SYM-AM-23-099

Excerpt from the Proceedings of the Twentieth Annual Acquisition Research Symposium

Acquisition Research:
Creating Synergy for Informed Change

May 10–11, 2023

Published: April 30, 2023

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

Towards an Enterprise All-Domain M&S Environment for T&E: Overcoming M&S Challenges Within the DoD

Jeremy Werner—was appointed DOT&E's Chief Scientist in December 2021 after initially starting at DOT&E as an Action Officer for Naval Warfare in August 2021. Before then, Werner founded a data science-oriented military operations research team at JHU/APL that transformed the analytics of an ongoing military mission. Werner previously served at IDA supporting DOT&E in the assessment of a variety of systems. Werner earned a PhD in physics from Princeton University where he was an integral contributor to the Compact Muon Solenoid collaboration in the experimental discovery of the Higgs boson at the Large Hadron Collider, CERN, Geneva, Switzerland. [jeremy.s.werner.civ@mail.mil]

Abstract

Our earlier paper Data Driven Modeling and Simulation to Test the Internet of War Things described how (Werner, 2023):

The Director, Operational Test and Evaluation's Strategic Initiatives, Policy, and Emerging Technologies division (DOT&E SIPET) is shaping the test and evaluation (T&E) of future multi-domain warfighting.

and how:

Comprehensive Live testing of multi-domain capabilities currently under development is not possible due to environmental, fiscal, safety, classification, and ethical constraints, and so our evaluations will become more dependent on modeling and simulation (M&S) to test the efficacy and interoperability of our systems.

That paper was written for a general audience but still explored the following technical challenges:

- Architecting M&S and live tests to engender a "predict, live test, refine" feedback loop to improve M&S accuracy over systems' life cycles
- Ensuring integration across all warfighting domains and digital capabilities
- Fielding M&S as a service so that the skillset required to operate it and understand its outputs mirrors the skills required of warfighters in the real world
- Implementing an environment with real-time analysis and accurate results for T&E and operational decisions

The present paper complements the earlier one by targeting a DoD technical audience and addressing several more challenges we see within the DoD:

- Using M&S to credibly extrapolate outside of the operational envelope covered in live test
- Rigorous life cycle approaches to V&V that are centered around quantitative estimates of uncertainty
- Accelerating M&S processing times
- The risk that we do all of this rigorous work and our models still turn out to be wrong
- Providing policy, guidance, best practices, executable examples, and training an M&S V&V/Uncertainty Quantification (UQ) workforce within the DoD.

that we must overcome to outpace our adversaries' capabilities.

Introduction

The need to maximally leverage modern M&S solutions for the T&E of multi-domain warfighting capabilities is manifest, as comprehensive live testing of these capabilities is not possible due to environmental, fiscal, safety, classification, and ethical constraints. Nevertheless, we must overcome a multitude of challenges to most effectively utilize M&S

for the T&E of these capabilities. Our earlier article *Data Driven Modeling and Simulation to Test the Internet of War Things* discussed means for overcoming several of these challenges while being addressed to a general audience; the article at hand complements the earlier one by targeting a DoD technical audience to overcome several more challenges we see within our department:

- Providing policy, guidance, best practices, executable examples, and training an M&S V&V workforce within the DoD
- Using M&S to credibly extrapolate outside of the operational envelope covered in live test
- Rigorous life cycle approaches to V&V that are centered around quantitative estimates of uncertainty
- Accelerating M&S processing times
- The risk that we do all of this rigorous work and our models still turn out to be wrong
- Providing policy, guidance, best practices, executable examples, and training an M&S V&V/Uncertainty Quantification (UQ) workforce within the DoD.

Using M&S to Credibly Extrapolate Outside of the Operational Envelope Covered in Live Test

Deducing models that accurately characterize phenomena far beyond and more generally than just those limited input observations from which they were originally derived has been a cornerstone of science, engineering, and technology for millennia. The Antikythera mechanism built by the ancient Greeks is the earliest known mechanical computer and was able to accurately predict astronomical positions and eclipses decades in advance. Isaac Newton was able to glean the mechanics underlying the motion of all bodies in the universe from the limited number of terrestrial-based observations available to him (that is, at least, until one looks at either the quantum or relativistic regimes—more on that later).

In general, the fact that a small handful of physical laws can be used to accurately characterize and predict phenomena across a vast—even infinite—set of input conditions is a crowning achievement of science. And one with strong implications for T&E since it means:

> It is possible to apply firmly-established physical laws joined to a limited number of live test observations to credibly assess system performance in regions of the operational envelope not directly covered in test.

