

APPLYING NASA-STD-7009 STANDARD FOR MODELS AND SIMULATIONS TO SURROGATE AND OTHER STATISTICAL MODELS

K.L. Johnson

NASA Engineering and Safety Center (NESC)
Marshall Space Flight Center, AL

ABSTRACT

NASA-STD-7009A, Standard for Models and Simulations, contains a worthy and insightfully-crafted credibility assessment. That set of metrics is intended to accumulate and communicate information that can be critical for users and their customers in evaluating findings due to a model or simulation. Was the model designed for the present use? Is it validated in the domain it's being applied to? And so on.

The Standard was originally focused on physics-based models, but it can be applied more generally. It can – and should – be applied to high-impact statistical models, including surrogate models based on larger models and sims and regression models based directly on data. A key issue with many of such models is their relative simplicity compared to most M&S despite their use in high risk engineering decisions. This report will make a case for using this Standard for a wider class of models, and how it might actually add value in excess of its cost for not only stakeholders, but modelers themselves. A rough example showing compliance for a statistical model is included.

INTRODUCTION

Excellent managers depend on excellent information provided by subject matter experts (SMEs). Information provided by an excellent team of SMEs should enable a manager to make a decision without having to perform their own investigations or analyses or even be an SME in all relevant topics themselves.

This, however, does not relieve a manager of the responsibility to weigh the quality of the inputs. The information that aerospace and defense engineering managers depend on can have extremely high impact. But the time, resources and even capacity required to fully understand each input is often simply out of reach. The manager must often rely on snippets of sometimes-imperfect substantiation offered in many forms and languages.

At the same time, responsible SMEs want their inputs to incur the proper amount of respect and understanding. Not too little: no competent SME wants a valuable input to be ignored. Yet not too much: a well-founded fear is that a good but still somewhat uncertain estimate can “grow legs” and be treated as sure or that a firm recommendation becomes associated with a condition for which it is invalid.

In short, managers want to and need to know what to believe and how much. SMEs want to and need to offer inputs attached with the right strength of certainty behind them. And everyone needs to perform this transaction completely, understandably and succinctly no matter what the subject matter being communicated.

NASA-STD-7009A Standard for Models and Simulations (to be called here the Standard or 7009A) [1] was instituted in large part to address these issues for the modeling and simulation (M&S) community. Acceptance has been sure but slow. The slowness may be in part because it may seem like “just another set of requirements” that just add complexity and cost.

The Standard's authors and contributors (including this writer) understand this, but mostly succeeded in creating a document and process that achieve the Standard's goals but with little overhead. While application does take some planning and time, applied judiciously 7009A should add value in excess of burden. Particularly for high-importance, high-risk models, the structural requirements within 7009A represents an excellent framework for constructing and exercising M&S, while the requirements pertaining to documentation and communication help M&S stakeholders, including future users of the M&S, understand what it is they have in their hands. The Standard's credibility scale, in particular, is a succinct and flexible metric tool for communicating key information useful to decisionmakers and others in a standard, visual format.

Originally, this paper was going to focus on adding statistical, probabilistic, stochastic, uncertainty quantification (UQ), probabilistic risk assessment (PRA), and similar models and simulations – all to be called “statistical models” in this paper for brevity – to what 7009A should apply. The reason for this was that some of the original authors of 7009A – including this one – had the opinion that statistical models were not covered in the Standard, and that 7009A may even need to be changed to be able to cover statistical M&S. In fact, statistical models, particularly metamodels (aka surrogate models, emulators, and other terms), are specifically included in 7009A. The focus of this writing thus shifted into explaining how 7009A could be simply understood and fairly simply applied by modelers unfamiliar with it and why its application can be a net benefit to both modelers and their stakeholders.

To date, 7009A appears to have been applied to statistical models particularly rarely. The reason, gleaned from experience as an applied statistical modeler for many NASA and community P/P and explicitly argued by a surprising number of people, appears to be that statistical models are often not considered to be engineering models by managers, engineers and other stakeholders. Statistical modelers themselves often consider there to be a difference between what they call models and a computational fluid dynamicist's code or a vehicle landing simulation.

By 7009A's definition of M&S, this difference is arbitrary. From NASA-HDBK-7009A, the handbook companion to the Standard:

Model: A description or representation of a system, entity, phenomena, or process¹.

Simulation: The imitation of the behavioral characteristics of a system, entity, phenomena, or process.

These definitions describe nearly all statistical M&S. Ignoring this allows anyone holding another view to have a reason to think of statistical methods as something outside of modeling, even outside of science and engineering, which is not helpful for understanding or acceptance of statistical methods by nonstatisticians and vice versa. Misunderstanding this results in an arbitrary distinction without a difference and stands in the way of communities sharing tools, methods, efficiencies and ideas. It is not helpful.

True, many if not most statistical models are nearly or purely empirical in nature rather based on first-principles physics. Surrogate models in particular mimic but are not directly linked to physical inputs and natural behavior. But direct linkage through physics is not a requirement in 7009A. If anything, inclusion and assessment of uncertainty may connect a statistical model to

¹ It gives a further note: “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 go into a model are considered part of the model.”

the real world system (RWS) it represents with a different yet usefully closer way than a purely deterministic model can, depending on what is required of the M&S. The best choice of M&S methodology to be applied depends on the problem that needs to be solved, not a particular preconception of what a model or sim must be.

It is important that any model be rigorously constructed by knowledgeable practitioners. Given the often counterintuitive nature of statistical analysis, practitioner expertise can be even more crucial. Statistical models are often opaque to many engineers and managers who never took, let alone enjoyed, a statistics course. Yet statistical models can certainly be powerful engineering modeling and predictive tools with high utility. They are used as part of common and critical engineering decisions: PRA certainly falls in this bin. The statistical underpinnings of reliability, factors of safety, material allowables and other engineering constructs are often forgotten. Even calculation of averages of test data or estimates of instrument measurement uncertainty can be considered critical modeling efforts that can be well- or poorly-performed and based on data of varying heritage.

Models' strengths and shortcomings should be communicated in understandable terms to managers and stakeholders. They need to know the level of expertise of the modeler supplying the information relayed to them. And it is most crucial that the findings and recommendations that follow from them be able to be evaluated for credibility and compared on equal footing with other models. 7009A gives a fairly clear and common structure for key pieces of this evaluation and communication chain. Whether or not the Standard is required and to the extent that any modeler wants their model understood and appropriately accepted as credible by their stakeholders, working through 7009A can help achieve those goals.

THE PROBLEM 7009A ADDRESSES – BOTH RIGOROUSLY AND FLEXIBLY.

