INSTITUTE FOR DEFENSE ANALYSES

Quantifying Uncertainty to Keep Astronauts and Warfighters Safe

John T. Haman, Project Leader

John W. Dennis III
James E. Warner

OED Draft

June 2024

Distribution Statement A.

Approved for public release:
distribution is unlimited

IDA Product 3001907

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task BD-9-2299(90), "Methods Development," for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The IDA Technical Review Committee was chaired by Dr. Heather M. Wojton and consisted of Dr. Daniel Pechkis, Dr. Adam Miller, Dr. Gregory Chesterton, and Dr. Kelly Avery from the Operational Evaluation Division and Dr. David Tate from the Cost Analysis and Research Division.

For more information:
Dr. John T. Haman, Project Leader
jhaman@ida.org • 703/845-2132

Dr. Heather M. Wojton, Director, Operational Evaluation Division
hwojton@ida.org • (703) 845-6811

Copyright Notice
© 2024 Institute for Defense Analyses
730 East Glebe Road, Alexandria, Virginia 22305 • (703) 845-2000

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 [Feb. 2014].

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Product 3001907

Quantifying Uncertainty to Keep Astronauts and Warfighters Safe

John T. Haman, Project Leader

John W. Dennis III
James E. Warner

Executive Summary

Since February 2014, an interagency agreement between the Director, Operational Test and Evaluation (DOT&E) in the Office of the Secretary of Defense and the Chief Engineer of NASA has supported collaboration on methodological improvements in test design and statistical engineering.

Both NASA and DOT&E increasingly rely on computer models to supplement data collection while employing statistical distributions to quantify the uncertainty in these models, in part, to ensure that decision-makers are equipped with the most accurate information about system performance and model fitness.

This article provides a high-level overview of uncertainty quantification (UQ) using a hypothetical example of an assessment of the reliability of a new spacesuit system. Our intention with this article is to convey for the general readership of *Significance Magazine* the importance and relevance of statistics for the defense and aerospace communities.

Quantifying Uncertainty to Keep Astronauts and Warfighters Safe

Consider an astronaut exploring the cratered lunar surface surrounded by a vast and desolate gray landscape and protected from the unforgiving vacuum and lethal radiation by a shield of synthetic polymers. The brilliant blue light of Earth in the distance piercing the vast stillness is a stark and constant reminder that home is hundreds of thousands of miles away and that, should anything go wrong, the options for remedy will be limited and the consequences of failure deadly.

Astronauts are among the most daring and significant adventurers of our time. A staggering level of resources and research go into the planning, engineering, and testing required to launch them into space, ensure their safety, and get them home. An astronaut must be assured that, should an accident take place, such as an elbow scraping against a panel during a repair or a tumble onto the lunar surface during a moonwalk, the spacesuit will continue to function properly. The basis of this assurance is the rigorous engineering and statistical approaches that engineers and scientists employ in the design and evaluation of an astronauts' equipment.

Nevertheless, NASA and DOD both face problems related to certifying the safety and effectiveness of hard-to-test systems – systems for which the operational conditions are challenging to emulate. To maximize test information and control risks, NASA and DOD collaborate on methodological improvements and the annual Defense and Aerospace Test and Analysis Workshop (DATAWorks) to improve the statistical rigor of our assessments.

So, what goes on behind the scenes? Spacesuits undergo rigorous testing and analysis to ensure they meet acceptable quality standards. Their design and performance must strike a delicate balance between the competing objectives of being lightweight and flexible enough to allow astronauts to complete their missions and being sturdy and robust enough to protect them from leakage or pressure loss. To ensure robustness, requirements are introduced in the certification process that are based on *reliability* – the probability that the suit will function without failure under operational conditions and for a typical mission duration. For example, the spacesuit design may be required to demonstrate that its reliability is 99.9 percent or better for 20 hours of lunar surface exploration.


NASA image

At NASA and DOD, verifying reliability requirements such as this one is a team effort involving engineers and statisticians alike. While NASA assesses the reliability of spacesuits, spacecraft, and other aerospace systems, DOD assesses the reliability of aircraft, missile-defense systems, earth-orbiting satellites, and other defense systems. For example, aircraft have a mission-reliability requirement – a probability that the system will perform all mission-essential functions for a period of time under specific conditions. Demonstrating the performance of systems against such requirements takes an extraordinary amount of planning and resources. So, teams at NASA and DOD have a shared interest in developing and applying statistical methods that can deliver accurate and robust analyses.

In an ideal world, reliability assessments would rely exclusively on physical testing. For example, spacesuit testers would collect impact data from operationally realistic experiments to estimate suit reliability. In practice, however, this is often infeasible. Using traditional methods to test whether a system satisfies a failure rate as low as 1 in 1,000 would require evaluating more than 10,000 independent 20-hour trials![1] An operationally realistic experimental setup would also involve damaging many spacesuits many times, either in space or in an environment that effectively mimics space – all of which would make for a challenging (not to mention expensive) proposition.

