NASA TECHNICAL STANDARD

NASA-STD-7009

National Aeronautics and Space Administration
Washington, DC 20546-0001

Approved: 07-11-2008

Expiration Date: 07-10-2013

Superseding NASA-STD-(I)-7009

STANDARD FOR MODELS AND SIMULATIONS

MEASUREMENT SYSTEM IDENTIFICATION: NOT MEASUREMENT SENSITIVE

APPROVED FOR PUBLIC RELEASE — DISTRIBUTION IS UNLIMITED

DOCUMENT HISTORY LOG

Status	Document Revision	Approval Date	Description
Interim		12-01-2006	Interim Baseline Release
Baseline		07-11-2008	Baseline Release

FOREWORD

This standard is published by the National Aeronautics and Space Administration (NASA) to provide uniform engineering and technical requirements for processes, procedures, practices, and methods that have been endorsed as standard for models and simulations (M&S) developed and used in NASA programs and projects, including requirements for selection, application, and design criteria of an item. This standard was specifically developed to respond to Action 4 from the 2004 report "A Renewed Commitment to Excellence," with consideration also given to related findings as identified in the Columbia Accident Investigation Board (CAIB) Report.

This standard is approved for use by NASA Headquarters and NASA Centers, including Component Facilities.

This standard covers the development and operation (or execution) of M&S, as well as the analysis and presentation of the results from M&S. This also includes the proper training of M&S practitioners and the identification of recommended practices, while ensuring the credibility of the results from M&S is assessed and properly conveyed to those making critical decisions.

Requests for information, corrections, or additions to this standard should be submitted via "Feedback" in the NASA Technical Standards System at http://standards.nasa.gov.

Original Signed By July 11, 2008

Michael G. Ryschkewitsch
NASA Chief Engineer

Approval Date

TABLE OF CONTENTS

TABLE OF CONTENTS, continued

LIST OF FIGURES

LIST OF TABLES

STANDARD FOR MODELS AND SIMULATIONS

1. SCOPE

1.1 Purpose

This standard was developed in response to Action 4 from the 2004 report "A Renewed Commitment to Excellence," which stated the following:

"Develop a standard for the development, documentation, and operation of models and simulations

a. 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),
b. Develop process for tool verification and validation, certification, reverification, revalidation, and recertification based on operational data and trending,
c. Develop standard for documentation, configuration management, and quality assurance,
d. Identify any training or certification requirements to ensure proper operational capabilities,
e. 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,
f. Develop a process for user feedback when results appear unrealistic or defy explanation."

Subsequently, in 2006, the NASA Chief Engineer provided the following further guidance:

g. "Include a standard method to assess the credibility of the models and simulations presented to the decision maker when making critical decisions (i.e., decisions that effect human safety or mission success) using results from models and simulations,
h. Assure that the credibility of models and simulations meet the project requirements."

Each of the requirements and recommendations in this standard can be traced to one or more of the eight objectives listed above. The traceability matrix of the requirements in this standard to the eight objectives can be found online upon accessing this standard at URL http://standards.nasa.gov; refer to "Requirements Traceability Matrix." Some of these objectives are met by recommendations rather than by requirements as a result of either (a) the practical impossibility of satisfying the requirement in all cases, or (b) further guidance received from NASA Headquarters.

These eight objectives are encapsulated in the overall goal for this standard, which is to ensure that the credibility of the results from M&S is properly conveyed to those making critical decisions. 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. The intent is to reduce the risks associated with critical decisions. This standard covers the development and operation (or execution) of M&S as well as the processes of analysis and presentation of the results from the M&S.

This standard addresses aspects of M&S that are common across NASA activities. Discipline-specific details of M&S should be addressed in future documents, such as Recommended Practices (usually entitled "Handbooks" in the NASA document hierarchy), and are not included in this standard.

The scope of this standard covers the development and maintenance of models, the operation of simulations, the analysis of the results, training, recommended practices, the assessment of the M&S credibility, and the reporting of the M&S results. Some of the key features of this standard are requirements and recommendations for verification, validation, uncertainty quantification, training, credibility assessment, and reporting to decision makers; also included are the cross-cutting areas of documentation and configuration management (CM).

The requirements/recommendations in sections 4.7 and 4.8 are the culmination of the standard. The requirements/recommendations in sections 4.1 – 4.6 are intended to support the requirements in sections 4.7 and 4.8. This is accomplished by ensuring that sufficient details of the M&S process along with intermediate results are available to support the requirements in sections 4.7 and 4.8 and to respond to in-depth queries by the decision maker. Appendix A provides guidance for assessing the risk of using M&S in engineering decisions. Appendix B provides details related to some of the requirements/recommendations in sections 4.7 and 4.8. Appendix C contains a template for a compliance matrix.

1.2 Applicability

This standard applies to M&S used by NASA and its contractors for critical decisions in design, development, manufacturing, ground operations, and flight operations. (Guidance for determining which particular M&S are in scope is provided in section 4.1 and Appendix A.) This standard also applies to 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. Generally, for such M&S, particular attention may need to be paid to defining the limits of operation and to verification, validation, and uncertainty quantification. Programs and projects are encouraged to apply this standard to M&S, if the M&S results may impact future critical decisions.

This standard does not apply to M&S that are embedded in control software, emulation software, and stimulation environments. However, Center implementation plans for NPR 7150.2, NASA Software Engineering Requirements, should specifically cover embedded M&S, and address

such M&S-specific issues as numerical accuracy, uncertainty analysis, sensitivity analysis, M&S verification, and M&S validation.

This standard may be cited in contract, program, and other Agency documents as a technical requirement. Requirements are indicated by the word "shall"; explanatory or guidance text is indicated in italics.

1.2.1 Tailoring for application to a specific program or project shall be formally documented as part of program or project requirements and approved by the Technical Authority.

1.3 Focus

In general, standards may focus on engineering/technical requirements, processes, procedures, practices, or methods. This standard focuses on requirements and recommendations. Hence, this standard specifies *what* shall or should be done; it does not prescribe *how* the requirements are to be met, nor does it specify *who* is the responsible party for complying with the requirements.

2. APPLICABLE DOCUMENTS

2.1 General

The documents listed in this section contain provisions that constitute requirements of this standard as cited in the text of section 4.

2.1.1 The latest issuances of cited documents shall be used unless otherwise approved by the assigned Technical Authority.

The applicable documents are accessible via the NASA Online Directives Information System at http://nodis3.gsfc.nasa.gov and the NASA Technical Standards System at http://standards.nasa.gov, or may be obtained directly from the Standards Developing Organizations or other document distributors.

2.2 Government Documents

None.

2.3 Non-Government Documents

None.

2.4 Order of Precedence

This document establishes requirements and guidance for models and simulations but does not supersede nor waive established Agency requirements found in other documentation.

2.4.1 Conflicts between this standard and other requirements documents shall be resolved by the responsible Technical Authority.

3. ACRONYMS AND DEFINITIONS

3.1 Acronyms and Abbreviations

AIAA	American Institute of Aeronautics and Astronautics
ASME	American Society of Mechanical Engineers
CAIB	Columbia Accident Investigation Board
CAS	Credibility Assessment Scale
CM	Configuration Management
COTS	Commercial-Off-The-Shelf
CPIAC	Chemical Propulsion Information Analysis Center
DMSO	Defense Modeling and Simulation Office
GOTS	Government-Off-The-Shelf
IEEE	Institute of Electrical and Electronics Engineers
ISG	Implementation Study Group
ISO	International Organization for Standardization
JANNAF	Joint Army-Navy-NASA-Air Force
M&S	Models and Simulations
MOTS	Modified-Off-The-Shelf
NASA	National Aeronautics and Space Administration
NASTRAN	NASA Structural Analysis
NPR	NASA Procedural Requirements
PMBA	Primary Mirror Backplane Assembly
Req.	Requirement
RPG	Recommended Practices Guide
SISO	Simulation Interoperability Standards Organization
STD	Standard
V&V	Verification & Validation
VV&A	Verification, Validation, and Accreditation

3.2 Definitions

The definitions listed below are those used in this document. Wherever possible, these definitions have been taken from official NASA documents. In some cases, after reviewing definitions of interest in the International Organization for Standardization (ISO), the Defense Modeling and Simulation Office (DMSO), professional society publications, and English language dictionaries, some of these definitions were taken or adapted from these sources to achieve the goal and objectives stated in section 1.1. Some definitions may have alternate meanings in other documents and disciplines.

Abstraction: The process of selecting the essential aspects of a reference system to be represented in a model or simulation while ignoring those aspects that are not relevant to the purpose of the model or simulation (adapted from *Fidelity ISG Glossary, Vol. 3.0*).

Accuracy: The difference between a parameter or variable (or a set of parameters or variables) within a model, simulation, or experiment and the true value or the assumed true value.

Analysis: Any post-processing or interpretation of the individual values, arrays, files of data, or suites of executions resulting from a simulation.

Artifact: Any tangible product that is produced by the project team, i.e., requirements documents, help systems, code, executables, test documentation, test results, diagrams, etc.

Calibration: The process of adjusting numerical or modeling parameters in the model to improve agreement with a referent.

Computational Model: The numerical representation of the mathematical model.

Conceptual Model: The collection of abstractions, assumptions, and descriptions of physical processes representing the behavior of the reality of interest from which the mathematical model or validation experiments can be constructed (adapted from *ASME V&V 10*).

Configuration Management (CM): A management discipline applied over the product's life cycle to provide visibility into and to control changes to performance, functional, and physical characteristics (NPR 7120.5D, NASA Space Flight Program and Project Management Requirements).

Credibility: The quality to elicit belief or trust in M&S results.

Critical Decision: Those technical decisions related to design, development, manufacturing, ground, or flight operations that may impact human safety or mission success, as measured by program/project-defined criteria.

Emulation: The use of an M&S to imitate another system, so that the M&S behaves like or appears to be the other system.

Endorsement: A formal assurance that a product, process, or service conforms to specified characteristics. (Examples of endorsement include "accreditation"—the official acceptance of a model or simulation and its associated data to use for a specific purpose—and "certification," which is similar to accreditation, but often applies to a class of purposes or a general domain and generally implies an independent and/or third-party certifier.)

Human Safety: The condition of being protected from death, permanently disabling injury, severe injury, and several occupational illnesses. In the NASA context this refers to safety of the public, astronauts, pilots and the NASA workforce (adapted from NPR 8000.4 and the NASA Safety Hierarchy).

Limits of Operation: The boundary of the set of parameters for which an M&S result is acceptable based on the program/project-required outcomes of verification, validation, and uncertainty quantification.

Mathematical Model: The mathematical equations, boundary values, initial conditions, and modeling data needed to describe the conceptual model (*ASME V&V 10*).

Mission Success Criteria: Standards against which the program or project will be deemed a success. Mission success criteria may be both qualitative and quantitative, and may cover mission cost, schedule, and performance results as well as actual mission outcomes (NPR 7120.5C, NASA Program and Project Management Processes and Requirements).

Model: A description or representation of a system, entity, phenomena, or process (adapted from Banks, J., ed. (1998). *Handbook of Simulation*. New York: John Wiley & Sons). (A model may be constructed from multiple sub-models; the sub-models and the integrated sub-models are all considered models. Likewise, any data that goes into a model is considered part of the model. A model of a model (commonly called a metamodel), e.g., a response surface constructed from the results of M&S, is considered a model).

Referent: Data, information, knowledge, or theory against which simulation results can be compared (adapted from *ASME V&V 10*).

Risk: The combination of the probability that a program or project will experience an undesired event and the consequences, impact, or severity of the undesired event, if it were to occur. Both the probability and consequences may have associated uncertainties (adapted from NPR 7120.5D).

