires calibration at the SET level has more predictive capability than does the model that requires calibration at the IET level. This statement is understandable by noting that the model calibrated with SETs has demonstrated that it can yield the same prediction accuracy as the model calibrated with IETs, even though the SET-calibrated model must extrapolate further from its calibration domain than the IET-calibrated model.

General descriptions of the various levels of physics and material model fidelity follow.

- *Level 0*: The model is fully empirical, or the model form is not known. There is little or no coupling of models representing multiple functional elements of the system, and the coupling that does exist is not physics based. Confidence in the model is strictly based on the judgment and experience of the practitioner.
- *Level 1*: The model is semi-empirical in the sense that portions of the modeling are physics based; however, important features, capabilities, or parameters in the model are calibrated using data from very closely related physical systems. The coupling of functional elements or components is minimal or ad hoc, and not based on detailed physics.
- *Level 2*: All important physical process models and material models are physics based. Calibration of important model parameters is necessary, using data from SETs and IETs. All model calibration procedures are implemented on the model input parameters, not on the SRQs. Important physical processes are coupled using physics-based models with couplings in one direction. Some level of peer review, such as informal or internal reviews, of the physics and material models has been conducted.
- *Level 3*: All models are physics based with minimal need for calibration using SETs and IETs. Where extrapolation of these models is required, the extrapolation is based on well understood and well accepted physical principles. All physical processes are coupled in terms of physics-based models with two-way coupling and physical process effects on physical and material parameters, BCs, geometry, ICs, and forcing functions. Independent peer review of the physics and material models have been conducted, e.g., formal review by the simulation effort customer or by reviewers external to the organization conducting the simulation.

#### *15.2.3.3 Code verification*

This attribute focuses on the following:

- correctness and fidelity of the numerical algorithms used in the code relative to the mathematical model, e.g., the PDEs;
- correctness of source code;
- configuration management, control, and testing of the software through SQE practices.

The correctness and fidelity of the numerical algorithms and the correctness of the source code are primarily determined by conducting various types of tests on the code. The primary type of test we advocate compares the numerical solution results from the code with highly accurate solutions, which are usually referred to as benchmark numerical solutions (Oberkampf and Trucano, 2008). The most rigorous benchmark solutions are manufactured and analytical solutions discussed in detail in Chapter 6, Exact solutions.

Comparisons between the results of the code being tested and the benchmark solutions typically yield two types of numerical accuracy information. First, the error in the SRQs from the code being tested is evaluated by using the benchmark result as the exact solution. Although this is useful, it provides no information concerning the numerical convergence characteristics of the code being tested. Second, using two or more solutions with uniformly refined discretizations for the code being tested, and using the benchmark as the exact solution, one can compute the observed order of convergence of the numerical algorithms in the code being tested. Observed order of convergence is a much more definitive statement of code verification.

The maturity of the SQE practices should measure the scope and rigor of configuration management and software control. Chapter 4, Software engineering gave a detailed discussion of this topic.

General descriptions of the levels of code verification are as follows.

- *Level 0*: Code verification is based almost entirely on the judgment and experience of the computational practitioners involved. There is little or no formal verification testing of the software elements. Little or no SQE practices are defined and practiced in the implementation, management, and use of the code.
- *Level 1*: Most associated software is implemented and managed with formal SQE practices. Unit and regression testing of the software is conducted regularly with a high percentage of line coverage attained. Verification test suites using benchmark solutions are minimal and only error measures are obtained in some SRQs.
- *Level 2*: All associated software is implemented and managed with formal SQE practices. Verification test suites are formally defined and systematically applied using benchmark solutions to compute the observed order of convergence of some numerical algorithms. Some features and capabilities (F&Cs), such as complex geometries, mesh generation, physics, and material models, have been tested with benchmark solutions. Some level of peer review, such as informal or internal reviews, of the code verification has been conducted.
- *Level 3*: All important algorithms have been tested using rigorous benchmark solutions to compute the observed order of convergence. All important F&Cs, such as two-way coupling of multiphysics processes, have been tested with rigorous benchmark solutions. Independent peer review of code verification has been conducted, e.g., formal review by the simulation effort customer or by reviewers external to the organization conducting the simulation.

