Session T3: V&V Research

Session T3 leaders:

Co-Chairs:

Bob Thomas (Sandia National Laboratories)
Hessam Sarjoughian (Arizona State University)

T3 Materials in Foundations '02 proceedings:

Papers

Formalization and Validation: An Iterative Process in Model Synthesis (18 pp)
Joerg Desel (Katholische Universität Eichstätt-Ingolstadt
Lehrstuhl für Angewandte Informatik, Eichstätt, Germany)

A Proposed Evolution of Validation Definition (7 pp)
Daniel Girardot (French Centre d'Analyse de Défense)
René Jacquart (ONERA, French research institution for Aeronautics & Space)

A Case Study of Verifying and Validating an Astrophysical Simulation Code (43 pp)
A. C. Calder (U. of Chicago)
B. Fryxell (U. of Chicago)
T. Plewa (U. of Chicago & Nicolaus Copernicus Astronomical Center, Poland)
R. Rosner (U. of Chicago)
L. J. Dursi (U. of Chicago)
V. G. Weirs (U. of Chicago)
T. Dupont (U. of Chicago)
H. F. Robey (Lawrence Livermore National Laboratory)
J. O. Kane (Lawrence Livermore National Laboratory)
B. A. Remington (Lawrence Livermore National Laboratory)
R. P. Drake (U. of Michigan, Ann Arbor)
G. Dimonte (Lawrence Livermore National Laboratory)
M. Zingale (U. of Chicago & U. of California, Santa Cruz)
A. Siegel (U. of Chicago)
A. Caceres (U. of Chicago)
K. Riley (U. of Chicago)
N. Vladimirova (U. of Chicago)
P. Ricker (U. of Chicago & U. of Illinois)
F. X. Timmes (U. of Chicago)
K. Olson (U. of Chicago, UMBC, & NASA)
H. M. Tufo (U. of Chicago)

Model Validation Methodology from Validation Experiments to Systems Level Application (18 pp)
Richard G. Hills (New Mexico State U.)
Ian Leslie (New Mexico State U.)

Slides (may contain back-up materials and notes)

Formalization and Validation (23 slides) [T3B_desel1 in both pdf and ppt formats]
Joerg Desel (Katholische Universität Eichstätt-Ingolstadt
Lehrstuhl für Angewandte Informatik, Eichstätt, Germany)

Validating the Petri Net Model (11 slides) [T3B_desel2 in both pdf and ppt formats]
Joerg Desel (Katholische Universität Eichstätt-Ingolstadt
Lehrstuhl für Angewandte Informatik, Eichstätt, Germany)

Validating the Model (20 slides) [T3B_desel3 in both pdf and ppt formats]
Joerg Desel (Katholische Universität Eichstätt-Ingolstadt
Lehrstuhl für Angewandte Informatik, Eichstätt, Germany)

A Proposed Evolution of Validation Definition (19 slides) [T3B_girardot in both pdf and ppt formats]
Daniel Girardot (French Centre d'Analyse de Défense)
René Jacquart (ONERA, French research institution for Aeronautics & Space)

A Case Study of Verifying and Validating an Astrophysical Simulation Code (32 slides) [T3B_calder in both pdf and ppt formats]
A. C. Calder (U. of Chicago)
B. Fryxell (U. of Chicago)
T. Plewa (U. of Chicago & Nicolaus Copernicus Astronomical Center, Poland)
R. Rosner (U. of Chicago)
L. J. Dursi (U. of Chicago)
V. G. Weirs (U. of Chicago)
T. Dupont (U. of Chicago)
H. F. Robey (Lawrence Livermore National Laboratory)
J. O. Kane (Lawrence Livermore National Laboratory)
B. A. Remington (Lawrence Livermore National Laboratory)
R. P. Drake (U. of Michigan, Ann Arbor)
G. Dimonte (Lawrence Livermore National Laboratory)
M. Zingale (U. of Chicago & U. of California, Santa Cruz)
A. Siegel (U. of Chicago)
A. Caceres (U. of Chicago)
K. Riley (U. of Chicago)
N. Vladimirova (U. of Chicago)
P. Ricker (U. of Chicago & U. of Illinois)
F. X. Timmes (U. of Chicago)
K. Olson (U. of Chicago, UMBC, & NASA)
H. M. Tufo (U. of Chicago)

