SAND2011-1971C

Overview of Verification, Validation, Uncertainty Quantification, and QMU

Kenneth Alvin
Sandia National Laboratories*
Albuquerque, NM

Presentation at Sandia National Laboratories, ESP700 Training Class

22 March 2011

*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.

Outline

- Introduction
- Background on:
  - Verification & Validation (V&V)
  - Quantification of Margins and Uncertainties (QMU)
- Engineering Application
- V&V at Sandia Today
- Predictive Capability Maturity Model (PCMM)
- Summary

Sandia National Laboratories

Introduction: V&V and QMU at the NNSA Labs

- V&V was recognized early in the ASC Program as necessary to mission success and impact
  - We had to establish credibility for our predictions
- As our understanding and strategy has evolved, we have increasingly focused on how verified and validated simulations should be used in prediction
- QMU stands for "Quantification of Margins and Uncertainties"; it is a framework for using simulation in risk-informed decision-making
- Achieving QMU assessments for our target applications is the principal driver for our V&V and UQ strategies

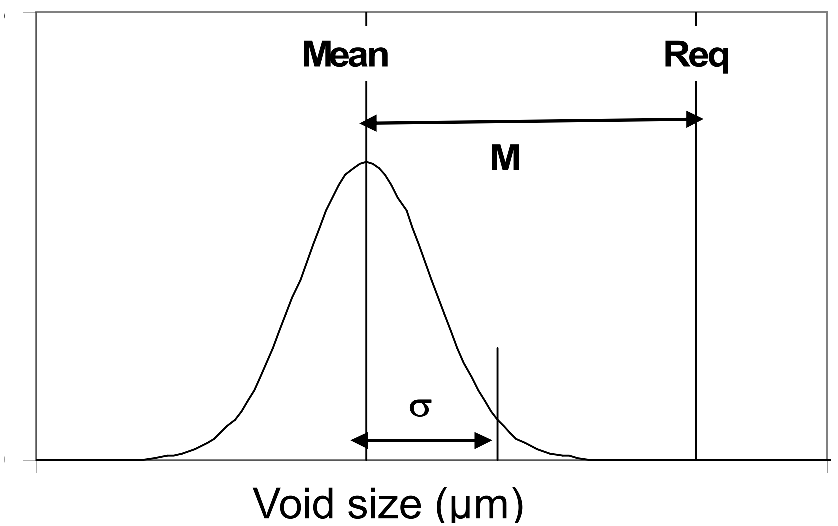
What is QMU?

- QMU: Quantification of Margins and Uncertainties
  - Margin - Difference between system's nominal/median performance vs. a do-not-exceed threshold.
- Sandia is employing QMU to:
  - Understand performance margins, uncertainties, & <u>changes with time</u>.
  - Provide higher confidence that our products meet requirements.
- NNSA and DoD customers have statistics-based requirements:
  - Probability of an Inadvertent Nuclear Detonation < 1×10^{-m} for normal environments.
  - Probability of an Inadvertent Nuclear Detonation < 1×10^{-n} for abnormal environments.
    - What is the probability of a welded joint failure if weapon Wxx is dropped from *k* feet?
- To answer these questions we need data <u>ensembles</u>:
  - Test data (both historical and new).
  - Simulation data (both medium- and high-fidelity).

Sandia National Laboratories

QMU Methods: A Quick Overview

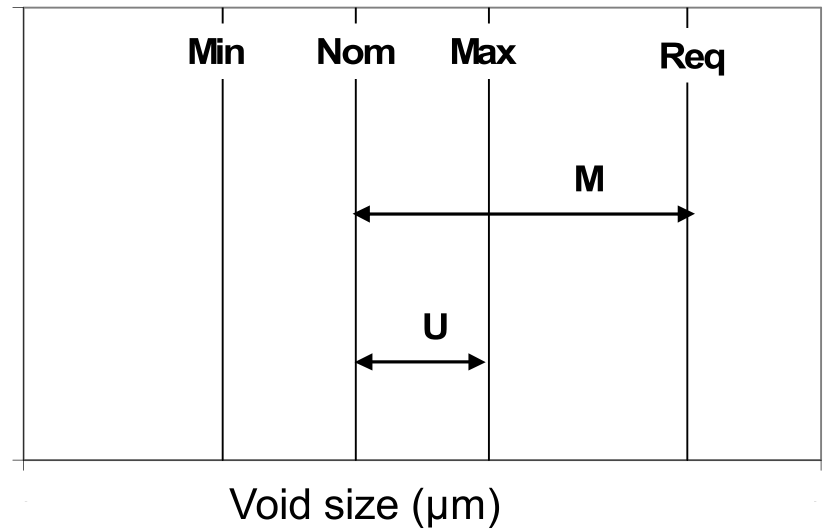
Aleatory (Probabilistic) Uncertainties



Hypothetical Example #1:

- Probability distribution curve obtained from many samples of void size in a material.
- "Req" = do-not-exceed void size.
- Margin (M), Standard Deviation (σ)
- M/σ is called the "k-factor"
- Can estimate Probability(void size > Req) and track M/σ changes w.r.t. time.

Epistemic (Lack of Knowledge) Uncertainties



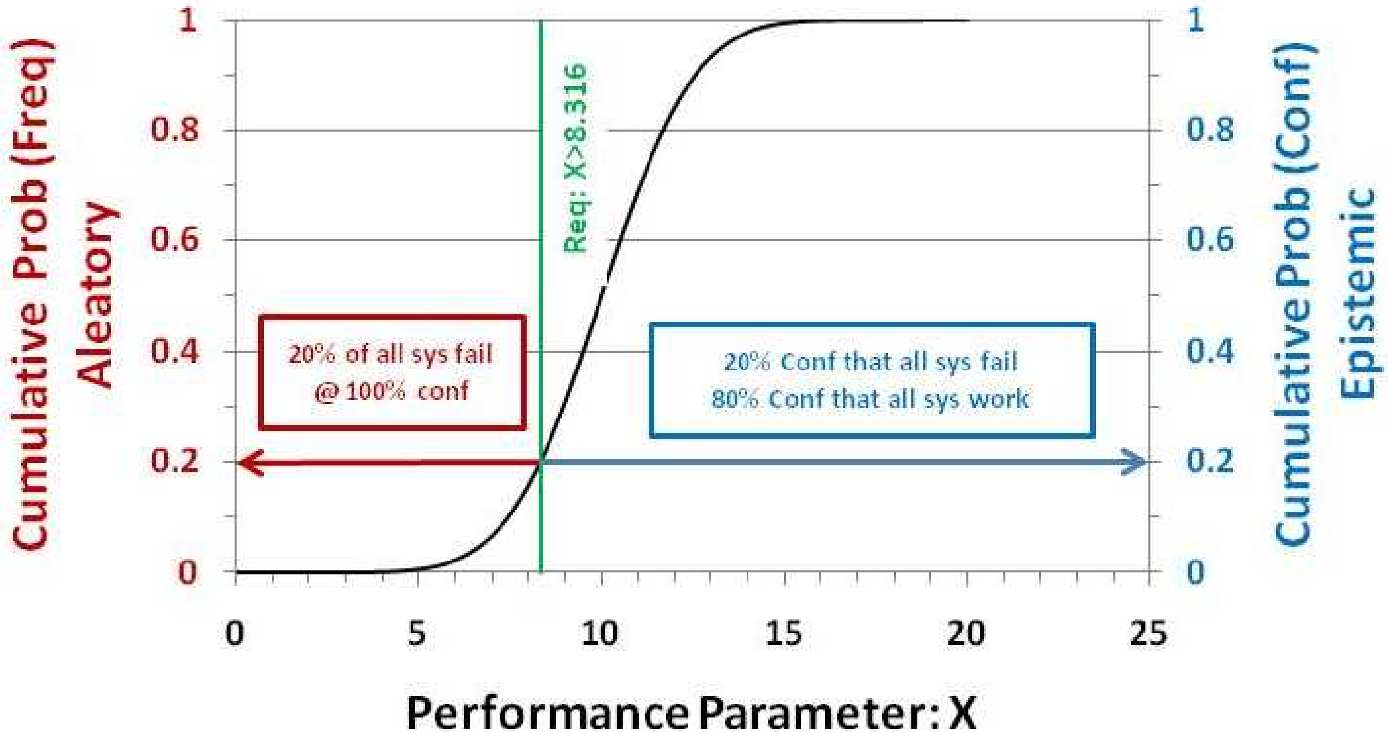
Hypothetical Example #2:

- Insufficient data to specify a probability distribution – only know min & max void size.
- Margin (M), Uncertainty (U)
- M/U is the "confidence ratio"
- Cannot estimate Prob(void size > Req), but can track M/U w.r.t. time.

Sandia National Laboratories

In QMU, the Type of Uncertainty Matters a Lot!

