SAND2007-2952C

Verification and Validation Benchmarks

William L. Oberkampf
wloberk@sandia.gov

Timothy G. Trucano
tgtruca@sandia.gov

Sandia National Laboratories
Albuquerque, New Mexico, USA

Seventh Biennial Tri-Laboratory Engineering Conference
Albuquerque, New Mexico
May 7-10, 2007

National Nuclear Security Administration

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Outline of the Presentation

- Background
- Review of Verification and Validation Processes
- Recommendations for Verification Benchmarks
- Recommendations for Validation Benchmarks
- Closing Remarks

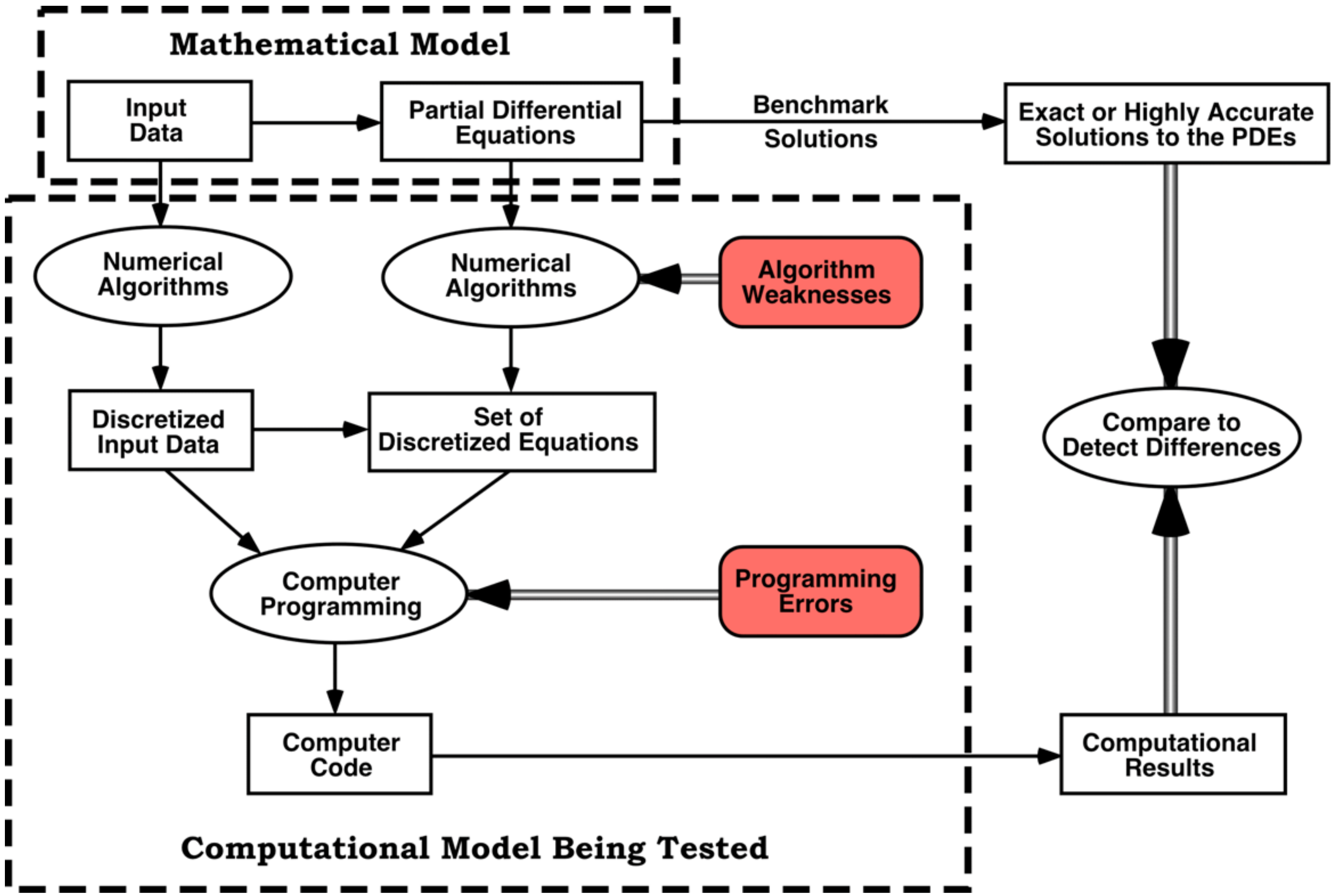
Status of Verification and Validation Benchmarks