Model Validation: Unit to System (26 slides) [T3B_hills in both pdf and ppt formats]
Richard G. Hills (New Mexico State U.)
Ian Leslie (New Mexico State U.)

Accreditation Issues for Verification and Validation of the Prototype Federation for the Joint Synthetic Battlespace (28 slides) [T3B_andrew in both pdf and ppt formats]

Lt Col Emily Andrew (USAF)
Gerald Pritchard (Dynetics)
Jeffrey Wallace (Envoy Tek)

Participants in this session are listed at the end of the Discussion Synopsis.

Discussion Synopsis (to provide perspective on papers & briefings identified above).

Introductory Note: The T3 Session was organized on a different principle from the other Foundation '02 sessions. In T3, there were several research papers, presented as one might expect at a usual professional conference. These papers were followed by a discussion of research needs, in general.

Presented Papers

T3-Session Research: Raw Notes

Jörg *Desel*

The presenter began by noting that there is a general lack of agreement between VV&A terms as used in the United States and in Europe. This observation would actually come up in several different discussions. In presenter's view, *verification* means *proof* but *testing* is decidedly not proof. He also pointed out that to test something, you must have an artifact to test. The general European/formal methods view is that validation starts with the writing down of the specification from the requirements.

The research reported here asks the question, "What is a formal theory of validation?" Here, of course, validation is taken in the sense above.

Validation must pay attention to the purposes for which the model or simulation is being developed. In this research, a control theoretic approach is used in which the *plant* is seen as "reality" and the *control* is seen as simulating the plant. In order to perform validation at this level, a model checking approach is used.

A very short digression into model checking ensued. In this research, the major focus is the use of Petri nets to simulate the logical conditions of the system so that the temporal logic conditions can be checked. There are three basic graphs used to consider the system:

1. Occurrence runs
2. Causal graph as a guide graph
3. A "time invariant graph" from the "occurrence graph by simulation order of execution."

The model can be explored using behavior information from simulation. Requirements filter possible scenarios so that only some outputs must be examined. The information gained can be fed back into the design, causing a spiral development of the system because this approach uncovers misunderstanding in the specifications by simulating use cases/scenarios..

The presenter showed an example wherein this technique was used in industry.

Outstanding research issues noted by the presenter are development methods for plant and control models, techniques for validating specifications and the relationship of the Petri net approach to other validation techniques.

Rene Jacquart

The focus of this paper is to document the evolution of a definition of validation.

The presenters adopted a systems engineering approach to modeling then searched for various definitions.

Considering the DMSO definitions, the presenters focused on "accurate representation" and "intended use" as items needing more investigation. The issues of intended use were resolved by taking the meaning to be *accessibility*. However, perhaps a better term might be *measurable*.

Continuing, the question comes about, "How does modeling enter into system development?" Modeling seems to be the process of fidelity adjustment versus simulation. From this, several criticisms of various approaches to definitions were put forth.

- Data poor environment
- Application domain limitations: Robin Miller DIS'96, Spring
- Technical limits
- Statistics: Kleijnen: WSC 2000. "There is an abyss between validation practice and statistical theory."
- DMSO's *Recommended Practice Guide* is about software engineering, not modeling or simulation.

The point of departure is to integrate the role of people in validation; i. e., organization versus problem. The *problem* leads to *organization and product processes*, where organization is taken to include agents, roles, responsibilities, obligations, resources, activities.

New paradigm was suggested as evidence-based reasoning approach to validation with a new emphasis on psychological/socio-technical issues.

