Development of NASA's Models and Simulations Standard

William J. Bertch
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109-8099
William.J.Bertch@jpl.nasa.gov

Thomas A. Zang
NASA Langley Research Center
Hampton, VA 23681-2199
Thomas.A.Zang@nasa.gov

Martin J. Steele, Ph.D.
IT Mission Support
Kennedy Space Center, FL 32899
Martin.J.Steele@nasa.gov

ABSTRACT: *From the Space Shuttle Columbia Accident Investigation, there were several NASA-wide actions that were initiated. One of these actions was to develop a standard for development, documentation, and operation of Models and Simulations. Over the course of two-and-a-half years, a team of NASA engineers, representing nine of the ten NASA Centers developed a Models and Simulation Standard to address this action.*

The standard consists of two parts. The first is the traditional requirements section addressing programmatics, development, documentation, verification, validation, and the reporting of results from both the M&S analysis and the examination of compliance with this standard. The second part is a scale for evaluating the credibility of model and simulation results using levels of merit associated with 8 key factors.

This paper provides an historical account of the challenges faced by and the processes used in this committee-based development effort. This account provides insights into how other agencies might approach similar developments. Furthermore, we discuss some specific applications of models and simulations used to assess the impact of this standard on future model and simulation activities.

1. Introduction

Figure 1 illustrates the genesis of the work summarized in this paper. Following the report of the Columbia Accident Investigation Board (CAIB) [ref. 1], the National Aeronautics and Space Administration (NASA) Administrator chartered an executive team (known as the Diaz Team) to identify those CAIB Report elements with Agency-wide applicability and develop measures to address each one. Action #4 from the January 30, 2004 Diaz Team report [ref. 2] called for NASA to: "Develop a standard for the development, documentation, and operation of models and simulations." There were six tasks associated with Action #4. These tasks are listed below and became objectives for development of the requirements for models and simulations (M&S).
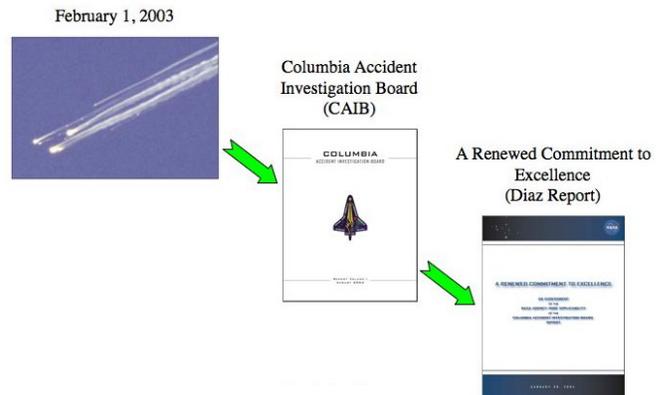

February 1, 2003

Columbia Accident Investigation Board (CAIB)

A Renewed Commitment to Excellence (Diaz Report)

Figure 1. Flowdown from Columbia Accident to M&S Standard

1. "Identify best practices to ensure that knowledge of operations is captured in the user interfaces (e.g. users are not able to enter parameters that are out of bounds).
2. Develop process for tool verification and validation, certification, re-verification, revalidation, and recertification based on operational data and trending.
3. Develop standard for documentation, configuration management, and quality assurance.
4. Identify any training or certification requirements to ensure proper operational capabilities.
5. Provide a plan for tool management, maintenance, and obsolescence consistent with modeling/simulation environments and the aging or changing of the modeled platform or system.
6. Develop a process for user feedback when results appear unrealistic or defy explanation."

Subsequently, the NASA Chief Engineer augmented this in an internal memo dated Sept. 1, 2006 with the additional expectations that "the M&S standard will:

1. Include a standard method to assess the credibility of the M&S presented to the decision maker when making critical decisions (i.e., decisions that effect human safety or mission success) using results from M&S.
2. Assure that the credibility of M&S meets the project requirements.
3. Establish M&S requirements and recommendations that will form a strong foundation for disciplined (structure, management, control) development, validation and use of M&S within NASA and its contractor community."

This paper reports the development process, the fundamental decisions, and key features for the M&S Standard that has resulted from this charge.

A thorough review of available guidance for M&S at the inception of the response to Diaz Action #4 revealed that although there is considerable guidance for the development, operations (use), and management of M&S, there is no formal, requirements-focused standard.

The scope of the NASA M&S Standard [ref. 3] is based on guidance from NASA Headquarters (HQ), and its final articulation read "This standard applies to M&S used by NASA and its contractors for critical decisions in design, development, manufacturing, ground operations, and flight operations. This standard also applies to the use of legacy as well as commercial-off-the-shelf (COTS), government-off-the-shelf (GOTS) and modified-off-the-shelf (MOTS) M&S to support critical decisions. This standard does not apply to M&S that are embedded in control software, emulation software, and stimulation environments."

The key phrase "critical decisions" is explained thusly [ref. 3]: "Critical decisions based on M&S results, as defined by this standard, are those technical decisions related to design, development, manufacturing, ground, or flight operations that may impact human safety or program/project-defined mission success criteria."