Sensitivity Analysis: The study of how the variation in the output of a model can be apportioned to different sources of variation in the model input and parameters (adapted from Saltelli and others, 2000).

Simulation: The imitation of the characteristics of a system, entity, phenomena, or process using a computational model.

Stimulation: The description of a type of simulation whereby artificially generated signals are provided to real equipment in order to trigger it to produce the result required for verification of a real-world system, training, maintenance, or for research and development.

Subject Matter Expert: An individual having education, training, or experience in a particular technical or operational discipline, system, or process and who participates in an aspect of M&S requiring his/her expertise.

Tailoring: The documentation and approval of the adaptation of the processes and approach to complying with requirements according to the purpose, complexity, and scope of a NASA program or project. (NPR 7123.1A, NASA Systems Engineering Processes and Requirements).

Uncertainty: (1) The estimated amount or percentage by which an observed or calculated value may differ from the true value (The American Heritage Dictionary of the English Language, 4th ed.). (2) A broad and general term used to describe an imperfect state of knowledge or a variability resulting from a variety of factors including, but not limited to, lack of knowledge, applicability of information, physical variation, randomness or stochastic behavior, indeterminacy, judgment, and approximation (adapted from NPR 8715.3B, NASA General Safety Program Requirements).

Uncertainty Quantification: The process of identifying all relevant sources of uncertainties, characterizing them in all models, experiments, and comparisons of M&S results and experiments, and of quantifying uncertainties in all relevant inputs and outputs of the simulation or experiment.

Validation: 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.

Verification: The process of determining that a computational model accurately represents the underlying mathematical model and its solution from the perspective of the intended uses of M&S.

Waiver: A documented authorization intentionally releasing a program or project from meeting a requirement (NPR 7120.5D). Deviations and exceptions are considered special cases of waivers.

4. REQUIREMENTS

This standard establishes a minimum set of requirements and recommendations for the use of M&S to support critical decisions.

For decisions based on results from M&S, the risk assumed by the decision maker is often misestimated due to inadequate assessment of uncertainties within M&S development, verification, validation, execution, analysis, and reporting. This standard establishes practices to enable a more accurate assessment of this risk by making M&S credibility more apparent to the decision maker. This standard emphasizes documentation and CM of M&S to enforce transparency, repeatability, and traceability; and it requires that key M&S personnel receive appropriate training.

The requirements and recommendations are generic in nature because of their broad applicability to all types of M&S. Implementation details of the M&S requirements should be addressed in discipline-specific Recommended Practices, project/program management plans, etc.

The following organizational structure is employed in this standard:

4.1 Programmatics

4.2 Models

4.3 Simulations and Analyses

4.4 Verification, Validation, and Uncertainty Quantification

4.5 Identification and Use of Recommended Practices

4.6 Training

4.7 Assessing the Credibility of M&S Results

4.8 Reporting Results to Decision Makers

In many instances, the modeling, simulation, and analysis activities are interwoven, particularly during the development, verification, and validation phases. This standard is intended to be inclusive of all these possibilities.

Many of the requirements in this standard require documentation. With the exception of the documentation required for reports to decision makers (section 4.8), this documentation may consist of a reference to other existing documents, such as a journal article, a technical report, or a program/project document, provided that all the required details are contained in the referenced document(s).

4.1 Programmatics

Critical decisions that are based entirely or partially on M&S are usually made within the context of a program or project. Program and project management have the responsibility to identify and document the parties responsible for complying with the requirements in this standard. The actual person identified by program and project management to fulfill the role of the "responsible party" in specific requirements will likely vary depending upon the context of the requirement; for example, the responsible party might be the lead, or another supporting person associated with the model development, operation, analysis, and/or reporting of results to decision makers.

Program and project management in collaboration with the Technical Authority have the responsibility to identify and document the extent and level of formality of documentation needed to meet the documentation requirements in this standard. Some requirements, in particular, 4.1.5, 4.2.6, 4.2.8, 4.3.6, 4.4.1, 4.4.2, 4.4.4, 4.4.5, 4.4.6, 4.4.7, 4.4.8, and 4.4.9, are to be interpreted as meaning that the activity in question is not required per se, but that whatever was done is to be documented, and if nothing was done a clear statement to that effect is to be documented.

Program and project management in collaboration with the Technical Authority have the responsibility to identify and document the critical decisions to be addressed with M&S and to determine which M&S are in scope. The latter determination should be based upon the risk posed by the anticipated use of the M&S. Appendix A describes a representative M&S risk assessment matrix for this purpose. Furthermore, the Technical Authority has the particular responsibility to assure appropriate outcomes of Req. 4.1.3.

The responsible party performs the following:

Req. 4.1.1 – Shall document the risk assessment for any M&S used in critical decisions,

Req. 4.1.2 – Shall identify and document those M&S that are in scope.

Req. 4.1.3 – Shall define the objectives and requirements for M&S products including the following:

a. The acceptance criteria for M&S products, including any endorsement for the M&S.

b. The rationale for the weights used for the subfactors in the Credibility Assessment Scale (see Appendix B.4).

c. Intended use.

d. Metrics (programmatic and technical).

e. Verification, validation, and uncertainty quantification (see section 4.4).

f. Reporting of M&S information for critical decisions (see section 4.8).

g. CM (artifacts, timeframe, processes) of M&S.

(The acceptance criteria in 4.1.3 (a) includes specification of what constitutes a favorable comparison for the Verification Evidence, Validation Evidence, Input Pedigree Evidence, and Use History level definitions in the Credibility Assessment Scale (see Appendix B).)

Req. 4.1.4– Shall develop a plan (including identifying the responsible organization(s)) for the acquisition, development, operation, maintenance, and/or retirement of the M&S.

Req. 4.1.5 – Shall document any technical reviews performed in the areas of Verification, Validation, Input Pedigree, Results Uncertainty, and Results Robustness (see Appendix B).

Req. 4.1.6 – Shall document M&S waiver processes.

Req. 4.1.7 – Shall document the extent to which an M&S effort exhibits the characteristics of work product management, process definition, process measurement, process control, process change, and continuous improvement, including CM and M&S support and maintenance.

4.2 Models

The processes of developing conceptual, mathematical, or computational models are all considered to be modeling activities. Empirically adjusting the results of a simulation in an attempt to improve correlation is considered a modeling activity.

For models, the responsible party performs the following:

Req. 4.2.1 – Shall document the assumptions and abstractions underlying the conceptual model, including their rationales.

Req. 4.2.2 – Shall document the basic structure and mathematics of the model (e.g., reality modeled, equations solved, behaviors modeled, conceptual models).

(For COTS, GOTS, MOTS, and legacy M&S, some of the documentation required in 4.2.1 and 4.2.2 may be available in published user guides; a reference to the user guides will suffice for this part of the documentation.)

Req. 4.2.3 – Shall document data sets and any supporting software used in model development and input preparation.

Req. 4.2.4 – Shall document required units and vector coordinate frames (where applicable) for all input/output variables in the M&S.

Req. 4.2.5 – Shall document the limits of operation of models.

Req. 4.2.6 – Shall document any methods of uncertainty quantification and the uncertainty in any data used to develop the model or incorporated into the model.

Req. 4.2.7 – Shall document guidance on proper use of the model.

(Guidance on proper use of a model includes descriptions of appropriate practices for set-up, execution, and analysis of results.)

Req. 4.2.8 – Shall document any parameter calibrations and the domain of calibration.

Req. 4.2.9 – Shall document updates of models (e.g., solution adjustment, change of parameters, calibration, and test cases) and assign unique version identifier, description, and the justification for the updates.

Req. 4.2.10 – Shall document obsolescence criteria and obsolescence date of the model.

(Obsolescence refers to situations where changes to the real system invalidate the model—see item (e) of Diaz Action #4.)

Req. 4.2.11 – Shall provide a feedback mechanism for users to report unusual results to model developers or maintainers.

Req. 4.2.12 – Shall maintain (conceptual, mathematical, and computational) models and associated documentation in a controlled CM system**.**

Req. 4.2.13 – Shall maintain the data sets and supporting software referenced in Req. 4.2.3 and the associated documentation in a controlled CM system**.**

4.3 Simulations and Analyses

The execution (operation) of a computational model and the processing of the results from the simulation are simulation and analysis activities, respectively.

For simulations and analysis, the responsible party performs the following:

Req. 4.3.1 – Shall do either of the following:

a. Ensure that simulations are conducted within the limits of operation of the models, or

b. Placard the simulation and analysis results with a warning that the simulation may have been conducted outside the limits of operation and include the type of limit that may have been exceeded, the extent that the limit might have been exceeded, and an assessment of the consequences of this action on the M&S results.

Req. 4.3.2 – Shall document and explain any observed warning and error messages resulting from the execution of the computational model.

Req. 4.3.3 – Shall document which computational models were used (including revision numbers) in the simulation.

Req. 4.3.4 – Shall document the versions of M&S results.

Req. 4.3.5 – Shall document data used as input to the simulation, including its pedigree (see Appendix B).

Req. 4.3.6 – Shall document any unique computational requirements (e.g., support software, main memory, disk capacities, processor, and compilation options).

Req. 4.3.7 – Shall document the processes for conducting simulations and analyses for generating results reported to decision makers.

Req. 4.3.8 – Shall document any use history of M&S, in the same or similar applications, which are relevant for establishing the credibility of the current M&S application (see Appendix B).

Req. 4.3.9 – Shall document the assessment as to the appropriateness of the simulation and analysis relative to its intended use.

Req. 4.3.10 – Shall document the rationale for the setup and execution of the simulation and analysis.

Recommendations for simulations and analyses are:

a. The relevant characteristics of the system that is modeled should be documented.

b. CM records should contain test cases that span the limits of operation for the M&S defined by the program or project. "Test cases" are defined as benchmark input/output sets used to verify proper execution of the M&S.

c. The simulation should fail in a manner that prevents misuse and misleading results.

(1) The simulation should provide messages that detail the failure mode and point of failure.

(2) The analyst should document and explain all failure modes, points of failure, and messages indicating such failures.

4.4 Verification, Validation, and Uncertainty Quantification

The definitions of verification and validation used in this standard are those of the M&S community. The present definitions are consistent with those of DMSO's VV&A Recommended Practices Guide (RPG Build 3.0), the American Institute of Aeronautics and Astronautics (AIAA) Guide for Verification and Validation of Computational Fluid Dynamics Simulation (AIAA G-077), and Guide for Verification and Validation in Computational Solid Mechanics (ASME V&V 10-2006). These differ from the Institute of Electrical and Electronics Engineers (IEEE's) Standard Glossary of Software Engineering Terminology (IEEE 610.12-1990), and NPR 7123.1A.

Verification

For verification, the responsible party performs the following:

Req. 4.4.1 – Shall document any verification techniques used and any domain of verification (e.g., the conditions under which verification was conducted).

Req. 4.4.2 – Shall document any numerical error estimates (e.g., numerical approximations, insufficient discretization, insufficient iterative convergence, finite-precision arithmetic) for the results of the computational model.

Req. 4.4.3 – Shall document the verification status of (conceptual, mathematical, and computational) models.

Validation

For validation, the responsible party performs the following:

Req. 4.4.4 – Shall document any techniques used to validate the M&S for its intended use, including the experimental design and analysis, and the domain of validation.

Req. 4.4.5 – Shall document any validation metrics, referents, and data sets used for model validation.

Req. 4.4.6 – Shall document any studies conducted and results of model validation.

Uncertainty Quantification

For uncertainty quantification, the responsible party performs the following:

Req. 4.4.7 – Shall document any uncertainty quantification processes used for

a. The referent data.

b. The input data.

c. The M&S results.

d. The propagation of uncertainties.

e. The quantities derived from M&S results.

Req. 4.4.8 – Shall document any quantified uncertainties, both physical and numerical, for

a. The referent data.

b. The input data.

c. The M&S results.

d. The propagation of uncertainties.

e. The quantities derived from M&S results.