### *15.2.3.4 Solution verification*

This attribute deals with assessment of the following:

- numerical solution errors in the computed results;
- confidence in the computational results as they may be affected by human error.

Rigor and numerical solution accuracy are the dominant components of the assessment of this element. Numerical solution errors are any errors due to mapping the mathematical model to the discretized model and any errors due to solution of the discretized model on a computer. Of concern in this element are numerical solution errors due to spatial and temporal discretization of the PDEs or integral equations, and the iterative solution error due to a linearized solution approach to a set of nonlinear discretized equations. Additional numerical solution errors that should be addressed are the potential detrimental effects of numerical parameters in solution algorithms; errors due to approximate techniques used to solve nondeterministic systems, e.g., error due to a small number of samples used in a Monte Carlo sampling method; and round-off error due to finite precision on a computer. Human errors, i.e., blind uncertainties, are also a concern in the assessment of this element, such as those made in (a) preparing and assembling the elements of the discrete model; (b) executing the computational solution; and (c) post-processing, preparing, or interpreting the computational results.

General descriptions of the levels of solution verification are as follows.

- *Level 0*: No formal attempt is made to assess any of the possible sources of numerical error. Any statement about the impact of numerical error is based purely on the judgment and experience of the computational practitioner. No assessment about the correctness of software inputs or outputs has been conducted.
- *Level 1*: Some kind of formal method is used to assess the influence of numerical errors on some SRQs. This could include *a posteriori* error estimation of global norms, iterative convergence studies, or sensitivity studies to determine how sensitive certain SRQs are to changes in mesh or temporal discretization. A formal effort is made by the computational practitioners to check the correctness of input/output (I/O) data.
- *Level 2*: Quantitative error estimation methods are used to estimate numerical errors on some SRQs, and these estimates show that the errors are small for some conditions of the application of interest. I/O quantities have been verified by knowledgeable computational practitioners who have some level of independence from the simulation effort. Some level of peer review, such as informal or internal reviews, of the solution verification activities has been conducted.
- *Level 3*: Quantitative error estimation methods are used to estimate numerical errors on all important SRQs, and these estimates show that the errors are small over the entire range of conditions for the application of interest. Important simulations are reproduced, using the same software, by independent computational practitioners. Independent peer review of solution verification activities has been conducted, e.g., formal review by the simulation effort customer or by reviewers external to the organization conducting the simulation.

A subtle, but important, point should be stressed regarding the maturity levels of solution verification. It was pointed out that higher levels of maturity do not *necessarily* imply higher levels of accuracy of the simulation results. However, in the descriptions of maturity levels just given it is apparent that higher levels of maturity *require* increased solution accuracy. This apparent dichotomy is resolved by understanding that increased numerical solution accuracy is necessary to gain more confidence in the fidelity of the mapping of the mathematical model to the solution of the discrete model. We are *not* necessarily gaining confidence in the comparison of the computational results with experimental data. In other words, we require increased correctness and accuracy of the numerical solution, including code verification, so that when we compare computational results and experimental results we are confident that we are indeed comparing the physics as simulated by the mathematical model with nature's reflection of reality in experimental measurements. High levels of maturity in code verification and solution verification are an indication of how well the numerical results represent the physics in the mathematical model, as opposed to a contaminated mixture of physics, numerical error, and possibly human error. If we cannot have confidence in what we believe we are comparing, then we are dealing with a convolved mixture of physics modeling, physics modeling approximations (error), and numerical error, in which no bases for confidence can be made. It will be seen below that more accurate comparisons between the computational results and the experimental measurements, although desired, are not *necessarily* required to achieve higher maturity levels in model validation.

