

Research Issues in Verification and Validation from Foundations '02¹

D. E. Stevenson, Department of Computer Science, Clemson University, Clemson, SC 29634-0974
steve@cs.clemson.edu

ABSTRACT

The Foundations '02: Foundations for V&V in the 21st Century was held at Johns Hopkins Applied Physics Laboratory on 22-23 October 2002. I review the research findings from that conference.

Keywords

Verification, validation.

1. INTRODUCTION

Foundations '02 was a two day workshop held at Johns Hopkins University's Applied Physics Laboratory on 22-23 October 2002. There were 198 attendees from many different M&S communities, providing both breadth and depth to the discussions. The distribution of attendees was approximately:

- 40% from the U.S. Defense community.
- 15% from other U.S. government organizations.
- 25% from academia.
- 10% from other industry organizations
- 10% from outside the U.S.

The diversity of participants made a substantial beginning improve communications about V&V across community boundaries, facilitating the sharing of insights and best practices from outside one's immediate associations. Moreover, the ability to look at M&S V&V practices from such a broad spectrum of M&S experience allowed a more comprehensive assessment of the state of the V&V art than would have been possible otherwise.

These attendees were offered 19 different sessions to attend. These sessions covered most of the topics of interest in V&V. At the end of each conference day, the attendees were asked to share their ideas about the sessions they attended on that day. At the end of the conference, the organizers provided plenary sessions on the conference as a whole and research topics in specific. The purpose of this paper is to provide the broader community with a condensed version of the research topics. In section 2, we paraphrase the contents of the Executive Summary of the Proceedings. The Proceedings are available from the Society for Computer Simulation or from two on-line sites: The Defense Modeling and Simulation Agency site at <http://www.dmsomil> and a mirror site at Clemson University at <http://www.cs.clemson.edu/~found04>. The Clemson site is also the site for the follow on workshop, *Foundations '04*.

2. INITIAL STATEMENT OF FINDINGS

Each Workshop day ended with a plenary session in which attendees were invited to comment on the session(s) they had attended. One recurring question was, "how much V&V is enough?" It remains impossible to quantify, either technically or managerially, how many resources must be allocated to these tasks. This section paraphrases the Executive Summary.

2.1 Global Ideas

The recurring themes from the plenary sessions constitute the core of initial findings.

- The primary motivation for M&S V&V is risk reduction providing the benefit

¹ This work is partially funded by the Shodor Computational Science and Engineering Fund, Shodor Education Foundation, Durham, NC, and NSF grant DUE -0127488.

side for cost-benefit concerns about V&V.

- Effective communication remains a problem because of differences terminology, concepts, and V&V paradigms among various M&S communities. This is exacerbated by excessive use of acronyms making cross culture communications difficult.
- Advances in M&S framework/theory can enhance V&V capabilities, and is essential for increasing automated V&V techniques.
- Limitation in the availability of required data and detailed characterization of associated uncertainties and errors, simulation/software artifacts, etc required for effective V&V have to be faced and addressed.
- Cost and resource requirements for M&S V&V are not as well understood as they need to be because meaningful information about such is not widely shared within M&S communities, and much more information about cost and resource requirements needs to be collected and made available to facilitate development of more reliable estimation processes.
- Many areas of M&S V&V need to employ more formal (repeatable and rigorous) methods to facilitate better judgments about appropriateness of simulation capabilities for intended uses.

2.2 Research Issues

A synopsis of M&S V&V research needs issue 3 above) was developed from papers prepared for Foundations '02, their discussion, and suggestions from Foundations '02 participants. The M&S V&V community is faced by two very different kinds of challenges:

1. Managerial challenges: How to do what we know how to do.
2. Research challenges: What areas that must we understand better in order to find viable technical solutions.

2.2.1 Management Challenges.

Three implementation challenges were cited: **qualitative assessment**, appropriate and effective **use of formal assessment processes**, and M&S/V&V **costs/resources** (accounting, estimation, benefit). The challenge is how to ensure that “best practices” are employed where they exist and are pertinent.

2.2.1.1 Qualitative Assessment.

Qualitative assessment involves human judgment in assessment: “peer review,” “subject matter expert (SME)” evaluation, face validation, etc. The managerial challenge is to guarantee that people have appropriate credentials and/or that formal processes are in place.

2.2.1.2 Formal Assessment.

Formal assessment can be difficult to employ fully. The management challenge is to develop appropriate “light-weight” variants of the processes which can be more easily employed in M&S V&V to enhance the quality of formal assessments.

2.2.1.3 Costs/Resources.

Correct estimation of resources is a primary challenge in any M&S application. We lack adequate information for reliable estimation of M&S V&V costs/needed resources. The management challenge is to collect and organize appropriate cost and resource information from whatever sources to develop for M&S/V&V cost/resource estimation can be developed.

2.2.2 Research Challenges

Four research challenges noted: **inference**, coping with **adaptation**, **aggregation**, and **human involvement/representation**.

2.2.2.1 Inference.

Data availability to support assessment of simulation “predictions” is a fundamental problem, especially for the test and evaluation community on the operational side and the experimental community on the science side. Comparison of simulation results with the available data can be described statistically and data-simulation result relationships can be specified in terms of accuracy, error, resolution, etc. for the region of the application domain for which data exist.

Action. Develop scientifically rigorous methods for making inferences about relationships between simulation results (“predictions”) and elsewhere in the application domain.

