Session A2: Selected V&V (especially validation) Methods and Techniques

Session A2 Leaders:

Co-chairs:
Jim Cavendish (General Motors)
Kevin Greenaugh (DOE National Nuclear Security Administration)
Recorder: **Michael McKay** (Los Alamos Nartional Laboratory)

A2 Materials in Foundations '02 Proceedings:

Papers

Statistical Foundations for the Validation of Computer Models (28 pp) [A2_combo]
Robert G. Easterling (Statistical Consultant)
James Berger (Duke University & National Institute of Statistical Sciences)

Measuring Predictive Capability of Computational Models: Foam Degradation Case Study (20 pp) [A2_easterling]
Robert G. Easterling (Statistical Consultant)

A Framework for Validation of Computer Models (59 pp) [A2_berger]
M. J. Bayarri (National Institute of Statistical Sciences)
James Berger (Duke University & National Institute of Statistical Sciences)
D. Higdon (National Institute of Statistical Sciences)
M. C. Kennedy (National Institute of Statistical Sciences)
A. Kottas (National Institute of Statistical Sciences)
R. Paulo (National Institute of Statistical Sciences)
J. Sacks (National Institute of Statistical Sciences)
J. A. Cafeo (General Motors)
J. Cavendish (General Motors)
C. H. Lin (General Motors)
J. Tu (General Motors)

Slides (may contain back-up materials and notes)

Statistical Foundations for the Validation of Computer Models (54 slides) [A2B_easterling1 – both pdf & ppt formats]
Robert G. Easterling (Statistical Consultant)
James Berger (Duke University & National Institute of Statistical Sciences)

Extra Points (14 slides) [A2B_easterling2 – both pdf & ppt formats]
Robert G. Easterling (Statistical Consultant)

A Framework for Validation of Computer Models (36 slides) [A2B_berger – pdf format only]
M. J. Bayarri (National Institute of Statistical Sciences)
James Berger (Duke University & National Institute of Statistical Sciences)
D. Higdon (National Institute of Statistical Sciences)
M. C. Kennedy (National Institute of Statistical Sciences)
A. Kottas (National Institute of Statistical Sciences)

R. Paulo (National Institute of Statistical Sciences)
J. Sacks (National Institute of Statistical Sciences)
J. A. Cafeo (General Motors)
J. Cavendish (General Motors)
C. H. Lin (General Motors)
J. Tu (General Motors)

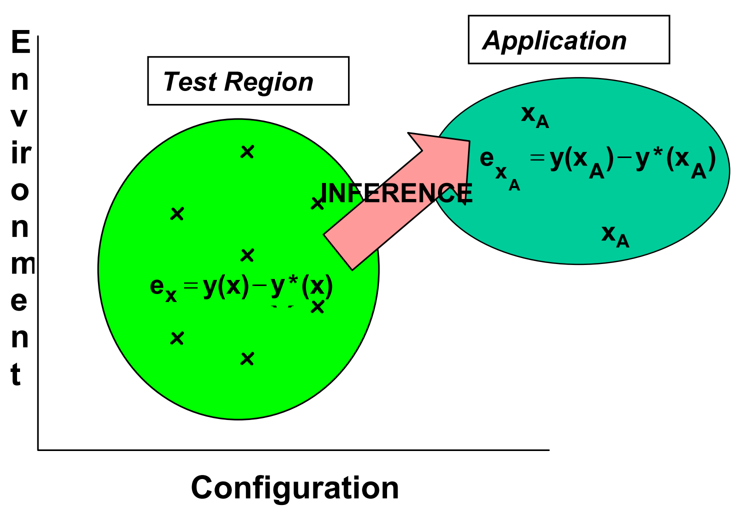
A2 Participants are identified at the end of this document

Discussion Synopsis (this material is to provide perspective on papers and briefings mentioned above – it should not be used without that context)

The starting point for the papers and discussion in this session is that model-validation means the comparison of field or experimental data to computational predictions of the outcomes of such events. This makes validation fundamentally a statistical endeavor: experimental design (or survey design), to generate pertinent data; statistical data analysis, to extract and communicated the information gleaned from the comparison. Thus, this session addressed statistical foundations of model-validation.

It is important that model-validation have a goal. This will drive the experimental design and data analysis (and provide a focus that seems often to be missing from model-validation efforts). The statistical goal advocated is to be able to attach credible, defensible, communicable "limits of error" to computational predictions of the outcomes of events, including events that have not, or cannot, be observed in an experimental context.

The following schematic illustrates the statistical model-validation process. Experiments and applications, in general, correspond to the imposition of some environment onto a system, or a representation of a system, with the objective of learning how the system will respond to that environment. Thus, the two meta-axes in the figure are configuration and environment. It may help to think in terms of systems or components, or simplified mock-ups of such devices being subjected to environments in which the system is supposed to operate successfully.