#### *15.2.3.5 Model validation*

This attribute focuses on the following:

- thoroughness and precision of the accuracy assessment of the computational results relative to the experimental measurements;
- completeness and precision of the characterization of the experimental conditions and measurements;
- relevancy of the experimental conditions, physical hardware, and measurements in the validation experiments compared to the application of interest.

The focus of model validation in the PCMM is on the precision and completeness of the process of the model accuracy assessment, *not* on the accuracy of the mathematical model itself. By *precision* of validation we mean (a) how carefully and accurately are the experimental uncertainties estimated and (b) how well understood and quantified are all the conditions of the experiment that are required as inputs for the mathematical model? By *completeness* of validation we mean: how well do the conditions (geometry, BCs, ICs, and system excitation) and actual physical hardware of the validation experiments conducted relate to the actual conditions and hardware of the application of interest?

For SETs, it is expected that there will be many dissimilarities between the SET experiments and the actual application of interest because they occur at lower levels in the validation hierarchy. For IETs, however, there should be a close relationship between the IET experiments and the application of interest, particularly with respect to the experimental hardware and the coupled physical phenomena occurring in each. For a more complete discussion of the concepts behind and the construction of a validation hierarchy, see Chapter 10, Model validation fundamentals.

As discussed earlier, the correctness and credibility of model validation fundamentally relies on assumptions that the numerical algorithms are reliable, that the computer program is correct, that no human procedural errors have been made in the simulation, and that the numerical solution error is small. These are major assumptions that we, and many others, have discovered are commonly unfounded. Consequently, to properly inform the user of the information in the PCMM table about the veracity of these assumptions, we require that the maturity level of the elements model validation and uncertainty quantification and sensitivity analysis can be *no higher than two levels above* the maturity levels of the minimum of code verification and solution verification. This requirement places further restrictions on conducting the PCMM assessment and means that *the maturity levels of code verification and solution verification must be assessed before the maturity levels of model validation and of uncertainty quantification and sensitivity analysis are assessed.* As an example of the dependencies between elements, assume that, as discussed in Table 15.2, code verification and solution verification were at levels 1 and 0, respectively. Consequently, the maximum maturity level that the model validation element and the uncertainty quantification and sensitivity analysis element could be is level 2, even if the assessor(s) were to independently judge either or both of these elements at a level higher than 2.

General descriptions of the various levels of model validation are as follows.

- *Level 0*: Accuracy assessment of the model is based almost entirely on judgment and experience. Few, if any, comparisons have been made between computational results and experimental measurements of similar systems of interest.
- *Level 1*: Limited quantitative comparisons are made between computational results and experimental results. Either comparisons for SRQs have been made that are not directly relevant to the application of interest, or the experimental conditions are not directly relevant to the application of interest. Experimental uncertainties, either in the SRQs and/or in the characterization of the conditions of the experiment, are largely undetermined, unmeasured, or based on experience.
- *Level 2*: Quantitative comparisons between computational results and experimental results have been made for some key SRQs from SET experiments and limited IET experiments. Experimental uncertainties are well characterized (a) for most SRQs of interest and (b) for experimental conditions for the SETs conducted. However, the experimental uncertainties are not well characterized for the IETs. Some level of peer review, such as informal or internal reviews, of the model validation activities has been conducted.
- *Level 3*: Quantitative comparisons between computational and experimental results have been made for all important SRQs from an extensive database of both SET and IET experiments. The conditions of the SETs should be relevant to the application of interest; and the conditions, hardware, and coupled physics of the IETs should be similar to the application of interest. Some of the SET computational predictions and most of the IET predictions should be *blind*. Experimental uncertainties and conditions are well characterized for SRQs in both the SET and IET experiments. Independent peer review of the model validation activities has been conducted, e.g., formal review by the simulation effort customer or by reviewers external to the organization conducting the simulation.

#### *15.2.3.6 Uncertainty quantification and sensitivity analysis*

This attribute focuses on the following:

- thoroughness and soundness of the uncertainty quantification (UQ) effort, including identification and characterization of all plausible sources of uncertainty;
- accuracy and correctness of propagating uncertainties through a mathematical model and interpreting uncertainties in the SRQs of interest;
- thoroughness and precision of a sensitivity analysis to determine the most important contributors to uncertainty in system responses.

