SAND2009-2190P

NEAMS

Overview of Verification

Jim Stewart
Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY | Nuclear Energy

NEAMS VU Workshop
April 7, 2009

Verification: Some Definitions

- **Definition used by AIAA**

  The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model.

- **Definition used by ASME**

  The process of determining that a computational model accurately represents the underlying mathematical model and its solution.

- **Definition used by DoD M&S Coordination Office**

  1. The process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications. 2. The process of determining that a model or simulation faithfully represents the developer's conceptual description and specifications. Verification evaluates the extent to which the model or simulation has been developed using sound and established software and system engineering techniques.

Verification answers the questions:
"Are we solving the equations correctly?"
"Are we solving the equations to sufficient accuracy?"
"Did I program what I thought I did?"

Verification Quiz

- **True or False:**

Good agreement with experiment means the equations are being solved correctly.

(Corollary: Comparing calculations with experiments does not address verification)

Verification has Two Primary Components

- **Code Verification**
  - Software Quality Engineering (SQE)
    - *Necessary ingredient, but not itself sufficient*
  - Numerical Algorithm Verification
    - *Verification testing (Order-of-Accuracy Tests)*
      - Eliminate code bugs
      - Eliminate inadequate algorithms
    - *Application-specific Verification Test Suite (VERTS) coverage analysis*

- **Solution Verification**
  - Assess adequacy of spatial and temporal discretization
    - *Mesh sensitivity studies*
    - *A Posteriori error estimation*
    - *Formal mesh refinement (e.g., Richardson extrapolation)*
  - Assure correctness of user-input algorithm parameters

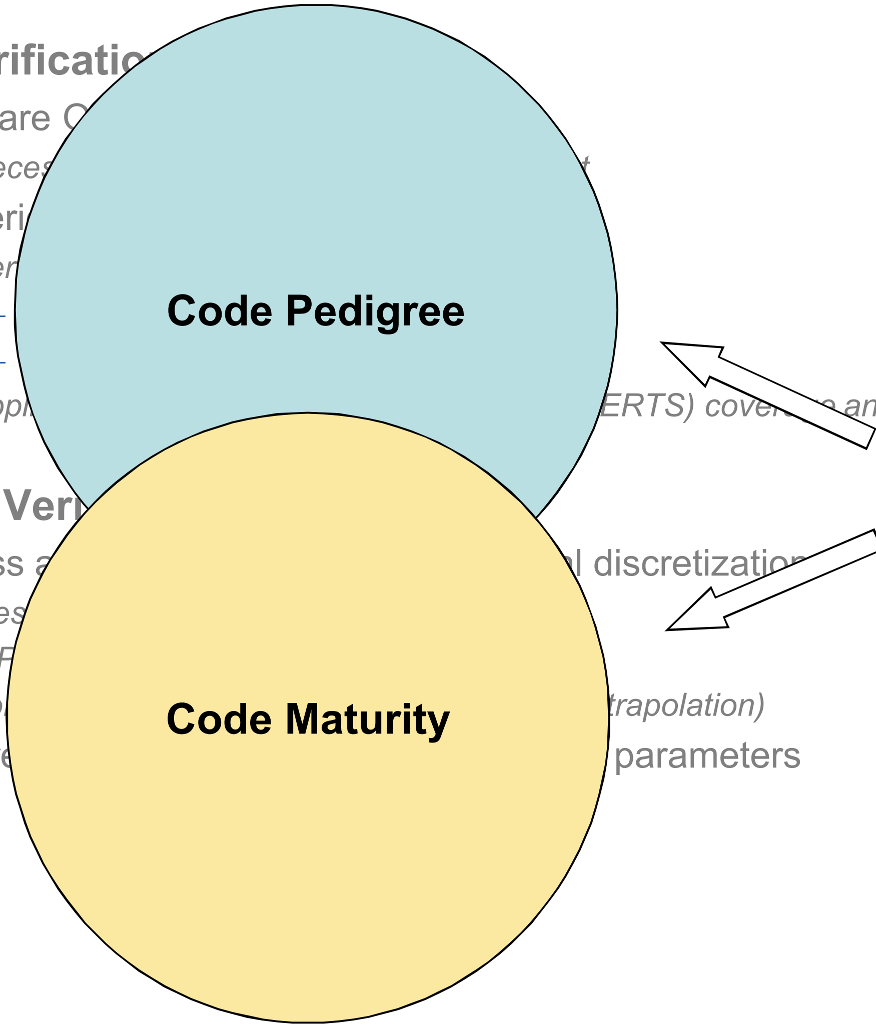
More on Solution Verification

Solution verification addresses the following questions:

- **In the context of *model validation*:**
  – Are numerical errors obscuring or undermining comparisons of calculations with experimental data?
- **In the context of *predictive simulation*:**
  – Is the solution accuracy adequate for the intended application?

