General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

Dimensions of Credibility in Models and Simulations

Martin J. Steele, Ph.D.
National Aeronautics and Space Administration (NASA)
IT-C1 / Kennedy Space Center, FL 32899
Martin.J.Steele@nasa.gov

Abstract

Based on the National Aeronautics and Space Administration's (NASA's) work in developing a standard for models and simulations (M&S), the subject of credibility in M&S became a distinct focus. This is an indirect result from the Space Shuttle Columbia Accident Investigation Board (CAIB), which eventually resulted in an action, among others, to improve the rigor in NASA's M&S practices. The focus of this action came to mean a standardized method for assessing and reporting results from any type of M&S. As is typical in the standards development process, this necessarily developed into defining a common terminology base, common documentation requirements (especially for M&S used in critical decision making), and a method for assessing the credibility of M&S results. What surfaced in the development of the NASA Standard was the various dimensions credibility to consider when accepting the results from any model or simulation analysis. The eight generally relevant factors of credibility chosen in the NASA Standard proved only one aspect in the dimensionality of M&S credibility. At the next level of detail, the full comprehension of some of the factors requires an understanding along a couple of dimensions as well. Included in this discussion are the prerequisites for the appropriate use of a given M&S, the choice of factors in credibility assessment with their inherent dimensionality, and minimum requirements for fully reporting M&S results.

Dimensions of Credibility in Models and Simulations

Martin J. Steele, Ph.D.
National Aeronautics and Space Administration (NASA)
IT-C1 / Kennedy Space Center, FL 32899
Martin.J.Steele@nasa.gov

Keywords: Credibility, Models, NASA, Validation, Verification, Simulations, Standard, Uncertainty, Robustness

Abstract
Based on the National Aeronautics and Space Administration's (NASA's) work in developing a standard for models and simulations (M&S), the subject of credibility in M&S became a distinct focus. This is an indirect result from the Space Shuttle Columbia Accident Investigation Board (CAIB), which eventually resulted in an action, among others, to improve the rigor in NASA's M&S practices. The focus of this action came to mean a standardized method for assessing and reporting results from any type of M&S. As is typical in the standards development process, this necessarily developed into defining a common terminology base, common documentation requirements (especially for M&S used in critical decision making), and a method for assessing the credibility of M&S results. What surfaced in the development of the NASA Standard was the various dimensions credibility to consider when accepting the results from any model or simulation analysis. The eight generally relevant factors of credibility chosen in the NASA Standard proved only one aspect in the dimensionality of M&S credibility. At the next level of detail, the full comprehension of some of the factors requires an understanding along a couple of dimensions as well. Included in this discussion are the prerequisites for the appropriate use of a given M&S, the choice of factors in credibility assessment with their inherent dimensionality, and minimum requirements for fully reporting M&S results.

1. Introduction

Credibility in models and simulations (M&S) is a complex topic and spans the range of understanding from the purely quantitative to the essentially qualitative, while touching the scientific and engineering practitioner/specialist as well as decision makers at the highest management level. This breadth of exposure, along these two linked dimensions alone (quantitative-qualitative, specialist-management), induces a level of complexity in the topic. This topic of credibility in M&S is a, more or less, direct outgrowth of the development of the National Aeronautics and Space Administration (NASA) Standard for Models and Simulations [M&S Standard], hereafter referred to as 'the Standard.' It is also the most contentious topic.

The origin of the NASA Standard stems from the Space Shuttle Columbia Accident Investigation Board (CAIB) [CAIB]. While the recommendations from that board are Shuttle Program centric, the subsequent review lead by Diaz [Diaz] looked at the CAIB findings and detailed intiatives/actions applicable across all of NASA. Action 4 called for the development of "a standard for the development, documentation, and operation of models and simulations" [Diaz]. Along with that action, the NASA Office of the Chief Engineer (OCE) directed the inclusion of "a standard method to assess the credibility of the M&S presented to the decision maker" [OCE letter]. The resulting development of the Standard started in April 2005, went through three iterations of development, and culminated in the proposed final version for NASA-wide concurrence in early 2008. Bertch (et al.) describe the development process of the Standard and the included Credibility Assessment Scale (CAS), along with some key lessons learned from that effort [SISO 08].