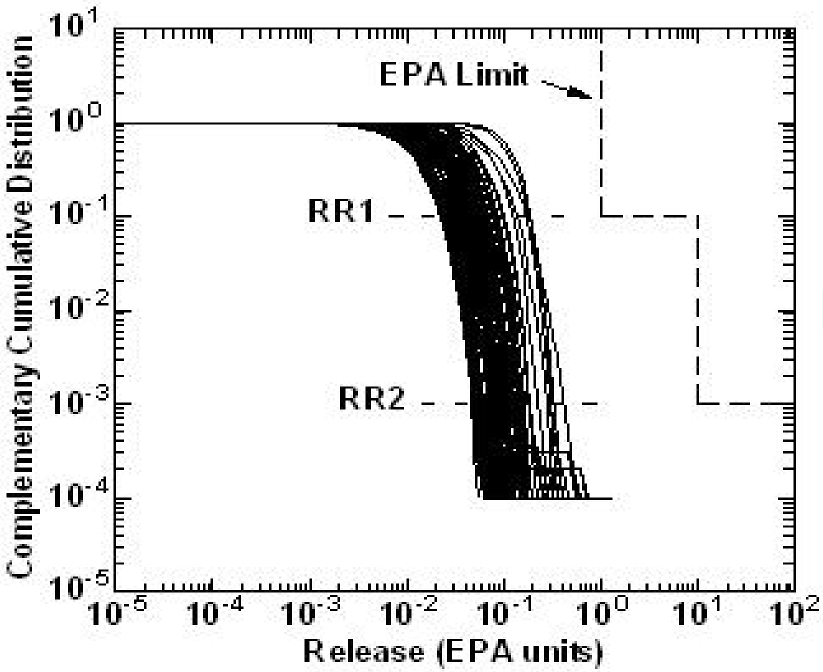
- **Aleatory uncertainty:** (perceived) randomness in the occurrence of future events **(frequency interpretation)**
- **Epistemic uncertainty:** Lack of knowledge wrt appropriate value to use for a quantity that has a fixed value in the context of a specific analysis **(confidence or belief interpretation)**

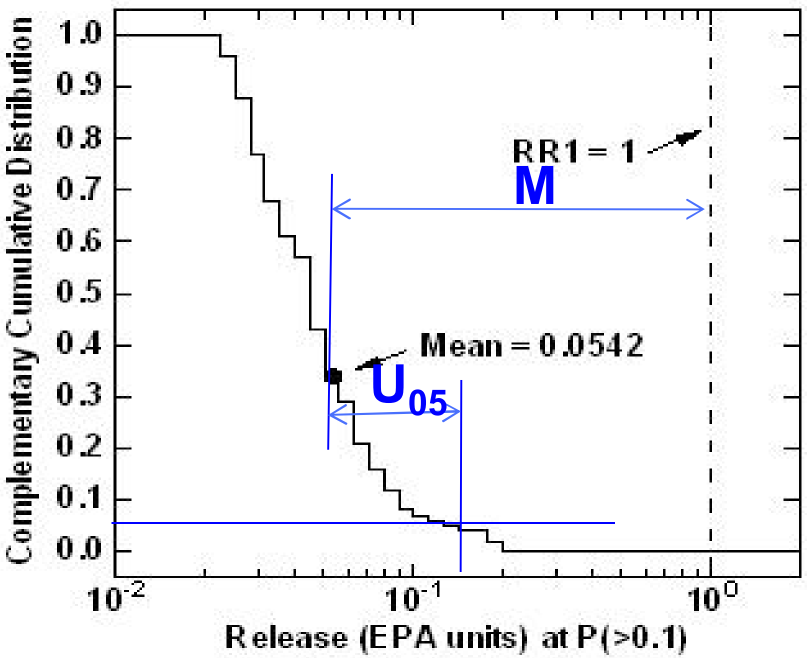


The distinction between aleatory uncertainties and epistemic uncertainties matters

Sandia National Laboratories

QMU Isn't New: WIPP Performance Assessment





M = Req – Mean = 1 - 0.0542 = 0.9458

U = U_{05} – Mean = 0.14 – 0.0542 = 0.0858

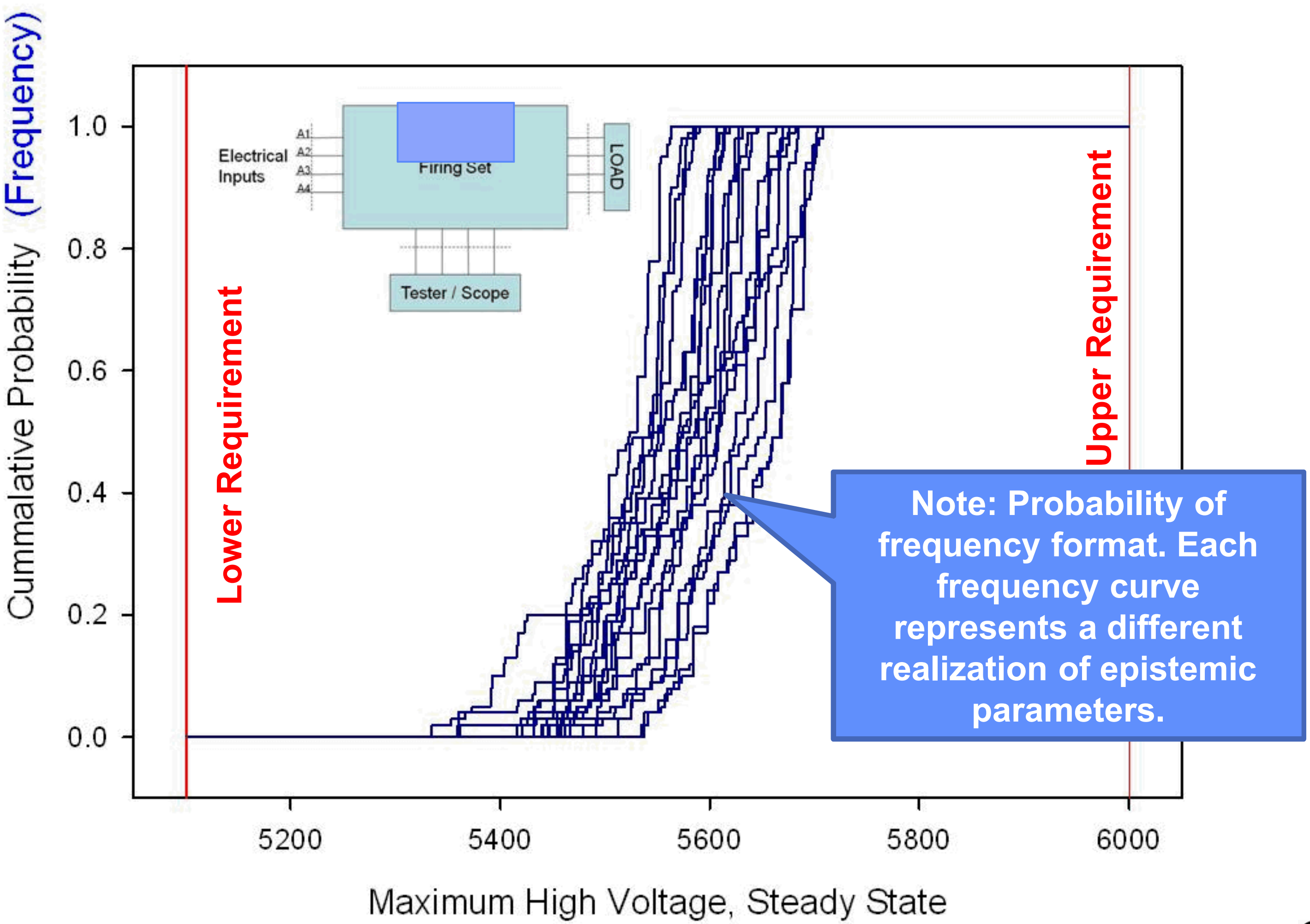
M/U = 0.9458/0.0858 = 11 (Definition is Not Unique)

A lot of information is lost in distilling QMU into a single number, M/U

Sandia National Laboratories

Often We Rely on M&S to Assess Performance

e.g., Environmental Extrapolation, Aging, etc



Note that QMU ≠ UQ …

QMU = PCMM(credibility)

and

UQ{ U_epistemic(Validation)
+ E_numerical(Verificaton)
+ U_aleatoric(Variabilities) }

Verification, Validation, and Uncertainty Quantification are the Science Behind QMU

- **Verification – "Are we solving the equations correctly?"**
  - Correctness of implemented mathematical algorithms.
  - Convergence to the correct answer, at the correct rate, as model is refined.

- **Validation – "Are we solving the right equations?"**
  - Correctness of physical models and sufficiency for the application.

- **Uncertainty Quantification (UQ):**
  - Statistical propagation of uncertainty through a simulation model, and statistical interpretation of model response.

- **Quantification of Margins and Uncertainties (QMU):**
  - Using the simulation model to make system performance predictions with quantified uncertainty, and with quantified margins with respect to system performance requirements.

Sandia National Laboratories

V&V Is a Tough Sell

- V&V is expected, but not well understood, by decision makers.
  - V&V is, in a nutshell, all about putting "correct" math methods and physics models in our codes.
  - We're <u>expected</u> to produce "correct" codes.
  - "If you haven't been doing V&V all along, then what have you been doing with my _____ money?"

- What's different now?
  - Computational simulation is different now than 10-20-30 years ago (e.g., auto industry, aircraft industry, nuclear weapons industry)
    - We're making million/billion $ decisions that are heavily influenced by comp. sim.
  - Definition of "correct codes/models" (see above) is now changing.
  - "Before I spend $M/$B on a decision, I want evidence of the correctness of your comp. sim. model and results."