Alan Calder

To the physical scientists, the real world is observations. One of the problems is that good, controlled experiments may be difficult to conceive and carry out. Unfortunately, simulation developers may forget experimental limitations. Furthermore, the value judgment of *good* is a subject matter expert issue. In the case, there are many different communities trying to define *good*. The presenter did a fine job of discussing and illustrating the difficulties in the physics.

On the computational side, the presenter noted that good software engineering practices and tools count, but that the astrophysics community seems largely unaware of these practices. He noted that the pairing of application and numerical algorithm may not be correct as the algorithm may be inappropriate to a particular use. The presenter also noted that in physical sciences, the *verification and validation* dicotomy may be unused or unappreciated.

Turning to V&V itself, the presenter noted that much of the effort in validating this code is attempting to measuring across uncertainty with qualitative techniques. That is, the simulation does not exactly match the microstructure but that microstructure is itself non-deterministic. Therefore, there seems to be good need for research into qualitative techniques in V&V. For example, what does “good agreement” mean? The presenter ended with the goal of considering various V&V techniques and how they might be applied to this system.

Rich Hills

The research focus of this paper is a method for extrapolating from component tests to global system concepts by combining results. The paper requires reading to completely appreciate the technique. This synopsis discusses generalities

The presenter began by developing a decision theoretic vocabulary with decision variables and then classifying uncertainty. He also discussed the difference between fitting parameters and independent parameters.

The questions to be ask deal with

- The combining data at unit level to get system data,
- The resolution of data at the component level, and
- The importance measures in the parameter space.

The presenter indicated that the method can be summarized by “the sensitivities are the glue binding the components to the aggregate.” Sensitivity is a risk management measure. There are practical limits in orders (first order, etc) of statistics. The key realization is that the measurement model used to define metrics is yet another model.

Statistics can help you determine if you’re measuring the correct things.

Conclusions:

- Validation is designed by models
- Effects of uncertainty must be considered in design
- First order techniques can be used.

LTC Emily Andrew

The project’s research focus is synthetic environments. The main goal of the research presented was to go through all the steps in verification, validation, and accreditation. The problem chosen for this effort was one to develop a synthetic environment from a live exercise. The environment had mixed resolution. Legacy codes were used initially to build up to the synthetic environment. The experiments were to determine the quality of detections on the ground.

There were many interesting findings. Because this was an *implementation,* much is to be gained on putting V&V into practice. The first realization is that V&V takes hard work in gaining measurements, with special difficulties in gaining measures of boundary effects. Measurements must be traced to models but this is very difficult. The presenter discussed several qualitative issues, such as comparing trends. The presenter emphasized that all this work amounts to an independent analysis of the entire system, not just the codes.

The presentation also highlights the difficulties in actually doing V&V. The most basic problems seems to be that of simply changing the mind sets. There is a problem of organizations doing V&V but then never reporting on it to the community.

On the technical front, one clear problem is the integration of science and engineering based systems with more human and behavioral systems. A clear research need is in this interface.

The presenter was asked about budgets. She stated10-12% of budget cost is needed for VV&A.

Discussion of Research Goals

This was a general discussion about general needs in the community.

Perceived Problems

- People don't want to put money into V&V. One possible solution is to do a return on investment study. Such a study may well exist.
- We need to remove the emotive reaction that VV&A is a punitive. One possible solution is similar to the above: show how V&V adds value.
- 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.
- Do we have data on cost of doing validation and not doing validation? Can we quantify costs. NOTE: this issue was subsequently addressed in the plenary session by Bob Lewis of Boeing. One possible route is to have Government agencies to provide data.
- Who's demanding credibility? What is the evidence and what is its relation to credibility. Who sets rules?

Perceived Research Questions

The below list is a summary of points raised in the technical paper sessions.

