Briefing slides in PDF Briefing slides in PowerPoint

Summary of V&V Research Needs

D. E. (Steve) Stevenson
November 1, 2002

This document seeks to condense the research findings of Foundations '02. Such a condensation necessarily omits details and many good points raised during the Workshop. I do not claim this document is a finished product; rather it should be viewed as a work in process – raw material in the middle of transformation to a useful product.

I begin with our perspective. Then I present the overarching issues that establish the context in which M&S V&V research needs should be viewed. Next I separate the implementation (management) issues out (i.e., the need to be better at doing what we know how to do in M&S V&V) and related public relation issues (how to get M&S V&V appropriately understood by the community so that needed resources and emphasis will be given to it by M&S sponsors, developers, and users) from the true research issues (those topics where our knowledge and methods have to be advanced in order to permit viable technical solutions to V&V problems). The ideas in this document have been stimulated by research suggestions from Foundations '02 papers and participants (I acknowledge our debt to them), but the responsibility for the statement of those ideas in this document rests only on the author identified. I hope that others will find my thoughts compelling and help to develop them more completely so that funding for the needed V&V research can be secured and the issues identified in this document addressed. I should add that an important area is not included in this document: how to get academia in include M&S V&V significantly in their programs. This is an important problem because without M&S V&V receiving adequate attention in academic programs, those entering the work place will be ill-equipped to deal with this important aspect of modeling and simulation.

Perspective

It is essential that thorough concepts, theories, and methods be developed for modeling and simulation (M&S) verification and validation (V&V), for they represent the fundamental enablers for V&V. These will only arise by studying model abstraction, heterogeneity, and composability of models and simulations. I expect these ideas to come from collecting patterns of V&V experience and application. Generic, higher level M&S frameworks with capabilities to enable and support specialized uses/domains must be researched and moved to practice. The problem solving environments (PSAs) being researched by the Department of Energy's Accelerated Strategic Computing Initiative (ASCI) program are an example of such technology. The PSAs must advance beyond being modeling only environments and incorporate V&V processes as well.

One suggested approach to removing the reluctance of organizations to use V&V appropriately would be to develop positive processes relating "development" (intended behavior), "diagnostic" (unintended behavior) and "operational" capabilities.

There seems to be general agreement that we must develop appropriate M&S V&V metrics — technical (complexity, scalability) and management measures (what and when to measure, how to measure and evaluate).

I continue below with a listing of *overarching* issues — those issues that are considered to be pervasive in V&V theory and practice. In organizing the remainder of this document, I adopted the suggestion that there be a subsection called "Implementation" for those areas in which research results were known but those results had not been moved to practice (i.e., we don't normally do what we know to do as well as we should). Similarly, issues were raised that spoke not about V&V but the *management* of V&V. A subsection addresses those issues. There was recognition that V&V needs to "market" itself, not only to professional organizations where V&V does not now have adequate visibility but everywhere appropriate professional awareness is crucial.

The subsection on research is itself subdivided. It has been suggested that Foundations '02 blurred the distinction between *model* and *software simulation* to the detriment of understanding the unique issues in each. That distinction is re-introduced. It was felt that distinctive issues arise in quality assurance, subject matter experts, and computational technology.

Overarching Issues

This section contains points that are considered to be pervasive in both M&S and V&V. No attempt has been made to prioritize these points.

For true verification to occur, it is necessary that *all* expectations and constraints be expressly stated. This includes such nebulous items as the purpose to be achieved with a given model. All stakeholders must be represented.

A disciplined approach must be followed so that the basis for the development of the system is rigorous, clearly communicated, and traceable to the purpose, with justifiable decisions on the characteristics of the system. This disciplined approach must be based on usable formal methods so that precise and rigorous reasoning can be applied throughout the development cycle.

Harkening to the comment by George Box, "All models are wrong. Some models are useful," we must understand the limitations of formal — and informal, for that matter — methods. This means that we must focus on understanding uncertainty in the physical system and variability in the model. Ultimately, a decision must be made that the system is "close enough." With that in mind, I provide the following list of perceived, overarching research needs.

- It is clear that there is no common detailed vocabulary in V&V for concepts and methods that all M&S communities share. In fact, even the fundamental terms *verification* and *validation* seem not to be uniformly defined across all M&S communities, though these terms have common specific connotations throughout the U. S. Defense community. The matter of terminology must be addressed if the field is to coalesce. In this context, it is important to

understand that *model*, *simulation*, *verification*, and *validation* are not absolute terms but are relative to a particular use.