- <u>Issues</u>: Code correctness is expected, quantified evidence of correctness (via V&V) takes extra effort beyond traditional code development work, you can't V&V every aspect of a code/model/project, and it's hard to retrofit V&V into a study that is already completed.

What Does Sell?

- Using comp. sim. results to aid decision making sells.
  - Decision making is based on knowing the tradeoffs for competing objectives, due to variations in designer-controllable parameters.
  - Quantities of interest: cost & performance
  - This sells (re: facility design hardness study):
    - "If you increase factor1 by A% and lower factor2 by B%, you reduce cost by X% and decrease the probability of kill by Y%."
    - "By the way, here is the evidence (tucked away in a report appendix) for the validity of predictions A, B, X, and Y."
  - This also sells:
    - "If were going to perform a comp. sim. study that influences a $M/$B decision, then let's carve out $m to run a V&V study to make sure we're getting good data, and $n to perform an adequate sensitivity/uncertainty analysis."

- Punch Line:
  - V&V doesn't sell for it's own sake.
    - Decision makers don't care about the rate of convergence of an iterative mathematical method, or % line coverage of tests.
    - For $M/$B issues, decision makers do care that you got the right answer and they expect a technical pedigree (aka "provenance") for your work.
  - V&V sells when it is included as an aid to decision making.
    - i.e., when V&V provides supporting evidence (provenance) to sensitivity analysis and UQ results on relevant technical/financial issues.

NNSA and Sandia Now Have Policies on QMU Deployment

- **NNSA draft policy (May 2007):**

  Nuclear Weapon Assessments Using Quantification of Margins and Uncertainties Methodologies:

  > **"Design agency assessments <u>shall incorporate QMU methodologies</u> as an essential part of the framework necessary for the evaluation of the performance of warhead and warhead components."**

- **Sandia directive (April 2007):**

  Steve Rottler, Vice President of Sandia Weapon Engineering:

  > **"We explicitly account for, monitor, and analyze margins and uncertainties throughout the warhead lifecycle using tools and a methodology collectively referred to as the <u>Quantification of Margins and Uncertainties</u>."**

M&S-Based QMU Results Being Used in High Consequence Decision-Making Should be Peer Reviewed

- Increased Objectivity
- Assurance of Evidence Basis for Predictive Capability Assessment
- Hedge against "unknown unknowns" that were actually "shoulda been knowns"

Sandia National Laboratories

Key Issues:
(1) Most analysts do these activities formally/informally.
(2) Amount of formal V&V needed is driven by customer needs.

Overview of the Sandia V&V Process

1 **DP Application**

Understand the application and requirements

PIRT Tables

2 **Planning**

Assess capabilities, identify gaps, & prioritize work

Validation Experiment Centered Elements

4 **Experiment Design, Execution & Analysis**

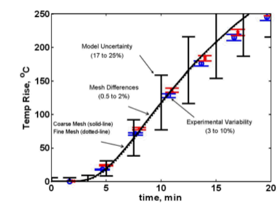
3 **Code Verification**

5 **Metrics**

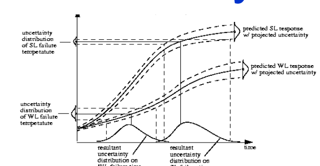
3 **Calculation Verification**

6 **Assessment**

Compare to test data



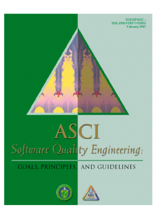
Predictions w/ quantified uncertainty

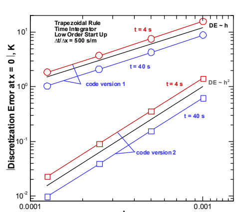


7 **Prediction & Credibility**

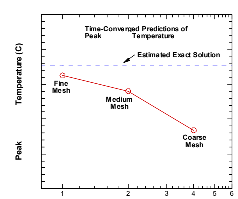
Document

Software quality practices & numerical accuracy checks on test problems





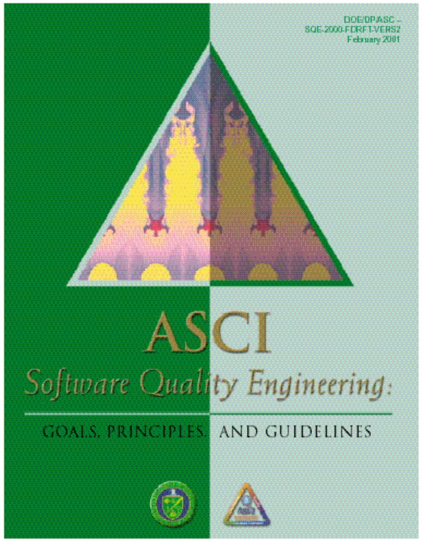
Convergence checks on engineering application



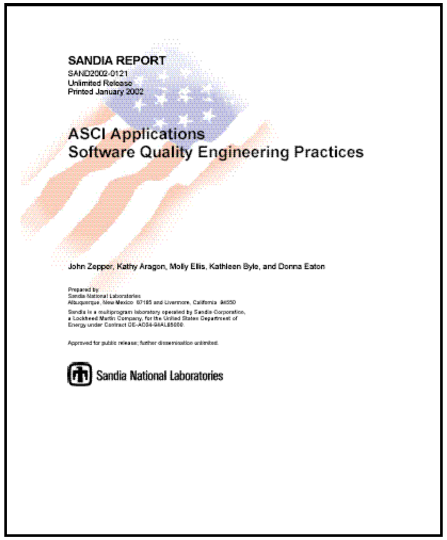
Sandia National Laboratories

Software Quality Engineering: Requirements and Assessments

Tri-Lab Guidelines



Lab-Specific Requirements



SANDIA REPORT
SAND2002-0121
Unlimited Release
Printed January 2002

ASCI Applications
Software Quality Engineering Practices

John Zepper, Kathy Aragon, Molly Ellis, Kathleen Byle, and Donna Eaton

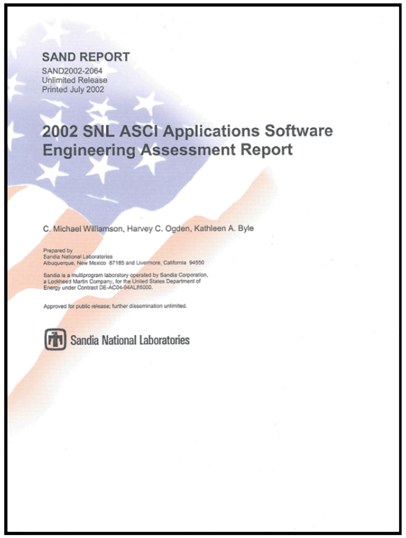
Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Indep. Internal Assessment



SAND REPORT
SAND2002-2064
Unlimited Release
Printed July 2002

2002 SNL ASCI Applications Software Engineering Assessment Report

C. Michael Williamson, Harvey C. Ogden, Kathleen A. Byle

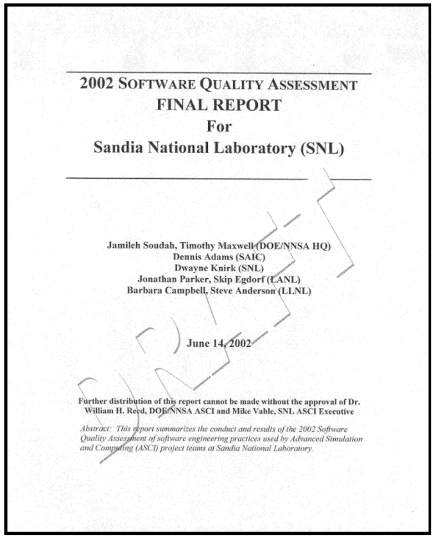
Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Indep. External Assessment



2002 SOFTWARE QUALITY ASSESSMENT
FINAL REPORT
For
Sandia National Laboratory (SNL)

Jamileh Soudah, Timothy Maxwell (DOE/NNSA HQ)
Dennis Adams (SAIC)
Dwayne Knirk (SNL)
Jonathan Parker, Skip Egdorf (LANL)
Barbara Campbell, Steve Anderson (LLNL)