2. Development Process

The principle players in the development of the M&S Standard are illustrated in Figure 2. The **Development Team** (DT) consisted of 6 engineers from NASA Langley Research Center (LaRC). They developed the initial 3 versions of the M&S Standard, and provided recommendations to the Topic Working Group on changes to Version 3. The **Topic Working Group** (TWG) consisted of 9 Center representatives, the chair, and a representative from the **NASA Engineering and Safety Center** (NESC). They provided informal comments on Version 2, modified Version 3, approved Version 4, and decided disposition of the formal comments from across the agency on Version 4 resulting in the final, permanent NASA Standard [ref. 3]. The **Technical Standards Working Group** (TSWG) has representatives from all NASA centers and HQ. It has the general responsibility for development of all NASA standards and coordinates the formal reviews of proposed standards. **The Engineering Management Board** (EMB) consists of the directors
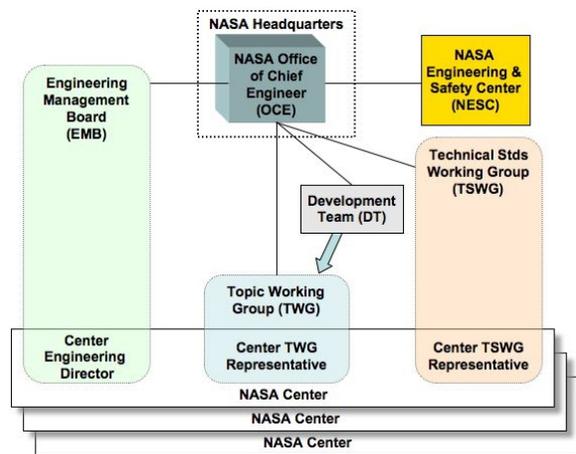

Figure 2. Participants in Development

of engineering at the various NASA centers. They render the official center concurrence (or objection) to all NASA standards. The **NASA Chief Engineer** chairs the EMB and formally issues standards; he happened to be the principal champion of this particular standard. The NESC, a functional arm of the Office of the Chief Engineer (OCE), provided oversight and logistical support.

This development process consisted of three phases, with the major activities illustrated in Figure 3. In Phase I (May 2005 – August 2006), the Development Team, formed at the behest of the OCE to respond to Diaz Action #4, performed background research, formulated the general approach, and developed the first three versions of the M&S Standard. An informal review of the second draft of the standard was conducted; this solicited comments from all NASA centers and was facilitated by members of the Topic Working Group. In June 2006 the NASA Chief Engineer determined (1) that the M&S Standard should be issued as a NASA interim standard in Fall 2006 (and then undergo the formal review process necessary for a permanent standard), and (2) that it must include a first cut at a credibility assessment scale (item #1 in his internal memo). Phase I ended when the Development Team submitted version 3 as their final deliverable in August 2006. Version 3 included the scale that appeared in Appendix A2 of the interim standard [ref. 4]

In Phase II (August 2006 –November 2007), the Topic Working Group, with membership from all NASA centers except for Dryden Flight Research Center (DFRC), revised the third draft of the standard into Version 4, which became the NASA Interim Standard for Models and Simulations [ref. 4]. This revision included an alternative scale, which appeared in Appendix A3. The Development Team served in a consulting role during this transitional phase.

In Phase III the Topic Working Group fostered the center pilot studies, facilitated the submittal of formal comments from their respective Centers on Version 4 during the NASA-wide Review of the Interim Standard, and revised the Interim Standard into the version submitted for EMB approval as the permanent standard. A parallel activity of the TWG involved developing a single credibility assessment scale with broad support among the TWG.

Phase I made extensive use of external consultants from the Dept. of Energy, the Dept. of Defense, NASA contractors and academia to determine the state-of-the-practice in M&S standards and to solicit comments on the first two versions of the standard. Phase III included outreach presentations of the standard at meetings of the AIAA, the ASME, JANNAF, and SISO.

The voting members of Topic Working Group were the nine Center representatives and they were responsible for representing their Center's perspectives on the issues. The quorum of voting members required for a TWG decision was 6 of the 9 voting members. The TWG made most of its decisions by a supermajority rule, e.g., if 8 voting members were present then a 6-2 vote was decisive, but a 5-3 (or 4-4) vote required more discussion.
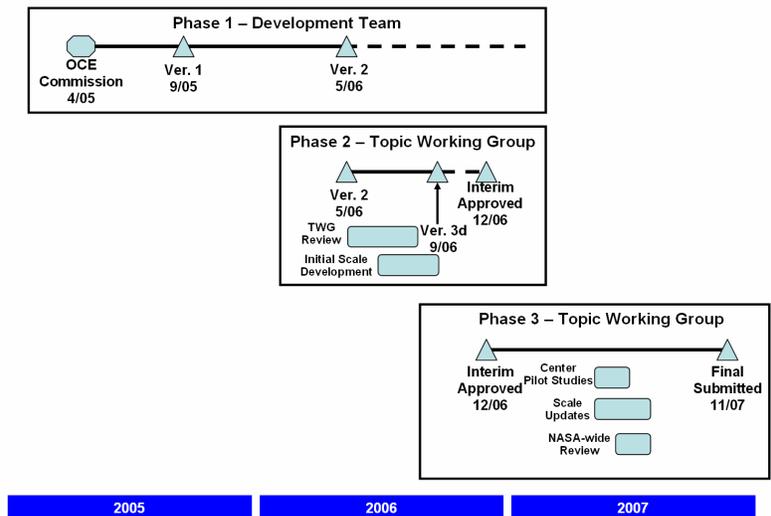

Figure 3. M&S Standard Development Phases

Phase II utilized weekly web-based meetings, as well as five face-to-face workshops to resolve key issues on the scale and the disposition of formal comments. The first three of these workshops utilized the services of a trained facilitator, which proved very useful in focusing and documenting the discussions and decisions.