National Aeronautics and Space Administration (NASA) Example[1]

NASA Langley's 14'×22' subsonic wind tunnel provides an excellent example. The wind tunnel is used to assess conventional performance for low-speed tests of powered and unpowered models of various fixed- and rotary-wing civil and military aircraft over a wide range of takeoff, landing, cruise, and high-angle-of-attack conditions. The 14'×22' wind tunnel is ideally suited for low-speed tests to determine high-lift stability and control, aerodynamic performance, rotorcraft acoustics, turboprop performance, and basic-wake and flow-field surveys (NASA Aeronautics Test Program, 2009). Small-scale models of aircraft are tested in the tunnel and the results then scaled to the full-size platforms using a thoroughly vetted and continuously-validated computational fluid dynamics (CFD) model.

[1] Special thanks to James Warner, NASA, for his help in formulating this example.

The ongoing feedback loop between live data and the CFD model is the key to success: Wind tunnel data is used to both calibrate and validate the model, while the model can then be used to provide accurate results scaled to the full-size platform as well as help identify important design constraints, air flows, and the like which in turn can then be tested in the wind tunnel. Of course, once the full-scale platform is eventually built and tested, then it's measured performance characteristics can also be compared to the design predictions and used to refine the prediction process as needed.

National Nuclear Security Administration (NNSA) Example[2]

The Department of Energy's (DOE) NNSA Office of Defense Projects mission is to maintain a safe, secure, and effective nuclear weapons stockpile for our nation; its Advanced Simulation and Computing (ASC) program is not only vital to this mission but provides another excellent example for the DoD to follow. The ASC has its origins in the nation's ongoing need to maintain and assess the readiness of our stockpile following the discontinuation of live, underground nuclear explosion tests in the 1990s. Over the span of nearly three decades, the ASC has successfully developed and validated advanced simulation capabilities based on well-known physics of magnetohydrodynamics, inertial confinement fusion, structural dynamics, etc.—and then calibrated and validated them against current experiments and historical data from live underground testing—to credibly assure our nuclear weapon stockpile in the absence of live nuclear explosion tests.

> We have provided both tactical (NASA) and strategic (NNSA) examples from our partner government agencies outside of the DoD showing that it is possible to apply firmly-established physical laws joined to a limited number of live test observations to credibly assess system performance in regions of the operational envelope not directly covered in test. DOT&E will seek collaboration and knowledge exchange with NASA and the NNSA, including national labs such as Sandia, Los Alamos, and Lawrence Livermore, so that we may adapt their M&S V&V methods to our mission.

Rigorous Life Cycle Approaches to V&V that are Centered Around Quantitative Estimates of Uncertainty

Our previous article discussed architecting M&S and live tests to engender a "predict, live test, refine" feedback loop to improve M&S accuracy over systems' life cycles. This is aligned with DoD policy; in particular DODI 5001.61 (DoD, 2009) describes how it is DoD policy that, "Models, simulations, and associated data used to support DoD processes, products, and decisions shall undergo verification and validation (V&V) throughout their life cycles."

Furthermore, the National Academies 2012 report *Assessing the Reliability of Complex Models* (National Academies, 2012) recognizes "the ubiquity of uncertainty in computational estimates of reality and the necessity for its quantification."

That report also contains a treasure trove of examples and methods for our DoD M&S community to learn from and apply.

We agree with the National Academies on the criticality of uncertainty quantification (UQ) and view UQ as the principal essence of rigorous V&V. After all, even simple measurements are subject to imperfections caused by stochastic variation, calibration

---

[2] Special thanks to Thuc Hoang, Director, Office of Advanced Simulation and Computing, NNSA for reviewing this example and all other references to NNSA throughout this article.

tolerances, etc., and so all measurement results must be associated with an uncertainty estimating these effects.

DOT&E Example

The 2022 DOT&E article *Uncertainty Analysis Demonstration: A Missile Case Study* (Werner, 2022) provides the reader a primer on uncertainty, including a succinct description of the different types of uncertainty and their causes brought to life using the measurement of the mean weight of basketballs approved for NBA games as an example. From there, the article presents a case study of a notional missile performance analysis to demonstrate how statistical uncertainty in live test data—due to the limited number of samples from which it was generated—can be reinterpreted as a systematic uncertainty in the simulation. The article is available on DOT&E's website and includes a link to download the executable code to reproduce all of the charts and findings it contains with a single command; in this way, the study is packaged so that it can be easily used and adapted to our community's applications.