Problem statement: **When faced with a critical decision, stakeholders may be presented with results based on an M&S they know little about. Managers see only a top-level summary. It's often hard for them differentiate between recommendations due to a solidly grounded model reflecting the real world system (RWS) the model is intended to represent well versus results from a tentative or rough one. When this understanding is lacking, decisional risk may be increased unknowingly and unnecessarily.**

To address this set of problems, 7009A begins by outlining and encouraging foundational best practices. It “establishes *uniform* practices in M&S to ensure essential requirements are applied to their design, development and use, while ensuring acceptance criteria are defined by the P/P and approved by the responsible Technical Authority (TA), ... provides an *approved* set of requirements, recommendations, and criteria with which M&S may be developed, accepted, and used in support of NASA activities (and) provides a *basic* set of best practices applicable to any M&S.” [1]

7009A is *required* only in particular circumstances. Specifically, the P/P must include it in its list of required applicable standards. The P/P TA defines a threshold of risk or criticality over which an M&S must employ the Standard and its requirements, so only a subset of models will *require* adherence.

7009A is comprehensive, covering the life span of an M&S. “This NASA Technical Standard (establishes) a minimum set of requirements for the key elements related to M&S, including development, maintenance, operation, results analysis, training, credibility assessment, and reporting.” A life cycle-based worksheet is included with the Standard (Table 22, also executed as an Excel worksheet) to help make this clear. A logical story about the particular model's generation process is written as a result.

7009A tries hard not to be overly prescriptive. It specifically encourages the TA to tailor the Standard from the initial stages of P/P planning onward. (Because of this intention, all examples and recommendations listed in this paper are provisional to P/P tailoring.) 7009A “covers what needs to be accomplished and communicated, not how it is to be done. Discipline-specific recommended practice guides should be consulted for specific applications or methodologies.” This phrasing not only encourages expertise and promotes use of established, proven methods, it also discourages shortcuts and works to expose inexperienced practice.

“In the development of NASA-STD-7009A, the decision was to focus on the product of models and simulations, rather than on the process of modeling and simulating.” This gives tremendous freedom to the modeler. What it means, though, is that whatever process is chosen, the modeler be responsible for communicating how well the model does its job. The Standard is all about what was accomplished vs not accomplished in terms of answering the question it was designed to answer. It correctly puts focus on solving the problem rather than on making a model. This is freeing, as there are often many ways to answer a question well.

WHAT IF 7009A ISN'T REQUIRED – WHY SHOULD I USE IT?

7009A can look hard to apply. Particularly for new users, the Standard can look daunting and the Handbook detailing its application is thick. Following it does require additional overhead in planning and documentation over on-the-fly work. Many modelers might prefer to ignore it away. The Standard's authors understood this, hence the only *requirement* to application is that it be applied to M&S affecting design, verification, risk assessment and similar decisions involving high risk. There is a case, however, for working to its intent using its basic structure for M&S involved with lesser risk as well. It doesn't have to be difficult to use, and statistical modelers are likely to already be familiar with and already use the structure the Standard asks for. The key is in understanding, planning and appropriate brevity.

A computational fluid dynamics (CFD) modeler may be used to presenting a number of graphical displays including animations, background mathematics, and maybe a step-through of physics assumptions going into their highly complex model along with engineering predictions and main sections of a team report. A typical expectation of a statistical modeler might be a few values in a table, some graphs, and a report section, no matter the complexity of the model behind the findings. To the modeler, this might seem both a relief and unfair. A statistical modeler can be forgiven for resisting applying a complex standard to what appears to be a relatively simple model. However, where this occurs, a number of opportunities in communicating and understanding are missed and model credibility might be incorrectly appreciated.

There are a handful of reasons for expanding the understanding of any modeling effort using the structure of 7009A that benefit the modeler and their stakeholders.

1. *Requirement.* If the TA requires adherence to the Standard, it has to be done.
2. *Communication.* It benefits all when a model, its results and the intended takeaways are understood intentionally at an appropriate level by all. Even simple models can be misunderstood. Use of 7009A's standard reporting format, particularly the credibility scale, can be powerful. It also doesn't hurt that filling in the Standard's worksheet makes report writing simple.
3. *Rigor.* Models and modelers benefit from using best practices. Establishing rigor makes for good models.
4. *Trust.* Communicating that level of rigor to stakeholders in language they understand leads to appropriate trust of the model and its results. It also goes to establish trust in the

modeler. This can be especially important when stakeholders are unknowledgeable or mistrustful of statistical methods in general.

The first point was covered previously. The third, rigor, should already be valued by the M&S practitioner; 7009A simply places a premium on, advertises and rewards full rigor. Trust takes time and understanding, the latter which 7009A can only help with. 7009A's authors knew that communicating all the important steps involving rigor and recording adherence to a solid set of requirements will go far in building trust. Communication is central.

As always with communication, language can ironically cause an issue. As an obvious example, statistical (e.g. linear) regression mentioned in the Standard. So is regression testing, a method for verifying modeling code after changes are made. Likewise, what are often known as use cases or parameter sets are called scenarios in the Standard. Care must be taken to ensure no one is confused by regression scenarios – or any other potentially confusing topics.

AN APPLICATION EXAMPLE

An example of use of the Standard based on a relatively simple modeling effort will be given here. The required reporting will be given first, followed by a rough story using the M&S Life-Cycle Worksheet (LCW) attached to 7009A. For this fairly simple example, filling in this form – the first time this author has done so in a couple of years, using an example from a decade ago – and creating the reporting took around 6 hours. Note that quite a number of requirements are missing or incompletely represented in the worksheet, but the author has found it to be useful as a recording, communication and reporting device.

There is also a very useful expanded timeline spreadsheet that illustrates some of the relationships between requirements along with when each might be planned, revised and fulfilled. Neither worksheet will match every modeling effort perfectly. They are provided as a guideline to be adapted by the P/P and the modeler. Final reporting guidance is fairly well-explained in the Standard. This reviewer has not found the associated reporting tables 11 – 13 in the Handbook helpful, so the example given here will use prose responses to the relevant requirements instead.

Please remember that this report is intended as a guide only. Users of 7009A are themselves responsible for correct interpretation and application to their own models and results.

The example is from a real aerospace application highly disguised as a test of puncture resistance in a multi-ply tire on an experimental aircraft. Picture the tire as consisting of a fairly thin balloon-like pressure vessel covered with armored yet porous exterior plies. The tire is stowed deflated during flight to save space and inflated quickly after landing gear deployment. During a previous otherwise nominal flight, foreign object debris (FOD) believed generated within the novel high-velocity inflation device was found embedded in the inner surface of the tire. While not punctured, the sharp particle cut partially through that interior pressure bladder. Models showed that if cut deeply enough, the tire would catastrophically burst on landing with no warning and potentially fatal results. Design mitigation would be costly and take time. The test's goal was to estimate under what conditions a puncture might occur to find out whether the current ground hold on the aircraft should be continued. An earlier test of 44 impacts implied high variation in penetration depth, but the data from the seat-of-the-pants one-factor-at-a-time test matrix was unanalyzable and inconclusive. A statistical modeler was asked to apply design of experiments (DOE) principles in designing a new test and perform statistical analysis on the test results.

