# A Credibility Assessment Scoring (CAS) Process for Mission Risk Management.

*Silvano P. Colombano*
NASA Ames Research Center
MS 269-2 Moffett Field CA 94035
650 604-4380 silvano.p.colombano@nasa.gov

*Wei A. Lin*
NASA Ames Research Center
MS 213-13 Moffett Field CA 94035
650 604-4072 W.Lin@nasa.gov

*Michael R. Lowry*
MS 269-2 Moffett Field CA 94035
650 604-3369 michael.r.lowry@nasa.gov

Keywords:
Verification and Validation, Credibility Assessment, Risk Analysis

**ABSTRACT:** *The Credibility Assessment (CA) process is designed to give project decision makers a detailed picture of the quality of the results obtained from simulations from several possibly crucial points of view. This allows them to better understand the risks that might be associated with the use of those results in the context of design and/ or process decisions. The motivation for this extra step, beyond typical V&V processes, is NASA's experience in rare but costly mission failures, both in terms of hardware and lost lives. Often mission decisions had been made on the basis of results obtained from models and simulations that had undergone traditional V&V but whose risks and limitations had not been clearly understood by decision makers. The CA process assigns a CA Score (CAS) in each of the following 8 areas: Verification, Validation, Input Pedigree, Results Uncertainty, Results Robustness, Use History, M&S Management and People Qualifications. The User/Analyst assesses and documents the credibility of the M&S results in each one of the above areas in accordance with NASA-STD-7009. Depending on the intended use, a required Degree of Confidence (DOC), ranging from 0 to 4,is assigned* a priori *in each area. User/analysts will assign the DOC that was achieved by the M&S process. The process has already been performed for the CEV Aerosciences Project M&S Database (CAP-Aero), and is being planned for other Orion M&S projects.*

## 1. Introduction

In this paper we review the historical background and technical issues that led to the establishment of a NASA standard for Modeling and Simulation (M&S). We then focus on the concept of "credibility" as brought forth by the standard and on the related process of assessing and quantifying the credibility of models and simulations with respect to intended use (sections 4-5). This process is identified as "Credibility Assessment" and its quantification is the Credibility Assessment Score (CAS). The Orion program (Crew Exploration Vehicle) is the first to adopt this process for its M&S products, and the Crew Exploration Vehicle Aerosciences Project (CAP) Aerodynamic database was the first M&S upon which a Credibility Assessment was performed. Section 6 describes the experience with the CA applied to this database. Finally, in section 7-8 we have a discussion on process issues encountered and conclusions.

## 2. Historical Background

One of the aftermaths of the Space Shuttle Columbia accident on February 1, 2003 was the establishment of the Columbia Accident Investigation Board (CAIB) [1], for the purpose of understanding the causes of the accident and providing recommendations. While their final report specifically addressed the shuttle program, a subsequent NASA management review led by Diaz [2] on the basis of added input from the entire agency determined that some of the problems were systemic and affected all NASA development projects. Specifically, Action 4 of the report recommended that a standard for the for the development, documentation, and use of Modeling and Simulation (M&S) should be established, for the purpose of lessening the likelihood of inadequate quality or inappropriate use of simulation results for crucial decision points.

One important aspect of the standard was the need to address the "credibility" of results and to insure that decision makers would have a complete understanding of the limitations of these results and their implications for the decision making process. The novelty of this recommendation was in the position that traditional V&V processes had not been adequate in capturing all considerations that were necessary for insuring the appropriate use of M&S results.

NASA Standard 7009 was established [3] to address all M&S development issues including that of results "credibility". This paper focuses exclusively on the issue of "credibility" as addressed by the Credibility Assessment Scale (CAS) defined in the 7009 Standard and on its application to the Crew Exploration Vehicle Aerosciences Project (CAP) Aerodynamic database. The CAS was also reviewed in a paper by Babula et al. [5].

## 3. Problem Statement

The problem can simply be stated as follows: "How can we be sure that the results produced by the M&S can be relied upon to make the decisions that are required?" A corollary of this question is "How can we quantify the risks involved in the use of any particular M&S?" Here are the reasons why this is a major problem.

