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 initiatives/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. Berte (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
organizations, a given industry, and national and international bodies. In general, the purposes for standards, are to facilitate interoperability, re-use, share a common vocabulary, document or benchmark best practices, and ensure a product or service is fit for their purpose [Coallier][Steele-SCSC07]. The overall purpose for the NASA Standard for M&S is to provide a common vocabulary for discussing and a common methodology for assessing the credibility of all types of M&S, so as to ensure that a specific M&S is fit for a particular purpose. These definitions and intents are completely consistent with those of the American National Standards Institute (ANSI) [ANSI] and the International Standards Organization (ISO) [ISO]. In the development of this standard, the discussion of terminology definitions was recurrent and occupied a significant amount of time. As the composition of the NASA M&S Standard Topic Working Group (TWG) included personnel from a variety of M&S specialties and from nine of the ten major NASA centers, coming to a common agreement (understanding) was not easy, but essential to the task. This difficulty alone is argument enough to justify the development of an M&S standard, at least on the basis of having a common set of terms to anchor discussions between the variety of M&S practitioners along with having a consistent discussion with decision makers.

The development process for the NASA M&S Standard consisted of three distinct phases: initial development, interim version development, and proposed final version development [SISO 08]. In each phase, review of the sourcing documents and M&S literature (Figure a) was central to progress in development of the Standard. A few other non-NASA and NASA related efforts proved instructive to this task.

In NASA internal efforts, decision maker interviews and pilot studies using the Standard were conducted. Some key elements of credibility identified by NASA directors include understanding sources of information, how an M&S is V&V’d, measures of uncertainty, and use history. The pilot studies acted as an initial proving ground for the interim version of the Standard and provided value in the re-work of the CAS.

In the third and final phase of development, interviews with decision makers as to what aspects of M&S analysis credibility to incorporate in the CAS and collaboration with the modeling and simulation community at the Sandia National Labs proved valuable.

As relayed earlier, the NASA Standard is applicable to all types of M&S and all phases of the M&S process, ultimately to ensure the usability of the analytical results. The area of M&S embedded in control software, emulation software, and stimulation environments was originally included, but eventually removed for inclusion in the NASA Software Engineering Requirements [NPR 7150.2]. A stimulation environment refers to “artificially generated signals are provided to real equipment in order to trigger it to produce the result required” [M&S Standard].

Some believe that abiding by software standards are adequate enough for M&S. While a good and growing

---

Figure a – Development Process of the NASA M&S Standard

In non-NASA efforts, four groups proved very informative to the development of this Standard and CAS. First, the Predictive Capability Maturity Model from the Department of Energy’s (DoE) Sandia National Laboratories [SAND Report] assesses, as the name implies, “the level of maturity of computational M&S efforts.” It specifically focuses on partial differential equation (pde) type simulations and the maturity in the ability to produce predictive computational models. Second, the U.S. Department of Defense (DoD) sponsored two separate efforts in a similar regard, one focusing again on maturity of M&S validation [Harmon & Youngblood], and another effort more broadly directed towards M&S quality [Balci]. Additionally, the American Institute for Aeronautics and Astronautics (AIAA) published a guide for the V&V of computational fluid dynamics [AIAA] and the American Society of Mechanical Engineers (ASME) published a guide for V&V in computational solid dynamics [ASME]. While these efforts definitely overlap with the intent of the NASA Standard, they also extend beyond the focus in some respects, while also not wholly represent the focus in others. For example, concept of quality embodied by Balci includes ease of model maintenance, which is beneficial, but not necessarily required to produce good results. Also, a perfectly valid model can be used in a manner that produces incorrect results. The NASA Standard, on the other hand, describes some important factors in the production and discussion of M&S results credibility.
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>
<thead>
<tr>
<th>Table z</th>
<th>Software</th>
<th>M&amp;S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>Perform a task within a system</td>
<td>Representing (the behavior of) a system</td>
</tr>
<tr>
<td>Purpose</td>
<td>Performance of tasks/functions for a system</td>
<td>Analysis &amp; Understanding of a system for insight or behavioral mimicking of a system for training/gaming</td>
</tr>
<tr>
<td>Requirements</td>
<td>Discrete functions to perform</td>
<td>Behaviors the simulation model to exhibit</td>
</tr>
<tr>
<td>Assumptions &amp; Abstractions</td>
<td>Typically, none</td>
<td>Always</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>Typically, none</td>
<td>Almost always</td>
</tr>
</tbody>
</table>

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](image)

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,
only 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
implementation. The distinction becomes apparent in that actual verification of the implemented model is accomplished relative to the details of the ICM rather than the generalities of the CM.