- NAFEMS has roughly 30 formal benchmarks:
  - Almost all are verification benchmarks
  - Primarily in solid mechanics, some in CFD
- Commercial software benchmarks:
  - Almost all are verification benchmarks in solid mechanics
  - Goal of their benchmarks is to demonstrate "engineering accuracy" of the codes, not precise error assessment
- Validation databases:
  - NPARC Alliance in the U.S.
  - ERCOFTAC and QNET-CFD in Europe
  - Most of the benchmarks are for "industrial applications" i.e., relatively complex flows
- Nuclear weapons laboratories:
  - JOWOG activities involving LANL, LLNL, and AWE
  - SNL, LANL, and LLNL beginning database for engineering issues

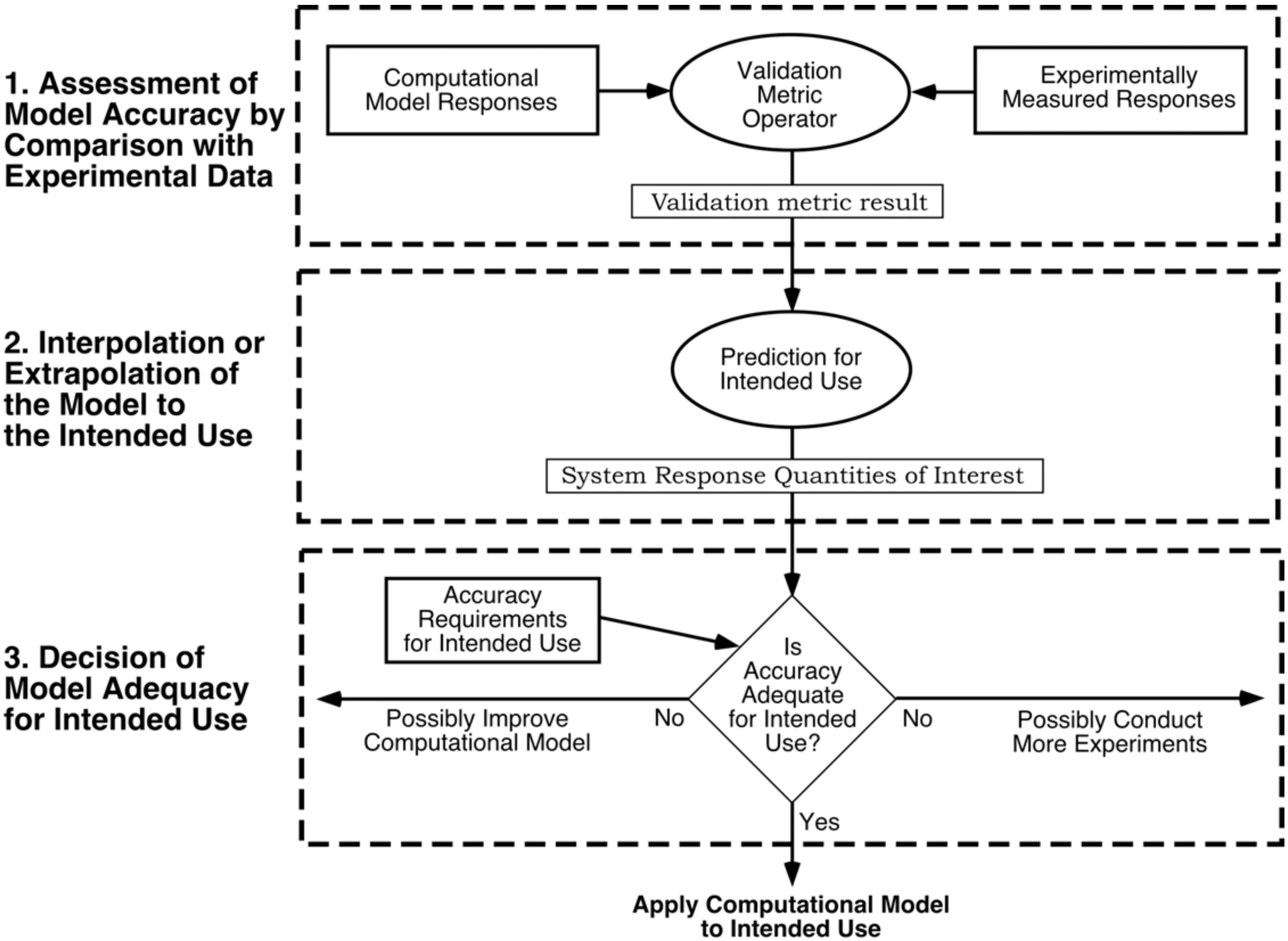
Review of Verification Processes

- **Definition used by AIAA, ASME and ASC program:**
  - **Verification: The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model.**
- **Code Verification: Verification activities directed toward:**
  - **Finding and removing mistakes in the source code**
  - **Finding and removing errors in numerical algorithms**
  - **Improved software reliability using software quality engineering practices**
- **Solution Verification: Verification activities directed toward:**
  - **Assuring the correctness of input and output data for the problem of interest**
  - **Estimating the numerical solution error (discretization error and iterative solution error)**

Method to Detect Sources of Errors in Code Verification



Three Aspects of Validation



Sandia National Laboratories

Benchmarks for Verification and Validation

- We recommend development of strong sense benchmarks (SSBs) for both verification and validation
- SSBs would be developed according to disciplines, e.g., fluid dynamics, solid dynamics, structural dynamics, etc
- The goal of SSBs is focused on accuracy assessment of results from candidate codes, not speed, robustness, etc.
- Characteristics for the construction of SSBs are similar for verification and validation, but notable differences exist
- SSBs should attempt to attain the status analogous to international measurement standards
  - This will be more easily attainable for verification than validation
- We recommend that comparison results of candidate codes with SSBs not be included in the SSB

Recommendations for Verification Benchmarks