Recognition of uncertainties refers to the activity of identifying and understanding all possible uncertainties within the system of interest (e.g., physical parametric uncertainty and uncertainties in the geometry), in the surroundings (e.g., BCs and system excitation), and in the environment (e.g., normal, abnormal, and hostile). Characterization of model predictive uncertainty primarily deals with the proper estimation and representation of all uncertainties that could exist as part of the prediction for the system of interest. A key aspect of characterization is the segregation of uncertainties into aleatory and epistemic elements. This segregation has been discussed at length in this book.

A sensitivity analysis (SA) provides additional important information to the user of the computational simulation analysis beyond what is typically considered a part of an uncertainty quantification analysis. SA was briefly discussed in Chapter 13, Predictive capability, but for a detailed discussion, see Helton *et al.* (2006) and Saltelli *et al.* (2008). A SA is typically directed at two closely related goals. First, one may be interested in determining how outputs locally change as a function of inputs. This is usually referred to as a local SA. The information obtained from a local SA is commonly used for system design and optimization, as well as for determination of the most advantageous operational conditions for maximizing system performance. Second, one may be interested in determining how the uncertainty structure of all of the inputs maps to the uncertainty structure of each of the outputs. This is usually referred to as a global SA. The information from a global SA may be used, for example, to determine which manufacturing variabilities contribute most to variability in certain SRQs, or to determine what physical experiments should be conducted to most reduce the epistemic uncertainty that is due to poorly understood coupled-physics phenomena.

As discussed with regard to model validation, the maturity level of the UQ and SA elements can be *no higher than two levels above the maturity levels of the minimum of code verification and solution verification*.

General descriptions of the various levels of UQ and SA are as follows.

- *Level 0*: Judgment and experience are dominant forms of uncertainty assessment. Only deterministic analyses were conducted for the system of interest. Informal *spot checks* or *what-if* studies for various conditions were conducted to determine their effect.
- *Level 1*: Uncertainties in the system of interest are identified, represented, and propagated through the mathematical model, but they are not segregated with respect to whether the uncertainties are aleatory or epistemic. Sensitivity of some system responses to some system uncertainties and environmental condition uncertainties was investigated, but the sensitivity analysis was primarily informal or exploratory rather than systematic. Many strong assumptions are made with respect to the UQ and SA; for example, most probability density functions are characterized as Gaussian, and uncertain parameters are assumed to be independent of all other parameters.
- *Level 2*: Uncertainties in the system of interest are characterized as either aleatory or epistemic. The uncertainties are propagated through the computational model, while their character is kept segregated both in the input and in the SRQs. Quantitative SAs were conducted for most system parameters, while segregating aleatory and epistemic uncertainties. Numerical approximation or sampling errors due to propagation of uncertainties through the model are estimated, and the

effect of these errors on the UQ and SA results is understood and/or qualitatively estimated. Some strong UQ and SA assumptions were made, but qualitative results suggest that the effect of these assumptions is not significant. Some level of peer review, such as informal or internal reviews, of the uncertainty quantification and sensitivity analyses has been conducted.

- *Level 3*: Aleatory and epistemic uncertainties are comprehensively treated, and their segregation in the interpretation of the results is strictly maintained. Detailed investigations were conducted to determine the effect of uncertainty introduced due to model extrapolations (if required) to the conditions of the system of interest. A comprehensive SA was conducted for both parametric uncertainty and model uncertainty. Numerical approximation or sampling errors due to propagation of uncertainties through the model are carefully estimated, and their effect on the UQ and SA results is demonstrated to be small. No significant UQ and SA assumptions were made. Independent peer review of UQ and SA have been conducted, e.g., formal review by the simulation effort customer or by reviewers external to the organization conducting the simulation.