June 14, 2002

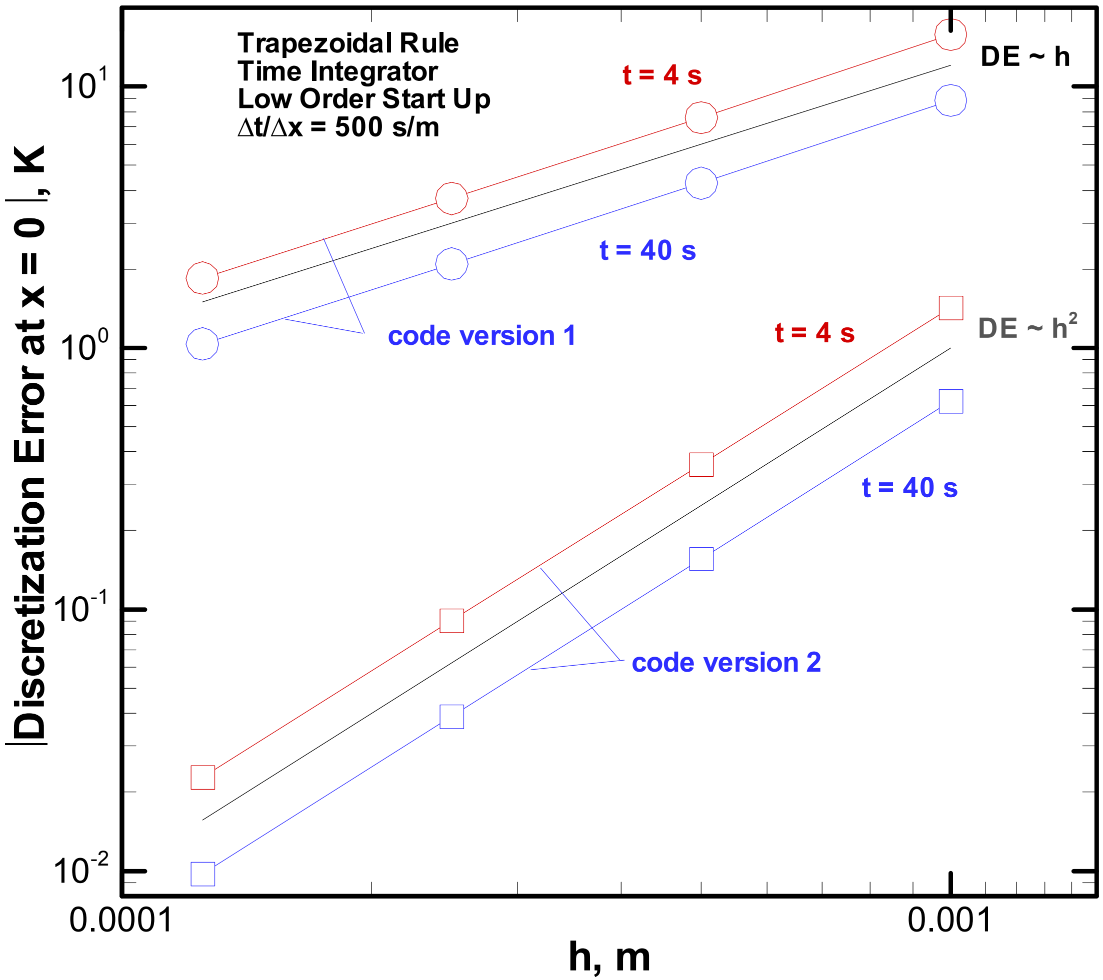
Further distribution of this report cannot be made without the approval of Dr. William H. Reed, DOE/NNSA ASCI and Mike Vahle, SNL ASCI Executive

Abstract: This report summarizes the conduct and results of the 2002 Software Quality Assessment of software engineering practices used by Advanced Simulation and Computing (ASCI) project teams at Sandia National Laboratory.

SNL/ASC Quality Management Council (AQMC)

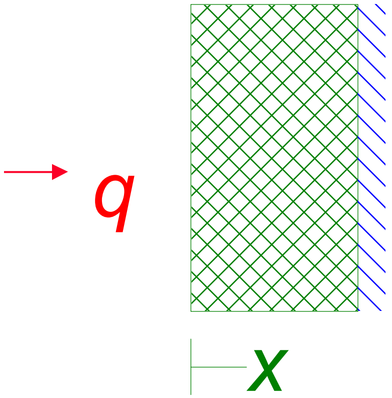
Sandia National Laboratories

Code Verification Example: Demonstrating the Correct Rate of Spatial Convergence on Analytic Tests



- VERTS mapped to required physics and code capabilities
- Code Verification of Calore (thermal response code) found bugs
- Code bug discovered and fixed based on priority and resource availability. Status tracked in code issue log, which can be accessed by analysts

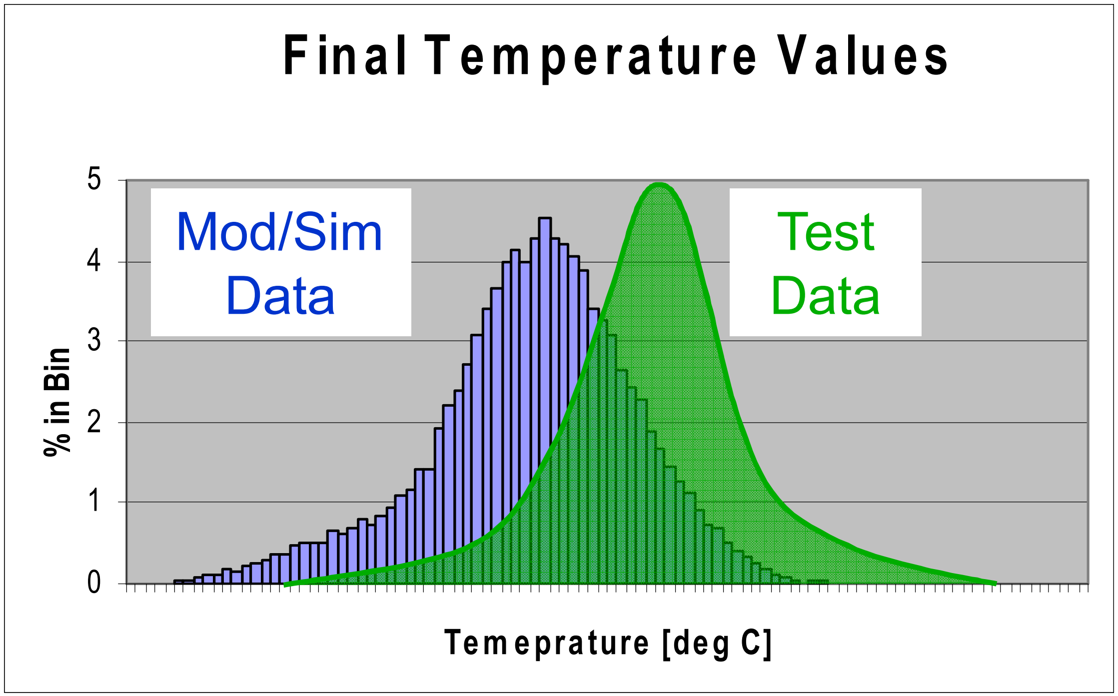
- Transient response of planar 1-D slab to constant flux



Modeled as full 3-D object

Sandia National Laboratories

Example of Uncertainty Quantification and Validation



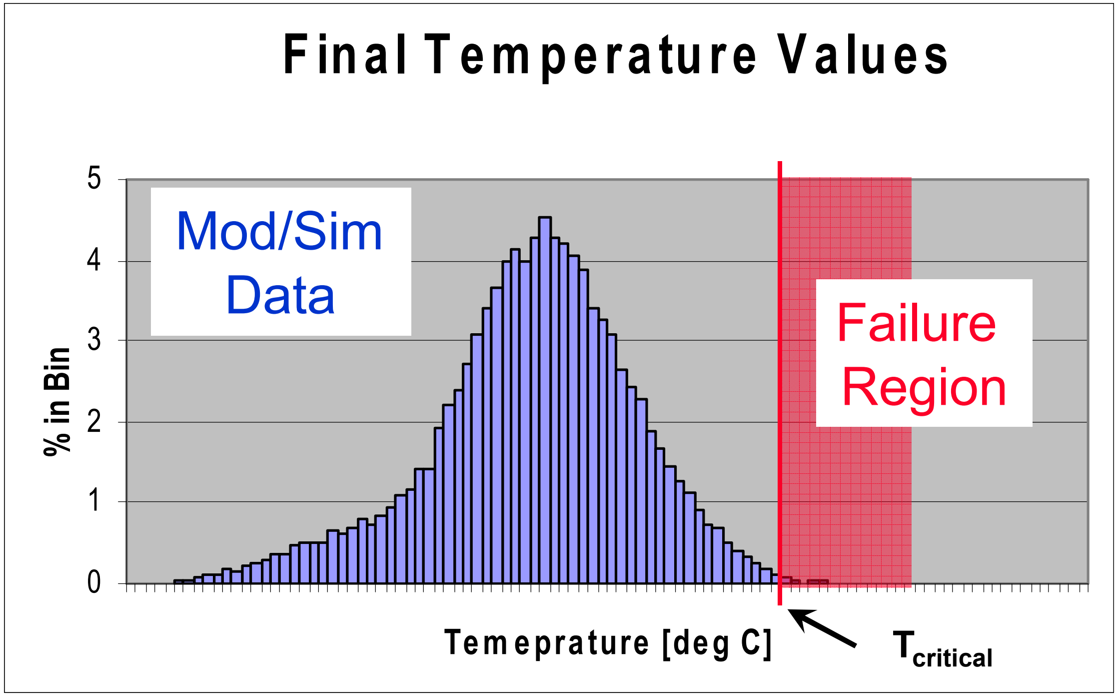
Uncertainty Quantification:
- UQ methods generate ensemble of mod/sim data.

Validation:
- Compare simulation data histogram to a test data histogram.
  - Quantify amount of "overlap" between histograms.
  - Assess sufficiency of overlap.