Req. 4.4.9 – Shall document the extent and results of any sensitivity analyses performed with the M&S.

Recommendations for verification, validation, and uncertainty quantification of M&S are that the responsible party should document the following:

a. Any aspects of M&S that have not been verified.

b. Any aspects of M&S that have not been validated.

c. If any significant physical processes, effects, scenarios, or environments have not been considered in the uncertainty quantification analysis.

4.5 Identification and Use of Recommended Practices

This standard addresses general issues with respect to the use of M&S. Implementation details that are specific to individual programs, projects, disciplines, or processes are not addressed in this standard. The implementation details are addressed in Recommended Practices. This section describes a requirement and recommendations for the identification of Recommended Practices (e.g., guidelines developed by professional societies such as AIAA G-077 and ASME V&V 10, best practices documented for specific simulation codes, and NASA Handbooks and Guidebooks).

For the use of Recommended Practices, the responsible party performs the following:

Req. 4.5.1 – Shall identify and document any Recommended Practices that apply to M&S for the program/project.

Recommendations for this section are that Recommended Practices for the following should be identified:

a. Input data verification and validation.

b. A quantified method of tracking adherence to Recommended Practices.

c. The purposes and objectives for the M&S and their pedigree.

d. Verification and validation processes for the M&S.

e. Uncertainty quantification methods for the M&S

f. Understanding of the disciplines incorporated in the M&S.

g. Analyzing and interpreting the M&S results including documentation of inference guidelines and statistical processes used.

h. Recognizing and capturing the need for any changes or improvements in the M&S.

i. Reporting procedures for results.

j. Identify best practices for user interface design to constrain the operation of the simulation to within its limits of operations.

4.6 Training

Training refers to the process of providing instruction on the proper use of M&S so that an individual can develop, operate, or analyze the relevant M&S.

For training, the responsible party performs the following:

Req. 4.6.1 – Shall determine the depth of required training for developers, operators, and analysts.

Req. 4.6.2 – Shall document the following:

a. Training topics required for developers, operators, and analysts of M&S.

b. Process and criteria for verifying that training requirements are met.

Req. 4.6.3 – Shall determine the qualifications for developers, operators, and analysts.

Recommended training topics for developers, operators, and analysts of M&S include:

a. The intended use of limits of operation for models.

b. CM requirements.

c. Documentation requirements as specified in sections 4.2, 4.3, and 4.4 of this standard.

d. How to recognize unrealistic results from simulations.

e. Feedback processes to improve M&S processes and results, including providing feedback for results that are not credible, are unrealistic, or defy explanation.

f. Sensitivity analysis.

g. Uncertainty quantification.

h. Verification and validation.

i. How to report simulation results to decision makers.

j. Statistics and probability.

k. Discipline-specific recommended practices. Other applicable Agency policy, procedural requirements, and standards.

l. Basic structures, mathematics, assumptions, and abstractions.

4.7 Assessing the Credibility of M&S Results

In order to ensure that NASA decision makers are informed about the credibility of M&S results in terms of a common process and a common language, M&S results used in critical decisions will be assessed using the CAS in accordance with the requirements in this section. Since credibility is a subjective attribute, the CAS does not purport to determine credibility, but merely to assess key factors that (a) contribute to a decision-maker's assessment of credibility and (b) are sensibly assessed on a graduated scale. The CAS provides the decision maker with an assessment of the M&S results against key factors. This assessment is focused on the results produced by the M&S, and by the quality of outputs associated with the processes that produced the M&S results.

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/her 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 rigor of the processes used to produce them, against key factors that affect the credibility judgment. The fundamental premise of this approach is that as a general rule, the more rigorous the key processes used for generating the M&S results, the greater the credibility of the M&S results, all else (including the estimated uncertainty) being equal.

The reporting requirements in the following section include the results of the CAS along with several other contributing factors that are either standalone data (the uncertainty statement) or significant caveats.

The details of the CAS are provided in Appendix B. In summary, the scale is comprised of eight factors. Each factor is divided into levels ranging from 0 to 4, with level definitions describing the evidence necessary for achieving that particular level. Level 0 corresponds to insufficient or no evidence, i.e., the M&S produces results, but there is insufficient evidence to warrant even level 1 assessment on that factor. A single, summary score on the CAS is determined by the minimum of the eight factor scores, which produces a single number between 0 and 4.

This standard itself levies no requirements with respect to what levels must be achieved (the sufficiency threshold levels—see Appendix B.5), merely that the achieved levels must be determined and reported.

The requirements for the assessment of the credibility of the M&S results that support critical decisions are that the responsible party performs the following:

Req. 4.7.1 – Shall assess the credibility of M&S results for each of the eight factors in the CAS described in Appendices B.2 and B.3.

Req. 4.7.2 – Shall justify and document the credibility assessment for each of the eight factors referenced in Req. 4.7.1.

Req. 4.7.3 – Shall perform the roll-up to an overall score according to the process described in Appendix B.4.

A recommendation is to gain additional insight into the credibility of M&S results by applying the process in Appendix B.5 to calculate and report any gaps between the achieved scores and the program/project-defined threshold scores for each of the factors.

4.8 Reporting Results to Decision Makers

This section differs from previous sections in that it defines requirements and recommendations for providing a high-level synopsis of M&S outcomes relevant to the intended use.

The requirements associated with reporting to decision makers are stated below as follows:

Req. 4.8.1 – Reports to decision makers shall include explicit warnings for any of the following occurrences, accompanied by at least a qualitative estimate of the impact of the occurrence:

a. Any unachieved acceptance criteria (as specified in Req. 4.1.3 (a)).

b. Violation of any assumptions of any model (as specified in Req. 4.2.1).

c. Violation of the limits of operation (as specified in Req. 4.2.5).

d. Execution warning and error messages (see Req. 4.3.2).

e. Unfavorable outcomes from the intended use and setup/execution assessments (described in Req. 4.3.9 and Req. 4.3.10).

f. Waivers to any of the requirements in this standard.

(*In the absence of documentation for any of the requirements referenced in (a)–(e), a warning must be provided.)*

Req. 4.8.2 – Reports to decision makers of M&S results shall include an estimate of their uncertainty and a description of any processes used to obtain this estimate as defined in Req. 4.4.7 and Req. 4.4.8.

a. Reported uncertainty estimates shall include one of the following:

(1) A quantitative estimate of the uncertainty in the M&S results, or
(2) A qualitative estimate of the uncertainty in the M&S results, or
(3) A clear statement that no quantitative or qualitative estimate of uncertainty is available.

Req. 4.8.3 – Reports to decision makers shall include the level of credibility for the M&S results and the subfactor weights, using the process specified in section 4.7.

For Req. 4.8.2, a complete quantitative uncertainty estimate would provide uncertainty intervals about the M&S results and confidence statements based on analysis, whereas a qualitative uncertainty estimate would be provided only in linguistic terms, e.g., small, medium, or large, rather than in numeric terms. Qualitative uncertainty estimates would still require justification, for example, by the descriptive phrasing of a subject matter expert or by resort to analogy with the quantified sensitivity of similar problems.

Recommendations for reporting are that the following be observed:

a. Reports to decision makers should include concluding remarks stating whether the M&S results are credible enough for the intended use.

b. Reports to decision makers should identify how to access more detailed backup material, including high-level descriptions of the models used, and key assumptions for limits of validity.

c. Reports to decision makers of M&S results should be placed in the CM system.

d. Reports to decision makers should summarize deviations from established recommended practices.

e. Reports to decision makers should include dissenting technical opinions regarding the credibility of the results or any recommended actions.

f. Developers and analysts should convey serious concerns about M&S to project managers (and decision makers, if appropriate) as soon as they are known.

5. GUIDANCE

5.1 Reference Documents

Document Number	Document Title
	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.
	Banks, J., ed. (1998). *Handbook of Simulation.* New York: John Wiley & Sons.
	Clemen, R.T. (1996). *Making Hard Decisions; an Introduction to Decision Analysis, Second Edition.* Pacific Grove, CA: Brooks/Cole.
	Cooke, R.M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. New York: Oxford University Press.
	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.
	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.
	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.
	Saltelli, A.; Chan, K.; Scott, E.M., eds. (2000). *Sensitivity Analysis.* Chichester, England: John Wiley & Sons.
AIAA-2005-4524	Lin, J.; West, J.S.; Williams, R.W.; Tucker, P.K. (2005). CFD Code Validation of Wall Heat Fluxes for a GO_2/GH_2 Single Element Combustor.

AIAA-2008-2156	Blattnig, S.R.; Green, L.L.; Luckring, J.M.; Morrison, J.H.; Tripathi, R.K.; Zang, T.A. (2008). Towards a Credibility Assessment of Models and Simulations.
AIAA G-077	*Guide for Verification and Validation of Computational Fluid Dynamics Simulation.* (1998). Reston, VA: AIAA.
ASME V&V 10	*Guide for Verification and Validation in Computational Solid Mechanics.* (2006). New York, NY: ASME.
CAIB	Columbia Accident Investigation Board Report. (August 2003). Vol. 1.
PB2005-100968	A Renewed Commitment to Excellence: An Assessment of the NASA Agency-wide Applicability of the Columbia Accident Investigation Board Report. (January 30, 2004).
Fidelity ISG Glossary	Simulation Interoperability Standards Organization (SISO). Fidelity Implementation Study Group (ISG). *Fidelity ISG Glossary*. (Dec. 1998). Vol. 3.0. http://www.sisostds.org/doclib/doclib.cfm?SISO_RID_1000789.
IEEE 610.12-1990	IEEE Standard Glossary of Software Engineering Terminology.
NASA-STD-8719.13	Software Safety Standard.
NPR 7120.5C	NASA Program and Project Management Processes and Requirements.
NPR 7120.5D	NASA Space Flight Program and Project Management Requirements.
NPR 7123.1A	NASA Systems Engineering Processes and Requirements.
NPR 7150.2	NASA Software Engineering Requirements.
NPR 8000.4	Risk Management Procedural Requirements
NPR 8715.3B	NASA General Safety Program Requirements.
RPG Build 3.0	*VV&A Recommended Practices Guide*, Defense Modeling and Simulation Office. http://vva.dmso.mil/.

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

5.2 Key Word Listing

Analysis
Credibility
Model
Sensitivity Analysis
Simulation
Uncertainty Quantification
Validation
Verification

APPENDIX A
M&S RISK ASSESSMENT

The determination of which M&S fall within the scope of this standard should be based upon an assessment of the risk posed by the potential use of the M&S. Such M&S risk assessments consider (1) the consequences to human safety or mission success criteria if a decision proves incorrect, and (2) the degree to which M&S results influence a decision. This appendix provides a sample M&S risk assessment matrix to support this determination. Program and projects may, of course, adapt this to their particular consequence definitions (and number of consequence levels). The sample provided here is based upon NPR 8000.4.

A.1 Decision Consequence

Consequence classifications assess the impact of a decision that proves incorrect. The number of Consequence levels and most of the language is taken from NPR 8000.4. The last item in each class description has been added to address impact upon mission success criteria, such as science objectives.

a. Class IV - Negligible. A poor decision may result in the need for minor first aid treatment but would not adversely affect personal safety or health; damage to facilities, equipment, or flight hardware more than normal wear and tear level; internal schedule slip that does not impact internal development milestones; cost overrun less than 2 percent of planned cost; all mission success criteria met, with at worst minor performance degradations.

b. Class III - Moderate. A poor decision may result in minor injury or occupational illness, or minor property damage to facilities, systems, equipment, or flight hardware; internal schedule slip that does not impact launch date; cost overrun between 2 percent and not exceeding 15 percent of planned cost; a few (up to 25 percent) mission success criteria not met due to performance degradations.

c. Class II - Critical. A poor decision may result in severe injury or occupational illness, or major property damage to facilities, systems, equipment, or flight hardware; schedule slippage causing launch date to be missed; cost overrun between 15 percent and not exceeding 50 percent of planned; many (between 25 percent and 75 percent) mission success criteria not met due to substantial performance degradations.

d. Class I - Catastrophic. A poor decision may result in death or permanently disabling injury, facility destruction on the ground, or loss of crew, major systems, or vehicle during the mission; schedule slippage causing launch window to be missed; cost overrun greater than 50 percent of planned cost; most (more than 75 percent) mission success criteria not met due to severe performance degradations.