How Does Verification Support Licensing?

- **Code Verificatio**
  - Software C
    - *Neces ... t*
  - Numeri
    - *Ver*
      - –
      - –
    - *Appl ... ERTS) coverage analysis*
- **Solution Veri**
  - Assess a ... l discretization
    - *Mes*
    - *A F*
    - *Fo ... trapolation)*
  - Assure ... parameters



Jerry Brock
Verification Supports Code Maturity and Fuels Licensing

The Remaining Talks Address Specific Areas

- **Code Verification**
  - Software Quality Engineering (SQE)
    - *Necessary ingredient, but not itself sufficient*
  - Numerical Algorithm Verification
    - *Verification testing (Order-of-Accuracy Tests)*
      - Eliminate code bugs
      - Eliminate inadequate algorithms
    - *Application-specific Verification Test Suite (VERTS) coverage analysis*

Mike Eldred
Software Quality Engineering – A DAKOTA Assessment

Kambiz Salari
Code Verification – Beyond SQE

- **Solution Verification**
  - Assess adequacy of spatial and temporal discretization
    - *Mesh sensitivity studies*
    - *A Posteriori error estimation*
    - *Formal mesh refinement (e.g., Richardson extrapolation)*
  - Assure correctness of user-input algorithm parameters

Brian Carnes
Overview of Solution Verification

Verification: When and Who?

- **When is verification done?**
  - Code Verification is done *before* Solution Verification
  - …which is generally done *before* Uncertainty Quantification
- **Who has the primary responsibility?**
  - For code verification, the *code developers*
  - For solution verification, the *code users*
  - *To be effective, these activities are integrated, team efforts supporting the Born-Assessed framework*
- **What can/should the VU program do?**
  - Serve as the primary integrator of these groups
  - Develop and support verification processes within the Born-Assessed framework
    - *E.g., Verification testing, verification testing environments (scripts, etc.), software for solution feature extraction, archival of results, technical peer review, etc.*
  - Contribute to code development for needed verification functionality (e.g., manufactured solutions, adjoints, error estimators, etc.)
  - Perform needed R&D to close capability gaps *(see last two slides)*

New Tools to Assess the Maturity and Confidence of M&S Efforts

- **The Predictive Capability Maturity Model (PCMM)**
  - Version 1 published in 2007: Oberkampf, Pilch, and Trucano, "Predictive Capability Maturity Model for Computational Modeling and Simulation," Sandia Report SAND2007-5984, October 2007.
  - Version 2 appeared in 2008
  - *Goal:* To judge the usefulness, or *confidence*, in a predictive capability
- **NASA's Credibility Assessment Scale (CAS)**
  - Published in 2008: "Standard for Models and Simulations," NASA Technical Standard NASA-STD-7009.
  - Similar in scope and content to the PCMM (some of the same people were involved in the development of both)
- **Others…**

The *Predictive Capability Maturity Model* (PCMM)

(Version 1: Oberkampf, Pilch, and Trucano; 2007)