- Verification and validation are performed in an organizational context. Technical and social dynamics interact in strange ways. The role of the human decision maker cannot be ignored, yet we have almost no data on the human decision process in this context.

- Some criticism were stated of the V&V research goals as presented at the Workshop. One series of criticisms brought to light a perception that V&V has become associated only with simulations and not models. One clear research goal must be to determine those processes that can be used for all models and those which are specific to software-based simulations.

- The scientific and general engineering communities seem largely outside and not connected with the V&V community that has focused more on system level simulations. In scientific models, much of the effort in validating code is focused on measuring uncertainty with qualitative techniques. There must be an outreach program from the V&V community to its constituent technical areas.

- Efforts to validate scientific simulations show those simulations often do not exactly match the microstructure of the reality represented in the simulation because that microstructure is itself non-deterministic. This highlights a need for research into qualitative techniques in V&V as they relate to such quantitative aspects of validation. For example, what does "good agreement" mean?

- There is a clear need to understand the techniques for verifying and validating components that are then composed into a larger system so that the V&V effort for the larger system may be much less than that required by starting over.

- The issue of formal methods came up in several venues. Formal methods are mathematically based techniques that provide frameworks within which model developers can specify, reason, and verify about designs and simulations in a systematic manner. Since formal methods enable systematic demonstration of the correspondence of the model with respect to its specification, the conceptual model validation process can benefit from such sound inference and reasoning methods. The deductive apparatus of formal methods can help model developers to establish the correspondence of their conceptual models to system abstractions, theories, and properties depicted in a formal specification language. Developing specific properties of interest for certain application domains and introducing approaches that demonstrate their satisfiability by model conceptualizations such as event graphs and discrete event simulation (DEVS) formalisms are promising areas of research that will introduce formal methods to early phases of the simulation lifecycle.

- The distributed and collaborative engineering of simulation model conceptualizations and specifications introduces consistency and coherence problems. Hence, error reduction earlier in the simulation modeling lifecycle is needed. Model-based verification methodologies that could detect inconsistencies and behavioral defects at the conceptualization and model specification stage are likely to alleviate these problems.

- Reusing existing simulation components requires reasoning about their fitness and suitability in the new context. Syntactic interface compatibility is a necessary but insufficient condition for reuse and composable simulations. Semantic compatibility of simulation components is essential to alleviate differences in fidelity levels, representation granularity, time management, and data compatibility. Establishing semantic compatibility among diverse components under a new federation (i.e., HLA) requires meaningful and timely exchange of information. Furthermore, original objectives of a reused simulation component may not be in conformance with the needs and goals of a new simulation. Verifying and testing components to assess their fitness into a new formally specified context requires methods that detect mismatches among specification. Since complete matching is not a practical solution, fuzzy and approximate matching with formal decision methods are more likely to be useful. Building models for reuse introduces new issues such as general applicability and conformance to published specifications.

- Agent-based modeling and multi-agent simulations are promising emerging areas in simulation modeling. Agent theory introduces new abstraction, organization, and decomposition mechanisms for system simulation. The emphasis on collaboration, coordination, task distribution, and adaptive behavior requires agent-based simulations to deploy specific algorithms and methods. New paradigms arising from complex adaptive systems as well as artificial life theories are attracting the attention of model developers. V&V of distributed multi-agents systems that incorporate complexity through interactions, patterns of adaptive organizations, and processes requires new means to analyze and assess the accuracy of their emergent, mostly unpredictable behavior.

- As agencies shift and integrate their essential simulation services onto distributed software intensive message-oriented middleware platforms, failures as well as performance bottlenecks would result in unacceptable and ineffectual simulations. Configurable performance benchmarking environments would be valuable for large-scale Publish/Subscribe (P/S)-based distributed simulation infrastructures such as High Level Architecture (HLA) mandated by Defense Modeling and Simulation Office. The science of testing messaging middleware and frameworks is an uncharted territory as there are no industry-accepted benchmarks.

- It is a well-known fact that about two-thirds of software-intensive simulation development time is devoted to maintenance and evolution. In particular, the current trend towards using open standards supported by commercial off-the-shelf (COTS) software/components/ simulations in developing open distributed federations of simulation systems increases the likelihood of constant system modernization and evolution. As new or upgraded models (.e., sensor models) are inserted into the baselines, model developers spend enormous amount of time to instill back confidence in the legacy systems and models to assure backward compatibility. There is an urgent need for tools that would at least help manage instilling back confidence on the legacy models with regard to their compatibility to the upgraded or modified baseline systems.