A.2 M&S Influence

Influence estimates the degree to which M&S results influence program/project engineering decisions. (Engineering decisions include determination of whether design requirements have been verified.)

a. Influence 1 - Negligible. Results from the M&S are a negligible factor in engineering decisions. This includes research on M&S methods, and M&S used in research projects that have no direct bearing on program/project decisions (for NASA missions).

b. Influence 2 - Minor. M&S results are only a minor factor in any program/project decisions. Ample flight or test data for the real system in the real environment are available, and M&S results are used just as supplementary information.

c. Influence 3 - Moderate. M&S results are at most a moderate factor in any program/project decisions. Limited flight or test data for the real system in the real environment are available, but ample flight or test data for similar systems in similar environments are available.

c. Influence 4 - Significant. M&S results are a significant factor in some program/project decisions, but not the sole factor for any program/project decisions. Ample flight or test data for similar systems in similar environments are available.

d. Influence 5 - Controlling. M&S results are the controlling factor in some program/project decisions. Neither flight nor test data are available for essential aspects of the system and/or the environment.

A.3 M&S Risk Assessment Matrix and Scope

Those M&S that are judged to fall within the red (R) boxes in figure 1 are within the Scope of this document, and those that fall within the green (G) boxes are not in Scope. The M&S that are judged to fall within the yellow (Y) boxes may be deemed in Scope at the discretion of the program/project management in collaboration with the Technical Authority.

M&S Results Influence	**5: Controlling**	(G)	(Y)	(R)	(R)
	4: Significant	(G)	(Y)	(R)	(R)
	3: Moderate	(G)	(Y)	(Y)	(R)
	2: Minor	(G)	(G)	(Y)	(Y)
	1: Negligible	(G)	(G)	(G)	(G)
		IV: Negligible	**III: Marginal**	**II: Critical**	**I: Catastrophic**
		Decision Consequence			

Figure 1—Sample M&S Risk Assessment Matrix

APPENDIX B
CREDIBILITY ASSESSMENT SCALE

B.1 Introduction

The credibility assessment of the M&S results should be assessed using the CAS and the processes described in this appendix, which contains explanatory and guidance text. The assessment process involves evaluating the M&S results on each of eight factors, and then rolling up these eight factor results into a single number that represents the summary credibility assessment. The detailed explanation of the eight factors is given in sections B.2 and B.3; the roll-up process is given in section B.4; and the process for comparing the assessed scores with the program/project-defined sufficiency thresholds (see the recommendation at the end of section 4.7) is given in section B.5.

B.2 Overview of Credibility Factors, Subfactors, and Categories

This CAS consists of eight factors grouped into three categories, as illustrated in figure 2. The eight factors are Verification, Validation, Input Pedigree, Results Uncertainty, Results Robustness, Use History, M&S Management, and People Qualifications. The three categories are M&S Development (Verification, Validation); M&S Operations (Input Pedigree, Results Uncertainty, Results Robustness); and Supporting Evidence (Use History, M&S Management, People Qualifications). A five-level assessment of credibility is defined for each factor.

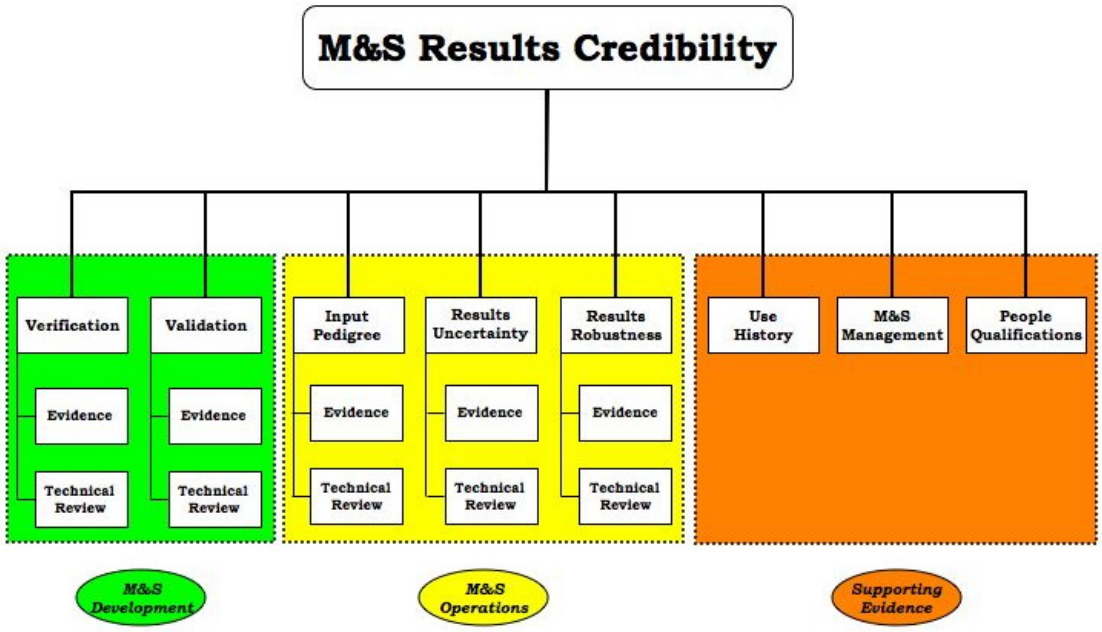

Figure 2—Credibility Assessment Scale

These eight factors were selected from a long list of factors that contribute to the credibility of M&S results because (a) individually they were judged to be the key factors in this list; (b) collectively they are nearly orthogonal, i.e., independent, factors; and (c) they can be assessed objectively. In short, the key aspects assessed by these eight factors are as follows:

a. M&S Development

(1) ***Verification****: Were the models implemented correctly, and what was the numerical error/uncertainty?*

(2) ***Validation****: Did the M&S results compare favorably to the referent data, and how close is the referent to the real-world system?*

b. M&S Operations

(1) ***Input Pedigree****: How confident are we of the current input data?*

(2) ***Results Uncertainty****: What is the uncertainty in the current M&S results?*

(3) ***Results Robustness****: How thoroughly are the sensitivities of the current M&S results known?*

c. Supporting Evidence

(1) ***Use History****: Have the current M&S been used successfully before?*

(2) ***M&S Management****: How well managed were the M&S processes?*

(3) ***People Qualifications****: How qualified were the personnel?*

The M&S Development category captures those aspects of the M&S that pertain to the general assessment of the credibility of the M&S for their broad intended use; the M&S Operations addresses the aspects relevant to the current application of the M&S to generate the particular M&S results under assessment; and the Supporting Evidence category addresses three cross-cutting factors.

Consider the case of M&S that are implemented using a general-purpose software package, e.g., a COTS structural analysis package, in the implementation of the computational model. All the verification and validation activities for the COTS tool are evaluated under M&S Development. The particular computational model that is assembled using the COTS tool is also evaluated in this category. However, the M&S Operations category deals with the credibility factors for the application of this particular computational model in the generation of the current M&S results. This includes the conduct of the present simulation and the analysis and reporting of the results. The use history of both the COTS tool in general and the particular computational model; the overall management of the M&S processes;

and the qualifications of the people involved in the development, operation, and analysis of the computational model, are covered in the Supporting Evidence category.

Table 1— Key Aspects of Credibility Assessment Levels
(Factors with a Technical Review subfactor are underlined)

Level	Verification	Validation	Input Pedigree	Results Uncertainty	Results Robustness	Use History	M&S Management	People Qualifications
4	Numerical errors small for all important features.	Results agree with real-world data.	Input data agree with real-world data.	Non-deterministic & numerical analysis.	Sensitivity known for most parameters; key sensitivities identified.	De facto standard.	Continual process improvement.	Extensive experience in and use of recommended practices for this particular M&S.
3	Formal numerical error estimation.	Results agree with experimental data for problems of interest.	Input data agree with experimental data for problems of interest.	Non-deterministic analysis.	Sensitivity known for many parameters.	Previous predictions were later validated by mission data.	Predictable process.	Advanced degree or extensive M&S experience, and recommended practice knowledge.
2	Unit and regression testing of key features.	Results agree with experimental data or other M&S on unit problems.	Input data traceable to formal documentation.	Deterministic analysis or expert opinion.	Sensitivity known for a few parameters.	Used before for critical decisions.	Established process.	Formal M&S training and experience, and recommended practice training.
1	Conceptual and mathematical models verified.	Conceptual and mathematical models agree with simple referents.	Input data traceable to informal documentation.	Qualitative estimates.	Qualitative estimates.	Passes simple tests.	Managed process.	Engineering or science degree.
0	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.
	M&S Development		**M&S Operations**			**Supporting Evidence**		

Table 1 gives a high-level summary of the evaluation criteria. These are explained in greater detail in section B.3. Table 1 by itself is not to be used in performing credibility assessments. Rather, the detailed level definitions in section B.3 are to be used.

The phrase *insufficient evidence* is uniformly used for all factors to characterize level 0. It means either that no evidence exists for that factor, or that the evidence that does exist does not meet even the level 1 criteria for that factor.

The word *favorable* as used in the level definitions for several subfactors or factors (Verification Evidence, Validation Evidence, Input Pedigree Evidence, and Use History) means that whatever relevant acceptance criteria have been deemed sufficient by the program/project in collaboration with the Technical Authority (see Req. 4.1.3 (a)) have been satisfied. If there is no

documentation for what constitutes favorable comparison, then those level definitions containing this phrase are deemed "not satisfied."

Other key phrases here, and in subsequent level definitions, are *real-world system*, *problems of interest,* and *unit problems*. The phrase *real-world system* refers to the real system operating in its real environment. A *problem of interest* refers to systems that are so close to the real-world system in its real environment that they capture most of the essential complexity of the real system and its environment (relevant to the current M&S application), and yet fall short of the real system in its real environment. This could be the real system in a similar environment, or a similar system in the real environment. The phrase *unit problem* refers to problems that capture one or more physical phenomena relevant to the current M&S application. (Some disciplines use the phrase "building block" for what is referred to here as a unit problem.) These terms are illustrated in the following two examples.

The first example is that of the Ares-1 launch vehicle. In this case the real-world system consists of the actual Ares-1 launch vehicle during an actual flight. The experimental data would be flight data taken at the relevant locations along the vehicle trajectory. The data would come either from sensors on the vehicle itself or from remote sensors capable of recording data from the flight.

A variety of problems of interest is conceivable, provided, of course, that the essential complexity is captured. One might be the flight of a similar launch vehicle, e.g., a Delta IV Medium vehicle, for basic performance data over the trajectory. Another might be wind-tunnel data (force, moment, and/or loads), when the essential geometric features and flight parameters (e.g., Reynolds number, Mach number, and angle of attack) of the Ares-1 were matched at some points along the trajectory. Yet another problem of interest might be the Solid Rocket Booster of the Shuttle for engine performance for the Ares-1 first stage.

For this example, some possible unit problems might be wind-tunnel experiments in which essential geometric features were missing and/or the Reynolds number was much lower than in flight, test stand experiments for a generic liquid rocket engine, or static and dynamic structural testing of an interface joint to assess running loads through a joint gap on Ares-1.