- **UQ methods provide statistical info on the code output data:**
  - **Probability distribution on Temperature, *given various $x_1,\ldots,x_N$ inputs.***
  - **Correlations (trends) of Temperature vs. $x_1,\ldots,x_N$.**
  - **Mean(T), StdDev(T), Probability($T > T_{critical}$)**

Sandia National Laboratories

Example of Uncertainty Quantification and QMU



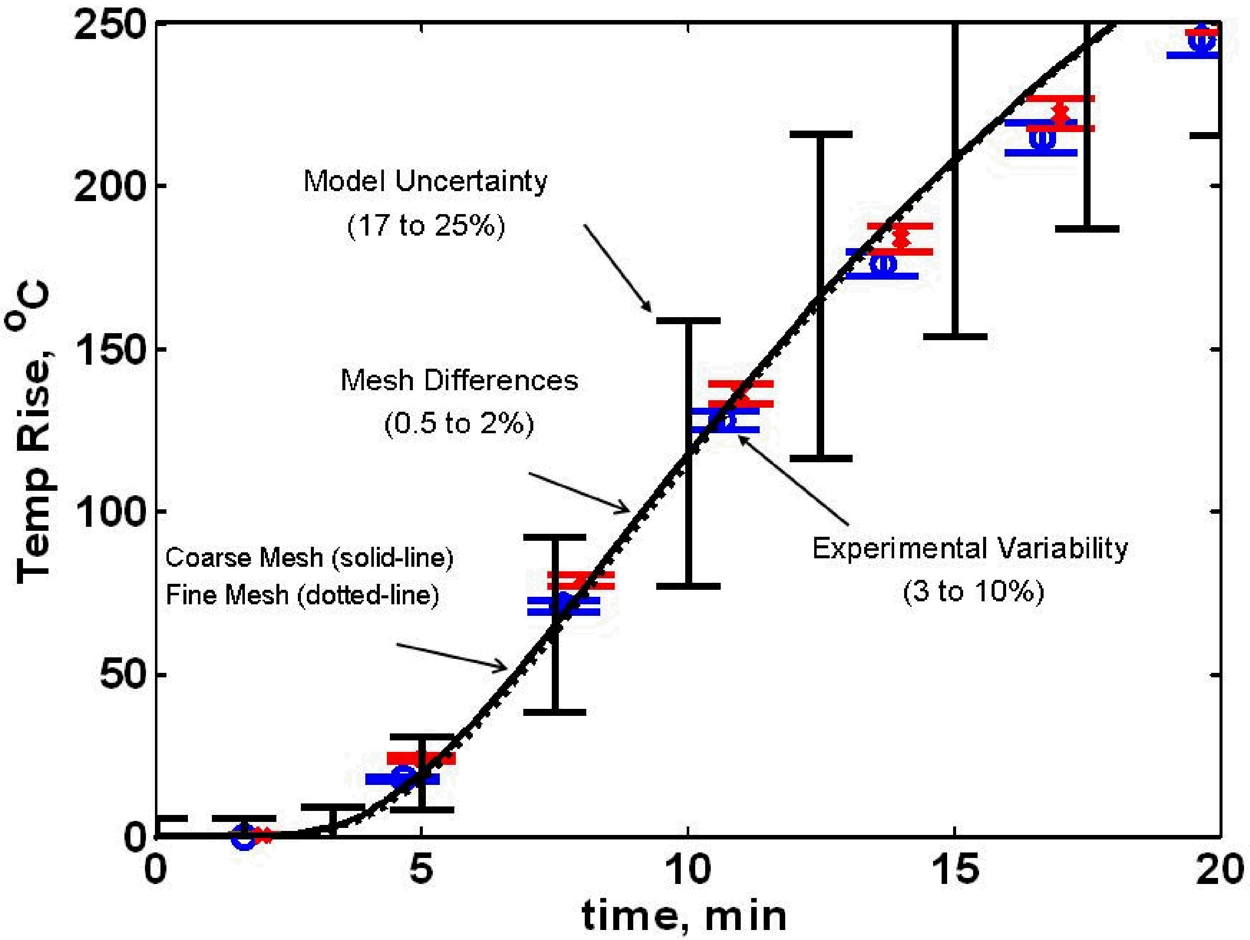
Uncertainty Quantification:
- UQ methods generate ensemble of mod/sim data.

Quantified Margins & Uncertainties:
- Estimate failure probability.
- Compare *estimated* failure probability to *allowable* failure probability, including all sources of uncertainty.

- **UQ methods provide statistical info on the code output data:**
  - **Probability distribution on Temperature, *given various $x_1,\ldots,x_N$ inputs.***
  - **Correlations (trends) of Temperature vs. $x_1,\ldots,x_N$.**
  - **Mean(T), StdDev(T), Probability($T > T_{critical}$)**

Sandia National Laboratories

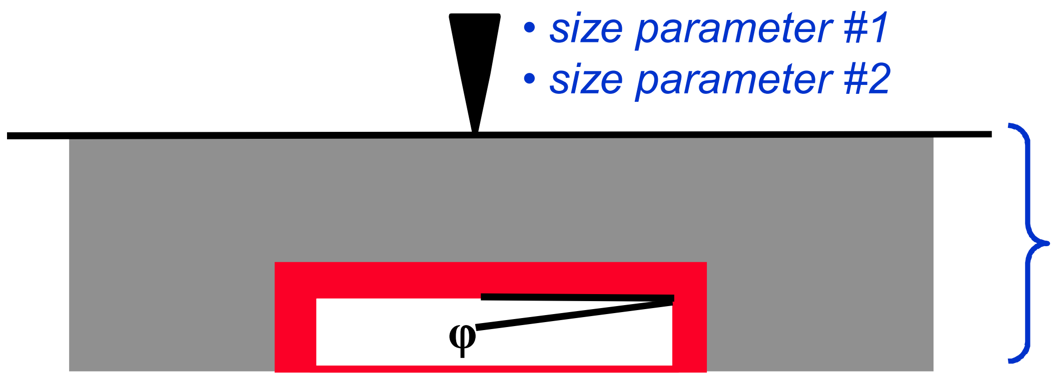
Close-up View of Validation Plot



Example: Underground Target Defeat/Protect Study

- **<u>Scenario</u>: underground target subject to an external threat**
- **<u>Goal</u>: Assess sensitivity/uncertainty in target response due to uncertainties in target construction and threat characteristics**
- **<u>Approach</u>: 9 parameters define uncertainties in threat and target**
  - **Each parameter has uncertainty specified by an interval**
- **<u>Metric</u>: deflection angle (φ) of target roof at mid-span (>critical angle=kill)**
- **<u>Tools</u>: CTH, DAKOTA, JMP statistical analysis software**

<u>Threat:</u>
- *size parameter #1*
- *size parameter #2*



<u>Target:</u>
- *Soil depth*
- *Materials*
- *Dimensions of structure – height, width, thickness, etc.*

<u>What Matters?</u>:
<u>Defense</u>: What design features (and how much $) are needed to protect my facility?
<u>Offense</u>: What facilities can I hold at risk, given my range of threat assets? What different assets (w/ different costs) I can employ to have high kill probability, given uncertainty in the target composition?

We Took a Nontraditional Approach in this Study, But Did Not Use V&V Best Practices

- Traditional approach:
  - Build computational model
  - Perform a user-guided sequence of "change one parameter at a time" runs of the computational model (~20-50 runs total - estimated)
  - Payoff: develop qualitative/quantitative parameter sensitivity and trend information for facility response (roof deflection) vs. 9 parameters
    - generate "local" performance information that looks like a derivative: e.g. d(deflection angle)/d(soil depth), etc.

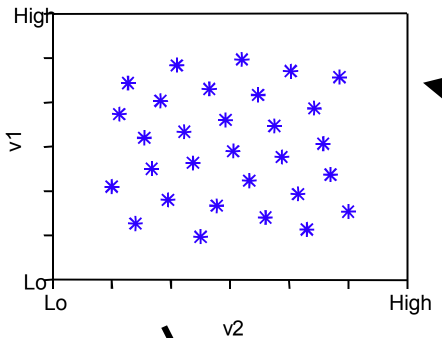
- Nontraditional approach with sensitivity/uncertainty quantification:
  - Build computational model
  - Use statistical data sampling methods to develop a "code run matrix" (initially 50 runs; 150 runs total); then run the code
  - Use statistical data analysis tools to sift the data
  - Payoff: (a) identify & quantify influence of 9 parameters, (b) produce global trend (tradeoff) models of facility response (roof deflection) vs. variations in the 9 parameters
    - you get quantified single- and multi-parameter sensitivity information
    - you get local and global tradeoff information
  - Repeat with new run matrix, if needed.
  - Cost: (a) one or more folks on the team need to know how to do the statistical sampling & analysis (training), (b) # code runs >= than traditional approach (% increase is problem dependant)

Sandia National Laboratories

Example Study: Underground Target Defeat/Protect

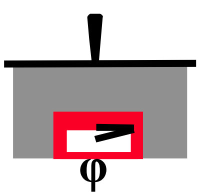
- **Example: Sensitivity analysis study with CTH and DAKOTA; simulations run on a Linux cluster; data analysis via JMP commercial statistics package**

Statistics-based "run matrix"