During the first half of 2007, one or more pilot studies were conducted by M&S teams at each NASA Center. Their purpose was to provide feedback to the TWG on practical experience with the Interim Standard and to ensure that the center comments during the subsequent formal review were informed by this practical experience. Collectively these pilot studies included a reasonably broad spectrum of M&S types (mostly continuous with some discrete and geometry) and applications. However, as this was mostly a time-constrained volunteer effort, it proved impossible to get as thorough a set of pilots as was desired.

Two types of questionnaires were submitted to the pilot teams. One of these focused just on the credibility assessment scales that were in the Interim Standard; the other covered the standard as a whole. For example, one question on the scale questionnaire was:

"How easy is Scale A2 to score for a typical individual model in your M&S?
  (a) The Level Definitions are very clear: I can readily assign a unique level in each category.
  (b) The Level Definitions are mostly clear: I can assign a unique level in most categories, but in a minority of the categories, I am uncertain about which of 2 adjacent levels to choose.
  (c) The Level Definitions are vague: In most categories I am unclear which level to choose.
  (d) The Level Definitions are meaningless: I have virtually no idea which level to choose in any category."

A second similar questionnaire asked questions about how well the goals of the Interim Standard (discussed in Section 3) are satisfied and what the cost impact of compliance with the Standard is.

From May to July 2007 the TSWG members collected comments on the interim standard from interested parties at their respective Centers, and furnished these to the TWG for resolution. The structured format for these comments called for the commenters to propose additions, deletions or changes to language in the document, and to provide

their rationale. In turn the TWG decided whether to concur, fully or partially, with the proposal (and make the change in the revised document), or provide their reasons for not concurring. Approximately 350 such comments were submitted. The TWG concurred (in full or in part) with 70% of them. Commenters objecting to the disposition of their comment(s) needed to persuade their engineering director to sustain their objection for this to be brought to the EMB for resolution. The TWG used its customary supermajority rule to decide all dispositions. This entailed rather lengthy deliberations at times in order to reach a decision.

This entire, sometimes contentious, process—Phases I–III—was conducted over a relatively short period of time because of the high-level, top-down push from agency senior management.

3. Overview of the Standard

The overall goal of this standard is to ensure that the credibility of the results from models and simulations is properly conveyed to those making critical decisions. That is, requirements are identified to ensure the development, operation, and documentation are properly addressed, but the critical requirements specify what is to be presented to the decision makers. Having requirements on the *presentation* is a rather unique approach for a standard, but gets to the very heart of the issues raised in the CAIB report where the information was not properly conveyed to the decision makers.

Since the M&S Standard evolved from an interim release that had some different TWG members, there were some interesting changes that occurred. These changes are noted in the discussion of the various requirements subsections below.

This standard consists of two parts as shown in Figure 4. The first part addresses the conventional set of requirements for M&S projects. The second part addresses the use of a credibility scale that was included to make the M&S credibility more apparent to the decision maker with the anticipation that this can expose the risk associated with M&S-based decisions. What follows is a discussion of the development of the requirements section. A discussion of the credibility scale development is in Section 4.

The requirements section consists of forty-nine requirements separated into eight subsections. The first six subsections provide the underlying activities

that support the credibility assessment requirements in subsection 7, and subsection 8 addresses the reporting of M&S results to the decision makers.

The introductory material for each requirements section includes a discussion of the intent of the requirements in that section. Thirty-five of these requirements start with the words "shall document," indicating that "the required documentation aspects for an activity that was not conducted may be simply satisfied by recording that the activity was not conducted." This wording is unique in that it provides a way to essentially not complete a requirement without requiring a formal waiver. The primary benefit of this approach is that it allows the project management team to identify the specific areas to be documented.

The first requirements subsection addresses *programmatic activities* performed by the project office of the M&S. These requirements require the project to 1) identify the M&S that may be used for critical decisions, 2) define the objectives and requirements for the M&S, and 3) develop a plan for the acquisition, development, operation, maintenance, and/or retirement of the M&S. This section has remained fairly stable since the interim release with minor cleanup to the requirements.

The second requirements subsection addresses the *requirements imposed on the model*, where model refers to the conceptual model, mathematical model, and computational model. The majority of these requirements address documentation for the assumptions, basic structure, mathematics, data sets, limits of operation, guidance in the proper use of the model, parameter calibrations, model updates, and methods for uncertainty quantification.

The third requirements subsection addresses the *requirements imposed on the simulation*. This includes requirements addressing the limits of operation, uncertainty/pedigree of the input data, processes for executing the simulations, processes for conducting analyses, appropriateness regarding intended use, and use history of the M&S. This section has remained fairly stable since the interim release, except for a few requirements moved to this area from other sections, and one requirement added to cover factors within the credibility scale.