One alternative to such a daunting approach involves augmenting traditional experimentation with modeling and simulation. A practice sometimes referred to as *certification by analysis*, modeling and simulation is being actively pursued by the defense and aerospace industries. With this approach, we could theoretically use a computer model that simulates impact damage to the suit to conduct virtual experiments under operational conditions that would be either too dangerous, too expensive, or too challenging to physically reproduce, while (potentially) saving ourselves time and money in the process.

But should we rely on predictions from a computer model when human lives are at stake? If we acknowledge that *all models are wrong, but some are useful* (to paraphrase British statistician George Box), then the critical challenge we face would be in determining the degree to which the model is wrong and then, with that understood, determining whether the model would be useful for our reliability evaluation.

[1] Assuming a one-sided lower tailed test of H_0: $p >= 0.999$ vs H_1: $p < 0.999$ with type 1 and 2 error rates of 10% percent, an effect size of 0.001, and 14 allowable failures.

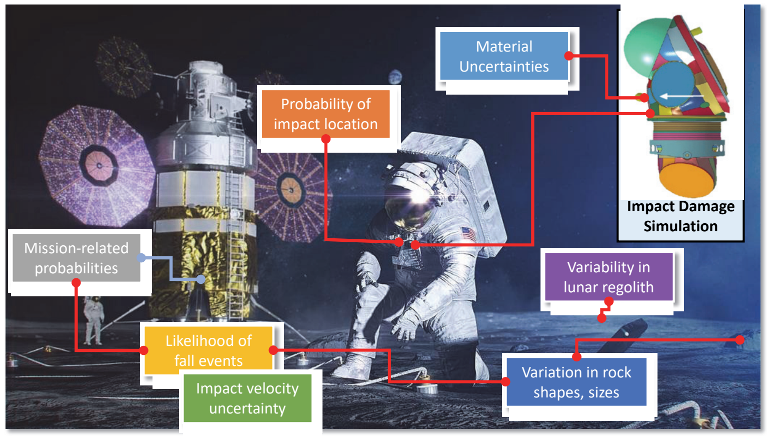
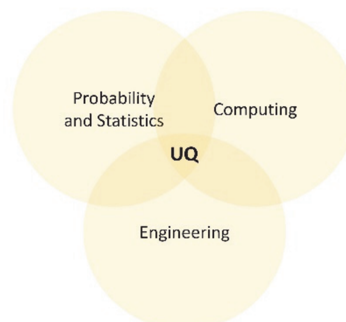


Figure 1. Factors Affecting Spacesuit Reliability (NASA image)

Spacesuit reliability is inherently probabilistic; many unknown factors and random events jointly contribute to the probability of suit failure. Properly assessing the probability of failure during lunar missions requires an understanding of these myriad, uncontrollable factors as well as the ways in which they may interact. What would the relative impact velocity be? What portion of the suit would be impacted? Should the astronaut fall, would the lunar regolith—the loose blanket of dust and rocks covering the bedrock—be sufficient to soften the blow? What would be the likelihood of the astronaut striking a rock? What would be the likely characteristics of that rock?



In modeling and simulation, uncertainty quantification[2] (UQ) provides us with tools we can use to assess the mismatch between model and reality and to quantitatively characterize real-world uncertainties and their impact on model predictions. The proper application of UQ requires a multidisciplinary team that includes statisticians to reason about uncertainties, engineers to develop physics-based models and conduct laboratory experiments, and, oftentimes, computer scientists to leverage high performance computing techniques for executing large numbers of computer experiments.

[2] According to American mathematician R.C. Smith, uncertainty quantification is "the science of identifying, quantifying, and reducing uncertainties associated with models, numerical algorithms, experiments, and predicted outcomes."

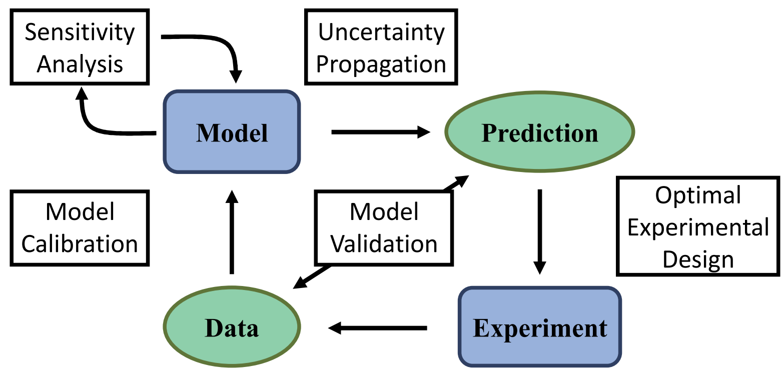


Figure 2. An Uncertainty Quantification Workflow (NASA image)[3]