The topics in the literature most closely related to credibility revolve around V&V. While this is indeed central to credibility, it does not adequately cover the full scope of the topic. In the forthcoming discussion (we'll see that), achieving credibility in M&S results requires more than verification and validation (V&V) of the M&S.

An added level of difficulty in developing and obtaining concurrence with this credibility assessment is that, by intent and design, it must apply to all types of M&S and all phases of the M&S process. While difficult, this broad applicability is the exact purpose for pursuing the path of a standard, rather than a recommended practice or handbook. Some additional background on the justification for a standard development is in the following background section.

2. Background

Standards come in a variety of fashions and purposes, and are levied by organizations, the market, professional portion of M&S are implemented in software, there are also some significant differences, which stem from their respective definitions. Generally, software is a program or code that allows a computer to perform a specific task, whereas a model or simulation is (possibly a computer program designed to mimic or imitate) an abstract representation/representation of the characteristics of a system. The difference between task performance and behavior mimicking is significant enough to consider a new standard. A comparison of software and M&S is shown in Table z. While software standards typically do not consider use of the software, an M&S standard must include the use as that is crucial in producing credible results.

Table z

	Software	M&S
Activity	Perform a task within a system	Representing (the behavior of) a system
Purpose	Performance of tasks/functions for a system	Analysis & Understanding of a system for insight or behavioral mimicking of a system for training/gaming
Requirements	Discrete functions to perform	Behaviors the simulation model to exhibit
Assumptions & Abstractions	Typically, none	Always
Uncertainty	Typically, none	Almost always

While developing some semblance of a credibility scale for one or two phases of the modeling and simulation cycle (e.g., V&V) or for one specific type of M&S can more adequately represent that particular segment or paradigm, it would not solve the problem of clearly and consistently communicating with management level decision makers. This is specifically why the development of a standard (though high level) methodology for M&S, and the reporting of results from M&S, is needed. It provides a common framework and terminology base across and between disciplines that makes communication more clear and consistent. This whole issue of clarity in the communication of technical data is central to the works of Edward Tufte, who provided such assessments to both the Challenger and Columbia investigations [Tufte]. This Standard provides at least the basis for necessary clarity and completeness in communication in the practice of and presentation of results from modeling and simulation.

3. Credibility

The specific term 'credibility' was not chosen quickly or without debate. Conceptually related terms, such as verification, validation, quality, rigor, and maturity surfaced to supplant it. While these concepts are valuable and even central to credibility, none of them sufficiently encompass the concept that leads to acting on the results from an M&S. The NASA Standard defines credibility as the quality to elicit belief or trust in M&S results.

A prescriptive approach to credibility across the broad spectrum of M&S is probably not possible, as each M&S type may approach the topic differently. This does not mean that some commonality in the discussion cannot occur. One instance of commonality is that all M&S should do some form of V&V. M&S literature contains much about V&V with the process traditionally represented as in Figure b [Sargent] [SCS].