The figure denotes a suite of experiments conducted in a test region. At these points a comparison of experimental outcomes to computational predictions provides information about prediction error in the test region. If that region does not cover the application space, then, as shown, there is an inference problem – to extrapolate from what is learned about prediction error in the test region to a statement about prediction error in the application region. The session addressed issues and methods for accomplishing a meaningful evaluation of predictive capability.

Discussion Points:

1. Communication of statistical concepts and results. Is either the Bayesian of Frequentist approach advantageous in that regard? Consensus was that larger issues of experimental design and objectives were more important than the particular machinery used to generated error bounds for predictions.

2. Bias. In order to properly characterize predictive capability, it is important to know that there are no appreciable systematic errors in experimental results.

3. Concern was expressed that modeling, and validation, have been oversold. A small number of "Admiral's tests" (full system tests) by themselves, does not provide adequate information about a model's predictive capability. When the results of such tests are combined with component tests, it may be possible to obtain a useful evaluation of predictive capability.

4. Sample size issues. How many experiments are needed? The answer to such questions depends on experimental objectives, such as how precisely characteristics of prediction error need to be estimated. Answers to such questions are always context-specific, but statistical methods can be used to frame objectives and determine sample sizes. There is, in general, a need to include some replication, but there are trade-offs to be made between exploration of a factor space and replication.

5. Model Tuning/calibration. Can valid prediction error limits be obtained when the validation experiments' results are used to estimate physical parameters in the predictive model? The consensus was that they can, under both Bayesian and Frequentist analyses, in some situations.

6. No data. With no data, statistical validation, as addressed in this session, cannot be done. Expert judgment, then, is sometimes used to make assessments of validity. Legacy data may be used to evaluate predictive capability, but use of "old" data could cause problems.

7. Extrapolation. Some sort of mathematical linkage is required to extend evaluations of predictive capability in the test region to a separate application region. The resulting bounds can be quite broad and it may be difficult to find suitable mathematical linkages.

A2 Participants (61)

Last Name	First Name	Company
Mark	Anderson	Los Alamos National Laboratory
Rosser	Baker	Marine Corps
M.J.	Bayarri	University of Valencia and NISS
James	Berger	Duke University
James	Bigelow	RAND
Pierre	Bouc	ALEKTO
John	Cafeo	GM R&D Center
James	Cavendish	General Motors R&D Center
Hugh	Coleman	Mechanical & Aerospace Eng.
James	Coolahan	Johns Hopkins University
Marcus	Coombs	Royal Military College of Science
Andrew	Cronk	Simulation Technologies Inc.
Kevin	Dowding	Sandia National Labs
Robert	Easterling	Sandia National Labs
Robert	Ebert	Sanderling Research Corporation
Joseph	Elmo	Atlantic Consulting Services, Inc.
Jerry	Feinberg	MSIAC
Timothy	Geipe	Johns Hopkins University
Robert	Gover	Naval Research Laboratory
Kevin	Greenaugh	NNSA/DOE
David	Gross	The Boeing Company
Shannon	Hall	Johns Hopkins University
Michael	Hemsch	NASA Langley Research Center
Peter	Hilgarth	
Susan	Hudson	Mississippi State University
Jesse	Kemp	Marine Corps
Michelle	Kilikauskas	NAWCWD (418000D)
Kristina	Langer	APRL/VASM
Ian	Leslie	New Mexico State University
Wu	Li	NASA Langley
James	Luckring	NASA Langley Research Center
Michael	Mandelberg	JHU/APL
Paul	Matthews	Telcordia Technologies
Michael	McKay	Los Alamos National Laboratory
Unmeel	Mehta	NASA
Tim	Menzies	WVU
Ron	Morishige	Lockheed Martin
Joseph	Morrison	NASA Langely Research Center
Thomas	Paez	Sandia National Laboratories
Lucia	Perkins	Arizona State University
Chris	Pettit	AFRL/VASD
Martin	Pilch	Sandia National Labs
John	Riley	Raytheon Company

Ollie	Rose	Mount Olive College
Brian	Rutherford	Sandia National Laboratories
Jerome	Sacks	NISS
Larry	Sanders	DTRA
Eric	Snyder	U.S. Army Research Laboratory
W. Glenn	Steele	Mississippi State University
Thomas	Taylor	Arizona State University
Ben	Thacker	Southwest Research Institute
Jian	Tu	General Motors
Duane	Veley	AFRL/VASD
Eric	Walker	NASA Langley Research Center
Sheldon	Wolk	NRL
Sheldon	Wolk	Naval Research Laboratory
Ren-jye	Yang	Ford