The second example is that of the James Webb Space Telescope. A NASA Structural Analysis (NASTRAN) model of the Primary Mirror Backplane Assembly (PMBA) is required in order to verify thermal-elastic stability requirements (changes in size, shape, and relative position/orientation) for critical optical components in response to on-orbit temperature variations. Successful execution of the project's model validation plan builds confidence in the as-built PMBA model, ultimately satisfying the criteria for Validation level 3.

A number of experiments are performed to validate models of unit problems (or building blocks)—small piece parts of low complexity but with critical features—of the cryogenic composite structure. The results of this validation activity provide confidence in the choices for basic NASTRAN elements and the representation of material properties as functions of temperature and orientation (anisotropy). These activities qualify as archetypical unit problems and as such Validation level 2 is achieved.

These experiments are followed by one designed to validate a model of an assembly of such building blocks, representing a significant fraction of the full flight structure but still lacking several key features, principally the hinges and latches that allow the backplane to fold up in the launch vehicle and then deploy once in space. In fact, a test of a more complex assembly actually precedes the unit-problem experiments, one that features hinges and latches but of a different design than what eventually is flown. That test article will also have different materials and geometry than what is flown. Both of these validation activities demonstrate the ability to aggregate large numbers of the appropriate NASTRAN elements, to capture complex and representative geometry, to capture spatial variation in material properties over large scales, and to model critical features not present in the unit problems. These experiments represented a gray area. The referent data is taken in a similar environment, provided by a small thermal-vacuum chamber. However, the degree of similarity of the test articles to the flight hardware (the real-world system) is subjective. An argument can be made that level 3 was achieved as a result of the combination of these two activities, but probably not either one on its own.

Eventually, the flight hardware is built, modeled, and tested in a large thermal-vacuum chamber. One of the test procedures is designed to validate the thermal-elastic stability model of the full, as-built backplane in a representative—but not flight—environment. The test environment cannot be made as thermally stable as space (at the Earth-Sun L2 Lagrange point); and furthermore, it will be necessary to "overdrive" the structure to get a response that could even be measured with the available metrology. Linearity is assumed down to the expected level of flight temperature variations. So while the referent data comes from the real-world system, it does not come from the real environment and, therefore, will only satisfy the criteria for Validation level 3 (assuming, of course, that favorable agreement is obtained).

Level 4 sets a high bar: there is minimal chance of model implementation errors, numerical errors are insignificant, validation has been accomplished against the real system in its real environment, input data are validated against data from the real system in its real environment, uncertainty estimates are nondeterministic and based upon real data, sensitivities to the most sensitive variables are known, the M&S has a strong track record, the management processes are strict, and the personnel are highly qualified.

Achieving a level 4 rating on the scale may be technically feasible, albeit difficult. Obviously, level 4 can only be achieved across the board for a system that is in the operations phase of the life-cycle. Lower levels on the scale are more appropriate targets for earlier phases of the life-cycle. Level 3 can only be achieved if several key factors (Validation, Input Pedigree, Results Uncertainty) are based on experimental data. Level 2 gives credit to comparison with referent data from expert opinion or other M&S. Level 1 is the minimal evidence that must be available for any credit to be given.

All of the factors in the M&S Development and M&S Operations categories have a Technical Review component, or subfactor, in addition to the regular component. However, although a Technical Review is a key contributor to credibility, as technical review is used in NASA, it is

treated better here as a component of other factors than as a standalone factor. Each of the factors that are included in the M&S Operations and M&S Development categories covers a broad range of aspects. In order to keep the CAS relatively simple, only the one or two most important aspects of each factor are singled out for evaluation in the first subfactor—the "Evidence" subfactor. The Technical Review subfactor should cover as many of these other aspects as feasible.

The components of CAS are arranged in a tiered hierarchy. The overall assessment is tier 1, the factors are tier 2, and the subfactors are tier 3. Since reporting of the assessment for the categories is not required, they are not considered a tier unto themselves, but can clearly be treated as such, if credibility assessments for the individual categories are desired; hence, categories are designated as tier 1.5.

Each subfactor (or factor when there are no subfactors) is divided into integer levels ranging from 0 to 4, with level definitions describing the evidence necessary for achieving that particular level. Level 0 corresponds to no or insufficient evidence; i.e., the M&S produce results, but there is insufficient evidence to warrant even the achievement of level 1 on that subfactor. The level definitions are summarized in table 1, and more detailed explanation is provided in the following subsections.

One might think that the overriding contributor to the credibility of the M&S results is their "accuracy." In the present context, accuracy is the difference between the M&S outputs of interest and their true value (or their assumed true value). However, the true value is virtually never known. In the M&S context, one can determine the accuracy vis-à-vis some referent value. The types of referents of interest to M&S are (a) a highly accurate numerical result (the assumed true value for assessing numerical error), and (b) a validation referent. One rarely is able to use as validation referent the true value or the assumed true value, i.e., for the real system in its real environment (the intended use of the M&S), because one does not have this data (otherwise, one would just use it in lieu of the M&S results). The CAS does assess accuracy in several places—with respect to a highly accurate numerical result in Verification and Results Uncertainty, and with respect to the assumed true value from a referent in Validation.

Two other scales (termed maturity matrices by their authors), both of which have related but somewhat different goals from the CAS, are those by Harmon & Youngblood (2005) and by Oberkampf and others (2007).

B.3 Level Definitions

In this subsection, a detailed explanation is provided for the various level definitions in the CAS. In many of the level definitions, multiple conditions are stated. These are to be treated as "and" conditions, unless, of course a condition is explicitly preceded by "or." In order to qualify for a rating at a given level, all of the "and" conditions must be satisfied. There is no partial credit given—the score derived from the level definitions must be an integer between 0 and 4. (The

weighted-average scores resulting from the roll-up described in section B.4 are not constrained to be integers.)

B.3.1 M&S Development Category

The focus of the two factors in the M&S Development category is an assessment of verification and validation of the M&S. Each of these factors has a Technical Review subfactor. The overall score for each factor is a weighted sum of the two subfactors per the roll-up process defined in section B.4. The level definitions for the Evidence subfactors are given in table 2. (The Technical Review subfactors are covered in section B.3.4.)

Table 2—Level Definitions for Evidence Subfactors in the M&S Development Category

Level	Verification Evidence	Validation Evidence
4	Reliable error estimation methods are used to quantitatively assess numerical errors. These estimates show that the errors are small from test suites, which exercise all important algorithms, all important features and capabilities, and all important couplings (physics, modules, etc.) of the full computational model.	M&S results compare favorably for the real-world system at validation points by comparison of M&S results to an acceptable referent, which is measurements on the real-world system.
3	Some formal method is used to assess numerical errors associated with unit testing with significant coverage of the code.	M&S results compare favorably for problems of interest at validation points by comparison of M&S results to an acceptable referent, which is experimental measurements on problems of interest.
2	Favorable results from unit and regression testing of key features of the computational model.	M&S results compare favorably for unit problems at validation points by comparison of M&S results to an acceptable referent, which is either experimental measurements or higher-fidelity M&S results.
1	Favorable evidence of verification for conceptual and mathematical models.	M&S conceptual and mathematical models compare favorably with “general problem” and “textbook” referents.
0	Insufficient evidence.	Insufficient evidence.

B.3.1.1 Verification Evidence Subfactor

Verification is the process of determining that a computational model accurately represents the underlying mathematical model and its solution from the perspective of the intended uses of the M&S. At its most elementary levels this involves assurance that the conceptual and mathematical models are correct.

Beyond that, there are two different aspects of the verification of the computational model: (a) is it coded correctly (code verification) and (b) are the numerical errors small (calculation verification), i.e., what is the numerical accuracy? In the case of mathematical models based upon differential equations (ordinary or partial), a detailed discussion of the distinction between these two aspects is provided in *ASME V&V 10* and Oberkampf and others, 2007.

Key aspects of code verification are regression testing and unit testing. The former is the process of testing changes to computational models to make sure that the older aspects still work with the new changes. The latter is a procedure used to validate that individual units of computational models are working properly. (In this context, a unit is the smallest testable part of a computational model.)

At least for mathematical models based upon differential equations, the computational model inevitably suffers from some level of numerical error (see Req. 4.4.2 for some examples). Levels 3 and 4 require an assessment of the size of the numerical error. Some examples of formal methods for assessing numerical error are heuristic truncation error estimates using Taylor expansions (in the case of finite-difference approximations) and rigorous error bounds (in the case of finite-element approximations). Level 3 requires demonstrating that the estimated numerical error is acceptably small for unit problems. Level 4 requires more comprehensive demonstration that the numerical error is acceptably small.

For the Verification Evidence subfactor, an assessment of level n ($n \geq 2$) requires that all conditions for levels 1, …, n-1 be satisfied in addition to the conditions at level n, i.e., a level 3 rating requires that the conditions at levels 1 and 2 also be satisfied.

B.3.1.2 Validation Evidence Subfactor

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. It involves both the accuracy of the results—the magnitude of the numerical difference between the mean of the M&S result and the mean of the referent data—and the associated uncertainty—the spread about the means. Both of these are assessed at the validation

point, i.e., the points at which the comparison is made. A favorable comparison between the M&S results and the referent data requires at a minimum that there be some overlap between the uncertainty intervals around the means.

An auxiliary aspect of validation involves the sensitivities of both the referent and the M&S results with respect to changes from the nominal input variables and parameters. While it is certainly recommended that the sensitivity of the M&S results be similar to that of the referent, experimental sensitivity data are rarely available. Accordingly, this aspect is not assessed by the Validation Evidence subfactor.

The "favorable comparision with the real-world system" phrase in level 4 includes ensuring that the data for the real-world system corresponds adequately to the present state (including any aging or changing) of the system, i.e., the last part of item (e) of Diaz Action #4 (see p. 7) must be satisfied.

For the Validation Evidence subfactor, an assessment of level n ($n \geq 2$), requires that all conditions for levels 1, …, n-1 be satisfied in addition to the conditions at level n, i.e., a level 3 rating requires that the conditions at levels 1 and 2 also be satisfied.

B.3.2 M&S Operations Category

The focus of the three factors in the M&S Operations category is an assessment of those M&S results that support the particular critical decision in question. Each of these factors has a Technical Review subfactor. The overall score for each factor is a weighted sum of the two subfactors per the roll-up process defined in section B.4. The level definitions for the Evidence subfactors are given in table 3.

Table 3—Level Definitions for Evidence Subfactors in the M&S Operations Category

Level	Input Pedigree Evidence	Results Uncertainty Evidence	Results Robustness Evidence
4	The input data compare favorably with measured data from the real-world system, or the input data came from M&S with a summary credibility rating above 3.5. Uncertainty associated with the input data is known.	Uncertainty estimates are quantitative and based upon nondeterministic and numerical analysis.	Sensitivity of the M&S results for the real-world system is quantitatively known for most of the variables and parameters, including all of the most sensitive variables and parameters.
3	The input data compare favorably with acceptable measured referent data from problems of interest, or the input data came from M&S with a summary credibility rating above 3.0. Uncertainty associated with the input data is known.	Uncertainty estimates are quantitative and based upon nondeterministic analysis.	Sensitivity of the M&S results for the real-world system is quantitatively known for many variables and parameters.
2	The input data is traceable to formal documentation, or the input data came from M&S with a summary credibility rating above 2.0.	Uncertainty estimates are quantitative and based upon deterministic analysis or expert opinion.	Sensitivity of the M&S results for the real-world system is quantitatively known for a few variables and parameters.
1	The input data is traceable to informal documentation, or the input data came from M&S with a summary credibility rating above 1.0.	Uncertainty estimates are qualitative.	Sensitivity of M&S results for the real-world system is estimated by analogy with the quantified sensitivity of similar problems of interest.
0	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.