- **Recommended characteristics for SSBs in verification:**
  - Purpose and scope of the benchmark should be clearly stated
  - Mathematical description of the benchmark should be precisely stated
  - Accuracy of the benchmark should be rigorously assessed
  - Benchmark should be carefully and formally documented
- **Accuracy requirements are distinctly different for the four different types of SSBs in verification:**
  - Type 1: Manufactured solutions to closely related PDEs
  - Type 2: Analytical solutions to PDEs
  - Type 3: Numerical solutions to ODEs that are analytically obtained from PDEs
  - Type 4: Numerical solutions to PDEs

Purpose and Scope of the Benchmark

- Elements of the purpose and scope that should be included:
  - Describe the general class of physical processes being modeled in the benchmark
  - State all of the initial conditions, boundary conditions, and any auxiliary conditions
  - Give examples of engineering applications that the benchmark is relevant to
  - State type of benchmark, i.e., 1, 2, 3, or 4
  - List what numerical algorithms, numerical accuracy issues, or software quality issues are being tested

Mathematical Description of the Benchmark

- Mathematical description should include:
  - All assumptions used to formulate the PDEs
  - The PDEs, ODEs, or integral equations being solved, including all sub-models
  - All initial conditions, boundary conditions, and auxiliary conditions as they apply to the differential or integral equations being solved
  - All system response quantities (SRQs) that are produced by the benchmark for comparison with a candidate code
  - If any quantities are uncertain, then a precise mathematical characterization of the uncertain quantity should be given
- Note that none of the mathematical description should be stated in discrete form
- There should be no ambiguity, or matter of opinion, remaining in the interpretation of the benchmark

Accuracy Assessment of the Benchmark

- A detailed and rigorous pedigree of the accuracy should be provided for each SRQ produced by the benchmark
- Accuracy of SRQs typically depend on:
  - Spatial coordinate
  - Temporal value
  - Parameters in the PDEs
- The difficulty of accuracy assessment greatly depends on the type of benchmark, i.e., 1, 2, 3, or 4
- The most common cause of failures of existing benchmarks is inadequate accuracy assessment