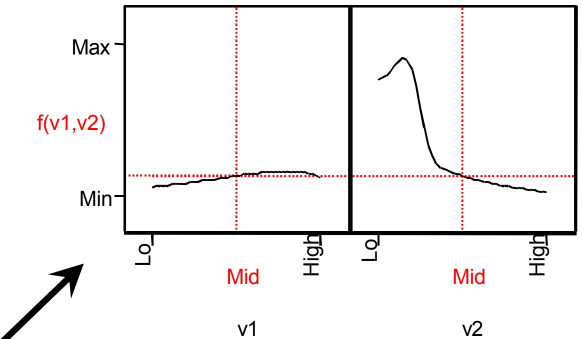
DAKOTA

CTH sims. run on a LINUX cluster



JMP
The Statistical Discovery Software

Graphics



Statistical analysis

Summary of Fit

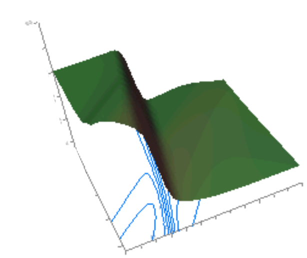
RSquare	0.866299
RSquare Adj	0.823758
Root Mean Square Error	#.#
Mean of Response	#.#
Observations (or Sum Wgts)	30

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	5343.9158	763.417	20.3638
Error	22	824.7562	37.489	Prob > F
C. Total	29	6168.6720		<.0001

Yields Qualitative and Quantitative Insight, e.g.:

- *Most important parameters*
- *Cost & Survivability tradeoff functions vs. parameter variations*



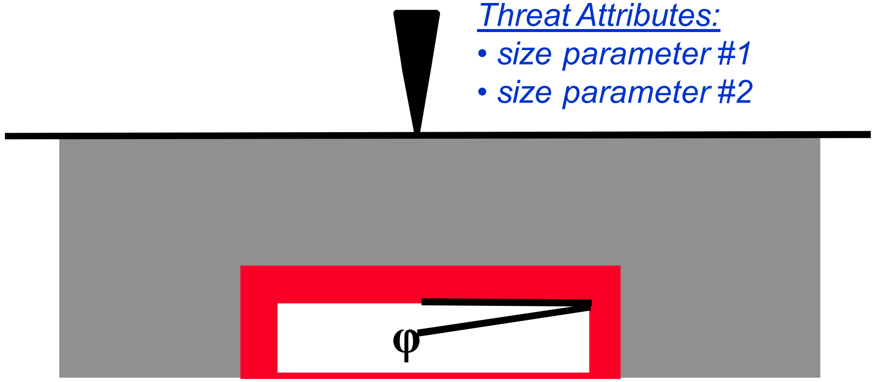
Sensitivity Analysis and UQ Study: Underground Target Defeat/Protect

Results:
- Identified 3 key target design attributes that most strongly affect facility cost & performance.

- Generated a statistical/mathematical model of roof deflection vs. 9 parameters -- useful as a planning tool for SNL and customer in considering various design tradeoff studies.

- Influenced the design of an expensive, large scale test.

Threat Attributes:
- *size parameter #1*
- *size parameter #2*



Target Attributes:
•Soil depth & type, structural materials, dimensions of structure, etc.

Where was V&V in this Study?

- Little/no formal V&V; some "historical" V&V.
  - Verification: some test problems with analytical solutions probably in the code test suite
  - Validation: considerable user experience, plus, broad testing of this code in Sandia/external community
  - Note: This study was done ~6 years ago.
- What would we do differently now?
  - Verification #1: Identify a few key analytic test problems to confirm that the code converges to (a) the right answer, and (b) at the right spatial/temporal rate.
  - Validation: Identify a few key experimental test data sets and confirm that "test data + uncertainty" and "sim data + uncertainty" agree sufficiently.
  - Verification #2: Perform some basic mesh convergence studies with the threat/target geometry to (a) confirm that the mesh is sufficiently refined that the results are converging to an answer, or (b) inform us that the mesh is not converged and that we need to be wary of the sensitivity/UQ results.
- Why would we do this extra work?
  - In almost every major V&V study that we've done, we found one or more serious errors in the math (numerics) and/or physics.
    - Best case – the code still converges to the right answer, slower than it should
    - Worst case #1 – the code doesn't converge to any answer
    - Worst case #2 – the code converges to the wrong answer, slowly

Sandia National Laboratories

Where is Sandia Now w.r.t. V&V?

- SNL nuclear weapon mission drivers:
  - annual assessment & certification that all weapon types are safe, secure & reliable
- Few/no tests at the full system level; few/some/no tests at subsystem/component level:
  - not allowed, and/or (radiation effects tests)
  - too expensive, and/or (crash impact tests)
  - too environmentally unfriendly, and/or (fuel/propellant fire tests)
  - too few units available (annual surveillance)
- In ~1996, Sandia entered the Stockpile Stewardship Program to develop comp-sim tools to (a) aid in decision making the absence/reduction of test data, and (b) improve the technical basis (i.e., understanding) of the basic physical processes that dictate weapon performance in all environments.
- As of ~2007, Sandia NW Engineering community is embracing comp. sim. (particularly high-fidelity comp. sim.) as an integral part of the NW design/analysis/qualification process.
  - Sandia NW Engineering is putting in place the policies, procedures, and peer reviews that essentially mandate V&V on all significant comp. sim. studies.

Sandia's V&V Strategic Components

- Capabilities to Enable QMU and PCMM
  - DAKOTA Optimization and UQ Toolkit
  - Code and Solution Verification
  - Software Quality Assurance
  - Methodologies for Aggregating Total Uncertainty in Prediction (PCAP)
- Intrinsic V&V: Integrating V&V and UQ into the Engineering and Computational Simulation Workflow
- Engineering V&V Assessments
  - Align and prioritize code development, phenomenological modeling, and V&V through ASC Focus Areas
  - Use PCMM to measure progress and guide investments
- Training and Outreach

A Quick Intro to DAKOTA:

(started in 1993, 3-6 staff per year, ~0.5-1M lines of code)

Design and Analysis toolKit for Optimization and Terascale Applications

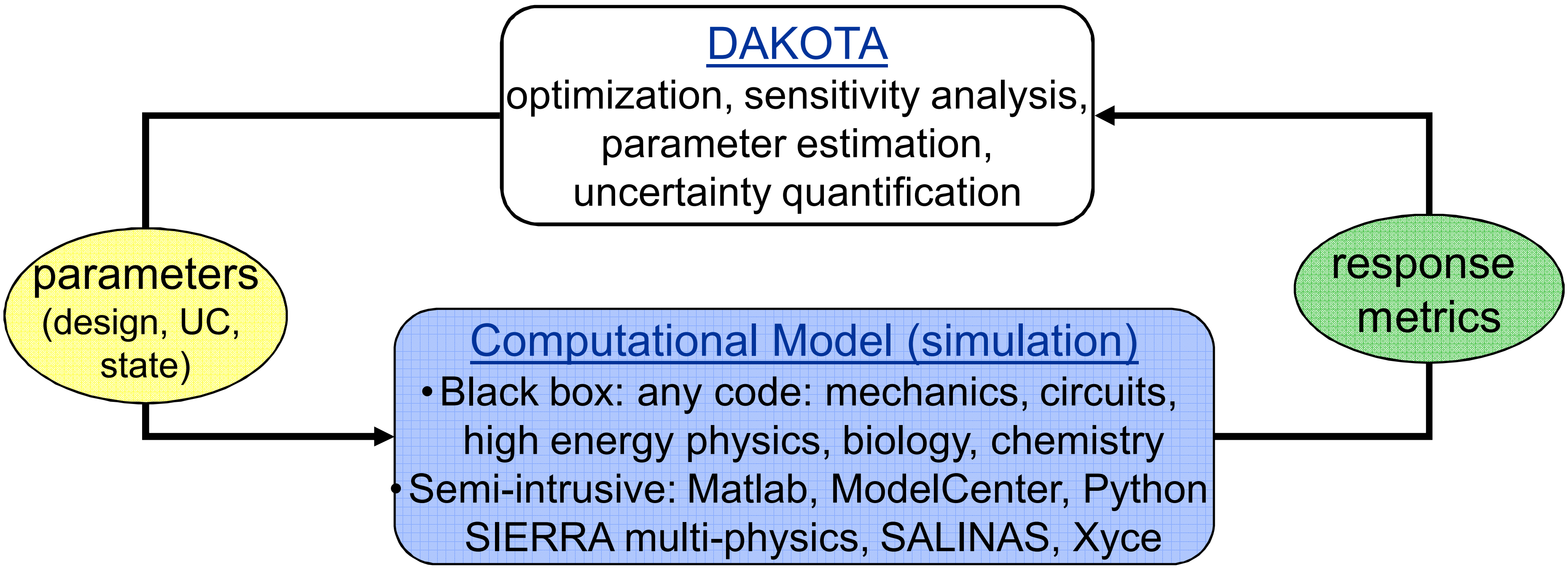
*Helps conduct "what if?" studies with computational models (simulations)**

- **What are the crucial factors/parameters and how do they affect key metrics? *(sensitivity)***
- **How safe, reliable, robust, or variable is my system? *(quantification of margins and uncertainty: QMU, UQ)***
- **What is the best performing design or control? *(optimization)***
- **What models and parameters best match experimental data? *(calibration)***
- ***All rely on iterative/ensemble analysis with a computational model for the phenomenon of interest***

Sandia National Laboratories

DAKOTA Provides an Automated Approach to Generate an Ensemble of Simulation Code "Runs"

Automate typical "parameter variation" studies with a generic interface and advanced methods



- Can support experimental testing: examine many scenarios/ conditions with computer models, then physically test a few worst-case conditions.