- Driven by fiscal constraints, increasing pressure exists to employ realistic training simulations for driving exploration into new, more effective methods for modeling activities. The synthetic natural environment (SNE) models, upon which operational system models impact, constitute the environmental ground truth. Data that satisfies the user requirements need to be composed to derive complete synthetic environments. Merging disparate SNE components needs sophisticated data schema V&V, interoperability analysis, diagnosis, and merge tools. The major programs that will be supported by this research are: (1) advance concept development and defense planning, (2) R&D, acquisition, and military doctrine development, and (3) training, exercise, and military operations domains.

Implementation Issues

This section and the two which follow it contain points relating to moving research results into practice and to improving V&V professionalism so that more often and more completely "best practices" are employed. The practice of V&V can be very difficult and resource consuming. We simply have inadequate data about how difficult or how effective V&V techniques are as they are applied with various kinds of M&S for diverse applications.

- We must find ways to track the impact of V&V techniques. We can profit by drawing an analogy to software development. The Software Engineering Institute (SEI) invented metrics that started practitioners counting defects and classifying them. With these metrics they could compare development methodologies, chief-programmer teams versus structured decomposition for example, and make arguments about which techniques were more cost effective. Today, the V&V community is just starting to produce techniques that compete with each other, such as lightweight proof of correctness versus automated test case evaluation. It is "easy to take data" but it is very difficult to interpret that data. We need development of metrics that are V&V related to compare methodologies.

- There are many difficulties in moving research results into practice. For example, quantitative/qualitative techniques with statistics are needed in transferring from theory to practice.

- The most basic problems seem to be that of simply changing the mind sets. It is suggested that lightweight verification and validation tools are essential to gain a foothold in practice for widespread use of formal methods and adaptive processes.

Public Relations and Marketing

- There is a problem of organizations doing V&V but then never reporting on it to the community. Ways to better disseminate insights effectively from V&V experiences to M&S/V&V communities are needed.

- One theme that arose in many Foundations '02 sessions was the necessity to have V&V portrayed in a positive light with developers. A thread in the Workshop is that of risk and the

belief that V&V lowers risk and raises benefits. The costs are (fairly) obvious. The benefits of risk reduction can be measured in higher reliability, longer lived systems, greater security, more predictable developments. Some suggestions were:

- Properly incorporating risk management into the simulation model development life cycle.
- Research into properly developing staff competency in V&V.
- A return on investment study, which may well exist, to document benefits.
- The technical communities and societies must be convinced to that better science, engineering, management, etc results from VV&A. One possible solution is for practitioners of V&V who also are active in societies should push VV&A onto the agenda.

Management

The management of V&V suffers from a lack of management tools and information. It is crucial that such be developed.

- Develop better cost estimation processes for M&S V&V that can be used to calibrate cost estimating relationships.

- Develop generally accepted standard work-breakdown structure for VV&A.

- Develop standardized test problems.

- Develop 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.

Research Issues

This section contains what might be characterized as "the usual" research objectives, i.e., they address areas where we must advance our knowledge and methods if we are to have viable technical capabilities identified as needed. In other words, we currently do not know how to do what needs to be done. The points raised here have been categorized for convenience of the reader. There are issues that cross boundaries which are not mentioned. For example, whatever we say about models should somehow carry over into digital, analog, or other simulations.

Modeling Methodology

- Modeling methodology development is in its infancy. The general directions of M&S theory is towards systems theory. Much more development is needed. There are several mathematical tools for this, including category theoretic concepts. This research must develop viable mathematical tools (ones which can be applied effectively within reasonable resource constraints, where those constraints include factors such as computational time required, level of personnel expertise, etc.) to establish semantic consistency and coherence of models and enable simulation implementers to track model issues throughout simulation development.

- Many more tools are needed to effectively deal with the overwhelming detail in large models and simulations. Along with this, we need to develop best practice guides for various classes of systems.

- Historically, models have tended to be deterministic. Modern systems are so complex and deal with such broad issues as to no longer be deterministic. Such systems must deal with both uncertainty and variability at a level that guarantees no one person can fully understand the large simulations that are becoming common. Methods are needed to deal with this level of non-determinism.

Simulation — Software Engineering

- There are issues of simulation development that relate to software simulations. It is not clear what is unique about software simulations vis-à-vis, say, analog simulations. Theoretical computer science has developed techniques similar to DEVS; for example, algebraic data types. On the other hand, systems theoretic treatments of software are not well-developed.