The fourth requirements subsection addresses the *verification, validation, and uncertainty quantification*. M&S practitioners typically understand the nuances of these requirements for their particular type of M&S; however, specific emphasis is given to communicating the domain in which the model is valid. This is specifically directed towards appropriately applying the model and understanding how uncertainties propagate throughout the run of the simulation. Considerable detail on these areas is addressed in Section 4 with regard to the associated factors in the credibility scale.
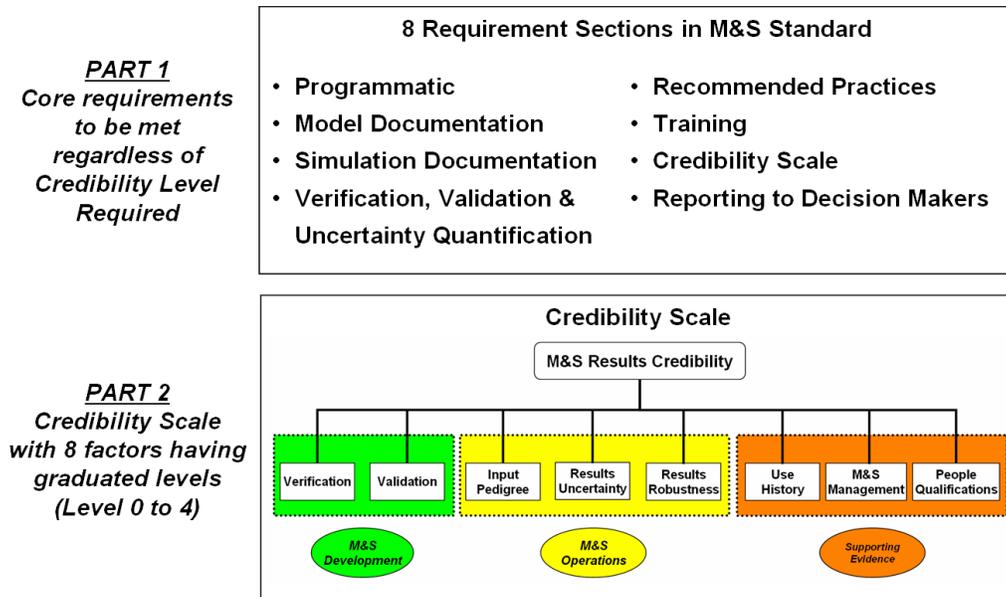

Figure 4. Two Parts of the M&S Standard

The fifth requirements subsection of the standard addresses the use of *recommended practices*. The use of recommended practices was one of the specific requirements specified in the Diaz Action #4. This section was totally revised from the interim standard with the final version containing a single requirement to identify and document applicable recommended practices. The Interim Standard contained nine requirements addressing the need to develop new recommended practices in areas where they didn't exist as well as updating existing recommended practices. The primary reason for removing these requirements was to avoid cost escalation in M&S projects and the anticipated waiver requests.

The sixth requirements subsection addresses training for developers, operators, and analysts. Although this section looks considerably different from the interim release, the content remains very similar. The primary changes consolidate the requirements and provide a listing of recommended training areas, rather than having a definitive list in the requirements.

The seventh requirements subsection addresses the *credibility scale*. The requirements specify an assessment of the M&S (results) using the scale for each M&S used for critical decision making as defined in the appendix. A detailed discussion of the credibility scale development is in Section 4.

The eighth and final requirements subsection addresses the *reporting of results* to decision makers. This is the key activity driven by this M&S Standard. The intent of these requirements is to inform the decision makers as to what analysis was performed, the results of that analysis, and what aspects of simulation modeling analysis was and was not accomplished (as embodied in the requirements of this Standard). Specific items to include in this reporting are uncertainty estimates, results from the credibility assessment and undesired results such as identifying violation of limits of operation, violation of assumptions, unfavorable outcomes, unachieved acceptance criteria, and any waivers to requirements.

In addition to the requirements section, there was also an extensive effort dedicated to the definition of M&S terms to ensure a common understanding within all TWG members. These definitions were extracted from NASA documents as the first source. Where terminology was not available from NASA documents, they were taken or adapted from other industry publications noted in the references.

4. Credibility Assessment

The operational concept of the credibility assessment scale is that the presentation of any results from M&S to a decision maker include (1) the best estimate of the results, (2) a statement on the uncertainty in the results, (3) the evaluation of the results on the credibility assessment scale, and (4) any explicit caveats that accompany the results. (An example of such a caveat would be use of the model in violation of its assumptions.) The decision maker then makes his own assessment of credibility based upon all four pieces of information in the context of the decision at hand. Just to emphasize this fundamental point, the credibility assessment scale does not purport to measure credibility; rather, it assesses the M&S results, and the processes used to produce them, against key factors that affect the credibility judgment. We stress that the goal of this scale is to assist in the assessment of the credibility of the particular results at hand, and not to assist in a broad certification decision for a class of uses of the M&S.