Key DAKOTA Capabilities

- Generic interface to simulations
- Time-tested and advanced algorithms to address simulations that are: nonsmooth, discontinuous, multimodal, expensive, mixed variable, failure-prone
- Supports scalable (multilevel) parallel computations on clusters
- Strategies to combine methods for advanced studies or improve efficiency with surrogates (meta-models)
- Object-oriented code; modern software quality practices
- Limited Windows interface (run via command prompt); graphical user interface and DART/DTA, SIERRA integration in progress

- Additional details: http://www.cs.sandia.gov/dakota
  - Extensive documentation, including a tutorial
  - Support mailing lists
  - Software downloads: stable releases and nightly builds (freely available worldwide via GNU GPL)

SIERRA Feature Coverage Tool

The FCT (Feature Coverage Tool) *is a software program that provides a systematic report on whether a set of test cases fully exercises the features for an intended purpose.*

Goal: enable the Thermal/Fluid and Solid Mechanics applications achieve PCMM Level 2 in Code Verification

How does it work?

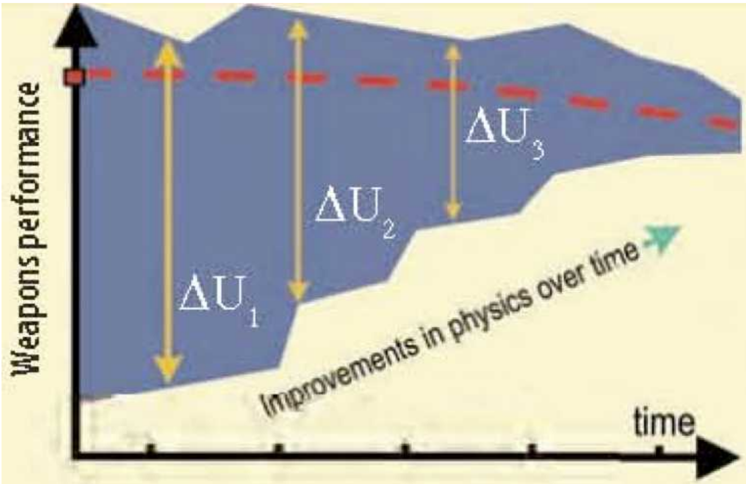
1. **The SIERRA applications are instrumented to provide:**
   a. ***input syntax*** **a complete list of all possible features in an application.**
   b. ***coverage log:*** **a list of the features used in a specific test, or in a specific user's model.**
2. **The FCT computes a *feature coverage certificate* the subset of all possible features that are exercised by the verification test suite.**
3. **Given one or more user model inputs, the FCT intersects them with the certificate, producing the coverage table: with entries for each input feature. Each entry in the table either shows a gap in coverage, or the list of tests that use that feature.**
4. **The FCT will also show two way coverage, the coupled testing of features/interaction.**

Sandia National Laboratories

Intrinsic V&V: Future initiatives

- Expand code verification coverage
  - Feature interactions, test formalism
- Sensitivity analysis
  - More extensive application coverage (leveraging AD and adjoint methods), Sensitivities to geometry
- Embedded UQ
  - Integrate DAKOTA with simulation codes
  - Leverage AD & adjoints, integrated RSM
  - Employ random fields/stochastic collocation
  - UQ for coupled systems (stochastic coupled operators)
- Solution verification
  - Estimation using adaptive refinement
  - Solution verification for multi-scale and coupled multi-physics
- User Interface/Experience: make these capabilities seamless

Predictive Capability Assessment Project (PCAP)



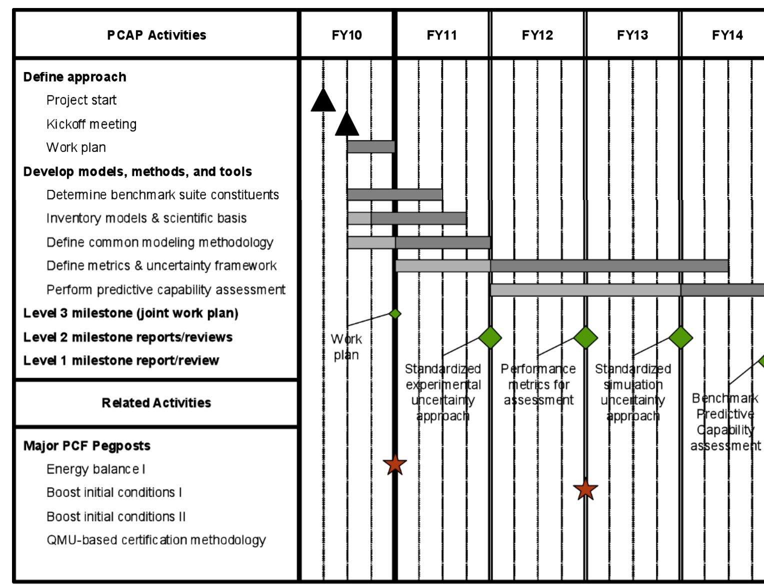
(d) Reduction in cumulative uncertainty in "un-tuned" scientific models over time

SNL Challenges:

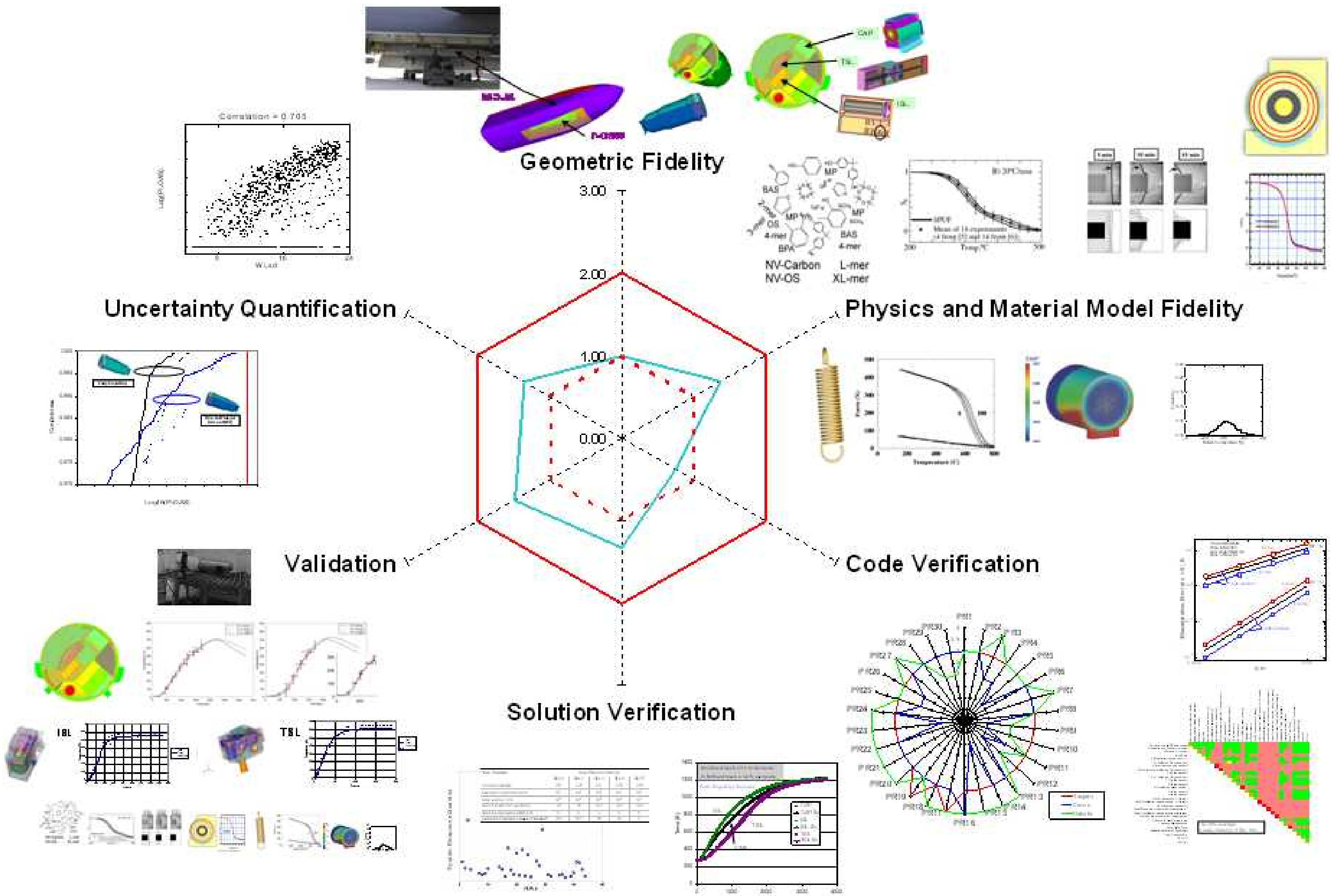
- **Broad range of application focus areas**
- **Limited quantitative test base**
- **Full systems models not developed and exercised across stockpile and development test base**

Background:

- **Based on nuclear performance metrics efforts at LANL and LLNL**
- **Designed to assess simulation predictive quality using "un-tuned" models against historical test base**



Credibility of the Modeling that Produces Simulation-based QMU Results Must be Measured and Communicated



Predictive Capability Maturity Model (PCMM)

Sandia National Laboratories

How Much is Enough?