As mentioned previously, the focus is for complete and clear communication of M&S results to decision makers, along with information surrounding the M&S relevant to its credibility, and to do so preferably in a standardized way. The goal is also to have a parsimonious set of information from which to assess that credibility. This set of information or factors comprising credibility, as developed by the NASA TWG, are the first set of dimensions. The attempt was to make them as orthogonal (i.e., non-overlapping) as possible, though the strictness of this is debated. What surfaced in defining each of these factors was some aspects (facets) important to their understanding. Ignoring a facet could compromise the full understanding of that particular factor, leading to decreased clarity and less credibility in the M&S results. These facets are the dimensions composing a given credibility factor. For the factors in the NASA developed CAS (Figure ggg), the dimensions are briefly discussed. Note that in the NASA Standard, technical review is included as a sub-factor to five of the eight high level CAS factors. Assessing these factors will likely vary between M&S types and implementation domains, but the general meaning remains in tact. Thus, this standard provides a framework for broad applicability. The remaining part of this discussion, therefore, is not to discuss the implementation of each of these dimensions for each type of M&S, but simply to enumerate the dimensions for consideration.

M&S verification is all about ensuring the structure, flow, and fidelity of an implemented (a coded) simulation model are correct with respect to its intended purpose (requirements) [Law reference possible – Pg 243]. Verification of the structure and/or flow is by code tracing, running the M&S through a series of ‘primitives tests’ or ‘minimal/maximal values tests,’ and comparison of the coded model with the CM. From the perspective of fidelity, all M&S are abstractions of the real-world system. The level of abstraction (or the level of detail to include in a M&S) is to some degree a matter of judgment. Additionally, information which is either not perfectly known about a system or certain details chosen not to include about a system in a M&S are the subject of assumptions and abstractions. While deftly intertwined upon consideration, both the structure and fidelity of a M&S are important considerations in verification.

Simply defined in the NASA M&S Standard, validation is “the process of determining the degree to which a model or a simulation is an accurate representation of the real world from the perspective of the intended uses of the model or the simulation” [M&S Std]. The real world representation, or referent, is not always forthcoming, particularly in the development of novel systems. In such cases analysis models must obtain data from some analogous system for validation purposes. Two dimensions (aspects) of a validation referent are important, and notionally depicted in Figure c. First, along the bottom axis, is the quality of the referent system. As one moves from left to right in this depiction, the quality of the system data from the referent improves. If no equal system exists and the data is from a similar system, the quality of similarity is important to consider and caution is warranted. Using another M&S as a referent also falls into this characterization. As the similarity of the referent approaches that of the target real-world system, credibility of the analysis correspondingly improves. However, this is not enough of a consideration. The environment of the referent system is, potentially, equally important. If the referent system is similar to the target system, but operates in an entirely different environment, then judgment of the suitability of the referent
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.

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.

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.
**Robustness** has several nuances, depending upon the context of use, and relates to sensitivity analysis in M&S and its implications to decision making. Generally, a robust system operates correctly under many conditions, so, the result from an M&S is robust if it is stable with respect to some level of changes in relevant input uncertainty. Understanding the ‘level of robustness’ is obtained by performing sensitivity analysis. The purpose of exploring the robustness of a solution is to understand the sensitivity of the real-world system to potential changes in the variables and parameters of the system [M&S Std]. What comes from an exploration of system sensitivity with respect to M&S, however, is the requirement to also know how well the sensitivity of the M&S matches that of the real-world system. As simulations aim to imitate the real world or a proposed real world, the goal is to have the M&S as similarly sensitive to the real-world system as possible. This is really a step beyond simple validation of the M&S. While the M&S may validate well with respect to certain point in the real-world system, the sensitivities to changes in input variables are considerations also. Understanding the sensitivities of a real-world system, and how they are manifest in the corresponding M&S results, are crucial pieces of information in accepting those results for informed decision making. Figure e depicts this comparison of M&S and real-world system (RWS) sensitivities. In the bottom left quadrant, the RWS and the M&S are similarly sensitive, as determined by validating a M&S. In the bottom right quadrant, validation studies show that the RWS is robust, but the M&S is not. The results from an M&S analysis are overly conservative in this case. Knowing this helps in the credibility assessment of the M&S results, and is also an issue in further refinement and validation of the M&S. On the other hand, in the upper left quadrant, the M&S is insensitive to changes in the input variables, while the RWS is not. The M&S results here indicate robustness not present in the RWS. This is the worst situation where the M&S is not so useful and an issue for M&S refinement and validation. Another acceptable situation occurs when the robustness of the M&S matches that of the RWS (i.e., the results of analysis are insensitive to changes in the input). While tied closely to validation in M&S Development, the performance of sensitivity analysis in the use of an M&S is also a crucial factor in credibility.