Documentation of the Benchmark

- Documentation should include all of the information discussed earlier
- In addition, sufficient information should be provided to:
  - Allow reproduction of the benchmark results by others
  - Identify any possible weaknesses in the accuracy of the benchmark results
- Also, information should be provided concerning:
  - Computer hardware used
  - Operating system version used
  - Compiler version used
  - Arithmetic precision used
  - Computer run time for each solution obtained

Comparing Candidate Code Results with Verification Benchmark Results

- Our recommendations are for formal comparisons of candidate codes with benchmarks
- Possible uses for results of formal comparisons:
  - Potential customers of commercial software
  - Governmental regulatory organizations assessing CSE software
  - Accident investigation committees examining CSE software
- Formal comparisons should contain much of the same information as the benchmark, particularly accuracy assessment of the candidate code:
  - For type 1 and 2 benchmarks, observed order of accuracy of the candidate
  - For type 3 and 4 benchmarks, difference in candidate and benchmark results as a function of mesh and temporal resolution
- We believe that comparison results of candidate codes with benchmarks should not be included in the benchmark

Recommendations for Validation Benchmarks

- Recommended characteristics for SSBs in validation:
  - Purpose and scope of the benchmark should be clearly stated
  - Description of the benchmark, experimental techniques, and facility used
  - Uncertainty quantification of the benchmark measurements
  - Benchmark should be carefully and formally documented
- Validation benchmarks must be more that a high quality traditional experiment
- The validation experiment is focused on the non-traditional customer: the model builder and the CSE analysts
- Validation experiments are much more feasible in the lower tiers of the validation hierarchy, than in the higher tiers

Purpose and Scope of the Benchmark

- Elements of the purpose and scope that should be included:
- Describe the primary types of physics, or coupled physics, that the benchmark is intended to test in the CSE model
- Describe the quantitative and qualitative SRQs measured in the experiment
- Give examples of engineering applications that the benchmark could be related to at higher tiers in the validation hierarchy

Description of the Benchmark, Experimental Techniques, and Facility

- Description of the benchmark should include:
  - Geometry of the actual hardware used in the experiment
  - Geometry of any supplementary experiments conducted
  - Experimental techniques used to measure all input quantities needed by the CSE analyst
  - Experimental procedures and operational characteristics of the facility being used
- Measurements of all input quantities needed by the CSE analyst, e.g.:
  - Initial conditions and boundary conditions
  - Material or surface properties
  - Imperfections in the test geometry or experimental facility
- Description of and measurement technique for all SRQs, e.g.:
  - Diagnostic techniques
  - Signal conditioning techniques

Uncertainty Quantification of the Benchmark Measurements

- **Describe the procedures used for:**
  - **Calibration of instruments, diagnostics, and facility operating conditions**
  - **Estimating the experimental uncertainty on all quantities, both CSE input quantities and SRQs**
- **Describe if an input quantity is either a controlled or uncontrolled quantity in the experiment**
- **If it is an uncontrolled quantity, then provide measurement of each experimental realization, if possible**
- **Provide estimates of both bias error and random error on CSE input quantities and SRQs and characterize as either:**
  - **Intervals**
  - **Imprecise probability distributions**
  - **Precise probability distribution**

Comparing Code Results with Validation Benchmark Results

- **Comparing code results and validation benchmark measurements do not result, inherently, in pass/fail outcomes**
- **Information that should be included in comparisons:**
  - **Code verification evidence**
  - **Solution verification results**
  - **Computation of SRQs, preferably nondeterministic SRQs which result from propagating uncontrolled inputs to SRQs**
  - **Computation of validation metrics**
  - **Calibration of any model input quantities that were not measured in the experiment**
  - **Global sensitivity analysis to determine the most important model input quantities**

Closing Remarks

- The most difficult implementation issues in construction of an SSB database in V&V will be:
  - Recognition by funding sources of the value added by the database
  - Open versus restricted use of the database
  - National and international competition
- We anticipate that the construction of an SSB database will be slow, difficult, and costly
- If CSE is to attain credibility for high-consequence decision making, then SSB must be constructed

Sandia National Laboratories