B.3.2.1 Input Pedigree Evidence Subfactor

Input Pedigree involves the evaluation of all data that is used as input for the current M&S results. It includes not only data that is unique to the model, but also data that is produced by other simulations. In the former case, the input data are assessed directly, whereas in the latter case the input data are assessed indirectly by the credibility assessment score(s) of the M&S that produced the data. Note that achievement of levels 3 and 4 requires that the uncertainty associated with the input data is known.

The direct assessment of the input data at levels 1 and 2 requires traceability to documentation of the input data. Informal documentation is merely written documentation that has not been independently reviewed. Formal documentation requires review by a third-party, and should be either a NASA STI-series document or a program/project-specific document.

B.3.2.2 Results Uncertainty Evidence Subfactor

Results Uncertainty is the quantification of the uncertainty in the current M&S results. Two important aspects of the uncertainty are (a) the size of the uncertainty, e.g., the size of the uncertainty interval; and (b) the confidence in or quality of the estimate of the uncertainty, e.g., a statistical confidence statement or the thoroughness used in the estimate. Depending upon the application, only one of these or both of these may be important. The responsible party will choose whether the uncertainty criteria consists of (a) only, (b) only, or both (a) and (b). The uncertainty in the M&S results is affected by the uncertainty in input variables and parameters, as well as by the numerical errors.

At level 1, the uncertainty is described only qualitatively, i.e., in linguistic terms as discussed in section 4.8. At level 2, the uncertainty is described in numeric terms, i.e., by statement of uncertainty intervals about the M&S results. The uncertainty interval may be based upon the opinion of a subject matter expert and/or by the quantified uncertainty at the most relevant validation point. At level 3, the uncertainty statement is numerical and makes use of nondeterministic analysis (including conventional probabilistic analysis as well as approaches such as interval analysis, evidence theory, fuzzy logic, and possibility theory); this level could be achieved by formal application of nondeterministic analysis to a set of Subject Matter Expert opinions (see Cooke (1991)); at this level, the uncertainty would most likely be stated without a measure of the confidence associated with the uncertainty interval. At level 4, the uncertainty statement makes use of nondeterministic analysis, but would also include a level of confidence associated with the uncertainty interval; that is, some numerical estimation of the uncertainty in the numerical uncertainty interval has been made.

B.3.2.3 Results Robustness Evidence Subfactor

Results Robustness is the determination of how thoroughly the sensitivities of the current M&S results (to the variables and parameters of the M&S) are known. The purpose of considering robustness is to garner an understanding of the sensitivity of the real-world system to potential changes in the variables and parameters of the system. Therefore, understanding how well the sensitivity of the M&S matches that of the real-world system is of particular interest. Simulations aim to imitate the real world or a proposed real world through execution of a computational model. Ideally, the imitated system behaves like the real-world system. That is, if the real-world system is sensitive to certain variables or parameters, then the M&S results should be similarly sensitive. The level of agreement between the real-world system and the M&S results falls within the domain of the Validation factor and not the Robustness factor.

What constitutes “few,” “many,” and “most” in levels 2, 3, and 4 cannot be specified precisely. As a guideline, “few” should mean that the sensitivity of, say, less than 20 percent of the potential variables and parameters is known; “many” should mean that the sensitivity of, say, between 20 and 50 percent is known; at level 4, “most” implies the majority (i.e., >50 percent) of

all parameters and variables is known including all of the most sensitive variables and parameters.

B.3.3 Supporting Evidence Category

The focus of the three factors in the Supporting Evidence Category is the assessment of three elements of the M&S process that may indirectly affect the credibility of the M&S results. The factors included are Use History, M&S Management, and People Qualifications. The Use History factor describes the extent of any prior use of the M&S in similar situations for critical decisions. The M&S Management factor assesses the level of formality applied by the program or project to the oversight of the M&S. The People Qualifications factor assesses the training and experience of the developers, operators, and analysts conducting the M&S activities. The level definitions for these three factors are given in table 4.

Table 4—Level Definitions for Factors in the Supporting Evidence Category

Level	Use History	M&S Management	People Qualifications
4	De facto standard.	Continuing Process Improvement: The M&S effort is using measurements on M&S processes to improve the repeatability of the M&S results.	Possesses an advanced engineering or science degree or extensive work experience in M&S, has extensive experience with the development and use of the M&S being reviewed, and has employed specific recommended practices relevant to current application.
3	Post-decision real-world events have been accurately represented in results (e.g., validated by mission data).	Predictable Process: The M&S effort is measuring repeatability of the M&S results generated by the M&S processes.	Possesses an advanced engineering or science degree or extensive work experience, has general M&S training, has specific experience with the M&S being reviewed, and has been trained on specific recommended practices relevant to the current application.
2	Used previously to perform analysis upon which critical decisions have been made.	Established Process: The M&S effort has established a documented process for M&S development and operations.	Possesses an engineering or science degree, has received formal training in formulation of M&S and generic training in recommended practices for M&S, and has developed M&S products.
1	Specific scenarios have been created to test application, or results compare favorably with outputs from other similar tools.	Managed Process: The M&S roles and responsibilities have been defined.	Possesses an engineering or science degree, has been introduced to the topic of M&S, and has been exposed to generic recommended practices in M&S.
0	Insufficient evidence.	Insufficient evidence.	Insufficient evidence.

B.3.3.1 Use History Factor

Use History assesses how extensively the M&S in the current application have been used in the same or similar applications in the past. This assesses the "heritage" of the M&S. The non-trivial levels range from favorable comparisons with other M&S (level 1) to M&S that have been used previously for critical decisions with post-flight confirmation and have become de facto standards (level 4).

For the Use History factor, a level 3 rating requires that the conditions at level 2 also be satisfied. In order to qualify at level 4 not only must the conditions at level 2 and level 3 be satisfied, but also the M&S must be widely used for the particular application.

B.3.3.2 M&S Management Factor

The term M&S Management is used to describe the extent to which an M&S effort exhibits the characteristics of work product management; process definition; process measurement; process control; process change; continuous improvement, including CM; and M&S support and maintenance. The levels are similar to those for most process maturity models. This factor assesses how rigorously item (e) of Diaz Action #4 (see p. 7) is addressed. Assessments at level 1 and higher require evidence addressing each of the topics in item (e) of Diaz Action #4.

For the M&S Management factor, an assessment of level n ($n \geq 2$), requires that all conditions for levels 1, …, n-1 be satisfied in addition to the conditions at level n, i.e., a level 3 rating requires that the conditions at levels 1 and 2 also be satisfied.

B.3.3.3 People Qualifications Factor

People Qualifications refers to the qualifications of the staff to use and/or interpret the results of M&S. Personnel qualifications are assessed in terms of general education, general experience, specific training for the subject M&S, and specific experience with the subject M&S. Education and experience include recommended practices in general and for the subject M&S.

If no recommended practices have been identified (see Req. 4.5.1), then the conditions pertaining to recommended practices in the level definitions do not apply; i.e., they have no influence on whether or not a particular level has been achieved. The adjective "extensive" in the definition for level 4 means that the individual (or team lead) has sufficient experience to mentor newcomers on the subject M&S without further technical oversight, and for level 3 it means that the individual (or team lead) has sufficient experience to mentor newcomers on the subject discipline (not necessarily the M&S) without further technical oversight. The term "formal training" means instructor-led classroom training of at least the depth of a semester-long university course at the advanced undergraduate or graduate level.

B.3.4 Technical Review Subfactors

Five of the eight factors have a Technical Review subfactor, which assesses the level of peer review that has been successfully completed relevant to that factor. By a "peer review" we mean an assessment that is conducted by one or more persons of equal technical standing to person(s) responsible for the work being reviewed. An "informal peer review" is one that is not conducted pursuant to a process established by the reviewed or reviewing organization, whereas a "formal peer review" is one that is sanctioned by the program/project and conducted in accordance with rules explicitly established by the reviewed or reviewing organization. Peer reviews are classified as "internal" or "external" depending upon whether or not the panel members are drawn primarily from within the lead Center for the project.

This subfactor appears in each of the following factors: Verification, Validation, Input Pedigree, Results Uncertainty, and Results Robustness. The reason for this repetition is that numerous reviews are generally constituted throughout the life of a program or project, with each review focusing on particular topics. It would not be uncommon for different subject matter experts to be involved in the various reviews, depending upon their availability at a particular time and their particular areas of expertise. Likewise, the same people may serve as reviewers in different capacities throughout the life of a program or project, again depending on their availability and expertise.

The level definitions for all the Technical Review subfactors are provided in table 5. These are the same at levels 0 to 3. At level 4, the italicized phrase should be replaced by the appropriate factor name. For example, level 4 for Verification Technical Review reads "Favorable external peer review accompanied by independent verification." In general, at level 4, the review would also include independent reproduction of the relevant findings by the external reviewers or their agents.

Table 5—Level Definitions for the Technical Review Subfactors

Level	Technical Review
4	Favorable external peer review accompanied by independent *factor evaluation.*
3	Favorable external peer review.
2	Favorable formal internal peer review.
1	Favorable informal internal peer review.
0	Insufficient evidence.

While the same language is used to describe the various levels within the Technical Review subfactor, it is assumed that the reviews for each factor focus solely on the material relevant to that factor. This is not to say, however, that reviews of the topics of several factors could not be considered within the same meeting of a panel of reviewers.

B.4 Credibility Assessment Roll-up Processes

B.4.1 Assignment of Weights for the Subfactors

The primary focus of the CAS is on the scores for the eight factors; and the secondary focus is on the overall score, which is the minimum of the scores for the eight factors. The five factors in the M&S Development and M&S Operations categories are weighted averages of the associated Evidence and Technical Review subfactors. Nevertheless, the emphasis is on the scores at the factor tier; the Technical Review subfactor just serves to tune the evidence subfactor by the results of internal and external assessments. The optional reporting at the category tier is covered at the end of this subsection.

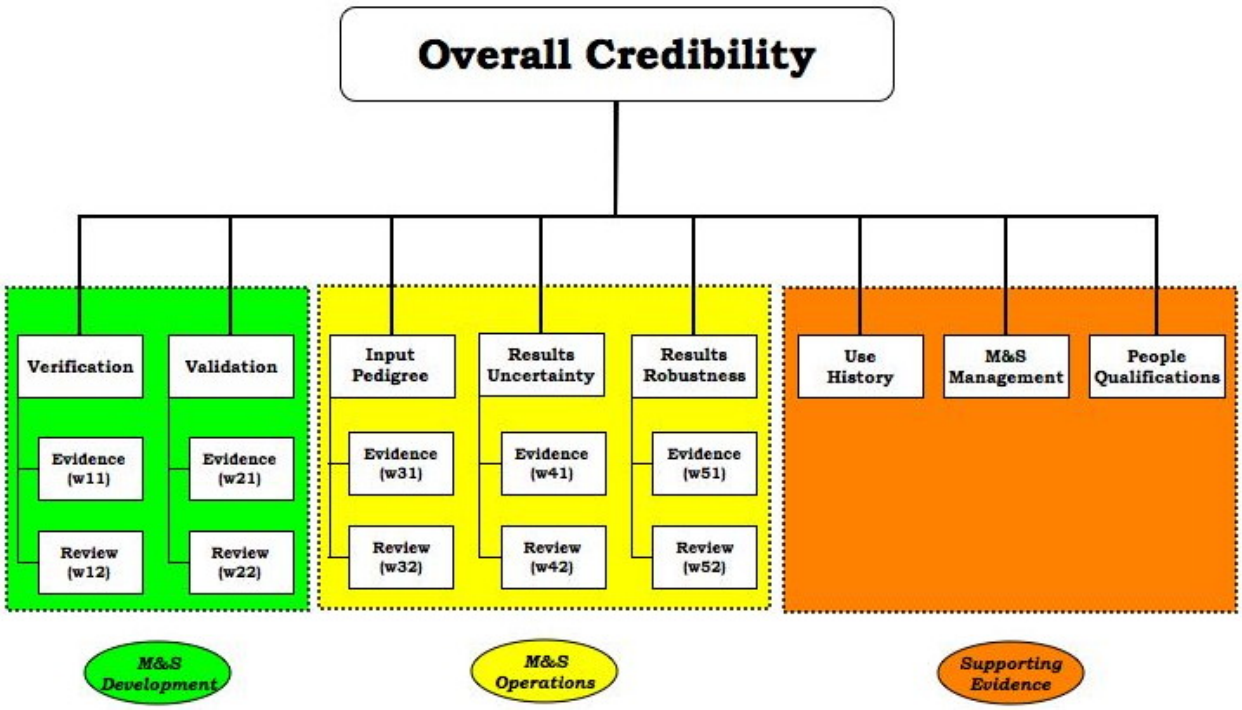

Figure 3—Subfactor Weights

Figure 3 illustrates the 10 weights that are needed for the roll-up from the subfactor to the factor tier. The constraints on these weights are as follows:

a. Each weight lies in the closed interval [0,1].
b. The sum of each subfactor pair, e.g., w11 and w12, is 1.
c. The subfactor weight for Technical Review is further constrained to be no more than 0.3.