**Figure e**

The supporting evidence category, though less technical, provides additional evidence lending credence to M&S results.

The use history factor describes the extent of any prior use of the M&S in similar situations for critical decisions. The two dimensions to consider for a specific M&S result is the time length of validated use of the M&S and the types of problems for which it is used. The central idea with regard to credibility is that the longer a given M&S is used and the closeness to the historical use the current use is, the more credible the results are. While this is not, by any means, a guarantee of good results, they are an indicator of the past successful trials of the M&S, and, therefore, an important point of discussion.

The easily misleading term M&S Management in this case refers to the manner in which the M&S and data are under configuration control and the status history of sustaining and/or maintaining the M&S. The first proposition is that a M&S under configuration control is more credible than one that is not. While this is not necessarily a ‘make or break’ evaluation, it is still something to consider. The second aspect of this factor relays the currency of the M&S, that is, is it maintained (kept in working condition) and sustained (kept up to date with system changes). Several things can affect the correct working of computer-based M&S, such as hardware and software platform changes for which regression testing is mandated. On the other hand, if the modeled system or environment changes from the time of M&S validation, similar upgrades to the M&S are required. This sustains the M&S for its original usefulness. The work performed in the management, control, maintenance, and improvement of the M&S supports credibility.

This final factor refers to the qualifications of the people developing, operating a M&S, and analyzing a system with the M&S. The education, training, experience with the
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.

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

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.
Continuing from the preceding discussion on the difference between software and M&S, there are further justifications for yet another standard. The ANSI and ISO exist solely for the development, integration, and proliferation of standards. Steele [SCSC 07] includes an introductory discussion of this topic. From an ISO perspective, Coallier [6th International Conference] states that standards facilitate interoperability and agreements on products, practices, and operations. This is directly the purpose of NASA’s Standard for Models and Simulations: to define a set of terminology applicable across all types of M&S, require a distinct set of requirements for reporting M&S results for critical decision making, and define a common method for discussing the credibility of M&S results. The purpose is to reduce the risk associated with M&S-based decision making.

In the time required to develop this Standard, existing operational M&S projects developed their own methods of communicating the acceptability of their own results. While necessary in the interim, if each M&S type and implementation domain developed their own unique method of defining and reporting acceptability, then much of the time and effort in making a decision from M&S results is in understanding that particular method. Again, this inefficiency and additional complexity alone justifies the development and use of standard for M&S.

Throughout the development of the Standard with a diverse team from the various NASA Field Center perspectives and M&S disciplines, coming to a common understanding of just the terminology was a difficult and iterative process. While some of this is justified due to the uniqueness of the various M&S types, rarely, if ever, are any one M&S isolated from other parts of the greater system or monolithic in perspective of the people involved in the results for the M&S. Coming to some general understanding of the commonly used terminology is beneficial. This is one of the benefits of the development of a standard across the broad domain of M&S, and is discussed in both national and international standards organizations literature [NASA-Gill] [ANSI] [ISO]. With the great variety of M&S in practice and the growing number of implementation domains, connotations in the M&S vernacular will not (always) match the denotations. The development and use of a M&S standard helps provide the common framework for discussion and understanding across this broad topic area.

5. Summary

As of this writing, NASA is currently in the approval process for the proposed permanent M&S Standard as a voluntary standard, which means it is available, but not prescribed, for implementation on NASA projects. Acceptance as a mandatory standard requires a second stage of approval. Before that, some further efforts are prudent.

The next steps to pursue follow three separate paths. First, rolling out and introducing the Standard to each NASA Center and its relevant contractors to use it, collect data on its use to assess its applicability across a variety of M&S disciplines, and making revisions to the Standard as necessary. Second, as the Standard is general in nature, it necessarily implements differently across the breadth of M&S disciplines and implementation domains. As such, a guide (recommended practice, handbook) relative to the uniqueness of each M&S type will enhance the quality and consistency in each discipline, thus further accomplishing the goals of the Standard and its originating impetus. Finally, increased collaboration with other governmental, academic, professional, and international organizations in M&S will help further refine and improve its contents, and disseminate its use. Beyond these three paths, the proliferation and increased dependence on M&S application software produces the need for a “clearing house” to vet these software applications. Without such a function, the M&S developer is required to V&V the M&S and the software application, which is more time-consuming and repetitive.

This is by no means the final or full word on the subject of credibility in M&S. This discussion is a framework for consideration in improving the credibility of M&S, which is an extension beyond M&S development (V&V), including operational use factors and supporting evidence factors of the NASA CAS. Each of the 8 (or 9, including Technical Review) factors and their inherent dimensions are topics for discussion in the acceptability (credibility) of M&S results to enhance informed decision making. The principles discussed provide the foundation for improving the discussion of credibility of M&S results.
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


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.