## 15.3 Additional uses of the PCMM

In this section, we suggest additional ways that the PCMM can be used and propose a method for the aggregation of scores in the PCMM table should that information be required. We also point out that the PCMM is only one of many factors that contribute to risk-informed decision making.

### *15.3.1 Requirements for modeling and simulation maturity*

After an objective assessment of M&S maturity has been made using the PCMM table, one can introduce the *project maturity requirements* for each element in the table. Six project maturity requirements can be specified, one for each element in the table. Project maturity requirements may be a result of, for example, system qualification or regulatory requirements, or they may simply be progress requirements for the development of a simulation effort. For this exercise, the essential question to ask for each element is: what should the appropriate level of maturity be for my intended use of the simulation activity? For example, a given element in the table has been assessed at a maturity level of 2. Is that an appropriate level for which the simulation information will be used, or should it be at a higher level? Although we have not discussed this issue, it is obvious that the costs, both in terms of time and resources, increase significantly as higher levels of maturity are attained. To determine the project maturity requirements, one uses the same descriptors in Table 15.3 that were used to complete the PCMM table. For this use of Table 15.3, we consider the descriptors to be project maturity requirements.

Table 15.4 depicts the results of specifying project maturity requirements for each of the assessed elements discussed earlier. The designator *Required* is used to indicate the project maturity requirement for each element. The scores for the project maturity requirements in this example are [2, 2, 1, 2, 2, 3].

Table 15.4 *Example of PCMM table assessment and project maturity requirements (Oberkampf* et al.*, 2007).*

| Element \ Maturity | Maturity level 0 | Maturity level 1 | Maturity level 2 | Maturity level 3 |
|---|---|---|---|---|
| Representation and geometric fidelity | | Assessed | Required | |
| Physics and material model fidelity | | | Assessed Required | |
| Code verification | | Assessed Required | | |
| Solution verification | Assessed | | Required | |
| Model validation | | Assessed | Required | |
| Uncertainty quantification and sensitivity analysis | Assessed | | | Required |

In an assessment such as Table 15.4, the values would be color coded and have the following meanings:

- green – the assessment meets or exceeds the requirement (rows 2 and 3);
- yellow – the assessment does not meet the requirement by one level or less (rows 1 and 5);
- pink – the assessment does not meet the requirement by two levels or less (row 4);
- red – the assessment does not meet the requirement by three levels or less (row 6).

Some examples of the useful benefits of comparisons of simulation maturity and simulation project maturity requirements, as shown in Table 15.4, follow.

- To construct Table 15.4, one must have already addressed the question: what are the project requirements for simulation maturity? In our experience, we have found that either (a) this question may not have been asked or (b) answering this question has proven difficult, but quite useful in its own right. If this question is asked, we have found that it initiates conversations not only within the simulation customer's organization (typically engineering design groups or decision makers) but also between the simulation developer, the customer, and the stakeholders. We have found that this conversation is particularly important when the simulation customer is not the source of funding for the simulation effort.
- Table 15.4 can be used as a project management tool to adjust resources for elements that are lagging in their progress to meet project schedule requirements. Note that some elements do not depend solely on computational or software issues. For example, the model validation element depends very heavily on capabilities and progress in experimental activities. We have found that one of the most common and damaging difficulties is the technical, scheduling, and/or funding disconnection between the computational and experimental activities in validation.

### 15.3.2 Aggregation of PCMM scores

The description of the PCMM has focused on the use of simulation for a particular engineering application. Situations can exist where PCMM scores will need to be aggregated into one score, such as the following.

- Suppose one has obtained a set of scores for multiple subsystems within a system, each subsystem represented by six scores. The desire is to aggregate all of the scores for the multiple subsystems into a single score for all of the subsystems.
- Suppose one has obtained a set of scores for multiple systems of different design, and each system is represented by six scores. The desire is to aggregate all of the scores for the multiple systems into one score that would represent, in some sense, a single score for the collection of systems or the system of systems.