Development of the credibility assessment scale was undoubtedly the most arduous task confronting the TWG. While the definition of credibility and list of synonyms are relatively succinct, the connotations of the term vary almost from discipline to discipline and from person to person. Terms related to the concept of credibility (such as verification, validation, quality, and maturity) are also rife with nuances that can either enhance or divert the quest undertaken. That quest is the development of a credibility assessment scale that is comprehensive, yet readily understandable and manageable to both M&S practitioners and decision makers alike. The development of the final scale started from a search of the literature, continuing through the distillation of (orthogonal) factors, to resolving the final subset of key *scalable* factors with a mechanism for reporting the credibility of M&S results (Figure 5).
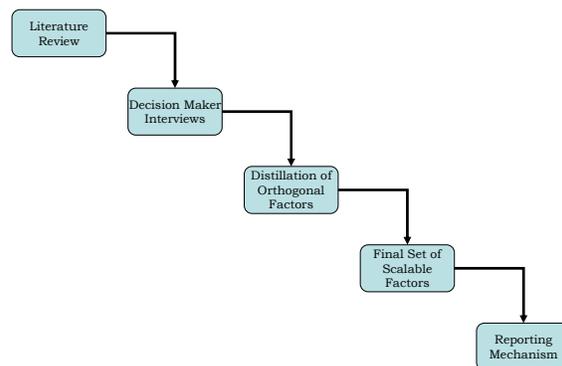

Figure 5. M&S Credibility Scale Development

One baseline position (or understanding) to which the TWG concluded early in the development is that credibility is not directly measurable. As such, there needs to be (other) more fundamental factors directly attributable to credibility. The search for these factors was rather arduous due to the broad range of M&S the NASA Standard and Scale needed to encompass, and the variety of viewpoints the making up the TWG. Several specific efforts in areas related to M&S standards and ensuring the rigor surrounding M&S development and operations (use). All of these are either restricted to a specific type of M&S (e.g., those with the mathematical model expressed as partial differential equations) or focus on a particular aspect of M&S (e.g., validation or quality). Specific direction from the NASA Chief Engineer, as related previously, kept this scale broadly applicable to all M&S types and all aspects of modeling and simulation. The work done in these other venues, primarily sponsored by the DoD and DoE, was very informative to this effort.

Even the specific term 'credibility' was not chosen quickly or lightly, as several related terms, such as verification, validation, quality, and maturity, continually resurfaced to supplant it. While these concepts are valuable, the TWG did not consider any one of them comprehensive enough for NASA's intensions, which focuses on the specific use of an M&S and its' results.

The first activity in the Phase III revision of the interim scales, was to examine related efforts. The Predictive Capability Maturity Model from Sandia [ref. 5] assesses, as the name implies, "the level of maturity of computational M&S efforts." It specifically focuses on pde-type simulations and the maturity in the ability to produce predictive computational models. The DoD sponsored two separate efforts in a similar regard, one focusing more narrowly on the maturity of just M&S validation [ref. 6] and another more broadly on the whole spectrum of M&S quality [ref. 7] NASA's Constellation Program is directed towards attaining M&S accreditation, along the path of rigorous verification and validation [ref. 8]. A fundamentally different approach was advocated by one of the TWG members [ref. 9].

This review of other scale approaches, along with a review of the findings of the CAIB related to M&S and interviews of some decision makers at each NASA center, provided some key insights into the makeup of a credibility assessment. This generated a list of over 100 factors to consider for inclusion in NASA's credibility assessment scale. In developing

an assessment scale, however, it is desirable to have *orthogonal factors* so as to prevent overlap and, thus, accentuating the affect of one factor over another. In this process of distilling so many factors down to something both manageable and comprehensive, the general make-up of *results credibility* spanned across the several specific areas discussed so far in the literature. From the influence of the Constellation Program's VV&A RPG and the Appendix A3 scale, a hierarchical arrangement of the factors of the final scale helped in providing a rollup to a single overall credibility rating. This arrangement carried through to the scale included in the permanent standard.

As mentioned previously, a series of weekly web-based and periodic face-to-face meetings provided the mechanism for development of the credibility scale. These started with understanding the published literature and brainstorming any credibility factor that might be useful to the intent of this standard. Subsequently, a sorting and culling of these factors reduced the list to something akin to an orthogonal set. This was by no means a simple amicable process, with 10 people from the different NASA centers and as different M&S backgrounds. It is, however, this diversity of perspective and experience that engendered the TWG's ability, eventually, to develop such a broadly applicable scale. A consolidated rough set of credibility factors emerged from the initial reviews and formed the basis for the final scale (Figure 6).

Over the course of several months, the TWG actively discussed the meaning and contents of each of these factors. As this scale developed, a more or less clear delineation between M&S development, M&S

- V&V
- Boundary Conditions of the M&S
- Understanding M&S Assumptions
- Developer, Operator, Analyst qualifications
- M&S Maintenance, Support, & Configuration Control
- M&S Use History
- Data Integrity (including currency, source validity, uncertainty, & sensitivity analysis)
- Supporting Documentation
- Accuracy
- Fidelity
- Peer Review
- Interfaces Between Models in a Federated (or Confederated) Model
- Use Assessment
- Results Traceability

Figure 6. Consolidated Rough Set of M&S Credibility Factors

operations (or use), and supporting activities (or evidence) surfaced, which eventually provided the basis for the resulting hierarchical structure. M&S *development* includes verification and validation, M&S *operations* (or use) includes input pedigree, results uncertainty, and results robustness, and *supporting evidence* includes use history, M&S management, and people qualifications. In these categories, the final eight high-level factors of results credibility surfaced as depicted in Figure 4.