NNSA Example

The National Nuclear Security Administration's (NNSA) 2022 Advanced Simulation and Computing (ASC) Simulation Strategy (Etim, 2022) delves into the ASC's mature uncertainty quantification capabilities and describes their uncertainty and margins framework. Central to the ASC's strategy is "addressing the demand of uncertainty quantification efforts being performed routinely and more quickly by making them more user friendly and more easily integrated into daily practice;" the need for the DoD to do the same is apparent.

> It is DoD policy that Models, simulations, and associated data used to support DoD processes, products, and decisions shall undergo V&V throughout their life cycles. The National Academies' 2012 report Assessing the Reliability of Complex Models recognizes the ubiquity of uncertainty in computational estimates and the necessity for its quantification, while providing a treasure trove examples and methods for our DoD M&S community to learn from and apply. DOT&E's 2022 article Uncertainty Analysis Demonstration: A Missile Case Study provides a primer on uncertainty and a case study demonstrating uncertainty quantification that is packaged with the executable code for our community to use and adapt. The NNSA's 2022 Advanced Simulation and Computing Simulation Strategy delves into their mature uncertainty quantification capabilities and framework. That strategy recognizes the need to integrate uncertainty quantification into their analysis workflows in a more routine and user friendlier way; although our community's uncertainty quantification capabilities are less mature than the NNSA's, the need for us to do the same is apparent.

Accelerating M&S Processing Times at the Hardware and Tactical Performance Levels

Achieving Machine Speeds Using Hardware

The fact that custom tactical hardware and their associated integrated circuit boards such as Field Programmable Gate Arrays (FPGA) and Application-Specific Integrated Circuits (ASIC) are much faster for their dedicated purposes than more generalized computer processors such as Central Processing Units (CPU) and Graphical Processing Units (CPU) presents a technical challenge to the development of an enterprise all-domain

M&S environment for T&E. This can be easily understood by highlighting that the duration between input data and output response from a given logical algorithm can be shorter than 1 microsecond when implemented on an FPGA but may take 50 microseconds or more on a CPU (van der Ploeg, 2018).

These speed constraints are fundamental to the device architectures; no amount of CPU/GPU parallelization or clever software engineering can overcome them. (After all, that's why our tactical platforms have integrated circuit boards in the first place.) Nevertheless, the need for achieving hardware speed in M&S is paramount to synchronously include operators in the loop and other complicating, real-world tactical effects. Therefore, achieving performant M&S at the low levels and high fidelity of hardware necessitates that CPU/GPU-based M&S capabilities be augmented with tactical hardware in the loop (HWIL). From there, these HWIL-integrated capabilities can be used to feed machine learning and other advanced methods to generate reduced high-level tactical performance models that can run in real time.

Scaling Across the Enterprise Using Parallelized and Distributed Computing

Fortunately, modern CPU- and GPU-based computing architectures do not present any fundamental limitations to the development of an enterprise all-domain M&S solution beyond the low level of hardware. Alternatively, modern computing architectures provide a wealth of solutions for us to exploit. CPUs, of course, are the primary engines of computer processing and modern hardware and software solutions enable CPU computations to be parallelized and scaled across vast numbers of cores; the large computational workloads and high throughput required of M&S can be accelerated by sharing the workload across a large number of CPU cores operating in parallel. Additionally, GPUs are natively massively parallelized and are faster and more energy efficient than CPUs for a variety of computational workloads; many industrial data centers are currently shifting their infrastructures to include more GPUs for these reasons. GPUs are particularly efficient at the rendering used in a variety of M&S solutions, and have enabled real time 3D ray tracing in video games and the production of movies that entirely use photo-realistic Computer Generated Imagery such as Disney's 2019 *The Lion King*. Finally, both CPU- and GPU-based parallelized computing capabilities can be distributed across large geographical areas and have disparate, asynchronous workflows integrated and brought into harmony using modern, enterprise software architectural solutions such as RESTful APIs (Amazon, n.d.).

> Achieving performant M&S at the low levels and high fidelity of hardware necessitates that CPU/GPU-based M&S capabilities be augmented with tactical hardware in the loop (HWIL). From there, these HWIL-integrated capabilities can be used to feed machine learning and other advanced methods to generate reduced high-level tactical performance models that can run in real time. Beyond the low level of hardware, the large computational workloads and high throughput required of M&S can be accelerated by sharing the workload across a large number of CPU and GPU cores operating in parallel. Furthermore, both CPU- and GPU-based parallelized computing capabilities can be distributed across large geographical areas and have disparate, asynchronous workflows integrated and brought into harmony using modern, enterprise software architectural solutions such as RESTful APIs.