There is a vast array of different types of models and simulations in use at NASA. They can be divided broadly in the following categories: Operations, Manufacturing, Assembly, Test and Evaluation, Design and Analysis, Natural Phenomena Predictions [4] For each of these categories models can be built for different purposes from initial "ballpark" estimates and approximations to detailed system behavior predictions that will guide design choices and mission decisions. The purpose for which a M&S is put to use needs to match the purpose for which it was built, for its validity to be established. Initial estimates may require just a coarse simulation. A more precise simulation would incur extra unnecessary costs. At the other end of the spectrum one would need to make sure that both accuracy and precision of given results are at the correct level of resolution needed for crucial design decisions, in addition to other factors we'll examine below. Here, settling for results of lesser quality for the sake of cost or expediency could spell disaster.

The problem is compounded by the obviously desirable policy of maintaining ever expanding libraries of models for re-use. Clearly there is the potential of huge cost savings in model reusability, as development costs are high. The drawback, however, is the possibility of using models for situations for which they were not intended, unless a clear understanding of model assumptions, caveats and other technical issues is clearly communicated to both analysts and decision makers.

Central, then, to the problem of reliability of M&S results is the concept of "credibility". Given the complexity of issues connected with the production of M&S results, how can we insure that the credibility of results will be sufficient for the task at hand, and how can we quantify the risks involved in relying on them for decision making? Crucial to this issue is also the need to establish processes that will be cost-effective.

## 4. Credibility

The NASA standard defines "credibility" as "the quality to elicit belief or trust in M&S results" [3]. The term was arrived at with some debate, with competing concepts such as "rigor", "quality", "maturity" and, still, Verification and Validation [6] being proposed as more apt. Indeed, the first question that comes to the mind of an M&S practitioner is "why is this concept necessary? Isn't the purpose of V&V to establish the credibility of M&S results?". And the answer is that of course traditional V&V is central in establishing the credibility of M&S results, but there other more ephemeral factors that enter into play when decision makers view M&S results. For instance, "what was the experience of the team that produced the results?" "Was this particular M&S used before?" etc. The concept of "credibility" intends to include these and other additional factors, and the CAS establishes the process of identification and quantification of these factors. To clarify, one can make the case that traditional M&S practice has implicitly included consideration of these factors, in addition to, or as part of, software V&V. The purpose of the CAS is to make this process more transparent and quantitative.

## 5. The Credibility Assessment Score (CAS), and Relevant Factors

Eight factors were identified by the 7009 Standard as playing a major role in establishing the credibility of M&S results. These are:

- Verification
- Validation
- Input Pedigree
- Results Uncertainty
- Results Robustness
- Use History
- M&S Management
- People Qualifications

They are explained in detail below. One important note is that, while these factors have been deemed to be common to all M&S applied to all categories listed in section 3, the degree to which each factor will be considered crucial to the credibility of results, as required by any particular project, will be left entirely up to the decision makers. The CAS consists in a score from 0 to 4, where 0 indicates that there is no evidence that the factor in question was even considered in the Credibility Assessment (CA), and 4 indicates that the highest possible confidence was achieved for that factor. It is left up to "responsible parties" to decide *a priori* what score will be expected for a given factor, depending on the project in question and the degree of risk that is considered acceptable. For instance, for a completely new M&S it would be obvious that no evidence of Use History, thus a score of 0, would be acceptable. The acceptable scores for each factor are defined as "thresholds" for that factor. Note also that different thresholds may be set for different phases of the project life-cycle; for example, at PDR, it may not be necessary or desirable (due to associated costs) to establish "results robustness" at the highest possible level of confidence (score of 4). The responsible parties may decide to set the threshold for that score at 2 instead of 4. A final report chart will indicate with colors to what extent the thresholds are satisfied. Figure 1 shows an example of this chart, with thresholds set for the PDR milestone. We now describe each factor, with the meaning of each score value for that factor.

### 5.1 Verification and Validation

Verification and Validation for M&S follow the same definitions generally applied to software V&V: where verification is "doing things right" and validation is "doing the right thing", but there is added complexity in both concepts for M&S, because, while software design, at least in principle, can verified if requirements are satisfied, a "model" needs to represent the external world