Although we recognize that arguments can be made to compute PCMM aggregate scores, we strongly recommend that this *not* be done. The score assessed for each of the six simulation elements is an ordinal scale – the four levels of maturity constitute a total order because each pair of levels can be simply ordered. However, the six elements *cannot be collectively ordered in any way*; they are apples and oranges. Each element is important and conceptually independent from each other element. If one argues that an average maturity of a simulation effort could be computed by simply taking the arithmetic mean of each of the six elements, the average value would have little meaning. The argument for using the average value would be analogous to someone claiming to compute the breaking strength of a chain by averaging the strength of each link in the chain. It is a fallacious argument.

Even though we argue against any type of aggregation method, our experience with using the PCMM has shown that pressure to condense information for decision makers can be irresistible. Given this reality, we recommend a simple procedure that would aid in maintaining some of the key information in the individual PCMM scores. We recommend that a set of three scores *always* be computed and presented to the user of the PCMM when *any* aggregation of PCMM scores is computed. The scores consist of the minimum over all of the elements being aggregated, the average of all the elements, and the maximum of all the elements. This aggregation triple can be written as:

$$\widehat{\mathrm{PCMM}} = \left[ \min_{i=1,2,\ldots,n} \mathrm{PCMM}_i,\ \frac{1}{n}\sum_{i=1}^{n} \mathrm{PCMM}_i,\ \max_{i=1,2,\ldots,n} \mathrm{PCMM}_i \right], \tag{15.1}$$

where $n$ is the total number of individual PCMM scores that are being aggregated. We believe that keeping the worst score of all aggregated scores will call attention to the situation so that the decision maker can pursue the issue in more depth if desired.

As an example, suppose that a system was made up of four subsystems. Assume each subsystem was assessed using the PCMM table discussed above, with the following result:

$$\mathrm{PCMM}_{\mathrm{subsystem1}} = \begin{bmatrix} 1 \\ 1.5 \\ 1 \\ 0 \\ 0.5 \\ 1 \end{bmatrix}, \quad \mathrm{PCMM}_{\mathrm{subsystem2}} = \begin{bmatrix} 1.5 \\ 1 \\ 0 \\ 0.5 \\ 1.5 \\ 0 \end{bmatrix},$$

$$\mathrm{PCMM}_{\mathrm{subsystem3}} = \begin{bmatrix} 2 \\ 1.5 \\ 0.5 \\ 1 \\ 1.5 \\ 1 \end{bmatrix}, \quad \mathrm{PCMM}_{\mathrm{subsystem4}} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0.5 \\ 1.5 \\ 1.5 \end{bmatrix}. \tag{15.2}$$

Using Eqs. (15.1) and (15.2), we compute the PCMM aggregate triple:

$$\widehat{\mathrm{PCMM}} = [0.0,\, 1.1,\, 2.0]\,. \tag{15.3}$$

This example demonstrates what we have observed in preliminary use of the PCMM: there is commonly a surprisingly wide range of scores uncovered in assessments. As an aside, it has also been found that sometimes there is resistance to conduct a PCMM assessment because it is suspected beforehand that the result will be similar to Eq. (15.3).

### *15.3.3 Use of the PCMM in risk-informed decision making*

As we have discussed, the PCMM is carefully and narrowly focused so that it can be properly understood and correctly used by computational practitioners, experimentalists, project managers, decision makers, and policy makers. Earlier, we suggested some ways in which the PCMM could be used to assess progress, used as a project-planning tool for both simulation and experimental activities, and used by consumers of the simulation information. In the larger context, however, the PCMM is only one factor that contributes to risk-informed decision making for engineering systems. Figure 15.2 depicts a number of factors that could affect the risk-informed decision making for an engineering system.

Figure 15.2 divides the factors into two major groups: technical issues and programmatic issues. Although not all factors are shown, the figure demonstrates that a number of diverse and complex factors are important in decision making. Sometimes individual technical factors are characterized fairly well. For example, required system performance and predicted system performance, say, for system reliability in normal operating conditions, might be mathematically characterized as a precisely known probability distribution. However, most of the factors in Figure 15.2, particularly programmatic issues, are not characterized well, or at all. For example, it is commonly very difficult to estimate the consequences of poor system reliability on financial liability and future business opportunities. As depicted, there are interactions and trade-offs between the two groups of issues and within each group.