A Graded Approach is Appropriate From the Perspective of Computational Simulation Stewardship

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
Representation and Geometric Fidelity (RGF) Are representation errors corrupting simulation conclusions?	Characterization (how close to as built are you representing the system)	• (unjustified) conceptual abstraction of the whole system	• Significant (unjustified) simplification or stylization of the system at the level of major elements	• Limited (unjustified) simplification or stylization of the system at the level of major and minor elements
	Computation Error (what impact does imperfect RGF have on computation results)	• Judgment only, numerical errors introduced because of imperfect RGF not addressed	• Sensitivity to imperfect RGF explored for some System Response Quant. (SRQs)	• Numerical errors estimated for imperfect RGF for relevant SRQs
	Verification (is what you represented really what was built)	• RGF not verified, RGF simply used without verification that it represents the actual system as built	• RGF verified only by the analysts	• RGF independently verified

Core Attribute | **Best Practices** | **Measured Against Standards Expressed in Terms of Increasing Rigor**

- **Level 0: Low consequence; minimal M&S impact (e.g., scoping studies)**
- **Level 1: Moderate consequence; some M&S impact (e.g., design support or qualification test support)**
- **Level 2: High consequence, high M&S impact (qualification decision support)**
- **Level 3: High consequence; decision making based predominately on M&S (dominant basis for qualification or certification)**

There are other ways to frame solutions to the need for a graded approach

Sandia National Laboratories

Predictive Capability Maturity Model (PCMM) Measures and Communicates Maturity of Mod/Sim Process

PREDICTIVIE ATTRIBUTE	Level 0 Low-Consequence M&S-Informed, e.g., Scoping or Res Activities Score=0	Level 1 Low-Consequence M&S-Informed, e.g., Design Support Score=2	Level 2 High-Consequence M&S-Informed, e.g., Qualification Support, Score=4	Level 3 High-Consequence M&S-Based, e.g., Qualification Score=6
Representation or Geometry Fidelity Are you overlooking important effects because of defeaturing or stylization	• Grossly defeatured or stylized representation based on judgment or practical considerations	• Significant defeaturing or stylization based on judgment or practical considerations • or lower fidelity representation justified w a significantly defeatured or stylized representation	• Limited defeaturing or stylization judged to retain the essential elements of "as built" • or appropriate lower fidelity representation justified w a slightly defeatured or stylized representation	• Highest fidelity representation "as is" w/o sig defeaturing or stylization • or appropriate lower fidelity representation justified w highest fidelity representation
Physics and Material Model Fidelity How science-based are the models?	• Unknown model form represented with ad hoc knob non-uniquely calibrated to IET • Empirical model applied w significant extrapolation, non-uniquely calibrated with IET	• Empirical model applied w/o significant extrapolation, uniquely calibrated with SET • Physics informed model applied w significant or unknown extrapolation, unique calibrations with SET • Physics-informed model applied w/o significant extrapolation, non-unique calibrations with IET	• Physics informed models applied w/o significant extrapolation, unique calibrations with SET • Physics-based model applied w significant or unknown extrapolation	• Well accepted physics-based model applied w/o significant extrapolation
Code Verification Are software errors or algorithm deficiencies corrupting simulation results?	• Judgment only	• Code managed to SQE standards • Sustained unit/regression testing w significant coverage of required Features and Capabilities (F&Cs)	• Code managed and assessed (internally) against SQE standards • Sustained verification test suite w significant coverage of required F&Cs	• Code managed and assessed (externally) against SQE standards • Sustained verification test suite w significant coverage of required F&Cs and their interactions
Solution Verification Are numerical errors corrupting simulation results?	• Judgment only • Sensitivity to discretization and algorithm parameters explored in SRQs not directly related to the decision context	• Sensitivity to discretization and algorithm parameters explored in SRQs directly related to the decision context • Numerical errors estimated in SRQs not directly related to decision context	• Numerical errors estimated in SRQs directly related to the decision context • Rigorous numerical error bounds quantified in SRQs not directly related to the decision context	• Rigorous numerical error bounds quantified in SRQs directly related to the decision context
Validation How accurate are the models?	• Judgment only • Qualitative accuracy w/o significant SET coverage	• Qualitative accuracy w significant SET coverage • Quantitative accuracy w/o assessment of unc and w/o significant SET coverage	• Quantitative accuracy w/o assessment of unc • w significant SET coverage and IETs	• Quantitative accuracy w assessment of unc • w significant SET coverage, IETs, and full system test
UQ and Sensitivities What is the impact of variabilities and uncertainties on performance and margins?	• Judgment only • Deterministic assessment of margins (e.g., bounding analyses) • Informal "what if" assessments of unc, margins, and sensitivity	• Aleatory and epistemic uncertainties represented and propagated w/o distinction • Sensitivity to uncertainties explored	• Aleatory and/or epistemic uncertainties represented separately and propagated w significant strong assumptions • Quantitative sensitivity analysis w significant strong assumptions • Sensitivity to numerical errors explored	• Aleatory and/or epistemic uncertainties represented separately and propagated w/o significant strong assumptions • Quantitative sensitivity analysis w/o significant strong assumptions • Numerical errors quantified

Predictive Capability Maturity Model (PCMM) Measures and Communicates Maturity of Mod/Sim Process

PREDICTIVIE ATTRIBUTE	Level 0 Low-Consequence M&S-Informed, e.g., Scoping or Res Activities	Level 1 Low-Consequence M&S-Informed, e.g., Design Support	Level 2 High-Consequence M&S-Informed, e.g., Qualification Support,	Level 3 High-Consequence M&S-Based, e.g., Qualification
Representation or Geometry Fidelity Are you overlooking important effects because of defeaturing or stylization				
Physics and Material Model Fidelity How science-based are the models?				
Code Verification Are software errors or algorithm deficiencies corrupting simulation results?				
Solution Verification Are numerical errors corrupting simulation results?				
Validation How accurate are the models?				
UQ and Sensitivities What is the impact of variabilities and uncertainties on performance and margins?				

- **Key issues:**
  - **Horizontal Axis – measures level of rigor in a mod/sim activity.**
  - **Vertical Axis – covers different aspects of mod/sim activity (geometric fidelity, physics fidelity, verification, validation, UQ, etc.).**
  - **PCMM provides a means to consistently document and communicate the status of a complex VV/UQ/QMU study to a non-ASC weapons customer.**
  - **Peer review is a critical component of PCMM (above level 0).**

PCMM is Still Evolving: e.g. PCMM (version 2) is now in use at Sandia

- **Status of PCMM within Sandia ASC:**
  - **PCMM self-assessments in use by ASC V&V project teams.**
  - **PCMM in use by some ASC V&V internal peer review panels.**
  - **ASC Program executives look for PCMM evidence for all Level-II milestones.**
    - **PCMM evidence now required for all ASC V&V-funded work.**

- **Other non-ASC programs at Sandia starting to investigate PCMM.**
  - **Nuclear weapons system engineering**
  - **DOD Work For Others projects using agent-based models**

- **We are revising PCMM in FY11 (to Rev. 3) to incorporate lessons learned from use over the last 2-3 years**

Summary

- **Sandia's science & engineering practices are evolving to include V&V, UQ, and QMU methods.**
  - **Enable risk-informed decision making on high-consequence applications.**
    - **Use sensitivity/UQ methods to get better, more complete answers**
    - **Employ V&V as an "evidence" tool for comp. sim.**

- **Science front:**
  - **Research, develop, and deploy new mathematical and statistical methods that improve V&V practices.**
  - **Leverage validation results to identify physics research topics.**

- **Engineering front:**
  - **Establish integrated teams: engineers, scientists, math/stats, etc.**
  - **Provide tools and training to facilitate use of V&V/UQ methods.**
  - **Select appropriate level/depth of V&V/QMU effort for each weapon application (via customer-desired PCMM "maturity" level).**
  - **Use V&V/UQ methods to deliver "best estimate + quantified uncertainty and margins" to customers.**

References: Sandia V&V/UQ/QMU Publications & Related Items

PCMM

W. L. Oberkampf, M. Pilch, and T. G. Trucano, "Predictive Capability Maturity Model for Computational Modeling and Simulation," Sandia Report SAND2007-5948, Sandia National Laboratories, Albuquerque, NM, 2007.

QMU

T. Trucano, M. Pilch, and J. Helton, "Ideas Underlying Quantification of Margins and Uncertainties (QMU): A White Paper," Sandia Report SAND2006-5001, 2006.

K. V. Diegert, S. E. Klenke, G. C. Novotny, R. A. Paulsen, Jr., M. Pilch, T. G. Trucano, "Toward a More Rigorous Application of Margins and Uncertainties Within the Nuclear Weapons Life Cycle: A Sandia Perspective" Sandia Report SAND2007-6219, Sandia National Laboratories, Albuquerque, NM, 2007.

(NEW) J. C. Helton, "Conceptual and Computational Basis for the Quantification of Margins and Uncertainty," Sandia Report SAND2009-3055, Sandia National Laboratories, Albuquerque, NM, 2009.

2008 National Academy of Sciences QMU review: "Evaluation of Quantification of Margins and Uncertainties Methodology for Assessing and Certifying the Reliability of the Nuclear Stockpile" (http://www.nap.edu/catalog.php?record_id=12531)

Sandia National Laboratories