With UQ, we can make real-world decisions in the face of the unknown by utilizing models and operational tests that each provide snapshots of reality through predictions and data, respectively, as shown in Figure 2. Included among the field's fundamental concepts are the following:

- *Uncertainty propagation* enables us to make probabilistic predictions using a sequence of probabilistic and deterministic models (given knowledge of input uncertainties).[4]
- *Optimal experimental design* can allow us to reduce uncertainty by designing tests that provide the maximum information gain about unknown parameters.
- *Model calibration* allows us to update the distributions of model parameters based on relevant test data.
- *Sensitivity analysis* allows us to understand the relative contribution of different input parameters to the total uncertainty in the model prediction.

At the core of the uncertainty quantification process, *model validation*[5] compares model predictions to experimental data to assess the credibility of the model (essentially characterizing how *wrong* the model is and whether it is *useful* for the scenario tested). In

---

3 Figure adapted with permission from the following: James E. Warner et al., "Assessing Next-Gen Spacesuit Reliability: A Probabilistic Analysis Case Study," (memorandum, NASA, Washington, DC, August 2021), https://ntrs.nasa.gov/api/citations/20210019495/downloads/NASA-TM-20210019495.pdf.

4 In a simple model, uncertainty propagation could be as simple as summing the variances of the constituent sub-scale models. However, in our work, we use Monte Carlo computational methods.

5 Model verification and validation is sometimes considered a subcomponent of the UQ field and sometimes as part of a combined topic, verification, validation, and uncertainty quantification (VVUQ). In defense research, verification and validation precede model accreditation as part of the verification, validation, and accreditation (VV&A) process.

principle, this procedure is straightforward: first we must conduct the validation experiment; then we can quantify the agreement between the simulated and experimental outcomes. In practice, however, it's rather more complicated because we must account for our inability to conduct testing in the ideal operationally realistic environments for validation. Fortunately, we can employ experimental designs to balance the realism of the test with constraints on time and budget.

Once the validation data are obtained, a researcher must face the difficult question: What is an acceptable level of disagreement between model and experiment? For M&S users and testers, statistical validation metrics with associated confidence levels are key for characterizing risks and making an informed decision about M&S adequacy.

As it can sometimes be infeasible to validate (or even evaluate) a full-scale model from a computational or experimental perspective, a common strategy is to focus instead on validating simpler sub-scale models and data. In this way, researchers can use the sub-scale components as building blocks to characterize uncertainty in system-level model parameters individually, thereby enabling validation and reliability analyses at the full-scale level.

When applying this approach to spacesuit reliability, for example, we can use sub-scale models and data to estimate a range of reasonable impact velocities, calculate variability in lunar regolith properties, characterize expected variation in lunar rock size and shape, and calibrate the empirical parameters of a material model used to predict impact damage.

When selecting sub-scale models to reduce complexity, a researcher must be mindful of the potential correlations that may exist in a complete, unified system. For example, would the astronaut be more likely to fall near areas where the rocks are irregularly sized and particularly jagged in shape, resulting in the hardest falls occurring near the sharpest rocks?

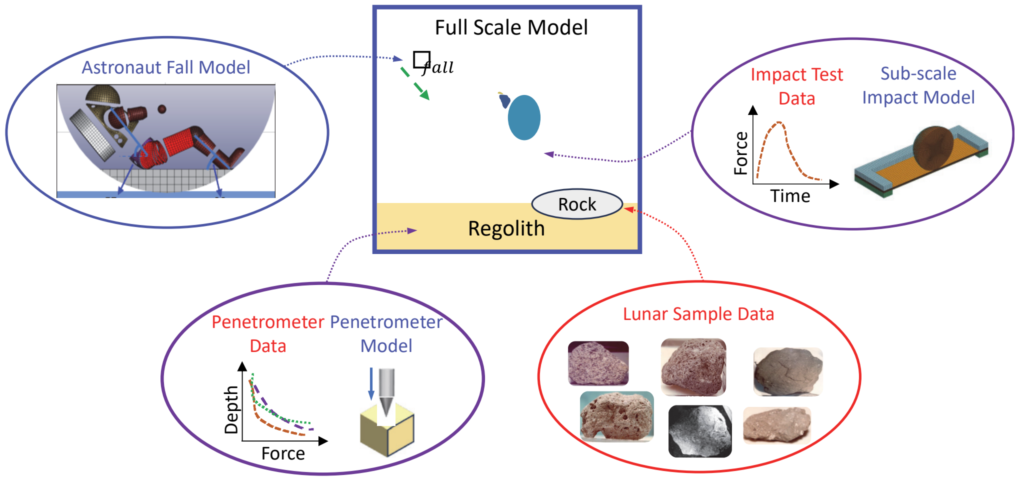


Figure 3. Building Blocks to the Full-scale Model of Spacesuit Reliability (NASA image)