7009A had been required during aircraft design. The Project had seen overall benefit from that decision, so decided to require it for key M&S during the investigation.

PREDECLARATIONS

Several items needed to be predeclared early in test and analysis planning, many negotiated between the modeler and test team or other stakeholders. These include the problem statement, model requirements like verification and validation criteria, and minimum credibility requirements per Appendix E in the Standard.

7009A Appendix A contains a form on which to record any tailoring for a particular modeling effort. The answers are best negotiated beforehand. Its completion will not be covered here.

Credibility requirements as predeclared are shown below on a copy of Table 3 in 7009A Appendix E by green shading. Verification of the analysis code was waived since modeling required relatively straightforward multiple linear regression analysis of data from a well-designed test matrix performed using widely-used commercial software. Verification of the model itself was performed during independent review. Validation will not be performed. Real world data was not available and funding and time constraints precluded running confirmation tests.

7009A Table 3— Predeclared Credibility Assessment Levels to Be Achieved

	M&S Development			M&S Use (Operations)			Supporting Evidence	
Level	Data Pedigree	Verification	Validation	Input Pedigree	Uncertainty Characterization	Results Robustness	M&S History	Process/ Product Management
4	All data known & traceable to RWS with acceptable accuracy, precision, & uncertainty.	Reliable practices applied to verify the end-to-end model; all model errors satisfy requirements.	All M&S outputs agree with data from the RWS over the full range of operation in its real operating environment.	All input data known & traceable to RWS with acceptable accuracy, precision, & uncertainty.	Statistical analysis of the output uncertainty after propagation of all known sources of uncertainty.	Sensitivities known for most parameters; most key sensitivities identified.	Nearly identical model <u>and</u> use.	Controlled processes are applied; measurements used for process improvement.
3	All data known & traced to sufficient referent. Significant data has acceptable accuracy, precision, & uncertainty.	Formal practices applied to verify the end-to-end model; all important errors satisfy requirements.	All key M&S outputs agree with data from the RWS operating in a representative environment.	All input data known & traced to sufficient referent. Significant input data has acceptable accuracy, precision, & uncertainty.	Uncertainty of results are provided quantitatively through propagation of all known uncertainty.	Sensitivities known for many parameters including many of the key sensitivities.	At most minor changes in model <u>and</u> at most minor differences in model use.	Controlled processes are applied; process compliance is measured.
2	Some data known & formally traceable with estimated uncertainties.	Documented practices applied to verify all model features; most important errors satisfy requirements.	Key M&S outputs agree with data from a sufficiently similar referent system.	Some input data known & formally traceable with estimated uncertainties.	Most sources of uncertainty identified, expressed quantitatively, and correctly classified. Propagation of the uncertainties is assessed.	Sensitivities known for a few parameters. Few or no key sensitivities identified.	At most moderate changes in model <u>and</u> at most moderate differences in model use.	Formal processes are applied.
1	Some data known and informally traceable.	Informal practices applied to verify some features of the model and assess errors.	Conceptual model addresses problem statement and agrees with available referents.	Some input data known and informally traceable.	Sources of uncertainty identified and qualitatively assessed.	Qualitative estimates only for sensitivities in M&S.	New model or major changes in model <u>or</u> major differences in model use; but model/ changes/ uses documented.	Informal processes are applied.

A lower level **0** indicates insufficient evidence to make a determination.

M&S History required some negotiation within the team. While the regression model itself would be entirely new, including test data, calling the model “new” – Level 1 – would not be particularly descriptive. The team and stakeholders decided to include the statistical analysis code producing the regression model as well as the regression model itself in this column. Since the standard commercial software was planned be used without modification and the well-designed experiment should require no liberties or approximations in analysis, this portion of the model would have strong history; the only change would be the new regression model itself.

Given what happened was what was planned, the results would rate a Level 3. If the data were problematic and new analysis code (in R or Matlab, say) would need to be used, a lower credibility level would be applied. Getting this straight during the planning phase might prevent longer and more tense discussion after the analysis was complete.

REPORTING

Reporting is relatively straightforward, outlined in the Standard using requirements M&S 32 – 39. The Handbook has three tables (11 – 13) that could be helpful for summarization, but this reviewer has not found them useful in their present form. A story of conformance to the requirements will be given instead. Blue text indicates the modeler's responses.

[M&S 32] Shall include explicit warnings for any of the following occurrences, accompanied by at least a qualitative estimate of the impact of the occurrence:

- (1) Any unachieved acceptance criteria (as specified in [M&S 8] (1)). *There were no significant deviations from acceptance criteria.*
- (2) Violation of any assumptions of any model (as specified in [M&S 11]). *Randomized run order was disregarded during test, presenting a challenge to independence of test data. This does not appear to have resulted in a significant effect to findings. Three of 18 tests were not run. This did not present a significant effect on analysis because of the clear and large relationship between inputs and measured test responses.*
- (3) Violation of the limits of operation (as specified in [M&S 13]). *No violations. Test matrix designed to preclude this.*
- (4) Execution warning and error messages (see [M&S 27]). *No significant warnings or errors. One point may have been an outlier, but constructing a model without this point returned a model that differed inconsequentially from model containing that point. There was no engineering reason to consider the point as anything other than a normal test run.*
- (5) Unfavorable outcomes from the proposed use assessments (described in [M&S 23]). *No unfavorable outcomes seen or expected.*
- (6) Unfavorable outcomes from any setup/execution assessments (described in [M&S 25]). *No unfavorable outcomes seen.*
- (7) Waivers to any of the requirements in this NASA Technical Standard. *Validation using RWS or independent data not required per investigation team, approved by Technical Authority. No waivers were required.*

[M&S 33] Shall include an estimate of results uncertainty, as defined in [M&S 29 (1)-(3)], in one of the following ways:

- (1) A quantitative estimate of the uncertainty in the M&S results, or

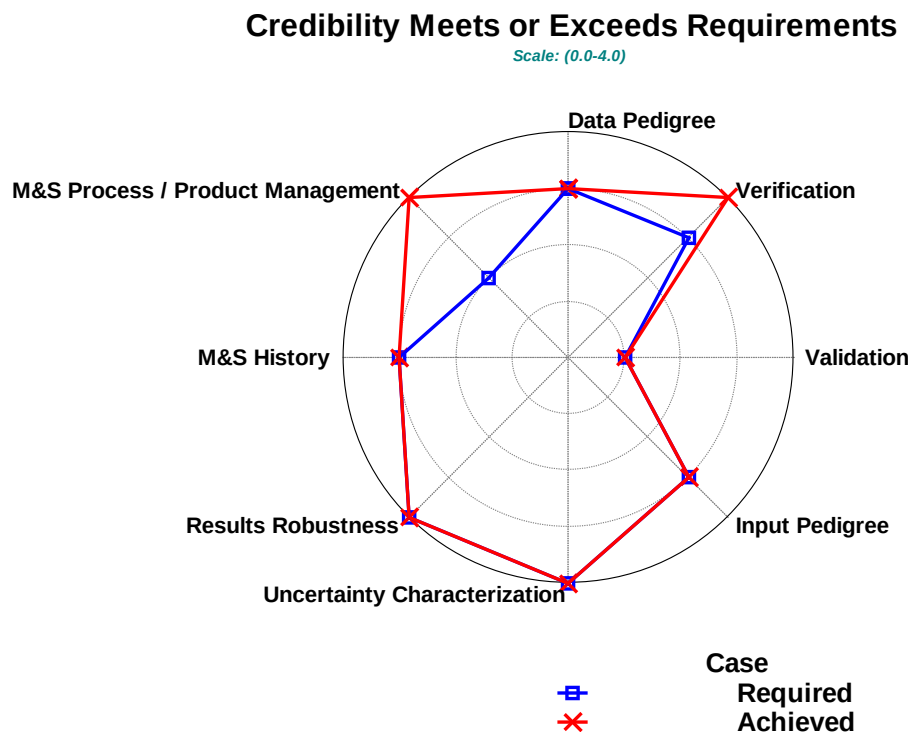
Random variation in the test data unexplained by the model (residual error) has a standard deviation of 0.0084 cm.

(2) A qualitative description of the uncertainty in the M&S results, or

(3) A clear statement that no quantitative estimate or qualitative description of uncertainty is available.

[M&S 34] Shall include a description of any processes used to obtain the estimate of uncertainty as defined in [M&S 28 (1)-(3)]. Largest source of model uncertainty, residual error, is produced during the modeling process as a result of linear regression analysis. Other uncertainties measured directly or through engineering calculation found to be inconsequential compared to this value. Since the model estimates a bounding case and test is shown to reflect RWS well, uncertainties between test data and RWS not likely of any consequence.

[M&S 35] Shall include the assessment of credibility for the M&S results for each factor specified in [M&S 31]. Graph of results below. Details in attached worksheet, to be expanded in investigation final report.



The star glyph above, here produced using commercial statistics suite Statgraphics, shows the comparison between predeclared requirements (blue) and as-met credibility. This is one of the suggested summary roll-up graphics suggested to display results. The graphic is available in most engineering and statistical software and in Excel. It may be called a radar plot.

Modelers should consider making such a graphic available when presenting model results. The first few times an audience sees the graphic may require explanation. It is hoped that the audience will come to expect the rating from others once the utility is explored and realized.

[M&S 36] Shall include the findings from any technical reviews accomplished in regard to the development, management (control), and use of the M&S. [Independent reviews performed by three statistical modelers and more than three engineering practitioners.](#) Findings by reviewers were all minor and either corrected or were deemed inconsequential to use of model results. The latter include nonrandom test run order, resulting in unknown error but judged not likely of any credible consequence by SME test engineers and completion of 15 of 18 planned tests, resulting in little discernible effect on results.

[M&S 37] Shall include the qualifications of the developers of the M&S and the users, operators, and/or analysts involved in producing the results from the M&S, including, but not limited to, their relevant education, training, and experience. [\(Will not be covered in this report.\)](#)

[M&S 38] Shall show what aspects of modeling and simulation are documented, as shown in Appendix A. [\(Will not be covered in this report.\)](#)

[M&S 39] Shall include an assessment of and rationale for the risks associated with the use of the M&S-based analysis. [Risk communicated by including model results in the next analysis step \(PRA\) will be increased due to higher-fidelity estimate of penetration depth from model replacing highly conservative current estimate. Due particularly to conservatism of bounding case supplied but also to high fidelity of test compared to RWS, analyzability of test data, straightforwardness of modeling analysis and other factors, added risk is expected to be low.](#)

Summary. The results of this test and modeling effort are offered with the credibility level required for use of the results in the investigation's risk model.

[Additional information on test planning, performance, data analysis and modeling will be made available in the investigation final report. In the interim, the modeler may be contacted directly. The data and model are under configuration control. They will be offered as a configuration controlled document into the NTRS and appropriate systems within NASA. No significant deviations from established practices are known to have occurred. There were no dissenting opinions recorded regarding this test and modeling effort or its credibility.](#)

(Signed)

WORKSHEET FOR EXAMPLE

The tables below were constructed by filling out the Worksheet 1 Excel file (currently [nasa-hdbk-7009a - worksheet.xlsx](#), available publicly) supplied with the Standard. The modeler's Responses are in blue text. Not all requirements are listed in the Worksheet. The author added a few, denoted by blue background. Note that for modelers' purposes, requirements begin with [M&S 6].

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Planning			
Criticality Assessment	[M&S 6]	What method was used to perform the criticality assessment for this M&S?	Apparent near-miss with no actual injury or damage to aircraft in a nominal flight indicates potential real issue. Existing probabilistic risk assessment (PRA) model revised and exercised using new assumptions due to near-miss investigation to date.
		What were the results of the Criticality Assessment?	Single point failure in credible scenarios. Current engineering model lacks sufficient fidelity to predict impact depth; testing required. Potential 1/300 probability of catastrophic failure in revised PRA, but conservative assumptions based on earlier inconclusive test used to estimate effects of particle impact on tire. Nonconservative estimate will result in probable loss of aircraft with significant likelihood of pilot loss. New testing required to improve PRA input. Craft currently grounded; excessive further delay may result in Project cancellation.
Covered by Standard	[M&S 7]	Is this M&S within the scope of 7009A?	Test data analyses required to be covered under 7009A.

The risk estimate is highly dependent on results of testing: high M&S results influence, in the words of the Standard. The Project also found that the decision held high consequence. The background and rationale are summarized briefly and should be readable as a complete description of why 7009A is being required.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Planning			
Life-Cycle Planning	4.1.2a	What are the plans for acquisition/ development of this M&S?	Using Design of Experiments (DOE) methods to design efficient test matrix expected to have sufficient number of samples (18) to achieve goals based on worst-case scenario and expected variability.
		What are the plans for use/ maintenance/ retirement of this M&S?	Expected deliverable will include: graphical display of test data; regression model relating FOD velocity and orientation with respect to tire to penetration depth; calculator which will be used to predict velocity required to achieve full penetration of test article for each orientation. Model including data will be archived with final report and configuration controlled as part of that report. The report will be available through NTRS.

While not required by the TA, these recommendations were followed. Note that a plan was made for configuration control, archive and future availability of the model.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Planning			
Best Practices	4.1.3 a, b	What best practices were identified for this M&S?	DOE test planning, execution and analysis methods to be employed.
		How were these best practices applied to this M&S?	Each factor tested at each condition twice (replicated design). Assumed there might be an increasing rate of depth of impact with increasing impact velocity (assumed potential velocity-squared term in model). Assumed impact depth increase with velocity for flat orientation of FOD might be different from rate of depth increase per unit velocity for knife-edge orientation (interaction effect). Test data analysis to use standard regression routines in widely-used commercial statistical analysis software designed for this purpose.
Technical Reviews	[M&S 9]	What technical reviews during development, control and use of the M&S were performed?	Reviews performed on completed model, calculator, results and reporting. No reviews were undertaken during development or control phases.