Two specific factors from the initial rough set, that are not included as top level factors in the final scale, require note. First, *technical review* made the short list in Figure 6 because it is such an integral part of NASA's engineering processes. (Having this as a factor would not be appropriate for organizations that do not have such a heavy reliance on reviews.) Nevertheless, it was removed as a separate top level factor because the results of technical reviews strongly affect manager's assessments of the factors in the M&S Development and Operations categories. A second factor included in an intermediate version of the scale, *use assessment*, was also removed from the final scale due to its intrinsic binary nature. That is, the use of an M&S is either correct or incorrect for a particular analytical use; therefore, it does not lend itself well to a graduated rating scale, as do the other included factors. So as not to forget the importance of *use assessment* in the consideration of M&S credibility, a requirement to perform such an assessment is in Section 4.3 of the Standard, and this is one of six caveats covered in the reporting requirements of Section 4.8..

One last point on the use of credibility assessment surrounds the rolling up of the results to a single score. The Standard permits flexibility in implementation of the rollup, where weighting of the factor scores and hierarchically calculating the scores within the three major categories of development, operations, and supporting evidence is a choice of the implementing organization. This was by no means an area where the TWG has consensus and warrants the cautions that the weighting of scores abstract basic information and the rolling up of scores hide it. While permitting the weighting of factor scores and requiring the reporting of a single rolled up score, the Standard also requires the clear reporting of the weights and the individual factor information from which the single score derives. However, that rollup may not be reported without the supporting background information. This supports the intended purpose to clearly and completely inform decision makers as to the credibility of the results obtained from M&S analyses.

5. Relationship to Other Standards

NASA requirements for software development (NPR 7150.2, Software Engineering Requirements), pre-dated the NASA M&S Standard. The TWG determined that there was very little overlap, and no outright contraction, between the two documents. That is, the NPR has one general requirement to "test, validate, and certify software models, simulations, and analysis tools [requirement SWE-070]", and it does not even mention uncertainty quantification. On the other hand, the M&S Standard has just a few software specific requirements such as providing version control and use of a configuration management system. Thus, the two documents are complementary, with the M&S standard providing requirements for all the aspects of M&S that have more to do with the scientific method than with software engineering. Discussions with the NASA official responsible for NPR 7150.2 led to inclusion of the following language in the M&S standard: "implementation plans for NPR 7150.2 … should … address such M&S-specific issues as numerical accuracy, uncertainty analysis, sensitivity analysis, M&S verification and M&S validation" to emphasize these M&S specific requirements.

6. Applications

The applications noted below culminate the assessment of the M&S Standard in three ways. First, they address results from the pilot studies used to assess the impact the Standard may have on future simulation activities. Second, they identify some typical NASA applications to assist other industries in correlating to the potential uses of this NASA M&S Standard. And, third, they close with two recent applications where projects felt the need to implement their own M&S management processes in anticipation of the release of this standard.

Case Study Evaluation of the M&S Standard
As noted above pilot studies were conducted by the NASA Centers to provide a notional assessment of the impact the standard might have if the Standard were used for development and operation of the simulation. Nine pilot studies were conducted, but not all Centers were able to support this activity. So, some Centers conducted more than one pilot study.

As a result of these various inputs, ten pilot studies were conducted in the areas listed below.

- Mars Exploration Rover Entry, Descent, and Landing Simulation
- Aerodynamic database supporting a Crew Exploration Vehicle-like atmospheric re-entry capsule
- Computational Fluid Dynamics Methane Technology Testbed model of a rocket thruster
- Thermal model of Mars Exploration Rover Cruise Stage
- Model of a Fine Guidance Sensor for the James Webb Space Telescope
- Simulation of the Kepler telescope, emphasizing detection of planet transients around the host star
- Model of an oceanographic sensor
- Matlab$^{@TM}$-based discrete event simulation for interplanetary logistics in building up and sustaining a lunar outpost
- Extend $^{@TM}$-based discrete event simulation to assess readiness and launch availability for the Crew Launch Vehicle
- Uncertainty analysis of historical hurricane data, in support of hurricane predictions.

Many of the comments from these applications were included in the formal comments submitted on the standard. These comments addressed documentation, cost drivers, use of the credibility scale, and benefits derived from following the standard as noted below.

Regarding documentation, there were two items noted. The first item involved comments from several pilots regarding the extensive documentation and associated costs. Two specific pilots noted that where they had hundreds to thousands of inputs, the documentation requirements would be impractical and challenged as to how much this documentation effort would contribute to conveying results to the decisions makers.

The second item involved the requirements defined in the standard with regard to simulations that have a short-term usage. That is, for models and simulations that have a long life and evolve over many years, having detailed documentation and compliance to the standard is readily justified. But, for short-term, single use simulations, maybe there should be some level of tailoring. Both of these documentation concerns can be addressed at the project level, because the project can specify the level of documentation required.

Regarding cost drivers, many of the pilots provided comments on cost drivers identifying areas such as documentation, uncertainty quantification, additional processes, validation, and others. There were cost increases noted for each pilot, but the maturity of the simulation being reviewed had such a significant affect on these costs, that the costs could not be normalized for generic reporting. The key element here is that only through monitoring of the initial applications of this standard can the cost impacts be assessed with more rigor.

Regarding the use of the credibility scale, there was one Center with a relatively flat organizational structure, where the decision makers were intimately familiar with the design and development of the simulation. Due to this, these decision makers wanted to see the actual results from the verification, validation, input pedigree, etc., rather than being presented a mapping to the factor levels of the credibility scale.