2.2.2.2 Adaptation.

Advances in technology have led to a new genre of computational programming, termed adaptive programming. Techniques employed in adaptive programs include artificial intelligence (AI), expert systems, genetic algorithms, fuzzy logic, machine learning, etc. As adaptive processes become more capable and more widely incorporated in M&S, the V&V challenge is clear: the M&S performance can differ from instance to instance and therefore presents fundamental challenges to the prediction and assessment of performance.

Action. Develop scientifically rigorous methods to ensure future M&S performance involving adaptive programming will meet the demands of the V&V environment.

2.2.2.3 Aggregation.

Elements and interactions of a simulation can be represented in varying levels of detail . As simulations become more complex, especially multi-resolution, better methods for determining the potential impact on simulation results from such variation in levels of detail are required to minimize potential misuse of simulation results.

Action. Develop supporting theory and assessment procedures.

2.2.2.4 Human Involvement/Representation.

The complexity of representing the variety of human behavior in an automated way that appropriately reflects impacts of the simulated situation on human decision making and performance is a major challenge. The critical stumbling block is uncertainty about influences of factors and processes involved for many kinds of simulation applications. Although better understanding exists about simulation V&V when people are involved for education/training purposes or to represent human behavior in the simulated situation, there are still many significant research issues concerning interactions among simulation characteristics, the people involved, and appropriate simulation uses.

Action. Develop representations of cognitive processes.

3. COMMENTS ON SPECIFIC TOPICS

3.1 Lingering Issues

Many of the general conclusions are not new. At the Workshop, there was some discussion on what were termed lingering issues. It is worthwhile briefly noting such issues:

1. How should V&V change with M&S size, type, and application and complexity?
2. How to develop better cost estimation processes for M&S V&V
3. How to make better use of visualization, especially to enhance SME reviews.
4. How to better connect statistical processes appropriately to SME validation reviews.
5. How to better disseminate insights from VV&A experiences to M&S/V&V communities.
6. How to better provide computer support (i. e., automation) for VV&A.

- a. Develop more automated support tools for VV&A.
- b. Adopt or adapt tools from the software industry.

Research recommendations were made to address these lingering issues.

3.2 Specific Recommendations

Clearly, more tools are needed. However, such tools must be more than *ad hoc* implementations of local standards. One approach is to develop formal methods based on the integration of concepts from systems-theory related disciplines. Development of such formal methods, perhaps based on the DEVS formulation, would lead to the development of standardized test problems within disciplines. Formal method research areas include

- Research into effective methods for generating complete coverage test cases from formal specifications. Development of standardized test problems is indicated here as well as the Method of Manufactured Solutions Foundations proposed by Oberkampf and Trucano.
- Research into formal verification, and “lightweight formal methods” approaches make it possible to formally analyze partial specifications and early requirements definitions.

It bears repeating: V&V is about risk management across the entire spectrum of research, development, and management. V&V cannot succeed unless we properly incorporate risk management throughout the cycle. This includes the development of staff competency in V&V. One place of specific concern is that development of subject matter experts (SME).

1. SME-related Knowledge Engineering

2. Research into methods of guaranteeing consistency in SME assessments
3. Capture in formal mechanisms of SME knowledge
4. What truly qualifies someone to fulfill the SME role?

One factor bearing on risk is the enormous complexity of some models and simulations. Management of this complexity is crucial. Oberkampf and Trucano suggested development of effective methods of using phenomenon identification and ranking tables (PIRT) for planning and assessment system

Formal methods are one avenue, but continued development of statistical methods for software quality assurance (SQA), M&S to establish the principles of predictable compositional modeling. Statistical methods are also central to the validation process.

Most models are dynamical in nature; we can ask fundamental questions of how well the computational world matches the “true” dynamics of the underlying equations. A simple truism in engineering holds here: one must have metrics to measure goodness. Research is needed to define validation metrics for a wide range of system-response measures. Development of effective methods for the construction and use of a validation hierarchy and the specification and use of quantitative assessment criteria for validation metrics at the different tiers of the hierarchy must be undertaken. We should formalize the propagation of validation metric information from the unit problem/benchmark tiers in the validation hierarchy to the subsystem/system tiers.

Hardware in the loop (HWIL) continues to present significant V&V challenges. There is a need to document conceptual models of components of HWIL and distributed simulation

systems, particularly in regard to model detail and semantic consistency. Some problems continue to exist from our inability to manage the communications latency in distributed systems and the need to manipulate and store dense environmental data for real time effectiveness. HWIL shares a problem with much of general computer science: research is needed to deal with non-determinism in parallel applications.

Finally, research into the cost effectiveness and the establishment of “best-practices” for investment in those systems.

An up and coming paradigm is that of adaptive processes in M&S in general and therefore for V&V in particular. The view presented at the Workshop is that of model-based development. However, it was noted that one modeling framework is unlikely to fit all circumstances. This is an evolving area in computer science as well as M&S. Some specific challenges were issued:

Challenge 1: Educate developers to jump between modeling frameworks

Challenge 2: Extend standard model-based methods such that executables and test-

harnesses are automatically generated from declarative models.

Challenge 3: Educate developers in goal-oriented thinking for their modeling.

Challenge 4: Develop better methods for understanding the range of possible behaviors that arise from a non-deterministic system.

Challenge 5: Improve our automatic program understanding methods such as static analysis, runtime verification, and model checking.

I end on a topic not normally discussed at technical conferences and workshops: education. There was a session on education, from which no firm research topics emerged. No special topics emerged for the simple reason that education is the foundation for the solution to all the issues noted here.

4. REFERENCES

Dale K. Pace, D. E. Stevenson, and Simone Youngblood. Executive Summary in *Foundations '02: Foundations for V&V in the 21st Century*. San Diego, CA: Society for Computer Simulation. 2002.