This contains an assertion that standard and best-practice methods would be used in planning and analysis and a rough statement of how. We'll assume the language was scrubbed with the test engineer to be understandable in engineering language but still show statistical rigor, and that any necessary deeper details will be reported elsewhere.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Planning			
Training	4.1.4 a, b	What training was required for M&S developers and operators?	Developer training is not required. Operators will be given a spreadsheet calculator allowing inputs within the range and domain of the test data and excluding other inputs and changes to the model.
		What training was accomplished for M&S developers and operators?	n/a.

It can pay to fill in blanks even where not necessary. Here, the modeler and test team are required to think about how difficult the model should be to run. The team settled on a simple-to-use calculator incorporating safeguards against misuse so that little user training will be necessary. Had training been necessary, figuring this out during planning could save time and money, potentially even guiding a different choice of model or model operation.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
M&S Initiation	[M&S 10]	What is the real world system?	Tire stowed deflated during flight pressurized during final approach. Innermost plies of tire not armored against damage. Pressurization system found to contain or produce occasional small metal flakes known to have been shot toward deflated tire surface during inflation. At least one penetrated a tire to a depth of 0.08 cm. Portion of RWS to be modeled consists only of FOD of a given energy/ orientation and tire interior plies.
		What is the RWS environment?	Temperature/ humidity controlled during tire stowage to 50°C/ 65% RH max.
		What is the RWS problem/ decision/ situation for the M&S?	Impingement of FOD on interior of this aircraft tire from inflation system operation resulting in penetration through strengthening plies to some depth during final approach. Need to know the probability that a worst-case particle at worst-case relative velocity to the tire will penetrate the tire. Worst case is radial impact of 0.2 gram steel chip at velocity of 1000 kph at 50°C.
		What is the RWS intended use?	Model to be used to predict likelihood penetration of worst-case FOD to critical depth of 0.16 cm through strengthening plies.
	[M&S 8] (2)	What is the M&S intended use in relation to the RWS problem/ decision/ situation?	Model predictions to be used in probabilistic risk analysis (PRA) to estimate risk of loss of aircraft and pilot due to tire puncture.

This is where the problem statement is elucidated. Note that a DOE practitioner would likely answer these questions during the planning phase. The real world system (RWS) is just what it sounds like – the process, product or activity that the model is intended to mimic. Here, it is the process of a piece of FOD impinging on a tire.

The RWS here is considered to be the actual tire. The test and simplify the RWS by looking only at the interaction between the energetic FOD and the tire material. Constructing the problem statement, including deciding the reality and scope of the RWS being modeled, is both difficult and important. Deciding this before testing and modeling begins results in focus on solving the problem rather than on what tests and modeling will be performed, making it more likely that the test campaign and model will be parsimonious and lead to a solution.

From here on, it will be assumed that the test has been run and the data has been analyzed. Note that due to the clarity of the data, testing was suspended after 15 of the 18 planned tests were run.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Model Concept Dev.	[M&S 11], [M&S 12]	What is the M&S approach? (Is it known how to model what is to be modeled?)	Multiple linear regression employed is a standard statistical method. Modeler is expert in the technique. Test matrix was designed to return fully and easily analyzable data from test. No unusual methods were required.
	[M&S 12]	What's included in the M&S, including model environment influences?	Linear regression model includes terms quantifying effects of FOD particle velocity and orientation on penetration depth. A term adjusting predictions for greater increase in penetration depth at any particle velocity for knife-edge FOD orientation than for flatwise orientation (velocity x orientation interaction) is modeled as well.
	[M&S 11]	Is there anything significant to this analysis not included in the M&S or scenarios?	Completed only 15 of 18 tests. Due to clear relationship between inputs and penetration depth, impact to analysis was deemed by test team including statistical analyst to be minimal.
		What assumptions & abstractions are included in the M&S and Analysis?	Standard statistical assumptions 1) constant variability throughout the range of the response and 2) normality of the unexplained residual variability in the data were checked and found acceptable by standard methods. Randomized schedule of testing was incompletely performed, so assumption of independence of test runs cannot be assured; impact to results likely inconsequential in opinion of test team and statistical analyst. Test performance showed no anomalies. It is assumed that the test adequately reflects the RWS adequately for these purposes and that no significant untested or unquantified variables either affected the test or affect the RWS.

Answers to these questions are relatively straightforward for a regression model of test data.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Model Design	[M&S 19]	How is referent data uncertainty characterized?	No RWS data is available. Referent data, deemed sufficient, is test data. Velocity was measured by high speed video as well as laser rangefinder; uncertainty is inconsequential compared to the domain of the velocity parameter. Video showed all orientations differed from intention by less than 3 degrees and were worse for flat orientations. The resulting variation in impact depth was calculated to be small compared to the unexplained variation due to other sources.
	[M&S 28] (1)	How is model input uncertainty characterized?	Referent data is test data/ model input data.
	[M&S 20], [M&S 21]	What uncertainties are included in the M&S?	Test data uncertainty, including measurement system uncertainty, is included in the model but is mostly represented in estimate of total variation in the test data unexplained by the model (residual error). Residual error was quantified and will be delivered to the PRA analysis along with the model.
	[M&S 8]	What are the requirements or specifications for the model?	Requirements included 1) a predictive mathematical model relating particle velocity and orientation to penetration depth; 2) an estimate of variation due to unmeasured random sources; and 3) a prediction of the likelihood of penetration to a depth of 0.16 cm at 1000 kph. Verification was performed through independent inspection of the modeling process and the data. Validation using confirmation runs – tests specifically run but held back from modeling – was neither required nor run.
	[M&S 12]	What are the implementation mechanisms (e.g., math models)?	Modeling was performed using standard multiple linear regression in a well-verified commercial statistics package. Details can be made available.
	4.2.2 m	What is the implementation architecture (e.g. platform-software)?	The final model is implemented in an Excel spreadsheet. It is also available as text in the final report, available for implementation in any suitable analysis software.

These questions are typically straightforward for statistical models. Requirements [M&S 8] should be predeclared at least in rough form during planning, but in practice some may be reactive, contingent on data or analysis results. If there are e.g. requirements in statute or other NASA Standards, these should be listed here. Consider documenting compliance to these additional requirements in documentation regarding compliance to 7009A where appropriate.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Conceptual Validation	[M&S 17]	Have the model design and architecture been conceptually validated? (e.g. reviewed by SMEs, both RWS and M&S)	Conceptual validation of the model's engineering predictions performed through presentation to the test team and anomaly team leads and exercise of the model calculator in real time at team meetings. Team found that test performance mimicked the RWS adequately and the predictions by the model based on that data matched data and engineering knowledge well. Conceptual validation of the statistical modeling performed by presentation to an applied statistician and engineer familiar with the statistical methods. No problems were found.