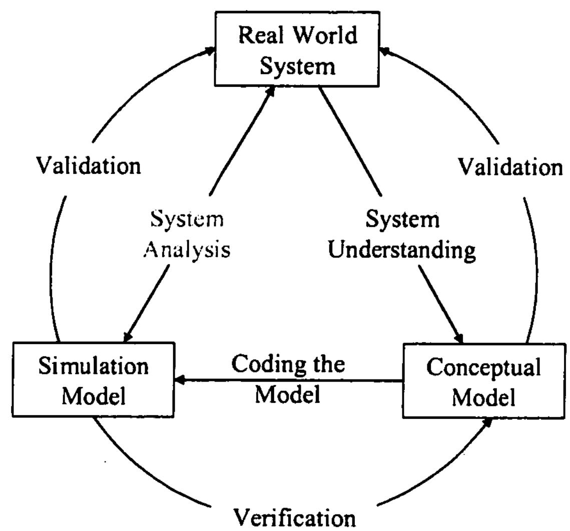

Figure b – Simplified Model of V&V ©Sargent & SCS

While the conceptual model (CM) in the traditional sense represents what is included in the system model, this is typically just the first step, which rarely includes enough information for detailed for implementation. Typically, once the CM is set, additional details are required to ensure the model has adequate representational fidelity to meet the requirements for the intended analyses with the model. This can include an additional level of detail from the initial CM, some 'business logic' required for the model to run or collect data as required for systems analysis, or detailed specification of the (computational) mathematics involved in a specific type of analysis. In the diagram of Figure c, this additional level of model specification is shown in the inserted step after the CM, termed the Implementable CM (ICM). The ICM is created from further understanding of the essential details of both the CM and the real-world system, and is thus validated by comparison to both. While the general CM concept includes this additional detail, many analysts add this step in practice prior to actual model data is important. The real-world environment (RWE) analogy is shown as the vertical axis on the left of Figure c, showing improvement in quality while moving up the axis. Both axes, though notional, depict a spectrum of possibilities, with full credibility achieved when an exactly matching referent system resides in the exact environment of the operational target system for the M&S analysis.

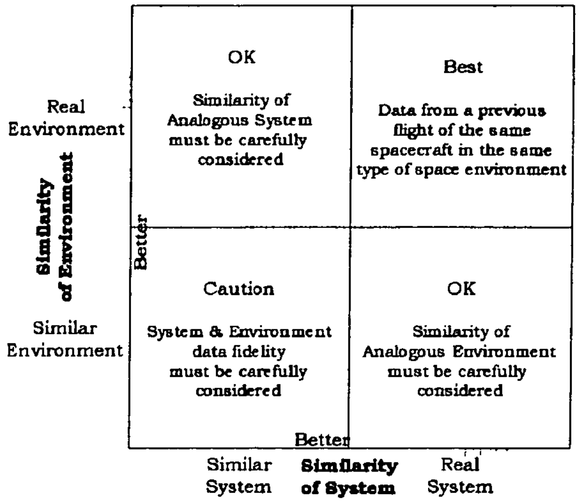

Figure c – Dimensions of a Validation Referent

The factors in the *M&S Operations* category relate to how a M&S is used for a specific analysis. Two things affect the level of uncertainty in the results of a M&S: the input and the methods in the computational system. Confidence in the M&S *input* more or less defines its pedigree, and is a product of the source of the data, the quality and quantity of the data used as the basis for M&S input, and the form of the M&S input. The input to a M&S can range from the purely notional to the rigorously derived stochastic, with the source playing a crucial part. With little or no real understanding of a given input, notional values can easily find their way into a M&S. Subject matter experts (SMEs) can lend credibility to input values from known point values (e.g., averages) to ranges of values (taking the form of uniform and triangular distributions). Beyond that, it is necessary to obtain real data from referent systems, and the more data that is available, the better the possibility of having fully representational input to the M&S. The quality of the input thus depends on the source and quantity of referent data. This is not the final word on input pedigree, however. What is done with this source data to transform it into the best form of M&S input is also key to improving results credibility. Even with lots of data in hand, it is readily reducible to any of several deterministic values (e.g., minimum, average, maximum). While running these values is certainly instructive, it by no means is a solution unto itself. Depending on the type of M&S and the specific input under consideration, either iteratively running the M&S with several values for the variables or a stochastic run is possible. This is where the form of the input becomes relevant. Deterministic runs are relatively simple to perform and analyze, while stochastic runs, with probability distribution functions as input, require more preparation on the front end and more analysis on the back end.

The *uncertainty* in results from M&S is potentially one of the most esoteric subjects in M&S, possibly because each M&S type includes or discusses it in such varied ways. It almost goes without saying that uncertainty is one of the key factors in M&S credibility and is directly related to the risk in accepting the results from an M&S analysis. Aleatory uncertainty, considered as some form of inherent or stochastic uncertainty, and epistemic uncertainty, considered either as lack of or incomplete knowledge of the system modeled, are becoming clear distinctions in the risk assessment community [Oberkamf, et al. – Challenge Problems]. As such, it's possible to consider the chosen assumptions or abstractions by the modeler as sources of epistemic uncertainty. Given these sources of uncertainty, there are two qualities of an uncertainty estimate that are manifest in M&S results: the size and the confidence. For a developed model and its myriad of inputs, the output from the simulation run has a level of uncertainty associated with it. The acceptability of that size is dependent on the system modeled and the intended purpose of the analysis. Along with that, however, is the confidence inherent in the computational uncertainty. Figure d shows combinations of these two aspects of uncertainty with cautions in their combination.