Regarding benefits derived from the use of this standard, there were several comments across a wide spectrum. One manager wanted to take the standard and use it as guidance for his entire simulation group to develop more standardized products across all the projects. One Center indicated that following the standard would provide easier maintenance, updates, and training due to the improved documentation. Another Center felt that it would standardize the reporting allowing better comparisons between similar modeling efforts. Another felt that it would allow for improved repeatability and reduced project efforts. And, one final input, that had a mixed response, was that if a typical approach for developing a simulation received a low credibility rating that it may be perceived more negatively than previously believed leading to future activities requiring higher credibility levels at increased cost – although viewed negatively regarding cost, this is probably one of the desired results from using the credibility scale is for the decision makers to see the level that is achieved.

The above inputs from the pilot studies were used to update the interim standard to the version that is now in the approval process. Responses to these comments were addressed in the formal comments of the standard. It will now be up to the monitoring activities associated with implementation of the standard which issues ought to be re-addressed in the inevitable future revision.

Typical NASA Applications

To correlate the NASA activities associated with this M&S Standard with applications in other industries, specific NASA applications are noted below. NASA conducts two general categories of missions. The first category is crewed missions, such as the Space Shuttle and International Space Station, where critical decisions involving both human safety and mission success are key drivers identifying simulations that fall within the scope of this standard. The second category is robotic missions such as the Hubble Space Telescope, Mars Exploration Rovers, and the New Horizons mission to Pluto where mission success is the primary driver.

The key element that determines which simulations fall within the scope of this standard is the project's determination of the simulations supporting critical decisions. In one situation, the structural model of a specific component may not fall within the scope, but in another mission that component may be associated with a critical event that requires performance to ensure safety or mission success. Thus, as noted below, structural finite element models typically fall within the scope of the standard, but the model is typically associated with a "system-level" entity such as an entire spacecraft and not one specific component such as the mounting of an antenna.

Example simulation applications used within NASA:

- Crew training (ascent and entry simulations, payload simulation, landing simulation)
- Aerodynamics & aerothermodynamics [e.g. external fuel tank on Columbia, shuttle re-entry]
- Atmospheric models
- Guidance, Navigation, and Control models
- Propulsion models
- Structural/Thermal Finite Element Models
- Orbital Debris Models
- Shuttle and ISS operations models

One application area that is interesting to note is where the spacecraft returns to earth or lands on another planet. These events are treated as mission critical events because once they are initiated, they can't be aborted – impact mission success. As shown in Figure 7, simulations of this magnitude include many different models such as navigation models, aerodynamic models, atmospheric models, planet models, and parachute models all integrated into a Monte Carlo simulation to assess the probability of success.
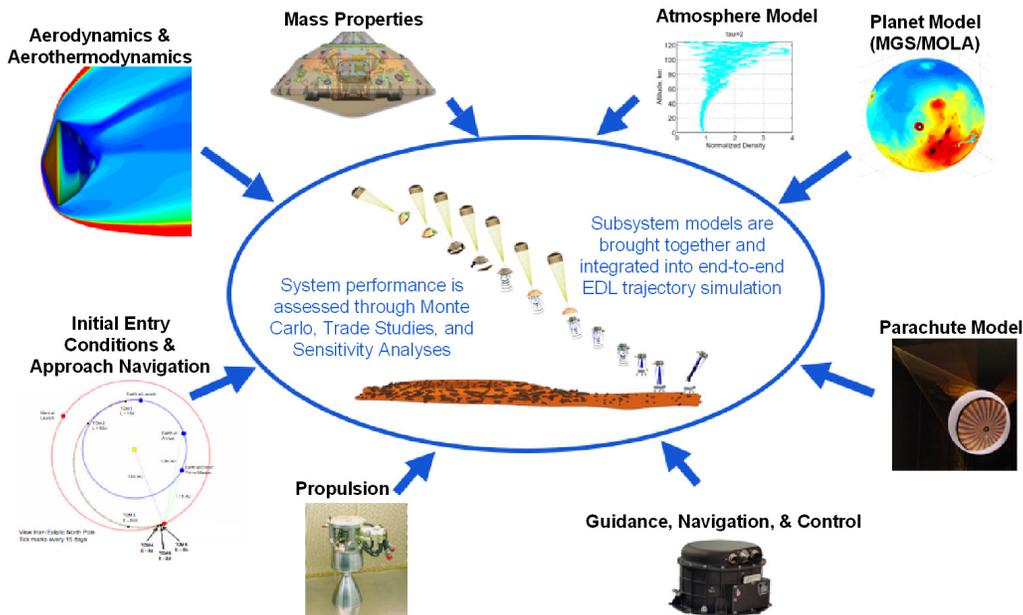

Figure 7. Typical Mars Entry, Descent, Landing Simulation Elements

Recent M&S Management Activities in Anticipation of M&S Standard Release

Several existing programs ventured out to develop their own guidance regarding management of the M&S activities to improve their processes in anticipation of the release of the M&S Standard. Two specific examples are the August 2007 Shuttle Endeavour (STS-118) tile damage evaluation and the Mars Science Laboratory (MSL) mission.