The model passes the engineering and statistical smell tests: the model makes sense and seems to predict the RWS and the data adequately. Unmentioned is the fact that since the test matrix was planned using DOE principles, it doesn't suffer from correlation of the input variables, a common mistake that could have made the model's parameters difficult to understand.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Implementation		What is the implementation status of the M&S?	The model is available to the team as an Excel calculator and will be available in text form in the final report. The model predictions have been forwarded for PRA analysis.

This question is a bit puzzling, but it appears to be included in case an interim model is released during development. This statement would then note its status. The authors want know if users find items such as this that might use improvement or explanation. Users are encouraged to submit feedback to <https://standards.nasa.gov>.

Note that there is no reference back to the Standard. This question only exists at this point in the worksheet.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Verification	4.1.2 a	What are the verification practices, methods, and requirements?	n/a
	4.1.3 b (3)	Has the model construction been verified? (i.e. Code Verification)	Commercial analysis code well-established. Details available on request. Duplicate analysis using competitive software performed by reviewer; resulting model is identical.
	4.2.4 a	Has the model operation and/or output been verified? (i.e. solution verification)	No. Not required; unlikely a problem due to simplicity of both model and analysis technique performed.
	[M&S 15]	What artifacts (evidence) of verification are available?	Duplicate analysis using software performed by reviewer; resulting model is identical. Visual analysis and model fit statistics indicate good fit to data; both are an indication of correct analysis.
	[M&S 16]	What is the domain of model verification?	The model has been verified for the domain of the input factors and factors held constant during test. Analysis showed that the factors held constant are worst-case, so model is expected to bound the RWS adequately for e.g. temperatures down to the specification limit of the tire and for smaller FOD.

Solution verification refers at top level to whether the accuracy of the sim model output is sufficient for its intended use. However, it doesn't usually apply to unmodeled variation with respect to the desired prediction; instead, it is usually meant to refer to mathematical uncertainty, e.g. rounding errors or convergence problems. In the case of regression, as long as a well-vetted analysis engine is used (Excel may not qualify), this may not be an issue. However, this may be a consideration for many analysis types (maximum likelihood methods, for instance). Because of the complexity of many statistical analyses, this can be an important question. This item is thus a recommendation that might bear being treated as a requirement.

In this particular case, the routine in the software used in analysis is verified and documentation is available. Note that commercial statistical software is not necessarily independently verified.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Empirical Validation	4.1.2 a 4.1.3 b (3) 4.2.6 a	What are the validation practices, methods, and requirements?	Validation directly to RWS or known standards not possible. Information in this section uses test data as referent. This was accepted by investigation team including model results' direct users.
	4.1.2 c	What artifacts (evidence) of validation are available? Who reviewed/ verified the RWS (referent system) data?	No model confirmation/ validation using independent data planned or performed. Previous test data judged not useful for this purpose. No RWS data available for validation.
	[M&S 8] (1), (3), (4)	What is the accuracy, precision, sensitivity, uncertainty, and bias of the model? Does it satisfy the requirements?	Imprecision in particle orientation at impingement for worst-case knife-edge orientation less than 3 degrees off vertical (limit of measurement accuracy) biases prediction low by an insignificant amount per engineering analysis. Lower impingement orientations shown to result in lower penetration depth. Accuracy of model predictions influenced by particle velocity measurement. Inaccuracy/ imprecision in that factor will result in imprecision in penetration depth estimate that can be calculated using the model itself. Velocity measurement during test known to be accurate and precise to very high level compared to effect size and other uncertainties. Model acceptance requirements set by the Project were met.
	[M&S 20], [M&S 21]	What uncertainties are characterized in the model and how do they compare to the RWS uncertainties?	Largest uncertainty in model is unexplained variation. Possible sources include location of ply cords vs impact location, roughness of material surface, air currents (all of which likely occur in the actual tire), and error in measuring penetration depth in tire (measurement believed conservative). Adding this uncertainty to the model prediction at 1000 kph/ knife-edge orientation (worst case)/ 0.2 g particle/ 50°C results in penetration highly unlikely to result in catastrophic penetration. Worst case particle mass and velocity calculated from existing FOD particle ballistic data.
	4.2.8 a	What uncertainties are not characterized in the model and what are those uncertainties in the RWS?	Tire lot-to-lot variability neither tested nor known, assumed inconsequential. Previous tire damage effects not included in modeling; deemed unlikely due to system design.
		What are the model sensitivities? How do they compare to sensitivities of the RWS?	Penetration depth highly dependent on particle orientation and velocity. Findings match understanding of basic physics of scenario.
	[M&S 18]	What is the domain of validation?	Model believed valid within domain of input factors and at or below values of factors controlled constant during test (within tire operational requirement limits).
Permissible Use	[M&S 14]	What are the permissible uses of the M&S?	Model specific to this tire and inflation system design. Model may not apply to redesigned inflation system and will not apply to any future tire redesign.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Development			
Model Release	4.1.2 c	What version of the model is released for use?	Final version released dated 10/14/2019 on password-protected Excel calculator. This was the version used to produce predictions for PRA use.
	4.1.2 c	Where are the model and its development artifacts archived?	Model will be permanently archived in final report. Excel calculator will be available at least through end of anomaly investigation from test team members.
	4.2.2 e	Was a user's guide produced and formally released with the model?	Excel calculator includes instructions for use and built-in warnings when inputs exceed domain of test parameters. Instructions will also be included in final report.

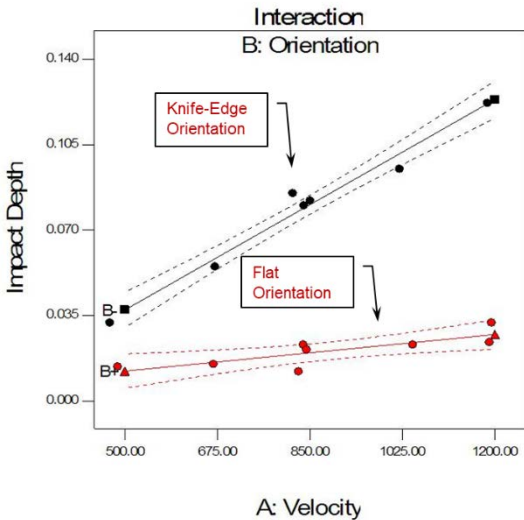
Validation using independent test data or real world data was not performed here, a conscious decision to accept a risk by the test team due to cost of testing and time constraints. This is fine if the team and stakeholders agree that this is acceptable, but that decision must be explicit in reporting. If it is not acceptable, this will result in a shortfall in the credibility rating. 7009A asks for a real accounting of what was performed and what wasn't, whether or not it was best practice, meets requirements or could embarrass the modeler or others. This is a good reason for understanding 7009A before modeling is begun and fully negotiating requirements with stakeholders during planning.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Use			
Use Processes	4.3.2 e	What are the processes for using the model?	Excel calculator requires – and allows – only inputs for orientation and velocity.
	4.2.2 e	Is a user's guide available?	User instructions available on Excel calculator.
Proposed Use	[M&S 22]	What is the proposed use for the model?	Proposed use is for calculation of risk of particle penetration to catastrophic depth into tire during flight.
Use Assessment	[M&S 23]	Does the M&S type & purpose of match the proposed use?	Yes.
		Do the modeling methods (abstractions, assumptions, mechanisms) provide the needed fidelity (level of detail, accuracy, precision, & uncertainty)?	Yes.
		Is the proposed range of use within the permissible model limits?	Yes. Testing performed at velocities in excess of expected maximum particle velocity.
		Are the expected M&S outputs (results) appropriate & within the needed accuracy, precision, and uncertainty?	Yes.