The achieved score at the lowest tier (factor or subfactor) is based on the objective assessment of the documented evidence against the level definition. In the M&S Development and M&S Operations categories, the achieved factor score is the Evidence score times the Evidence weight plus the Review score times the Review weight. Constraint c limits the amount by which Technical Review can increase or decrease the factor score with respect to the Evidence subfactor score. In the most extreme case, with an Evidence score of 0 and Technical Review score of 4.0, the factor score is 1.2.

The choice of the weights is necessarily subjective. As specified in Req. 4.1.3, the responsible party designated by program and project management will select the various weights, subject to approval by the program/project management and the Technical Authority (see section 4.1).

The roll-up of the 8 factor scores into the overall score is performed by taking the minimum of the 8 factor scores.

B.4.2 Roll-up Examples

An illustration of a subfactor-to-factor roll-up is given in table 6. In this case the primary subfactor score of 3 is slightly upgraded to 3.3 because of a strong Technical Review.

Table 6—Roll-up of Subfactor Scores to Factor Score

Subfactor	Subfactor Weight	Assessed Score	Weighted Score	Factor Score
Validation Evidence	0.7	3	2.1	**3.3**
Validation Technical Review	0.3	4	1.2	

An example of the roll-up from the factor tier to the overall score is provided in table 7. The overall score is the minimum of the factor scores. The scores are only reported to two significant digits, as carrying more than one digit past the decimal point would convey a false sense of precision.

Table 7— Roll-up of Factor Scores to Overall Score

Factor	Factor Score	Overall Score
Verification	3	**1.7**
Validation	3.3	
Input Pedigree	3.3	
Results Uncertainty	3	
Results Robustness	1.7	
Use History	4	
M&S Management	3	
People Qualifications	3	

Per Req. 4.7.1, Req. 4.7.3, and Req. 4.8.3, reporting of M&S results will be accompanied by reports of the eight factor scores and the single, overall score. Possible reporting formats for the factor scores are bar charts and radar plots, illustrated in figure 4.

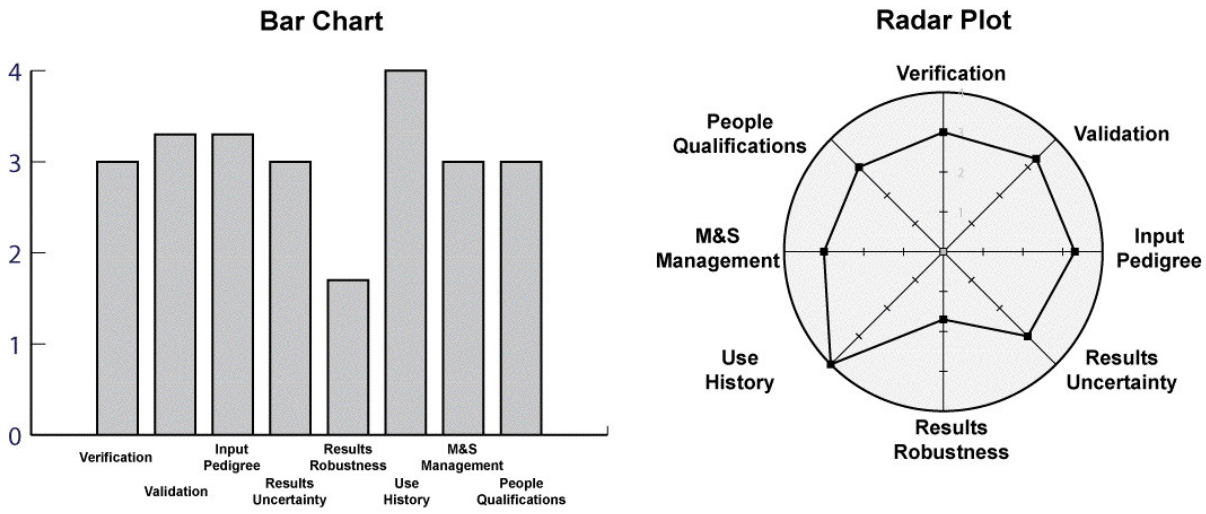

Figure 4—Bar Chart and Radar Plot for Factor Scores

Reporting of the scores at the category tier is optional. If this is performed, then it is performed by taking the minimum of all factor scores in the category. See table 8 for an illustration of the roll-up to the M&S Operations category, corresponding to the factor scores of the example in table 7.

Table 8—Roll-up of Factor Scores to Category Score

Factor	Factor Score	Category Score
Input Pedigree	3.3	**1.7**
Results Uncertainty	3	
Results Robustness	1.7	

B.5 Comparison with Sufficiency Thresholds

B.5.1 Introduction

Sections B.2 and B.3 introduced and described the CAS—its overall, category, factor, and subfactor tiers, along with the associated lowest-tier level definitions against which M&S results are assessed. Section B.4 described the methodology for rolling-up assessed scores from one tier to the next-higher tier. This section introduces and describes sufficiency or adequacy thresholds (aka "good-enough" bars), assigned by the responsible party, against which the assessed scores are compared for the intended purpose. The threshold concept is further augmented by using corresponding "deficiency/exceedance flags" to provide management-level highlights of any shortfalls/exceedances, compared to the threshold values, of an M&S credibility factor measurement. This framework provides simple "dashboard indicators" for M&S results credibility for top management, yet total traceability to the lower-tier credibility assessments necessary for providing technical feedback to developers and users.

B.5.2 Determination of Adequacy for Each Factor

The responsible party sets the desired thresholds for each of the factors in the CAS. This is done by considering what is "good enough" for that factor for the intended purpose. The threshold values should be integers between 0 and 4, inclusive. Further, the threshold levels should (a) change over the life-cycle of the project supported by the M&S, and (b) depend upon the severity of the risk associated with the decision.

The comparison of the factor threshold level set by the responsible party and the assessed score is used to determine the adequacy of the M&S application for each credibility factor. If the assessed score for the factor is equal to the threshold, the factor condition is met; and the deficiency/exceedance flag is green, meaning that the M&S has achieved a sufficient credibility level. If the score is more than one-half unit (>0.5) less than the threshold, a deficiency flag will be set, indicating that the credibility is not adequate for the factor for the M&S application. A deficiency gap greater than one-half unit (>0.5) but less than or equal to one unit (≤1.0) will

generate a yellow flag. A deficiency gap greater than one unit (>1.0) on the scale will generate a red flag (figure 5). If the score is greater than the threshold, a blue exceedance flag will be set, indicating that the credibility is more than adequate for the factor for the M&S application (figure 6). These flags are used to show deficiency/exceedance conditions all the way to the tier 1, overall score. An illustration of these flags at the factor tier is provided in figure 7.