The Risk That We Do All of this Rigorous Work and Our Models Still Turn Out to be Wrong

It took humanity until the 1680s to consolidate all of our observations dating from antiquity into three fundamental laws that describe the mechanics underlying the motion of all bodies in the universe. Or so we thought. Around 1890 hints of Newton's laws breaking down towards the speed of light began to appear. Then around 1900 more hints of Newton's laws breaking down—this time at microscopic distance scales—began to appear. But by then the scientific method had been fully institutionalized. Einstein's theory of relativity solved the speed of light problem in 1905, effectively extending our foundational laws of mechanics to their light speed limit. By the mid-1920s quantum mechanics had been fully formulated by Niels Bohr, Erwin Schrodinger, and a cast of many others to extend our understanding of mechanics to the microscopic scale.

The institutionalization of the scientific method meant that it only took a couple of decades to mend cracks in physical laws that took millennia to formulate in the first place. And then of course it took just a couple of more decades to realize the weaponization of these new physics through the advent of nuclear weapons and propulsion.

Now let's return to the topic at hand: Using M&S for the rigorous T&E of our military systems and future joint warfighting concepts, with the ultimate goal of reducing the risk posed to our warfighters. Warfare is wrought with risk and uncertainty; real-world combat data may prove our models wrong despite our best efforts. But by design—by institutionalizing the rigor, mechanisms, and processes discussed in this article, its preceding one, and many related efforts to follow—we will have most thoroughly prepared ourselves for this exact eventuality. Just as the scientists of the early 20th century were well-prepared to reformulate physics for entirely new and unexpected regimes at an incredibly rapid pace, so too must our community be prepared to quickly ingest data from all venues—including real-world combat operations—to rapidly adapt our models. But to truly be prepared for these critical risks, we need to move out with agility now to build and stress this enterprise all-domain M&S plumbing. The discoveries we make along the way could be surprising and profound for warfare; after all, they have been before.

> Warfare is wrought with risk and uncertainty; real-world combat data may prove our models wrong despite our best efforts. The institutionalization of the scientific method meant that it only took a couple of decades to mend cracks in physical laws that took millennia to formulate in the first place, and this eventually led to nuclear weapons and propulsion. By institutionalizing the rigor, mechanisms, and processes discussed in this article, its preceding one, and many related efforts to follow, our community will be best prepared to quickly ingest data from all venues—including real-world combat operations—to rapidly adapt our models as needed. To truly be prepared for these critical risks, we need to move out with agility now to build and stress this enterprise all-domain M&S environment. The discoveries we make along the way could be surprising and profound for warfare; after all, they have been before.

Providing Policy, Guidance, Best Practices, Examples, and Training an M&S V&V/Uncertainty Quantification (UQ) Workforce Within the DoD

DOT&E is taking action to provide updated policy, guidance, best practices, and executable examples for M&S V&V and UQ:

- We are in the process of updating our 2016 and 2017 policy memoranda (DOT&E, 2016, 2017) on M&S V&V and consolidating them into a single, expanded DoD Manual (DODM) that will be released later this year.
- Later this year, we will start developing a M&S V&V UQ Companion Guide that will include a wide array of best practices and examples; these examples will each be provided along with the executable code to fully reproduce them with a single command, inclusive of all charts, tables, and numerical results; in this way, each example will be packaged in such a way that it can be easily used and adapted to our community's applications. Many of these best practices and examples can already be found in IDA's *2019 Handbook on Statistical Design & Analysis Techniques for Modeling & Simulation Validation*, which also describes a rigorous methodology for planning tests that utilize both M&S and live test data for evaluation[3].

We further recognize the need to train a dedicated M&S V&V UQ workforce and to learn from V&V UQ communities outside of our department—NASA and the DOE NNSA enterprise for nuclear weapons stockpile assurance, in particular:

- DOT&E will seek collaboration and knowledge exchange with NASA and the NNSA, including national labs such as Sandia, Los Alamos, and Lawrence Livermore, so that we may adapt their M&S V&V UQ methods to our mission.
- DOT&E will pursue a training curriculum targeted at creating a dedicated M&S V&V UQ workforce for the T&E enterprise in conjunction with the Under Secretary of Defense for Research and Engineering and the Services.

  Preparing our nation for the next generation of T&E capabilities means providing updated M&S policy, guidance, best practices, executable examples, and the training of a dedicated M&S V&V UQ workforce; DOT&E is taking a number of actions to fulfill these needs and will pursue a training curriculum to create a dedicated M&S V&V UQ workforce.