This section may seem redundant to earlier questions. Remember, however, that this section is meant to be filled out when the model is used, where the previous sections refer to phases before the model is exercised to solve the problem at hand. The question on proposed use, for instance, is to gauge whether there is a change from the earlier question regarding intended use. In this case, the exercise of the model occurred at the same time as model construction was completed, so many of the questions really are duplicates.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Use			
Scenarios	[M&S 24]	What are the scenarios for model use?	Single scenario: evaluation of worst case condition. Scenario developed using design information, existing models and first-principles physics.
Setup	[M&S 25]	What are the model setups?	Worst-case scenario well within valid limits of model.
Execution		What is the execution status for the use?	Model fully executed.
Results			
Analysis	[M&S 28] (3)	How was the output (data) analyzed?	Output graphically compared to test data. Full standard statistical examination of results (residual analysis) performed. One potential outlier found but shown to have little effect on model, pulling it less than 0.003 cm in nonconservative direction. This point left in dataset during analysis.

Some questions may require considerable writing to fully answer. It is up to the modeler and their stakeholders to decide the level of detail to include. That negotiation should include whether any important information will be available elsewhere. In this case, the model will be used only during a brief period and will be associated with that full report. Had it been a standalone model that might be used in other work, for example, this documentation attached to the model may have been the only development and use information available. Giving a fuller explanation about how the original use scenario was constructed right here in the Scenario response might have been appropriate not only to satisfy the intent of the Standard, but also to show future users how to construct their scenarios with similar rigor.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links																		
M&S Use																					
M&S Results		What are the best-estimate results provided by the analysis?	Expected worst-case impact depth 0.098 cm; 99% reliability/ 95% confidence tolerance bound 0.128 cm. In preliminary analysis, it appears unlikely that a 0.16 cm penetration will occur in flight before redesigned systems can be deployed.																		
	[M&S 28] (2)	How were uncertainties in the results characterized?	Model results uncertainties are calculated directly as a result of regression modeling. No post-process analysis was performed on M&S results within the scope of this modeling effort.																		
		How well (how directly) do the analysis results address the problem statement?	The model results exceed expectations given earlier testing and are expected to be more than adequate for next-phase analysis. The problem statement will be more directly answered through PRA using this model as an input. The model dataset is probably too small by itself to directly quantify penetration likelihood less than 1% without other inputs to be provided in the PRA.																		
Uncertainty in Estimate	[M&S 29]	What are the uncertainties in the results of this analysis?	Random variation in the test data unexplained by the model has a standard deviation of 0.0084 cm. Within bounds and scope of the analysis, no significant additional uncertainties are anticipated. No significant model uncertainties are known to apply.																		
Sensitivities	[M&S 30]	What are the sensitivities in the results of this analysis?	Penetration depth is sensitive to both particle velocity and orientation at impact.  Interaction B: Orientation Impact Depth Knife-Edge Orientation Flat Orientation A: Velocity <table><caption>Approximate data points from the Interaction Plot</caption><thead><tr><th>Velocity (A)</th><th>Impact Depth (Knife-Edge)</th><th>Impact Depth (Flat)</th></tr></thead><tbody><tr><td>500.00</td><td>0.035</td><td>0.010</td></tr><tr><td>675.00</td><td>0.055</td><td>0.015</td></tr><tr><td>850.00</td><td>0.085</td><td>0.020</td></tr><tr><td>1025.00</td><td>0.105</td><td>0.025</td></tr><tr><td>1200.00</td><td>0.125</td><td>0.030</td></tr></tbody></table>	Velocity (A)	Impact Depth (Knife-Edge)	Impact Depth (Flat)	500.00	0.035	0.010	675.00	0.055	0.015	850.00	0.085	0.020	1025.00	0.105	0.025	1200.00	0.125	0.030
Velocity (A)	Impact Depth (Knife-Edge)	Impact Depth (Flat)																			
500.00	0.035	0.010																			
675.00	0.055	0.015																			
850.00	0.085	0.020																			
1025.00	0.105	0.025																			
1200.00	0.125	0.030																			

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Use			
Caveats	[M&S 32] (1)	Unachieved acceptance criteria	None.
	[M&S 32] (2)	Violation of assumptions	Randomized run order not achieved, resulting in unknown detriment to assumption of independence (no correlation between test runs). Impact to model and prediction judged to be inconsequential.
	[M&S 32] (3)	Violation of limits of operation	None.
	[M&S 32] (4) [M&S 27]	Warning or error messages	No warnings nor errors occurred. A possible outlier was found to have no significant effect on model.
	[M&S 32] (5)	Unfavorable proposed use assessments	None.
	[M&S 32] (6)	Unfavorable setup/ execution assessments	None.
	[M&S 26]	Was model exercised within limits of operation of the model?	Yes.
	[M&S 26]	Were results analyses performed within limits of operation of the model?	Yes.
	[M&S 32] (7)	Waivers to requirements	None.
Requirements Compliance	[M&S 32] (1)	Give details on non-compliances with all M&S requirements and their consequences.	None.

This is where to place cautions on use and interpretation of the model and results. Where caveats are noted, consider stating the expected effect on findings due to or credibility of the model. Also consider training stakeholders to look for and look carefully at this section whenever presented with a model. Among other things, this can prevent this list from reading like a list of grievances. Requirements compliance refers to requirements set by stakeholders for the modeling effort. If requirements from other standards, procedures, statutes etc. apply to the modeling effort, they should also be listed here.

The credibility scale evaluation will be considered last. The final set of questions to address involve a few odds and ends, some of which may be redundant to earlier answers.