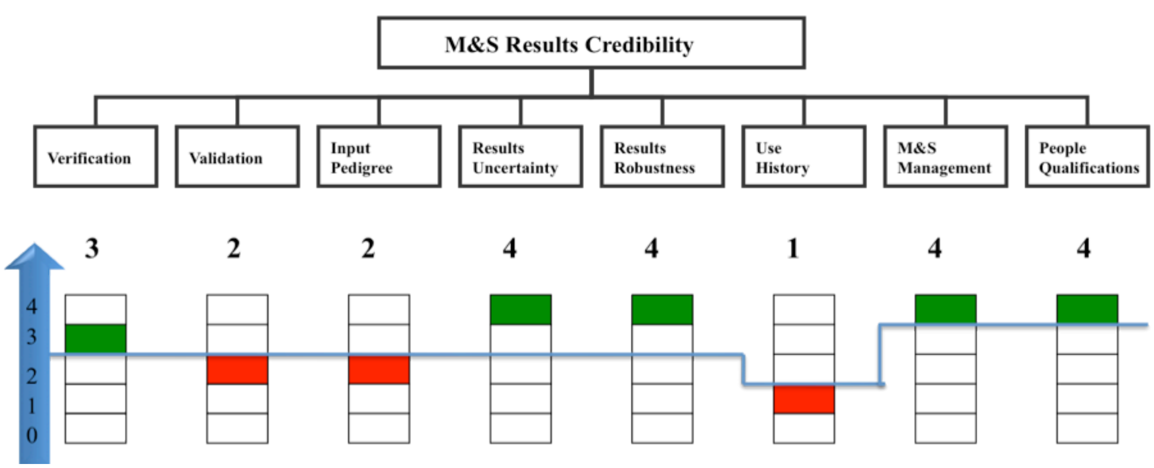


Figure 1. Example of CAS factors, with threshold bar (in blue) and color indications of compliance: green and red, above and below threshold, respectively.

in a way that is appropriate for the problem at hand. The translation from the external world to a “Conceptual Model” is a step that needs to also be verified and validated, together with the correct implementation of the Conceptual Model into simulation software. We will not delve into this added complexity, as approaches and techniques for M&S V&V are well established in the field. We will limit ourselves to showing the meanings assigned to each level for both factors. For Verification and Validation they are listed in Tables 1 and 2, respectively

| Score | Verification Evidence |
|---|---|
| 4 | Reliable estimation methods are used to quantitatively assess numerical errors. These estimates show that errors are small from test suites which exercise all important algorithms, all important features and algorithms and all important couplings (physics, modules, etc.) of the full computational model. |
| 3 | Some formal method is used to assess numerical errors associated with unit testing with significant coverage of the code. |
| 2 | Favorable results from unit and regression testing of key features of the computational model. |
| 1 | Favorable evidence of verification for conceptual and mathematical models. |
| 0 | Insufficient evidence |

Table 1: Evidence scoring for the Verification factor

| Score | Validation Evidence |
|---|---|
| 4 | 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 | 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 | 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 | M&S conceptual and mathematical models compare favorably with “general problem” and “textbook” referents. |
| 0 | Insufficient evidence |

Table 2: Evidence scoring for the Validation factor

**5.2 Input pedigree**

Input to M&S can consist of notional or rigorous data derived from referent systems. Data can be deterministic or statistical in nature. If deterministic, it is crucial that representative point values and ranges be chosen. Both quality and quantity of data from a referent system are crucial to the credibility of the simulation. If the data are stochastic then the probability distributions and the choice of runs necessary for adequate data analysis require careful preparation. The issue of the quality of input data can be understood with the following simple example: suppose a particular re-entry capsule design needs to be tested in a simulation to show that a proposed trajectory will be safe. The results may be satisfactory because indeed the design was right for the task. Unfortunately it could also appear to be safe because of an error in the M&S. So, merely obtaining a satisfactory result is no guarantee that that result is credible. We also need to have confidence in the quality of the input data. Subject matter experts (SMEs) are relied upon to make the necessary choices for the data needed for input to M&S systems. The issue of the quality of the input is categorized as “input pedigree” and is considered one of the crucial factors for results credibility. Table 3 shows the scoring criteria for input pedigree.

| Score | Input Pedigree Evidence |
|---|---|
| 4 | M&S results compare favorably with measured data for 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. |
| 3 | M&S results compare favorably with acceptable measured referent data from problems of interest, or the input came from M&S with a summary credibility rating above 3.0. Uncertainty associated with the input data is known. |
| 2 | Input data is traceable to formal |

| | |
|---|---|
| | documentation, or the input data came from M&S with summary credibility rating above 2.0. |
| 1 | Input data is traceable to informal documentation, or the input data came from M&S with summary credibility rating above 1.0. |
| 0 | Insufficient evidence |