MATURITY / ELEMENT	**Maturity Level 0** Low Consequence, Minimal M&S Impact, e.g., Scoping Studies	**Maturity Level 1** Moderate Consequence, Some M&S Impact, e.g., Design Support	**Maturity Level 2** High-Consequence, High M&S Impact, e.g., Qualification Support	**Maturity Level 3** High-Consequence, Decision Making Based on M&S, e.g.. Qualification or Certification
Representation and Geometric Fidelity What features are neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or geometric fidelity for the system and boundary conditions (BCs)	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components is defined	• Limited simplification or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components • Some peer review conducted	• Essentially no simplification or stylizationof components in the system and BCs • Geometry or representation of all components is at the detail of “as built,” e.g., gaps, material interfaces, fasteners • Independent peer review conducted
Physics and Material Model Fidelity How fundamental are the physics and material models and what is the level of model calibration?	• Judgment only • Model forms are either unknown or fully empirical • Few, if any, physics-informed models • No coupling of models	• Some models are physics based and are calibrated using data from related systems • Minimal or ad hoc coupling of models	• Physics-based models for all important processes • Significant calibration needed using separate-effects tests (SETs) and integral-effects tests (IETs) • One-way coupling of models • Some peer review conducted	• All models are physics based • Minimal need for calibration using SETs and IETs • Sound physical basis for extrapolation and coupling of models • Full, two-way coupling of models • Independent peer review conducted
Code Verification Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results?	• Judgment only • Minimal testing of any software elements • Little or no SQE procedures specified or followed	• Code is managed by SQE procedures • Unit and regression testing conducted • Some comparisons made with benchmarks	• Some algorithms are tested to determine the observed order of numerical convergence • Some features & capabilities (F&Cs) are tested with benchmark solutions • Some peer review conducted	• All important algorithms are tested to determine the observed order of numerical convergence • All important F&Cs are tested with rigorous benchmark solutions • Independent peer review conducted
Solution Verification Are numerical solution errors and human procedural errors corrupting the simulation results?	• Judgment only • Numerical errors have unknown or large effect on simulation results	• Numerical effects on relevant SRQs are qualitatively estimated • Input/output (I/O) verified only by the analysts	• Numerical effects are quantitatively estimated to be small on some SRQs • I/O independently verified • Some peer review conducted	• Numerical effects are determined to be small on all important SRQs • Important simulations are independently reproduced • Independent peer review conducted
Model Validation How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?	• Judgment only • Few, if any, comparisons with measurements from similar systems or applications	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown exper-imental uncertainties	• Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs • Some peer review conducted	• Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs • Independent peer review conducted
Uncertainty Quantification and Sensitivity Analysis How thoroughly are uncertainties and sensitivities characterized and propagated?	• Judgment only • Only deterministic analyses are conducted • Uncertainties and sensitivities are not addressed	• Aleatory and epistemic (A&E) uncertainties propagated, but without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• A&E uncertainties segregated, propagated, and identified in SRQs • Quantitative sensitivity analyses conducted for most parameters • Numerical propagation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• A&E uncertainties comprehensively treated and properly interpreted • Comprehensive SAs conducted for parameters and models • Numerical propagation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted

The *Predictive Capability Maturity Model* (PCMM)

(Version 1: Oberkampf, Pilch, and Trucano; 2007)

MATURITY / ELEMENT	Maturity Level 0 Low Consequence, Minimal M&S Impact, e.g., Scoping Studies	Maturity Level 1 Moderate Consequence, Some M&S Impact, e.g., Design Support	Maturity Level 2 High-Consequence, High M&S Impact, e.g., Qualification Support	Maturity Level 3 High-Consequence, Decision Making Based on M&S, e.g.. Qualification or Certification
Representation and Geometric Fidelity What features are neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or geometric fidelity for the system and boundary conditions (BCs)	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components is defined	• Limited simplification or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components • Some peer review conducted	• Essentially no simplification or stylizationof components in the system and BCs • Geometry or representation of all components is at the detail of "as built," e.g., gaps, material interfaces, fasteners • Independent peer review conducted

Remarks on the PCMM:

- *The PCMM is now required for all nuclear weapons V&V assessments at Sandia*
- *The model is applied to a particular application (not just a code)*
- *The maturity levels should be scored by a team, not just an individual*
  - *E.g., Code developers, analysts, customers, SME's, VU experts, etc.*
- *For each PCMM element, two scores are assigned:*
  1. *Required*
  2. *Current assessment*

Model Validation How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy?	• Few, if any, comparisons with measurements from similar systems or applications	accuracy of SRQs not directly relevant to the application of interest • Large or unknown experimental uncertainties	accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs • Some peer review conducted	accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs • Independent peer review conducted
Uncertainty Quantification and Sensitivity Analysis How thoroughly are uncertainties and sensitivities characterized and propagated?	• Judgment only • Only deterministic analyses are conducted • Uncertainties and sensitivities are not addressed	• Aleatory and epistemic (A&E) uncertainties propagated, but without distinction • Informal sensitivity studies conducted • Many strong UQ/SA assumptions made	• A&E uncertainties segregated, propagated, and identified in SRQs • Quantitative sensitivity analyses conducted for most parameters • Numerical propagation errors are estimated and their effect known • Some strong assumptions made • Some peer review conducted	• A&E uncertainties comprehensively treated and properly interpreted • Comprehensive SAs conducted for parameters and models • Numerical propagation errors are demonstrated to be small • No significant UQ/SA assumptions made • Independent peer review conducted

Verification Research and Implementation Issues*

- **Develop manufactured solutions for a wide range of physics and engineering disciplines for order of accuracy testing**
- **Develop improved measures of code coverage in testing software; line coverage in regression testing is inadequate**
- **Develop less expensive and more robust methods for estimating spatial and temporal discretization error**
- **Develop numerical error estimators for nonlinear parabolic and hyperbolic PDEs (including multi-physics problems)**
- **Develop methods to integrate verification into UQ assessments**
- **Require improved code verification evidence from code developers**

"I've already refined the mesh
down to the microstructure of the metal!"