In the evaluation of the Shuttle Endeavor tile damage; there were several simulation activities that were performed. In this effort, three categories for simulation maturity were identified as shown in Figure 8 and then the simulations used in the evaluation were assessed to determine the appropriate category. This information was combined with peer review results and presented to the decision makers to assess the confidence in the resulting analysis as shown in Figure 9.

For the Mars Science Laboratory (MSL) mission, an extensive checklist was developed that is used to evaluate the simulation results and assess its applicability for the intended use. This checklist includes sections on topics such as, 1) compliance to requirements and design documentation, 2) results of verification and validation activities, 3) status of documentation, and 4) the culminating statement that the simulation is ready for its intended use. (Note how prominently technical review figures into this assessment.)

	Model Category	Peer Review
Aeroheating	A	✓
Thermal	A	✓
Stress	A	✓
Analysis Confidence	High – standard process	

Figure 9. Example Simulation Evaluation Using Above Categories

7. Summary

Some of the lessons learned from this process were:

1. Developing an M&S standard that covers all types of models and simulations and all phases of the modeling and simulation process is extremely challenging.
2. Trained facilitation was extremely useful in containing the passionate "discussions" about the scale
3. Once a decision is made, the temptation to revisit that decision is contained by a firm rule requiring a formal motion accompanied by a second to even to begin the discussion
4. Pilot studies are very important in bringing practical experience to bear on the development of a new standard
5. The supermajority rule for final decisions is critical to ensuring that the final product had consensus support from the TWG.
6. Dedicated funding (as opposed to a volunteer activity) and involvement of practitioners was extremely beneficial to ensuring a feasible standard that would be accepted by the M&S community
7. A high-level champion, in this case, the OCE, was indispensable to overcoming barriers

This issuance of the NASA M&S Standard is just the "end of the beginning" of bringing more rigor to NASA's use of M&S. Of the many actions that are needed in the near future, we highlight two. First, the agency should actively monitor its use to collect data and suggestions for future improvements. Second, the agency should underwrite the development of recommended practices guides by the various M&S communities in order to improve the state of the practice of M&S in the agency.

- **Category A**
  - Baselined model used within model limitations or intended use
  - Requires no additional reporting
- **Category B**
  - Baselined model with minor modifications
  - Uncorrelated models which are baselined
  - New environments developed with standard processes
  - Requires identification of minor modifications
    - The amount of information will vary based on the scenario
- **Category C**
  - Unbaselined models
  - Models developed during real-time or baselined model with extensive modifications
  - Baselined model used outside model limitations or intended use
  - In-flight test
  - New environments developed with non-standard processes
  - Requires detailed explanation of model utilization

Figure 8. Simulation Categories Used in STS-118 Shuttle Tile Evaluation

8. References

[1] Columbia Accident Investigation Board Report. (August 2003). Vol. 1. [http://caib.nasa.gov/]

[2] A Renewed Commitment to Excellence: An Assessment of the NASA Agency-wide Applicability of the Columbia Accident Investigation Board Report. (January 30, 2004) [http://www.nasa.gov/pdf/55691main_Diaz_020204.pdf]

[3] Standard for Models and Simulations, NASA-STD-7009, in final management review

[4] Standard for Models and Simulations, NASA-STD-(I)-7009, Dec. 1, 2006

[5] Oberkampf, W.L.; Pilch, M.; Trucano, T.G. (October 2007). Predictive Capability Maturity Model for Computational Modeling and Simulation, SAND2007-5948, Sandia National Laboratories

[6] Harmon, S.Y.; Youngblood, S.M. (2005). A Proposed Model for Simulation Validation Process Maturity, J. Defense Modeling & Simulation. Vol. 2, No. 4, pp. 179-190

[7] Balci, O. (2004). Quality Assessment, Verification, and Validation of Modeling and Simulation Applications. Proceedings of the 2004 Winter Simulation Conference. R.G. Ingalls; M.D. Rossetti; J.S. Smith; B.A. Peters, eds. Dec. 5-8. Piscataway, NJ: IEEE. pp. 122-129

[8] Hale, J.P.; Hartway, B.L.; Thomas, D.A. (2007). A Common M&S Credibility Criteria-set Supports Multiple Problem Domains. The 5th Joint Army-Navy-NASA-Air Force (JANNAF) Modeling and Simulation Subcommittee Meeting, May, CDJSC 49. Columbia, MD: Johns Hopkins University

[9] Mehta, U.B. (2007). Simulation Credibility Level. The 5th Joint Army-Navy-NASA-Air Force (JANNAF) Modeling and Simulation Subcommittee Meeting, CDJSC 49, May, CPIAC. Columbia, MD: Johns Hopkins University

Author Biographies

WILLIAM J. BERTCH is a System Engineer at the Jet Propulsion Laboratory. Prior that he worked as a system engineer and software engineer supporting defense contracts and Dept. of Energy nuclear research activities.

THOMAS A. ZANG the Chief Technologist of the Systems Analysis and Concepts Directorate at the NASA Langley Research Center. Prior to that, from 1994-2002, he headed the Multidisciplinary Optimization Branch. These organizations conducted multidisciplinary analyses ranging from space exploration architecture studies to preliminary-design-level optimization. His personal M&S experience has focused on PDE-based models, especially on spectral methods for fluid dynamics, on which subject he has co-authored three textbooks. He chaired both the Development Team and the Topic Working Group during the development of the M&S Standard.

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.