Table 3: Evidence scoring for the Input Pedigree factor

**5.3 Results uncertainty**

There are two sources of uncertainty: stochastic and epistemic. Stochastic uncertainty is related to the quantity of inputs, the number and types of runs and the nature of the analysis conducted. Epistemic uncertainty is related to inherent lack or incompleteness of knowledge of the system being modeled and the assumption that are being made in building the appropriate conceptual and computational models. Techniques for dealing with statistical uncertainty are algorithmic and based on statistical theory. Epistemic uncertainty is being addressed by a growing body of work. See table 4 for scoring criteria.

| Score | Results Uncertainty Evidence |
|---|---|
| 4 | Uncertainty estimates are quantitative and based upon nondeterministic and numerical analysis. |
| 3 | Uncertainty estimates are quantitative and based upon nondeterministic analysis. |
| 2 | Uncertainty estimates are quantitative and based upon deterministic analysis or expert opinion. |
| 1 | Uncertainty estimates are qualitative |
| 0 | Insufficient evidence |

Table 4: Evidence scoring for the Results Uncertainty factor

**5.4 Results robustness**

Results robustness is the quality of results' stability with regard to input uncertainty. Confidence in results is clearly undermined when input variations that are within an acceptable range of uncertainty yield results differences that are of consequence in the decision making process. Robustness is determined by conducting a "sensitivity analysis" of the system, where inputs are systematically varied to determine the effect of these variations on the system's output. Another important aspect of robustness is how well the robustness of the M&S matches that of the real system. If the M&S is more robust, then there is some important aspect of the real system that is being missed. If it is less robust, then there are some possible problems with the computational model, and reliance of M&S results becomes questionable. Both aspects of robustness are captured in the robustness factor (Table 5).

| Score | Results Robustness Evidence |
|---|---|
| 4 | Sensitivity of the M&S results for the real-world system is quantitatively known for most of the variables and parameters. |
| 3 | Sensitivity of the M&S results for the real-world system is quantitatively known for many variables and parameters. |
| 2 | Sensitivity of the M&S results for the real-world system is quantitatively known for a few variables and parameters. |
| 1 | Sensitivity of the M&S results for the real-world system is estimated by analogy with the quantified sensitivity of similar problems of interest. |
| 0 | Insufficient evidence |

Table 5: Evidence scoring for the Results Robustess factor

**5.5 Use History**

The use history factor considers two dimensions of importance to the credibility of results: how long the M&S has been in validated use, and how similar the problems addressed have been to the problem at hand. Assuming past use has led to successful decisions, clearly the credibility of results is enhanced. Scoring criteria are shown in Table 6.

| Score | Use History Evidence |
|---|---|
| 4 | *De facto* standard |
| 3 | Post-decision real-world events have been accurately represented in results (e.g. validated by mission data). |
| 2 | Used previously to perform analysis upn which critical decisions have been made. |

| | |
|---|---|
| 1 | Specific scenarios have been created to test application, or results compare favorably with outputs from other similar tools. |
| 0 | Insufficient evidence |

Table 6: Evidence scoring for the Use History factor

**5.6 M&S Management**

The "Management" terms refers to three different aspects of the development and sustainment of the M&S. One is configuration management of the data and software that comprises the M&S. Clearly, decreased likelihood of version-related problems increases confidence in the results. The second aspect considered is how well the M&S has kept up with changes in systems (computer platforms or modeled systems) that may affect the running or the structure of M&S. Initial system validation may have to re-visited in such cases.
The third aspects deal with continuous improvement of all processes related the M&S, designed to improve the quality and repeatability of results. Overall management that demonstrates how the M&S has been sustained, kept up to date or improved, is scored in this factor is shown in table 7.

| Score | M&S Management Evidence |
|---|---|
| 4 | Continuing Process Improvement: The M&S effort is using measurements on M&S processes to improve the repeatability of the M&S results. |
| 3 | Predictable Process: The M&S effort is measuring repeatability of the M&S results generated by the M&S processes. |
| 2 | Established Process: The M&S effort has established a documented process for M&S development and operations. |
| 1 | Managed Process: The M&S roles and responsibilities have been defined |
| 0 | Insufficient evidence |