**Thanks to Bill Oberkampf for much of this slide*

Path Forward: A Balanced Approach

1. **Integration**
   - Ensure each IPSC team has a practical V&V plan that includes sufficient verification
   - Help establish uniform SQE guidelines and shared processes, coverage metrics, etc.
   - Develop and deploy shared verification software tools across the IPSC teams
   - Support born-assessed process
2. **IPSC development**
   - Work closely with the IPSC teams to design and implement appropriate verification methods (manufactured solutions, adjoints, error estimators, adaptive capabilities, etc.)
3. **Verification research to address capability gaps**
   - *See previous slide*
   - *We have written a short white paper on broader NEAMS VU capability needs*

Discussion?

Extra Slides

PCMM (Version 2)

		Maturity Level 0 Low Consequence, Minimal M&S Impact, e.g. Scoping Studies	Maturity Level 1 Moderate Consequence, Some M&S Impact, e.g. Design Support	Maturity Level 2 High-Consequence, High M&S Impact, e.g. Qualification Support	Maturity Level 3 High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification
Code Verification (CVER) Are software errors or algorithm deficiencies corrupting the simulation results?	Software Quality Engineering practices (SQE: how mature are the SQE practices)	• Judgment only, codes informally managed to SQE practices or no documented SQE process requirements • Software process is characterized as ad hoc, and occasionally even chaotic	• Codes managed to repeatable and defined SQE practices • Repeatable: Basic project management processes are established to track cost, schedule, and functionality. • Defined: The software process for both management and engineering activities is documented, standardized, and integrated into a standard process for the organization and applied in a graded manner.	• The SQE process in managed • Managed: Detailed measures of software process and product quality are collected. Both the software process and products are quantitatively understood and controlled.	• The SQE process is optimized • Optimized: Continuous process improvement is enabled by quantitative feedback from the process and from piloting innovative ideas and technologies.
	Software Quality Assessment (SQA: assurance that code development is managed to an appropriate level of process maturity)	• Judgment only, no assessment to SQE practices	• Self assessment and documentation of full or partial compliance to organizational SQE practices by code team • Self-assessments or formal assessments have identified compliance gaps	• Formal assessment and documentation of full compliance to organizational SQE practices by group external to the code development team	• Formal assessment and documentation of compliance to SQE practices and accreditation to an appropriate level of a nationally recognize set of SQE standards (e.g., CMMI, ISO9000, IEEE, etc) by team external to the code development team
	Test coverage (can the user be confident that the code is adequately tested for the intended application)	• Judgment only, minimal testing of any software elements	• Sustained unit and regression testing and/or limited scope Verification Test Suite (VERTS) routinely conducted with 75% coverage • Note: unit and regression problems track code drift and not necessarily code correctness • Here, VERTS address *comparison (not convergence)* to the *correct answer* • Coverage: Line, function, or feature and capability (F&C)	• Sustained VERTS re-run regularly w 75% F&C coverage and 75% coverage of all 2-way interactions of F&C • VERTS address *convergence behavior to the correct answer*	• Sustained VERTS re-run regularly w 75% coverage of F&C and all their interactions (2-way, 3-way, etc)

PCMM (Version 2)

PCMM Practice		Maturity Level 0 Low Consequence, Minimal M&S Impact, e.g. Scoping Studies	Maturity Level 1 Moderate Consequence, Some M&S Impact, e.g. Design Support	Maturity Level 2 High-Consequence, High M&S Impact, e.g. Qualification Support	Maturity Level 3 High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification
Solution Verification (SVER) Are human procedural errors or numerical solution errors corrupting simulation conclusions?	Numerical Solution Errors (what is the impact of numerical solution errors on relevant SRQs)	• Judgment only, numerical solution errors not addressed	• Sensitivity to discretization and algorithm parameters explored for some System Response Quantities (SRQs)	• Numerical errors estimated for discretization and algorithm parameters for relevant SRQs	• Numerical errors rigorously quantified for all relevant SRQs
	Input/Output Verification	• Input/output not verified	• Input/output verified only by the analysts	• Input/output data independently verified	• Input/output data independently verified, calculation results reproduced independently
	Technical Review (confirmation that the solution verification activities are relevant, adequate, and carried out in a quality manner)	• Judgment only, no technical review of the solution verification evidence	• Informal technical review or technical review from within the project team or stakeholder community only	• Formal technical review by Subject Matter Experts (SMEs) external to the project team or stakeholder community	• Formal technical review by SMEs external to the project team or stakeholder community • Formal technical review SMEs played an oversight and approval role of solution verification activities