- There are many techniques available in computer science that are not being used in M&S on a regular basis: static analysis, runtime verification, model checking, formal methods, etc. Other methods, such as automatic programming and PSAs are more speculative. There is a steep learning curve for many of these techniques precisely because they are formally grounded in mathematics. Such formally methods are not routinely taught in undergraduate — or even graduate — programs in the U. S. although they are common in Europe and Canada. Research into formal methods is crucial to V&V. In this regard, we need research into "lightweight formal methods" as well as "industrial-strength formal methods." Lightweight methods would make it possible to deal with partial information.

- Finally, a lingering question: How does simulation V&V change with M&S size, type, and application?

QA, SQA, and Testing

- How to better provide computer automated support for VV&A, including visualization techniques.

- Testing continues to be a difficult process. Along formal proof-based methods, we need to develop tools and techniques to guarantee efficient test cases for simulations. These methods should be model-based, not code-based. Development of standardized test problems within disciplines are essential.

- Development of statistical methods for SQA, M&S is urgent.

Subject Matter Experts (and related qualitative assessments)

- How to use visualization capabilities to enhance SME and other qualitative reviews.

- How to better connect statistical processes appropriately to SME and other qualitative validation reviews.

- SME-related Knowledge Engineering, i.e., how to ensure that appropriate knowledge is obtained from experts, ensure that the elicited knowledge is correctly represented, and ensure that the elicited knowledge is stated in ways that facilitates effective reuse of it and comparison of it with related knowledge.

- Research into methods of guaranteeing consistency in SME assessments and ways to use expert estimations as "quantitative" surrogates for real-world data when such does not exist.

- Capture in formal mechanisms of SME knowledge to enable automated comparisons and compilation of knowledge from experts.

- Better processes to determine what truly qualifies someone to fulfill the SME role?

Computational Technology

The size and complexity of software simulations can require very large distributed computer systems. This area has been an active area of research since the mid-1980s. It continues to be so today. However, little progress has been made in making distributed simulations "easy" to develop. All of the performance issues such as data layout and communications patterns have resisted a single paradigm such as the traditional von Neumann paradigm taught in traditional programming classes. Continued research is imperative.

Even for a single-computer simulation, the addition of live systems or hardware in the loop greatly increases the complexity and scope of the simulation. These mixed simulations must continue to be the focus of research.

Statement of M&S V&V Research Needs for Foundations '02 Executive Summary

The following was prepared as a statement of M&S V&V research needs for the Foundations '02 *Executive Summary*. This abbreviated description is compatible with the fuller description of research needs above, but is focused to address items expected to be of primary interest to those who might read only the *Executive Summary* and not examine other Foundations '02 materials.

A synopsis of M&S V&V research needs 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. One set of challenges relate to M&S management (or implementation): *how to do what we know how to do*. The other challenges have a research flavor: *areas that we need to understand better in order to find viable technical solutions*. We discuss both below.

Management Challenges: Three management (implementation) challenges stand out: ***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.

> ***Qualitative Assessment***. Qualitative assessment involves human judgment in assessment: "peer review," "subject matter expert (SME)" evaluation, face validation, etc. Often people involved in qualitative assessments are selected and perform their assessments without appropriate credentials and/or formal processes. Methods exist, which, if used, can increase qualitative assessment objectivity and consistency.
>
> ***Formal Assessment***. Formal assessment, whether statistical in nature or following some other rigorous mathematical approach, can be difficult to employ fully. The management challenge is development of 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.
>
> ***Costs/Resources***. Correct estimation of resources needed is a primary challenge in any M&S application. Inadequate information is currently available 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 case studies and other sources), making it available to the M&S community so robust methods for M&S/V&V cost/resource estimation can be developed.

Research Challenges: Four research challenges stand out: ***inference***, coping with ***adaptation***, ***aggregation***, and ***human involvement/representation***.

> ***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; *but no scientifically rigorous methods currently exist for making inferences about relationships between simulation results ("predictions") and elsewhere in the application domain.*
>
> ***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. *No scientifically rigorous methods currently exist to ensure future M&S performance involving adaptive programming will be as good as or better than past performance.*
>
> ***Aggregation***. Elements and interactions of a simulation can be represented in varying levels of detail . As simulations become more complex, especially in the case of distributed simulations which may use more than one level of resolution for the same kind of element or interaction, 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. *Present theory and assessment processes related to this topic are embryonic.*
>
> ***Human Involvement/Representation***. Representation of human behavior in simulations is widely recognized as being critical; 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.

The V&V research challenges for these four areas are to develop clear and coherent definition of expected M&S performance under various conditions. Other important research challenges for M&S V&V are also identified in the Foundations '02 proceedings, such as those involving the theoretical foundation for M&S V&V and aspects of visualization in M&S V&V.