Table 7: Evidence scoring for the M&S Management factor

**5.7 People Qualifications**

The final factor considers the academic qualifications, as well as experience, both generic on M&S and specific to the M&S under review, of the people involved in the project. These will include developers, users, and analysts. An understanding of the modeled system, as provided by Subject Matter Experts (SME) is also one the aspects relevant to this factor. The scoring criteria are shown in table 7.

| Score | People Qualifications Evidence |
|---|---|
| 4 | Possesses an advanced engineering or science degree or extensive work experience with the development and use of the M&S being reviewed, and has employed specific recommended practices. |
| 3 | 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 | 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 | 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 |

Table 7: Evidence scoring for the M&S Management factor

**5.8 Technical reviews**

Although not explicitly mentioned in the factors listed above, for the first five factors, the evidence produced requires a technical review, conducted by an appropriate team selected by the organization responsible for M&S development, V&V and data analysis. As expected, the factors requiring a technical review are those that require subject matter and/or process expertise: V&V, input pedigree and results uncertainty and robustness

## 6. Experience with the Crew Exploration Vehicle Aerosciences Project Aerodynamic Database

The Crew Exploration Vehicle Aerosciences Project (CAP) provides an Aerodynamic database for the Orion Program, including the Crew Module (CM), Launch Abort Vehicle (LAV), Service Module (SM), Crew and Service Module (CSM), Launch Abort Tower (LAT), and the associated separation environments. These are Government Furnished Equipment (GFE) products. It characterizes the aerodynamic performance and stability, and the aerodynamic surface pressures (air loads) for certain configurations. The database includes an Application Programming Interface (API) that provides analysts a method to access the CEV Aerodynamic Database. This is the first of number of GFE M&S products that are slated for a Credibility Assessment in the near future.

### 6.1 Identification of the organization and responsibilities

The general framework for the complete M&S life-cycle, from development to final CA is shown in figure 2. Note the difference between "Accreditation" and "Credibility Assessment". Accreditation is a step that confirms that use of a given M&S has been *authorized* for a given purposes. Accreditation takes place *before* simulation results are obtained and analyzed. The CA is performed instead *after* M&S results have been obtained and analyzed. In the case of the CAP-Aerodynamic Database, the framework was adhered to, except that the responsibilities of the Orion M&S Working Group (OMSWG) were split, with the responsibility for accreditation given to Orion Vehicle Integration Team (OVIT).

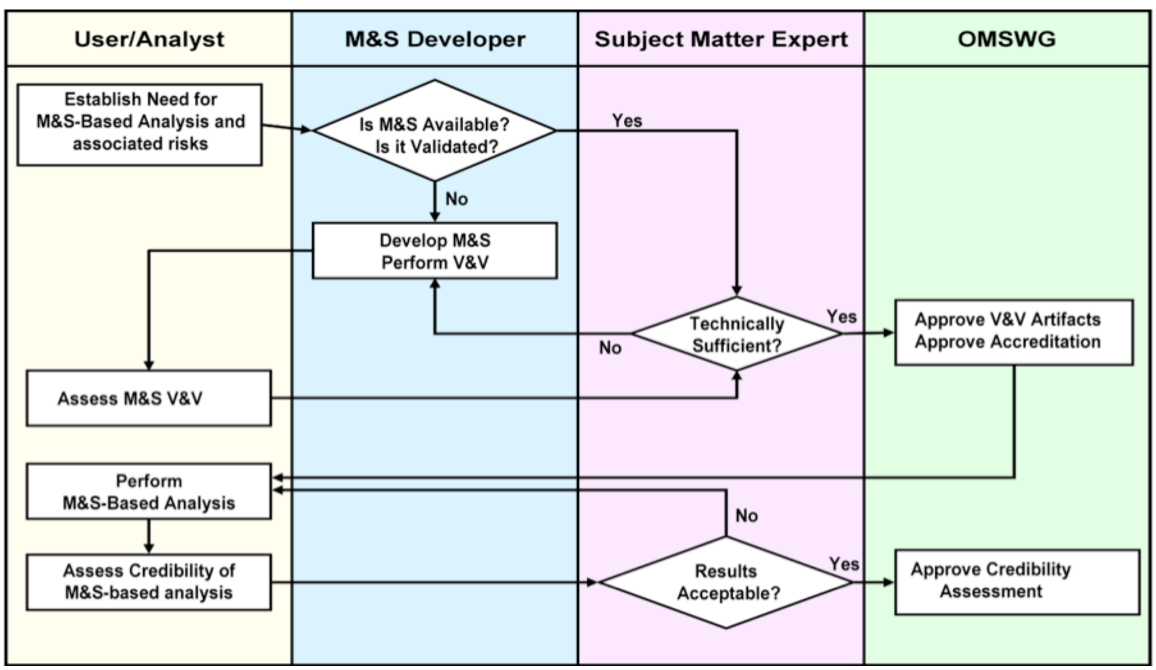


Figure 2: M&S life-cycle, from identification of need to final Credibility Assessment.