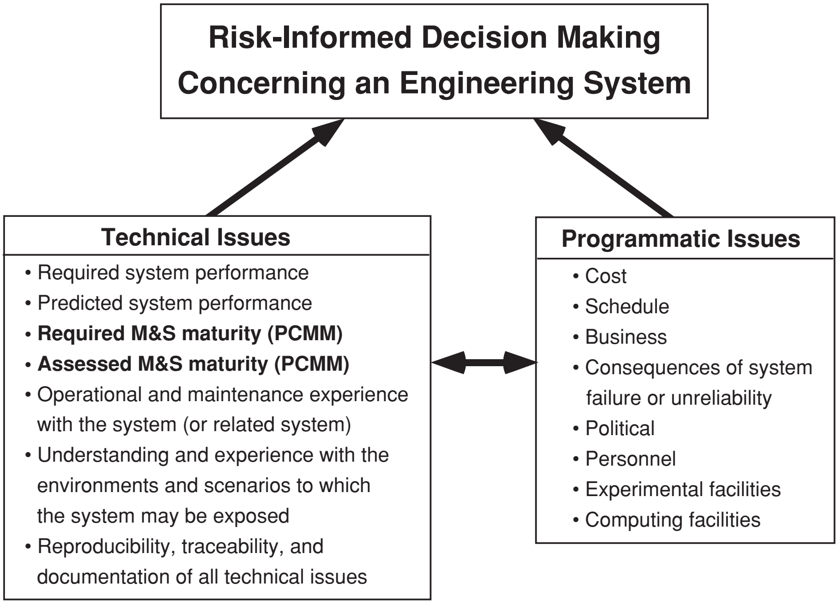


Figure 15.2 Factors influencing risk-informed decision making (Oberkampf *et al.*, 2007).

Managers and decision makers must weigh the importance of all factors, try to understand the complex interactions of factors, and decide on the trade-offs that must be made to optimize their view and their organization's view of success. Of course, success can mean widely varying things to the various participants and stakeholders involved.

Our purpose in constructing and discussing Figure 15.2 is to make it clear how the PCMM is but one factor in a complex set of factors. We have argued here that the assessment of simulation maturity is a relatively new factor that should be explicitly included in risk-informed decision making. In addition, we have argued that the assessment should be clearly separated from other important factors in decision making. If this is not done, there will be, at best, a convolution of factors causing confusion and miscommunication and, at worst, a contortion of factors intended to satisfy various agendas of individuals and organizations involved.

## 15.4 References

Ahern, D. M., A. Clouse, and R. Turner (2005). *CMMI Distilled: a Practical Introduction to Integrated Process Improvement*. 2nd edn., Boston, MA, Addison-Wesley.

Balci, O. (2004). Quality assessment, verification, and validation of modeling and simulation applications. *2004 Winter Simulation Conference*, 122–129.

Balci, O., R. J. Adams, D. S. Myers, and R. E. Nance (2002). A collaborative evaluation environment for credibility assessment of modeling and simulation applications. *2002 Winter Simulation Conference*, 214–220.

Bertch, W. J., T. A. Zang, and M. J. Steele (2008). Development of NASA's models and simulations standard. *2008 Spring Simulation Interoperability Workshop* Paper No. 08S-SIW-037, Providence, RI, Simulation Interoperability Standards Organization.

Blattnig, S. R., L. L. Green, J. M. Luckring, J. H. Morrison, R. K. Tripathi, and T. A. Zang (2008). Towards a credibility assessment of models and simulations. *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA 2008–2156, Schaumburg, IL, American Institute of Aeronautics and Astronautics.

Chrissis, M. B., M. Konrad, and S. Shrum (2007). *CMMI: Guidelines for Process Integration and Product Improvement*. 2nd edn., Boston, MA, Addison-Wesley.