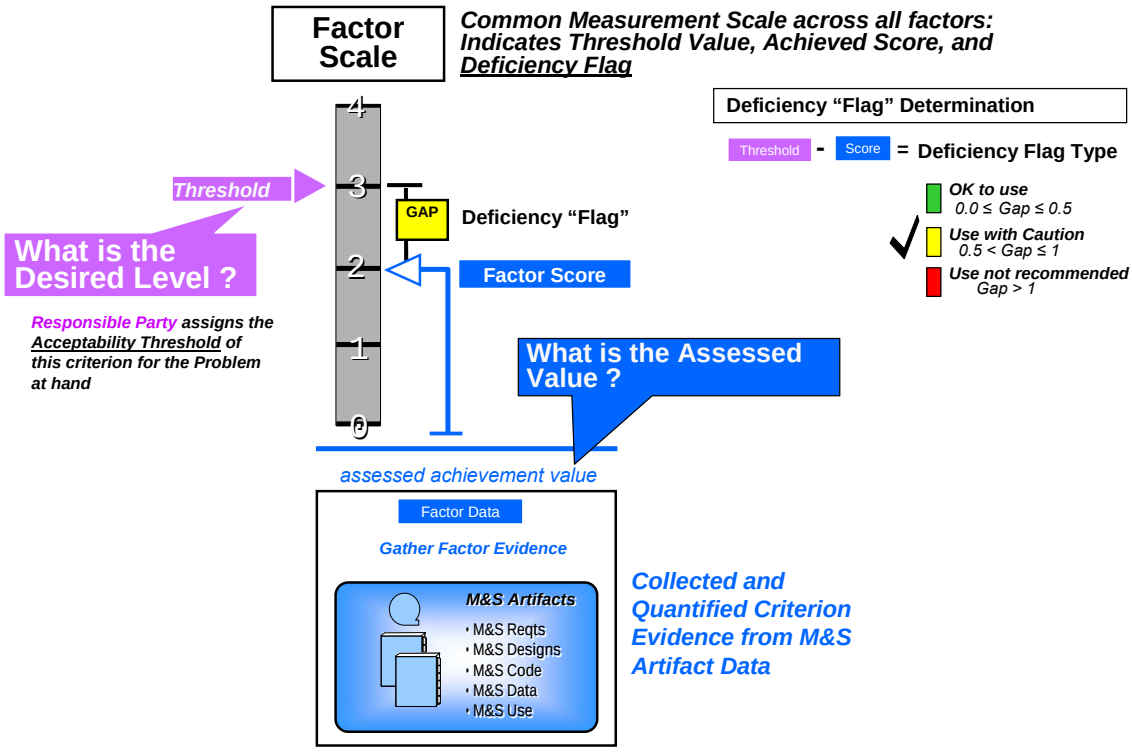

Figure 5—Deficiency Flag Illustration

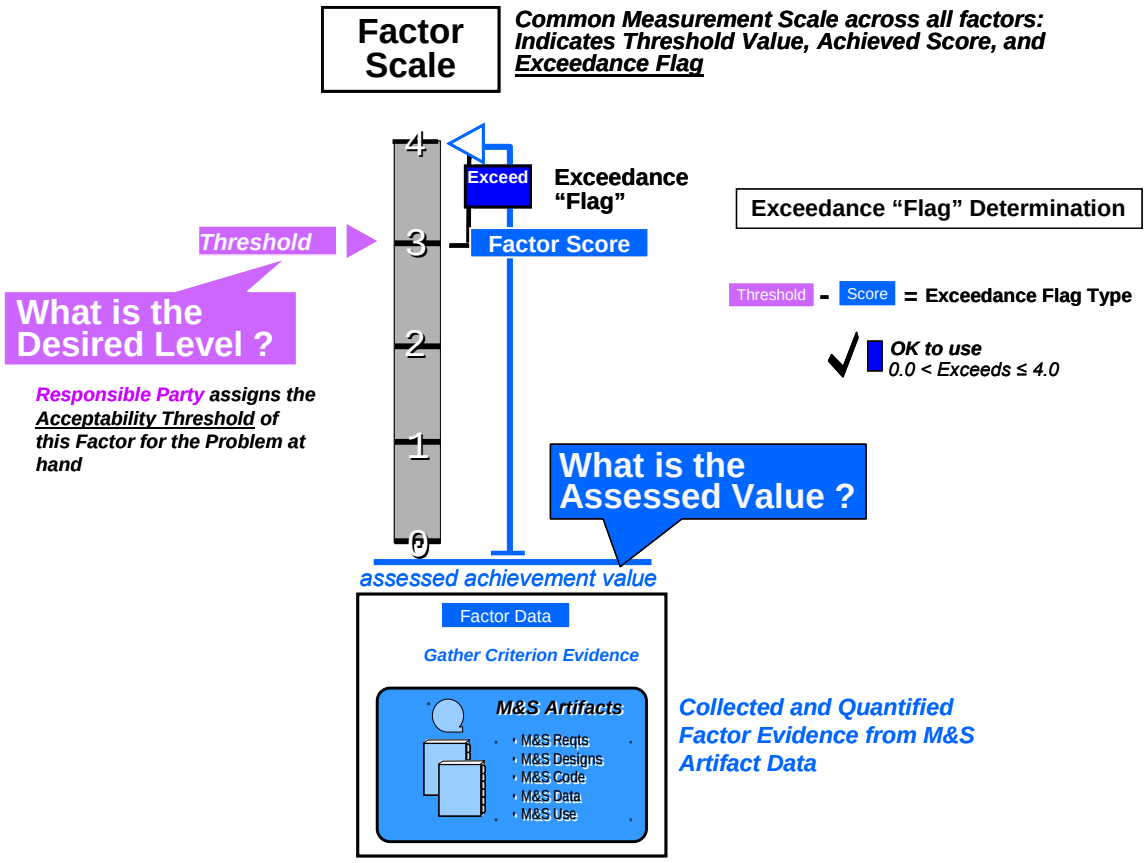

Figure 6—Exceedance Flag Illustration

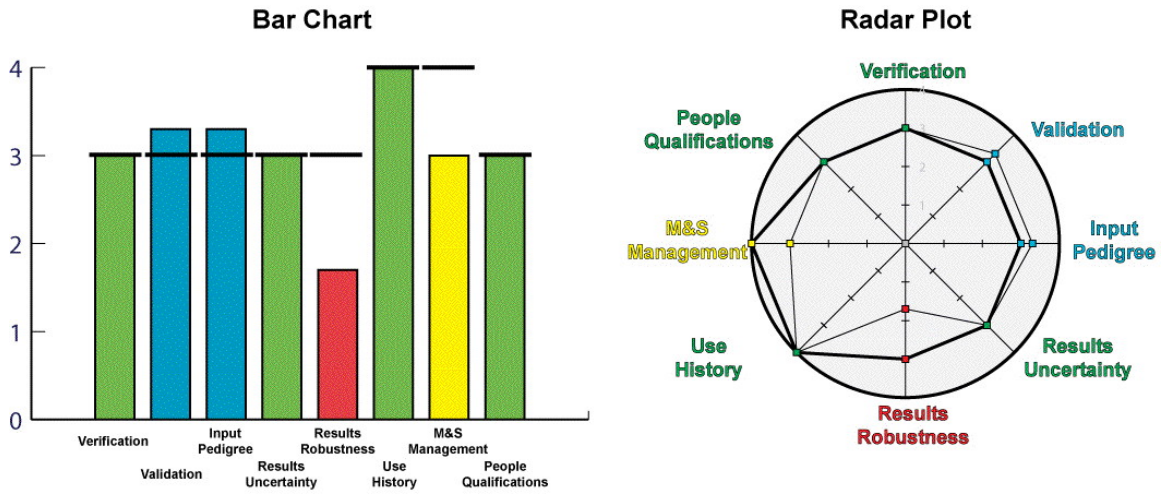

Figure 7—Sufficiency Thresholds and Color Coding on Bar Chart and Radar Plot for Factor Scores

APPENDIX C
COMPLIANCE MATRIX

The matrix below provides a template for assessing compliance with the requirements in this standard. The first column lists the requirements. The key for the second column is C–compliant, NC–not compliant, NA–not applicable. The third and fourth columns can be used to record the method of verifying compliance and the evidence of compliance, respectively.

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
Req. 4.1.1 – Shall document the risk assessment for any M&S used in critical decisions.			
Req. 4.1.2 – Shall identify and document those M&S that are in scope.			
Req. 4.1.3 – Shall define the objectives and requirements for M&S products including the following: a. The acceptance criteria for M&S products, including any endorsement for the M&S. b. The rationale for the weights used for the subfactors in the Credibility Assessment Scale (see Appendix B.4). c. Intended use. d. Metrics (programmatic and technical). e. Verification, validation, and uncertainty quantification (see section 4.4). f. Reporting of M&S information for critical decisions (see section 4.8). g. CM (artifacts, timeframe, processes) of M&S.			
Req. 4.1.4 – Shall develop a plan (including identifying the responsible organization(s)) for the acquisition,			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
development, operation, maintenance, and/or retirement of the M&S.			
Req. 4.1.5 – Shall document any technical reviews performed in the areas of Verification, Validation, Input Pedigree, Results Uncertainty, and Results Robustness (see Appendix B).			
Req. 4.1.6 – Shall document M&S waiver processes.			
Req. 4.1.7 – Shall document the extent to which an M&S effort exhibits the characteristics of work product management, process definition, process measurement, process control, process change, and continuous improvement, including CM and M&S support and maintenance.			
Req. 4.2.1 – Shall document the assumptions and abstractions underlying the conceptual model, including their rationales.			
Req. 4.2.2 – Shall document the basic structure and mathematics of the model (e.g., reality modeled, equations solved, behaviors modeled, conceptual models).			
Req. 4.2.3 – Shall document data sets and any supporting software used in model development and input preparation.			
Req. 4.2.4 – Shall document required units and vector coordinate frames (where applicable) for all input/output variables in the M&S.			
Req. 4.2.5 – Shall document the limits of operation of models.			
Req. 4.2.6 – Shall document any methods of uncertainty quantification and the			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
uncertainty in any data used to develop the model or incorporated into the model.			
Req. 4.2.7 – Shall document guidance on proper use of the model.			
Req. 4.2.8 – Shall document any parameter calibrations and the domain of calibration.			
Req. 4.2.9 – Shall document updates of models (e.g., solution adjustment, change of parameters, calibration, and test cases) and assign unique version identifier, description, and the justification for the updates.			
Req. 4.2.10 – Shall document obsolescence criteria and obsolescence date of the model.			
Req. 4.2.11 – Shall provide a feedback mechanism for users to report unusual results to model developers or maintainers.			
Req. 4.2.12 – Shall maintain (conceptual, mathematical, and computational) models and associated documentation in a controlled CM system.			
Req. 4.2.13 – Shall maintain the data sets and supporting software referenced in Req. 4.2.3 and the associated documentation in a controlled CM system.			
Req. 4.3.1 – Shall do either of the following: a. Ensure that simulations are conducted within the limits of operation of the models, or b. Placard the simulation and analysis results with a warning that the simulation may have been conducted outside the limits of operation and			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
include the type of limit that may have been exceeded, the extent that the limit might have been exceeded, and an assessment of the consequences of this action on the M&S results.			
Req. 4.3.2 – Shall document and explain any observed warning and error messages resulting from the execution of the computational model.			
Req. 4.3.3 – Shall document which computational models were used (including revision numbers) in the simulation.			
Req. 4.3.4 – Shall document the versions of M&S results.			
Req. 4.3.5 – Shall document data used as input to the simulation, including its pedigree (see Appendix B).			
Req. 4.3.6 – Shall document any unique computational requirements (e.g., support software, main memory, disk capacities, processor, and compilation options).			
Req. 4.3.7 – Shall document the processes for conducting simulations and analyses for generating results reported to decision makers.			
Req. 4.3.8 – Shall document any use history of M&S, in the same or similar applications, which are relevant for establishing the credibility of the current M&S application (see Appendix B).			
Req. 4.3.9 – Shall document the assessment as to the appropriateness of the simulation and analysis relative to its intended use.			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
Req. 4.3.10 – Shall document the rationale for the setup and execution of the simulation and analysis.			
Req. 4.4.1 – Shall document any verification techniques used and any domain of verification (e.g., the conditions under which verification was conducted).			
Req. 4.4.2 – Shall document any numerical error estimates (e.g., numerical approximations, insufficient discretization, insufficient iterative convergence, finite-precision arithmetic) for the results of the computational model.			
Req. 4.4.3 – Shall document the verification status of (conceptual, mathematical, and computational) models.			
Req. 4.4.4 – Shall document any techniques used to validate the M&S for its intended use, including the experimental design and analysis, and the domain of validation.			
Req. 4.4.5 – Shall document any validation metrics, referents, and data sets used for model validation.			
Req. 4.4.6 – Shall document any studies conducted and results of model validation.			
Req. 4.4.7 – Shall document any uncertainty quantification processes used for a. The referent data. b. The input data. c. The M&S results. d. The propagation of uncertainties. e. The quantities derived from M&S results.			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
Req. 4.4.8 – Shall document any quantified uncertainties, both physical and numerical, for a. The referent data. b. The input data. c. The M&S results. d. The propagated uncertainties. e. The quantities derived from M&S results.			
Req. 4.4.9 – Shall document the extent and results of any sensitivity analyses performed with the M&S.			
Req. 4.5.1 – Shall identify and document any Recommended Practices that apply to M&S for the program/project.			
Req. 4.6.1 – Shall determine the depth of required training for developers, operators, and analysts.			
Req. 4.6.2 – Shall document the following: a. Training topics required for developers, operators, and analysts of M&S. b. Process and criteria for verifying that training requirements are met.			
Req. 4.6.3 – Shall determine the qualifications for developers, operators, and analysts.			
Req. 4.7.1 – Shall assess the credibility of M&S results for each of the eight factors in the CAS described in Appendices B.2 and B.3.			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
Req. 4.7.2 – Shall justify and document the credibility assessment for each of the eight factors referenced in Req. 4.7.1.			
Req. 4.7.3 – Shall perform the roll-up to an overall score according to the process described in Appendix B.4.			
Req. 4.8.1 – Reports to decision makers shall include explicit warnings for any of the following occurrences, accompanied by at least a qualitative estimate of the impact of the occurrence: a. Any unachieved acceptance criteria (as specified in Req. 4.1.3 (a)). b. Violation of any assumptions of any model (as specified in Req. 4.2.1). c. Violation of the limits of operation (as specified in Req. 4.2.5). d. Execution warning and error messages (see Req. 4.3.2). e. Unfavorable outcomes from the intended use and setup/execution assessments (described in Req. 4.3.9 and Req. 4.3.10). f. Waivers to any of the requirements in this standard.			
Req. 4.8.2 – Reports to decision makers of M&S results shall include an estimate of their uncertainty and a description of any processes used to obtain this estimate as defined in Req. 4.4.7 and Req. 4.4.8. a. Reported uncertainty estimates shall include one of the following: (1) A quantitative estimate of the uncertainty in the M&S results, or			

Requirements	Compliance Status (C, NC, N/A)	Method	Evidence
(2) A qualitative estimate of the uncertainty in the M&S results, or (3) A clear statement that no quantitative or qualitative estimate of uncertainty is available.			
Req. 4.8.3 - Reports to decision makers shall include the level of credibility for the M&S results and the subfactor weights, using the process specified in section 4.7.			