### 6.2 Examination of the artifacts

CAP Aerodynamic Database artifacts were examined to make sure there were no issues. The artifacts included approved baseline documents, databooks, user's guide, test case files, and V&V plans and reports. Included in these reports were the votes of the Subject Matter Experts (SMEs) on their assessment of the technical sufficiency of the new database.

### 6.3 V&V and Accreditation

The Orion Vehicle Integration Team (OVIT) and the Vehicle Integration Control Board (VICB) approved (accreditation) the use of the CAP Aerodynamic Database. The database may go through additional boards, so a list of any further steps leading to final approval was elicited as well. Note that the simple "accreditation" step, as shown in Figure 2, may be a more complex process in different organizations.

### 6.4 Credibility Assessment interviews

In addition to examining the artifacts relevant to all eight factors, members of the user/analyst team were provided with additional detailed insight on the 7009 standard, so that they could better understand the scope of the CA. Then they were interviewed and asked to respond to a questionnaire that been prepared in advance.

### 6.5 Credibility Assessment briefing

Based on all the evidence and documentation provided for the V&V and accreditation process, and for the analysis of results, as applicable to the CAP Aero database, a briefing was prepared for presentation to management and technical leads. The key decision makers were invited and a request was made for CA results to be communicated to all responsible parties. The assessment for each factor, prepared on the basis of artifacts and responses (see 6.4 above) was presented and compared with predetermined threshold values. The CA scores (CAS) are not presented here, as it is not within the scope of this paper to evaluate the CAP Aerodynamic Database. The point of the briefing was to show which factors were as strong as required and which were not, and to elicit appropriate discussions on the confidence that could be placed in the database. A separate "lessons learned" briefing was prepared for the OMSWG to asses the CA process itself.

## 7. Discussion and Lessons Learned

### 7.1 Cost benefit analysis

One of the key questions that is immediately raised every time a new procedural standard is created and applied is that of its cost effectiveness. Unfortunately this is also one of the most difficult questions to answer. Is the effort of process implementation and adherence worth the associated costs? We have seen examples of process standards like ISO, and, lately CMMI, where the same issue is still being debated. It typically takes years and careful studies to reach any conclusions even when, at least in principle, as with CMMI, it should be possible to measure productivity gains. In the case of M&S the typical driver is not productivity, rather missions success and astronaut safety. What value can one associate with these factors in order to determine cost effectiveness?

In light of unproven added value, the immediate reaction of interested parties is typically one of resistance. The introduction of the CAS hasn't been immune from these issues. For this reason the approach taken by the M&S team that was assigned the task to oversee adherence to the CA process has been one of proactive action, where the M&S development and management teams were simply asked to provide what they had done in whatever form they had chosen to do it, with no addition of special tasks except for occasional Q&A sessions when absolutely needed to understand the processes that were followed. The M&S process team has been asked to keep track of their time spent on this process and that of the development and management teams. This is the start of data collection that will be needed for a thorough evaluation, but consideration needs to be given to the fact that there is a learning curve for all involved. The assumption that guides much of this effort is that, once established, it would actually streamline processes that are in fact already being performed, but on an ad hoc basis. Streamlining and clear identification of steps also makes measurements possible. Initial feedback will be anecdotal at the beginning and become quantitative as experience is gained.

The question of cost-benefit analysis will need to address the time and resource taken up by the CA process. A realistic assessment of time and resources is not possible at this point because of our learning curve and because this initial effort was not integrated into the current work of users/analysts. Work in integrating the CA in current practices, including its use by management, is separate from the CAS process itself. Note that the first five factors are already normally addressed within standard V&V best practices, and the last three (history, management and personnel) would normally be explicitly or implicitly considered by management or decision makers. The power of the CAS process is to provide a scheme for *quantification* of these factors, and thus enable decision makers to also quantify and buy down, as needed, any risks involved with the use of a particular M&S. Our expectation is that the added clarity provided by the CAS approach will help streamline current processes and contribute to risk reduction.