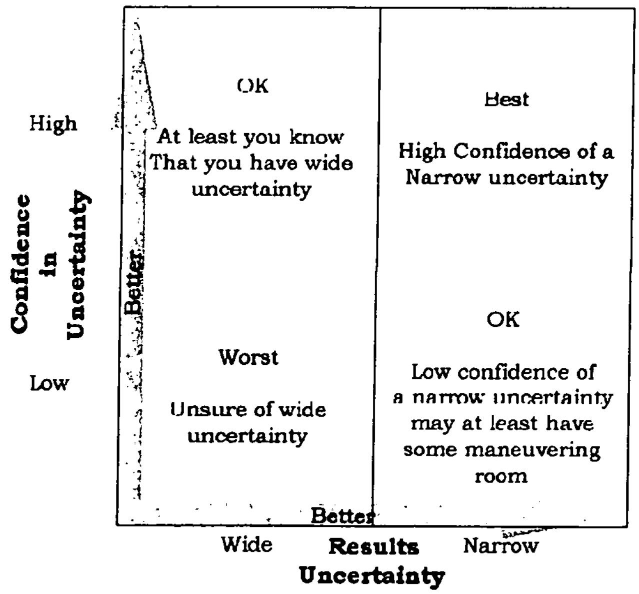

Figure d – Uncertainty

specific type of M&S, and the experience with and understanding of the modeled system of these people play a part in assessing the credibility of the M&S results.

Five of the eight factors in the CAS of the NASA M&S Standard include a required *technical review* sub-factor, which assesses the level of peer review successfully completed relevant to the parent factor. The idea is that a M&S (and/or modeling and simulation process) that is successfully peer-reviewed is more credible than one that is not. The level of independence, the qualifications of the members, and the level of formality of the peer-review group can also lend credibility to a particular M&S result. The formality of the review refers to conduct in accordance with rules explicitly established by the reviewed or reviewing organization.

4. Discussion

While many of the concepts for these credibility factors are general in nature, they provide a context for discussing critical aspects of M&S results, applicable to all types of M&S. Much in the literature is written on V&V, especially as applicable to a particular M&S discipline, or even more specifically one particular M&S discipline as applied to a particular study area. While firmly based on traditional V&V, the discussion of credibility is also more than that. In the approach developed by the NASA M&S Standard TWG, six additional factors contribute to a credibility assessment beyond V&V.

One prime purpose for the development of this standard is to not permit the presentation of just M&S results, since, in and of itself, a set of results is not a complete picture. To that end, the requirements in Section 8 of the Standard spell out the distinct reporting requirements for presenting M&S results for critical decision making. For NASA's purpose, a critical decision is one that impacts human safety or project-defined mission success criteria [NASA Standard]. Thus, the general reporting requirements when presenting M&S results are:

- An unfavorable *use assessment* of the M&S for the particular analysis
- The best estimate of the results
- A statement on the uncertainty in the results
- The evaluation of the results using the credibility assessment scale
- Any explicit caveats that accompany the results (e.g., errors or warnings occurring during a simulation run)

As most of these requirements are self-explanatory or already discussed, use assessment needs a little explanation. The concept here is to ensure the use of the M&S is within the known bounds of its verified and validated operation. The development of a M&S typically allow a wide range of inputs and tuning parameters, but only a smaller portion of the allowable domain is rigorously examined and accepted. When making critical decisions with M&S, the intent is to ensure the use of the M&S is within its verified and validated bounds.

In the final approval process for this Standard, there was a fair amount of objections raised to it, especially with respect to the CAS. Upon consideration, this is not dissimilar to the objections raised when the first software standards were in development. Table 1 lists the primary objections with rebuttal comments.