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Use			
Technical Review	[M&S 36]	Provide a summary of the technical reviews performed on this M&S/ analysis.	Statistical analysis, model and findings due to analysis reviewed by two NASA statisticians. No major findings; minor findings included incomplete randomization of runs, uncompleted runs, and an error in the calculator (repaired). Test performance, data, model and findings reviewed by team and independently by engineers. No significant findings were reported.
People Qualification	[M&S 37]	What are the qualifications & experience of the people developing, testing, and using this M&S?	Statistical analyst has a PhD in statistics and mathematics, a BA and MS in mechanical engineering and 22 years' experience with NASA in both engineering and applied statistics. Users of the model results hold a minimum of a MS in engineering with at least 10 years' experience in reliability and probabilistic risk assessment on similar systems. Statistical reviewers hold a minimum of a MS in applied statistics and 15 years at NASA; independent engineering reviewers hold a minimum of a PhD in engineering and experience with relevant systems.
M&S-based Analysis Risk	[M&S 39]	What are the risks of basing this decision on the M&S-based analysis?	Low going into the PRA. Testing represents RWS adequately and was performed with no issues of consequence. Model clearly describes test data well and predictions, including uncertainties, appear to be accurate, useful and the correct inputs required to complete the PRA.
Results Archiving	4.3.9 c	Were the M&S results (and related artifacts) archived?	The model, including the data, will be included in the final anomaly resolution report, expected to be made available to appropriate audiences through NTRS.

Much of this might be copied and pasted from earlier answers. The repeated information might be considered a roll-up or summary of what has been written earlier brought together in a different context.

The Standard asks that results of the credibility analysis outlined in Appendix E of the Standard be rolled into the analysis risk accounting here. That means that if there are any shortfalls in assessed vs. required credibility, the associated risk should be documented.

It is recommended that any accounting of compliance to NASA-STD-7009A be included with models when being archived; particularly, again, if other supporting documents are not available. This will help future users of the model with context and also help them should they need to comply with 7009A themselves. New NASA Office of the Chief Information Officer (OCIO) efforts are making archive of statistical models, including the data they are based on, more readily available and simpler. DoD has similar systems already in place. The benefit to posting data, models and full metadata and context to a searchable server is that the data and model may never need to be expensively recreated in future, enabling engineers, scientists and statisticians to focus on advancement rather than reinventing wheels.

THE FINAL CREDIBILITY ASSESSMENT

Item	STD Ref	Questions	Responses, Comments, Notes, References, Links
M&S Use			
Credibility Assessment	[M&S 31]; 4.3.7 a, b, c	Data Pedigree	3. Data used in constructing model from well-run test. RWS data not available.
		Verification	4. Standard reliable methods used to quantify model errors. Analysis code in widely-used statistical software verified independently. Model reviewed by independent statistical modelers, including duplication using independent commercial software. Unexplained residual error quantified.
		Validation	1. No validation runs planned. RWS data not available. Validation by visual analysis of test data vs model expectation. Problem statement clear; model describes test data well in visual analysis and standard statistical tests.
		Input Pedigree	3. Data from controlled test. RWS data not available. Measurement system accuracy and precision shown to have inconsequential impact on results.
		Uncertainty Characterization	4. Uncertainties fully characterized statistically.
		Results Robustness	4. Sensitivities captured in quantitative model, including for combination of factors (interaction).
		M&S History	3. Standard multiple linear regression method in standard vetted software used without changes as intended to be applied. Regression model itself is novel. Change to entire model including statistical software code rated as minor per planning phase discussion with test team and end users. Regression model is stable and configuration controlled.
		M&S Process / Product Management	4. Standard formal experimental design and data analysis process rigorously followed and documented in standard form. Statistical metrics used to select best model from list of potential candidates. All review comments addressed.

The credibility assessment is outlined in Appendix E of the Standard. Credibility refers to the credibility of the results coming from the model, not necessarily the model itself. This is important, yet in application seemingly difficult to remember.

The section here refers back to the table predeclared during model planning. This section allows justification as to why the modeler finds the model results warrant a particular credibility rating. It is recommended that the ratings be vetted by impartial reviewers to make sure nothing missed or misunderstood resulted in either a too-high or too-low rating.

NOTES: UNCERTAINTY CHARACTERIZATION AND PROPAGATION THROUGH A MODEL.

As representations of the mean and variation of a RWS, uncertainty distributions should be considered to be models. When used in e.g. an uncertainty quantification or Monte Carlo simulation, they might be considered sub-models, requiring their own evaluation. It is unlikely that this is widely followed. Indeed, [M&S 19] treats uncertainty in referent data as something other than a model, and [M&S 21] requires only that uncertainties incorporated into an M&S be “documented”.

The statistical modeler should carefully consider how to describe their work and assumptions. The number of individual distributions fed into a Monte Carlo simulation can be overwhelming if each needs to be fully documented as to their pedigree. On the other hand, the statistical modeler may be uniquely able to describe the foundational qualities of distributional assumptions.

Given this level of the modeler’s understanding along with the requirement that model sensitivities be explored suggests a resolution. Most distribution inputs to the model will either have little effect on the simulation output, thus will not themselves pass enough risk through the simulation to require any more than a slight mention of their development.

Only the vital few distributions with high impact on the simulation output may warrant any more dissection. In those cases, the modeler is given a choice by 7009A as to the depth of description. The choice should be driven by risk considerations. For distributions that drive a model’s output, if any of those cannot be explained simply, they must be grounded in a tenable story. For any subset of those where the story is one that is not already understood and accepted, and particularly if it relies on uncertain sources or data that, if another credible assumption is made, changes the model output materially, then that distributional assumption may need to be considered as its own model. For instance, a key distributional input may rely on sparse data from a process for which predictability (i.e. statistical control) is not evident. These likely rare cases should be considered for coverage under 7009A.

SUMMARY AND CONCLUSIONS

NASA-STD-7009A might be seen as a burden. This is unfortunate and doesn’t need to be true. This paper shows that with comparatively little effort beyond what is normally performed during an M&S task, compliance to the standard might not represent undue overhead. In fact, it adds value to an M&S by insisting on focus on the problem to be solved, requiring structure where necessary and sanctioning flexibility where it is not, suggesting helpful planning steps and handing the modeler an outline to document key modeling steps. The credibility scale provides a standard, reasonably clear overview of the context surrounding the recommendations presented to a decisionmaker and other stakeholders with regard to the quality, uncertainties and sensitivities of the model’s results.

The Standard is being required in some cases. It is recommended to the reader that they consider using it at least roughly for nearly any important modeling effort. The time taken in filling out the Life-Cycle Worksheet and displaying a thoughtfully-applied credibility scale when presenting results of modeling might help communication of important information beyond simple numbers. Communication of this information can help build trust in both the model results and the modeler, a desirable outcome.

ACKNOWLEDGMENTS

The author would like to appreciate Gary Mosier (NASA Goddard Space Flight Center) and Martin Steele (NASA Kennedy Space Center) for working with the author many times in applying the Standard – and for their leadership in being among the hard-working authors of 7009A and its associated documents. Peter Parker (NASA Langley Research Center) and Curtis Larsen (NASA Johnson Space Center retired) also deserve acknowledgement for planning and analysis of experiments included in the example.

REFERENCES

References

- [1] NASA, "NASA-STD-7009A Standard for Models and Simulations w/Change 1," NASA, Washington, DC, 2016.
- [2] NASA, "NASA Handbook for Models and Simulations: An Implementation Guide for NASA-STD-7009A," National Aeronautics and Space Administration, Washington, DC, 2019.