### 7.2 Decision support and risk assessment

Some of the lessons learned in this initial process are rather typical for process improvement practices. Usage decisions that in an ideal world would have come after the process had in fact already been made, and it's not clear that the CA scores had impact on further decision making. They can however have impact on the assessment of risk of continued use of the database, and on future releases. With regard to risk assessment, the likely policy will be that only M&S deemed to be in a high risk zone (high reliance placed on them, and catastrophic consequences of failure) will be required to conduct a CA; on the other hand it is not yet clear how the CA will affect continuing risk assessments of the M&S. This needs to be the object of further study.

### 7.3 Artifacts and threshold values

As would be true for any external assessment with regard to some particular technical or organizational or audit need (as might be the case with a CMMI appraisal) the artifacts were not necessarily organized in a fashion that made it easy to identify and evaluate them in the context of the CA. Assignment of threshold values for each factor was also done on the basis of assumptions made by the M&S process team, as it would have required an understanding of the scoring criteria that was not yet a prerogative of the decision makers. This is of no great consequence at this point, since the purpose of the thresholds at this point was simply to generate discussion on the relative importance of each credibility factor for the CAP Aerodynamic database. Presentation of these factors with proposed thresholds could be viewed in fact as initial training in the use of the CA scores. What is most relevant is the fact that a CA was recommended for all CAP Aerodynamic database products.

## 8. Conclusions

The Credibility Assessment and Scoring system captures the essence of the NASA 7009 Standard in providing potentially crucial information to decision makers on the credibility of Modeling and Simulation results with respect to intended use. As with all new process standards, actual implementation will undoubtedly result in recommendations for process changes, and we are just beginning to gather the experience that will be necessary for further improvement. The experience with the CAP Aerodynamic database indicates that the time spent on the process is reasonable and that this will be an important new tool for risk analysis and decision making. For the CAS to be widely adopted, however, both user/developers and decision makers will need to become better acquainted with both the requirements of the process and its potential impact on future missions. The CAS has already been recommended for future releases of the CAP Aerodynamic database and for the Guidance, Navigation and Control Low Resolution System (GN&C LRS) Chutes.

## 9. References

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

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

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

[4] Steele, M. J. (2007). The NASA Standard for Models and Simulations. Proceedings of the 2007 Summer Computer Simulation Conference, San Diego, CA, July 15 – 18, 2007, 5 – 10.

[5] Babula, M., Bertch, W. J., Green, L. L., Hale, J. P., Mosier, G. E., Steele, M. J., Woods, J. NASA Standard for Models and Simulations: Credibility Assessment Scale (AIAA 2009-1011). In 47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition, 5 – 8 January 2009, Orlando, FL.

[6] Steele, M. J. (2008). Dimensions of Credibility in Models and Simulations. Proceedings of the 2008 Summer Computer Simulation Conference, Edinburgh, Scotland, July 2008.

## Author Biographies

**SILVANO COLOMBANO** received an M.A. in Physics and a Ph.D. in Biophysical Sciences from the State University of New York at

Buffalo. He has spent most of his working career at NASA-Ames Research Center first as a researcher in Closed Ecological Life Support Systems (CELSS) simulations and later in Artificial Intelligence and Software Engineering. He has been working in support of the Constellation and Orion programs with different software and systems engineering responsibilities since 2006.

**WEI LIN** is a Software Systems Engineer in the Systems Engineering Technical Area at NASA Ames Research Center. She is responsible for Verification, Validation, and Accreditation (VV&A) in the Orion Level III Modeling & Simulation Office.

**MICHAEL LOWRY** is the NASA chief scientist for Reliable Software Engineering. After receiving his BS/MSfrom MIT and PhD from Stanford, all in computer science, he joined the Kestrel Institute as PI working on program synthesis. In 1993 he joined NASA Ames as group lead then area lead, and was promoted to chief scientist in 2008. Dr. Lowry is the editor of MIT Press "Automating Software Design" and serves on the editorial board of the journal Automated Software Engineering. He has published numerous papers principally on the topics of program synthesis and software V&V. He is currently the software production tools lead for NASA Orion, as well as the PI for NASA's research in advanced software engineering for exploration systems.