- Social and technical issues and how they interact should be explored.
- Validation can be a guide in designing both systems and experiments.
- Techniques are needed to take component level testing/validation information and then deriving validations of composed systems. A concomitant issue is scaleability.
- Positive V&V concepts that should be used in design loop.
- Real world is in the mind of the decision maker in mixed science-psychological models. Validation of physics is much different than validation for intended use/mind of the decision maker. Observable nature versus mind of the decision maker.
- Are there alternatives to "verification first, validation next" development schemata?
- We need methods for estimating numerical errors without grid refinement studies.
- Quantitative/qualitative techniques with statistics are needed in transferring from" from theory to practice.
- Lightweight verification and validation tools are essential to gain a foothold in practice.
- Can evidence-based methods be extrapolated?
- We need to understand uncertainty quantification and its relation to credibility.
- Uncertainty is not just one number. How do we handle ensembles of uncertainty quantification?
- We need to recognizing and quantifying variability and uncertainty. They are not the same thing.
- How much of V&V is implementation and how much is research?
- Can we develop techniques for verifying and validating components for classes of uses rather than one use?

Final Thoughts

It is essential that concepts, theories, and methods be developed, for they represent the fundamental enablers for VV&A. These will only arise by studying model abstraction, heterogeneity, and composability of models and simulations. We expect these ideas to come from collecting patterns of VV&A experience and application. For example, we should consider the "ilities" describing SQA:

1. high-reliability systems such as health-care, finance, nuclear energy and space exploration to low-reliability systems.
2. short- vs. long-life operation (e.g., weapon stockpile)
3. adaptiveness.
4. specialized vs. general domains – traditional, well-studied (aerospace, manufacturing) vs. innovative (homeland security) domains
5. Approaches to determine "multiple breakpoints" – success vs. failures given criteria such as adaptability (poorly vs. well known environments), operational mode (e.g., soft vs. hard real-time), security, cost, time to market.

There is a need for reusing VV&A technology. While DMSO has been a repository of VV&A information, we need to develop repository across areas outside DMSO's charter — finance and science, for example. In order to deal with such repositories, questions of ownership and accountability must be addressed.

Generic, higher level M&S frameworks with capabilities to enable and support specialized uses/domains must be researched and moved on to practice. For example, the so-called "problem solving environments" (PSA) being researched by ASCI. The PSA must advance beyond being modeling only environments and incorporate VV&A processes as well.

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

Finally, there seems to be general agreement that we must develop VV&A metrics — technical (complexity, scalability) and management measures (what and when to measure, how to measure and evaluate).

General Discussion of Needs

In general, we need to have a more global view of how V&V fits within the M&S development process, otherwise, improvements will at best be local and of limited applicability. Two suggestions are:

1. We need M&S frameworks that support both not only verification and validation but all the various underpinning relationships such as does a model even fit the intended purpose (e.g., as formalized by the applicability relation between experimental frame and model) and intra-model relations (such as morphisms between models).

2. We need to integrate such M&S frameworks into existing generic development processes such as the FEDEP -- more than just overlaying such processes, tests of relationships should guide and inform the development process.

T3 Session Participants (17)

First Name	Last Name	Organization
Emily	Andrew	USAF ESC/CXC
Joanna	Boyette	Johns Hopkins University
Alan	Calder	University of Chicago
Dan	Craigen	ORA Canada
Jörg	Desel	Katholische Universität Eichstätt
Kevin	Dowding	Sandia National Laboratories
Richard	Hills	New Mexico State University
Rene	Jacquart	ONERA-DTIM
Ken	Kennedy	Clemson University
Ian	Leslie	New Mexico State University
Anthony	Marthaler	Johns Hopkins University
Martin	Pilch	Sandia National Laboratories
Sean	Price	Cranfield University
Brian	Rutherford	Sandia National Laboratories
Hessam	Sarjoughian	Arizona State University
D. E. (Steve)	Stevenson	Clemson University
Robert	Thomas	Sandia National Laboratories