Another important distinction for modeling is whether a high-fidelity, physics-based representation or a low-fidelity, effects-based representation is adopted. Compared to their high-fidelity counterparts, low-fidelity models provide valuable computational speedup but with reduced accuracy, introducing additional uncertainty into the analysis that must be quantified. Alternatively, a researcher may choose to employ models of varying fidelity in an iterative fashion, with a low-fidelity model learning to mimic a high-fidelity representation. However, the best approach is often one that combines models of varying degrees of fidelity, separates out components based on expert knowledge of the system, and iterates modeling with experimentation and analysis.

Regardless of the modeling approach taken, we ultimately need confidence in the results that drive critical decisions. This confidence comes from understanding the uncertainty in our estimates. Reliability estimates that are biased or highly uncertain may indicate more risk than implied by the estimates alone. Further, we must estimate this uncertainty based on variability in our data (aleatoric) and assumptions that we make about our models (epistemic). How we estimate parameter uncertainty may provide overly narrow or overly broad bounds on the estimated reliability.

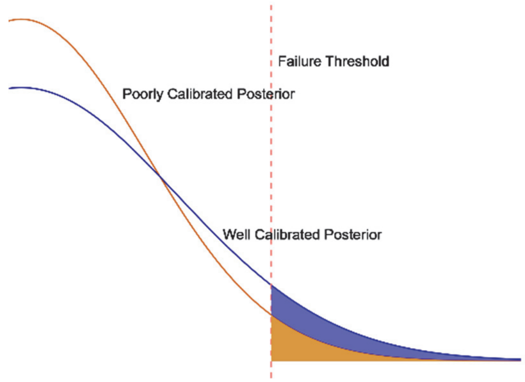


Figure 4. Example of Estimated Risk of Failure Being Lower than the True Risk of Failure

Rather than seeking the tightest bounds on uncertainty we should quantify uncertainty in a way that appropriately uses the available information. Specifically, we want our measure of uncertainty to encompass the values that are probable at a given level of confidence while nonetheless being as tight as possible. This is the well calibrated posterior of Figure 4. We can achieve a proper balance of uncertainty bounds using a UQ approach that judiciously combines modeling, testing, subject matter expertise, and rigorous statistical practices. In this way, we can achieve and deliver the meaningful insights that decision-makers require.

In the end, the spacesuit that shields our astronauts from the harsh lunar environment represents the culmination of extensive engineering and statistical analysis. Every detail of the suit design, from material selection to joint sizing, has been scrutinized to ensure that reliability requirements are met. Astronauts can therefore push the boundaries of human discovery with confidence knowing that a team of scientists, engineers, and statisticians back on Earth has leveraged the state-of-the-art in testing and analysis to certify their equipment.

To help keep both astronauts and warfighters safe, NASA and DOD collaborate on methodological improvements and knowledge exchanges to improve the rigor of our test design, data analysis, and uncertainty quantification. This article is one example of this collaboration.

Acknowledgments

This article was made possible by the NASA-DOT&E interagency agreement IA1-30831 and partly funded by the Strategic Initiatives, Policy and Emerging Technology division within the office of the Director, Operational Test and Evaluation (DOT&E).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
XX-05-2024

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Quantifying uncertainty to keep astronauts and warfighters safe

5a. CONTRACT NUMBER
HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jay Dennis, James Warner, and John Haman

5d. PROJECT NUMBER
BD-9-2299(90)

5e. TASK NUMBER
229990

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
730 East Glebe Road
Alexandria, Virginia 22305

8. PERFORMING ORGANIZATION REPORT NUMBER
3001907

3001907

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)
DOT&E

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release: distribution is unlimited

13. SUPPLEMENTARY NOTES
Project Leader: John Haman

14. ABSTRACT
Both NASA and DOT&E increasingly rely on computer models to supplement data collection, and utilize statistical distributions to quantify the uncertainty in models, so that decision-makers are equipped with the most accurate information about system performance and model fitness. This article provides a high-level overview of uncertainty quantification (UQ) through an example assessment for the reliability of a new space-suit system. The goal is to reach a more general audience in Significance Magazine, and convey the importance and relevance of statistics to the defense and aerospace communities.

15. SUBJECT TERMS
Uncertainty Quantification, Modeling and Simulation, Statistics, Reliability

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	SAR	12	John Haman
					19b. TELEPHONE NUMBER *(Include area code)* (703) 845-2132