Conclusion

DOT&E's SIPET division—Strategic Initiatives, Policy, and Emerging Technologies—is shaping the T&E of future multi-domain warfighting. We understand that comprehensive Live testing of multi-domain capabilities currently under development is not possible due to environmental, fiscal, safety, classification, and ethical constraints and so our evaluations will become more dependent on modeling and simulation (M&S) to test the efficacy and interoperability of our systems. Our earlier article *Data Driven Modeling and Simulation to Test the Internet of War Things* discussed means for overcoming the following M&S challenges while being addressed to a general audience:

- Architecting M&S and live tests to engender a "predict, live test, refine" feedback loop to improve M&S accuracy over systems' life cycles
- Ensuring integration across all warfighting domains and digital capabilities
- Fielding M&S as a service so that the skillset required to operate it and understand its outputs mirrors the skills required of warfighters in the real world
- Implementing an environment with real-time analysis and accurate results for T&E and operational decisions.

[3] Many of these best practices and examples can already be found in Wojton (2019).

The article at hand complemented the earlier one by targeting a DoD technical audience to overcome several more challenges we see within our department:

- Using M&S to credibly extrapolate outside of the operational envelope covered in live test
- Rigorous life cycle approaches to V&V that are centered around quantitative estimates of uncertainty
- Accelerating M&S processing times
- The risk that we do all of this rigorous work and our models still turn out to be wrong
- Providing policy, guidance, best practices, executable examples, and training an M&S V&V/UQ workforce within the DoD.

The exposition of these challenges was not academic; alternatively, the discussion was pragmatic and centered around already-mature or rapidly maturing technologies, advanced methods, and real-world use cases pertinent to the M&S required for the T&E of future joint warfighting concepts. DOT&E and our partners will soon have multiple R&D projects underway to advance our M&S and V&V/UQ capabilities for T&E and position us to meet this critical challenge; we have released our S&T strategy to the public (https://www.dote.osd.mil/News/News-Display/Article/3118739/dote-strategy-update-2022/) as well as our detailed implementation plan.

We will do our part by transforming M&S and its V&V/UQ for T&E to enable delivery of the world's most advanced warfighting capabilities at the speed of need. We seek your proposals to collaborate with us.

References

Amazon. (n.d.). *What is a RESTful API?.* https://aws.amazon.com/what-is/restful-api/#:~:text=RESTful%20API%20is%20an%20interface,applications%20to%20perform%20various%20tasks

DoD. (2009). *DODI 5000.61, incorporating a change from 2018, DoD modeling and simulation (M&S) verification, validation, and accreditation (VV&A).*

DOT&E. (2016). *Guidance on the validation of models and simulation used in operational test and live fire assessments.* https://www.dote.osd.mil/Portals/97/pub/policies/2016/20140314_Guidance_on_Valid_of_Mod_Sim_used_in_OT_and_LF_Assess_(10601).pdf?ver=2019-08-19-144201-107

DOT&E, 2017, Clarifications on Guidance on the Validation of Models and Simulation used in Operational Test and Live Fire Assessments, https://www.dote.osd.mil/Portals/97/pub/policies/2017/20170117_Clarification_on_Guidance_on_the_Validation_of_ModSim_used_in_OT_and_LF_Assess(15520).pdf?ver=2019-08-19-144121-123

Etim. (2022). *Advanced computing and simulation, NNSA, LLNL, simulation strategy.* https://asc.llnl.gov/file-download/download/public/2731

National Academies. (2012). *Assessing the reliability of complex models.* https://nap.nationalacademies.org/catalog/13395/assessing-the-reliability-of-complex-models-mathematical-and-statistical-foundations

NASA Aeronautics Test Program. (n.d.). *14- by 22-Foot Subsonic Tunnel.* https://www.nasa.gov/sites/default/files/atoms/files/m187003_14_22print_508.pdf

Stevens. (2022). *25 years of accomplishments.* Advanced Simulation Program, National Nuclear Security Administration. https://www.energy.gov/sites/default/files/2022-10/20221013%20ASC%2025-year%20Accomplishments%20Report.pdf

van der Ploeg. (2018). *Why use an FPGA instead of a CPU or GPU?.* https://blog.esciencecenter.nl/why-use-an-fpga-instead-of-a-cpu-or-gpu-b234cd4f309c

Werner. (2022). *Uncertainty analysis demonstration: A missile case study.* DOT&E. https://www.dote.osd.mil/News/What-DOT-Es-Saying/Saying-Display/Article/3254156/uncertainty-analysis-demonstration-a-missile-case-study/

Werner. (2023). Data driven modeling and simulation to test the internet of war things. *MODSIM World Proceedings.*

Wojton. (2019). *Handbook on statistical design & analysis techniques for modeling & simulation validation.* https://apps.dtic.mil/sti/pdfs/AD1122387.pdf

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net