Clay, R. L., S. J. Marburger, M. S. Shneider, and T. G. Trucano (2007). *Modeling and Simulation Technology Readiness Levels*. SAND2007–0570, Albuquerque, NM, Sandia National Laboratories.

DoD (2005). *Technology Readiness Assessment (TRA) Deskbook*. Washington, DC, Department of Defense.

GAO (1999). *Best Practices: Better Management of Technology Development Can Improve Weapon System Outcomes*. GAO/NSIAD-99–162, Washington, DC, U. S. Government Accounting Office.

Garcia, S. and R. Turner (2006). *CMMI Survival Guide: Just Enough Process Improvement*, Boston, MA, Addison-Wesley.

Green, L. L., S. R. Blattnig, J. M. Luckring, and R. K. Tripathi (2008). An uncertainty structure matrix for models and simulations. *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA 2008–2154, Schaumburg, IL, American Institute of Aeronautics and Astronautics.

Harmon, S. Y. and S. M. Youngblood (2003). A proposed model for simulation validation process maturity. *Simulation Interoperability Workshop*, Paper No. 03S-SIW-127, Orlando, FL, Simulation Interoperability Standards Organization.

Harmon, S. Y. and S. M. Youngblood (2005). A proposed model for simulation validation process maturity. *The Journal of Defense Modeling and Simulation*. **2**(4), 179–190.

Helton, J. C., J. D. Johnson, C. J. Sallaberry, and C. B. Storlie (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering and System Safety*. **91**(10–11), 1175–1209.

IEEE (1989). *IEEE Standard Glossary of Modeling and Simulation Terminology*. Std 610.3–1989, New York, IEEE.

Mankins, J. C. (1995). *Technology Readiness Levels*. Washington, DC, National Aeronautics and Space Administration.

NASA (2006). *Interim Technical Standard for Models and Simulations*. NASA-STD-(I)-7009, Washington, DC, National Aeronautics and Space Administration.

NASA (2008). *Standard for Models and Simulations*. NASA-STD-7009, Washington, DC, National Aeronautics and Space Administration.

Oberkampf, W. L., M. Pilch, and T. G. Trucano. (2007). *Predictive Capability Maturity Model for Computational Modeling and Simulation*. SAND2007–5948, Albuquerque, NM, Sandia National Laboratories.

Oberkampf, W. L. and T. G. Trucano (2008). Verification and validation benchmarks. *Nuclear Engineering and Design*. **238**(3), 716–743.

Pilch, M., T. G. Trucano, D. E. Peercy, A. L. Hodges, and G. K. Froehlich (2004). *Concepts for Stockpile Computing (OUO)*. SAND2004–2479 (Restricted Distribution, Official Use Only), Albuquerque, NM, Sandia National Laboratories.

Pilch, M., T. G. Trucano, and J. C. Helton (2006). *Ideas Underlying Quantification of Margins and Uncertainties (QMU): a White Paper*. SAND2006–5001, Albuquerque, NM, Sandia National Laboratories.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis: the Primer*, Hoboken, NJ, Wiley.

SEI (2006). *Software Engineering Institute: Capability Maturity Model Integration*. www.sei.cmu.edu/cmmi/.

Smith, J. (2004). *An Alternative to Technology Readiness Levels for Non-Developmental Item (NDI) Software*. CMU/SEI-2004-TR-013, ESC-TR-2004–013, Pittsburgh, PA, Carnegie Mellon, Software Engineering Institute.

Steele, M. J. (2008). Dimensions of credibility in models and simulations. *International Simulation Multi-Conference*, Paper No. 08E-SIW-076, Edinburgh, Scotland, Simulation Interoperability Standards Organization/The Society for Modeling and Simulation.

Wang, R. Y. and D. M. Strong (1996). Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems*. **12**(4), 5–34.

West, M. (2004). *Real Process Improvement Using the CMMI*, Boca Raton, FL, CRC Press.