[redacted]

Objection	Rebuttal
Not right for a particular M&S type	TWG was comprised of practitioners of a variety of M&S methods and problem domains
Assessments are subjective	Subjectivity is a part of all M&S, and also of decision making – the purpose of the CAS is to provide information to a decision maker to enhance their decision making
CAS is too complex	M&S is complex, which is precisely why a standardized approach to M&S development, operations, results reporting is needed
Already have software standards	There are many similarities, but the focus of software and M&S is different

In short, software performs tasks within a system, whereas, M&S mimic the behavior of a system. Even though M&S are often computer/software-based implementations, the distinction between performing a task and mimicking a behavior is key. Software standards address functions that are static and deterministic, that is, without uncertainty. Purely analytical models with deterministic inputs and outputs can also be included here. In simulations, however, the inclusion of dynamic and stochastic behaviors introduces uncertainty into the functioning of the system. Law and Kelton [3rd ed.] originally published a somewhat simpler version of Figure IV-a, which is augmented here with additional characteristics of various M&S types. Further experiential aspects of M&S are also included with the addition of visualization and other sensory components to enhance the immersive aspects of the simulated system.

6. Acknowledgments

The author wishes to acknowledge the members of the Topic Working Group from nine of the 10 NASA centers participating in the development of this NASA Standard: Unmeel Mehta (Ames Research Center), Maria Babula (Glenn Research Center), Andre Sylvester (Johnson Space Center), Gary Mosier (Goddard Spaceflight Center), Bill Bertch (Jet Propulsion Laboratory), Larry Green (Langley Research Center), Joe Hale, II (Marshall Spaceflight Center), and Jody Woods (Stennis Space Center), as well as Tom Zang (leader of the TWG), Hal Bell (NASA Headquarters Office of the Chief Engineer), and Ken Johnson (NASA Engineering and Safety Center). The NASA Office of the Chief Engineer sponsored the participation of all NASA centers in the development of this Standard. Participation by the Jet Propulsion Laboratory, California Institute of Technology, was funded under contract with NASA.

References

[1] Agnew, P. G. (1928). "Work of the American Engineering Standards Committee" in *Annals of the American Academy of Political and Social Science*, Vol. 137, Standards in Industry (May, 1928), pp. 13-16.

[2] American National Standards Institute, Consumer Affairs Overview, "What is a Standard" (n.d.). Retrieved June 22, 2007 from http://www.ansi.org/consumer_affairs/overview.aspx?menuid=5

[3] Columbia Accident Investigation Board (CAIB) Report (August 2003).

[4] A Renewed Commitment to Excellence – An Assessment of the NASA Agency-Wide Applicability of the Columbia Accident Investigation Board Report (January 2004).

[5] Final Report of the Return-to-Flight Task Group (July 2006).

[6] NASA Standard for Models and Simulations, NASA-STD-(I)-7009, (December 2006).

[7] NASA Procedural Requirements (NPR) for Software Engineering Requirements, NPR 7150.2.

[8] NASA Software Safety Standard (NASA-STD-8719.13).

[9] Osman Balci and William F. Ormsby (2006), "Quality Assessment of Modeling and Simulation of Network-Centric Military Systems," In *Modeling and Simulation Tools for Emerging Telecommunications Networks: Needs, Trends, Challenges and Solutions,* A.N. Ince and E. Topuz, Eds, Springer, New York, NY, Chapter 19, pp. 365-382.

[10] Banks, J. (Ed.). (1998). *Handbook of Simulation.* New York: John Wiley & Sons.

[11] DoD Modeling and Simulation Office (DMSO), DoD Modeling and Simulation Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide, Build 3.0, (September 2006).

[12] Oberkampf, W.L., Trucano, T.G., and Hirsch, C. (2004). Verification, Validation, and Predictive Capability in Computational Engineering and Physics. In Applied Mechanical Review, vol. 57, no.5 (September 2004), 345-384.

Author Biography

MARTIN J. STEELE is a simulation analyst in the Mission Support division of the Information Technology Directorate at NASA's Kennedy Space Center (KSC), FL. He has over 25 years of professional and military experience in space systems engineering and operations, primarily at KSC and Cape Canaveral Air Force Station (CCAS). He is currently leading the coordination and integration of discrete event simulation models in NASA's Constellation Program, as well as a topic working group member developing the NASA Standard for Models and Simulations.