An Introduction to Verification, Validation and Uncertainty Quantification

Josh Mullins

Validation, Verification, Uncertainty Quantification and Credibility Processes Dept.

SAND Number: 2014-2695P

Overall Course Outline

- Introduction to concepts and tools related to:
  - Solution and code verification (V)
  - Model Validation (V)
  - Uncertainty quantification (UQ)
  - Predictive Capability Maturity Model (PCMM)

- How to establish credibility in computational simulations using V&V/UQ

Schedule of Classes

	Class Name	Date	Instructor(s)
1	Overview of V&V/UQ Concepts	March 21, 2016 8:30 – 12	J. Mullins
2	Code and Solution Verification	March 21, 2016 1 – 4:30	B. Carnes
3	Uncertainty Quantification, Sensitivity Analysis, and DAKOTA intro	March 22, 2016 8:30 – 12	J. Winokur
4	Validation of Computational Models	March 23, 2016 8:30 – 12	K. Dowding
5	V&V/UQ/Credibility	March 24, 2016 8:30 – 10:30	J. Mullins

The Instructors

V&V/UQ Applications and Credibility

Josh Mullins, Dept. 1544
(505) 284-9169
jmullin@sandia.gov

Code and Solution Verification

Brian Carnes, Dept. 1544
(505) 284-1332
bcarnes@sandia.gov

The Instructors

Uncertainty Quantification and Sensitivity Analysis

Justin Winokur, Dept. 1544
(505) 844-0630
jgwinok@sandia.gov

Validation of Computational Models

Kevin Dowding, Dept. 1544
(505) 844-9699
kjdowdi@sandia.gov

SESSION 1:
OVERVIEW OF V&V AND UQ

Outline of Session 1

- Introduction and Motivation
- V&V/UQ terminology
- Introduction to the Example Problem
- Class exercise
- The V&V Process
- Summary

About this Presentation

- This is about:
  - A high level overview of the V&V/UQ processes that support the credibility of our computational simulations (CompSim).
  - Introduction to the common example problem

- This **is not** about:
  - An equation-rich description of how to do V&V and UQ – This will be covered in each class.

Introduction

- Our common ground:
  - The nation is making million/billion $ decisions that are strongly influenced by computational simulation – e.g., weapon life extensions, full scale tests, facility/infrastructure protection upgrades, etc.
  - How do we build/demonstrate confidence in our CompSim. results?

- My goals for this talk:
  - Influence your thinking about computational simulation, via role of:
    - verification & validation (V&V)
    - uncertainty quantification (UQ)
  - Motivate you to attend the rest of ESP700

V&V/UQ - Definitions

- Verification – "Are we solving the equations correctly?"
  - Code verification: Correctness of implemented mathematical algorithms.
  - Solution verification: Convergence to the correct answer, as model is refined → numerical error estimation
- Validation – "Are we solving the right equations?"
  - Correctness of physical models and sufficiency for the application.
  - Model Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model
- Uncertainty Quantification (UQ):
  - Statistical propagation of uncertainty through a simulation model, and statistical interpretation of model response.

V&V is expected but why?

- 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 and models which leads to "correct" results.
- In the past V&V was an after-though if thought of it at all. Sometimes consider a nuisance.

V&V is expected but why?

- What's different now?
  - CompSim 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 CompSim.
  - "Before I spend $M/$B on a decision, I want evidence of the correctness of your CompSim results."

- Issues:
  - Correctness is expected or implied, but isn't innate → requires extra effort to provide quantitative evidence (via V&V)
  - Due to resource constraints, you can't V&V every aspect of a code/model/project
  - It's hard to retrofit V&V into a study that is already completed.

V&V is expected but why?

- Using CompSim results to aid decision making is a good thing because:
  - Decision making is based on knowing the tradeoffs for competing objectives, due to variations in designer-controllable parameters.
  - Quantities of interest: cost & performance
  - This is good: (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 is also good:
    - "If were going to perform a CompSim 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."

V&V is expected but why?

- Punch Line:
  - V&V is not palatable 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 is palatable 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.

Where is SNL now w.r.t. CompSim & V&V?

Sandia National Laboratories

- 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)

Where is SNL now w.r.t. CompSim & V&V?

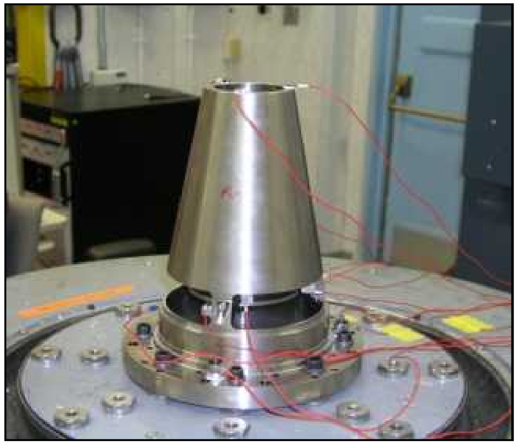
Sandia National Laboratories

- In ~1996, Sandia entered the Stockpile Stewardship Program to develop CompSim tools to
  - (a) aid in decision making in 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.

- In ~2007, Sandia NW Engineering community embraced CompSim (particularly high-fidelity CompSim.) 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 CompSim studies.

The Example

The Example



Need: To Quantify Margins of performance for a given environment in the presence of Uncertainty (i.e. do a QMU analysis)

Translation

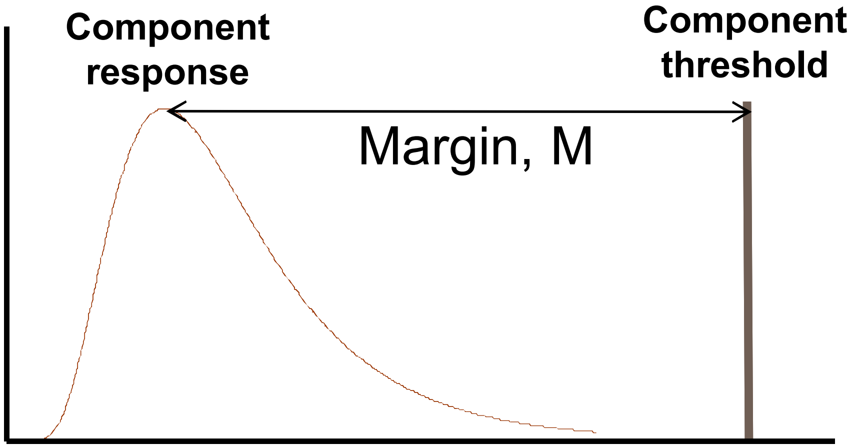
Given that:

We don’t know exactly how a component behaves and
We don’t know exactly what the environment (normal, abnormal, hostile) is

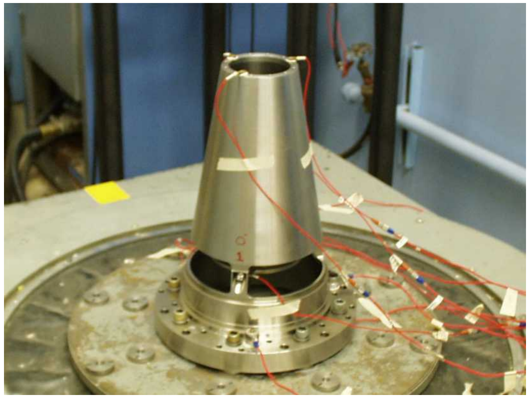
can we determine if:

The component will survive and
By how much

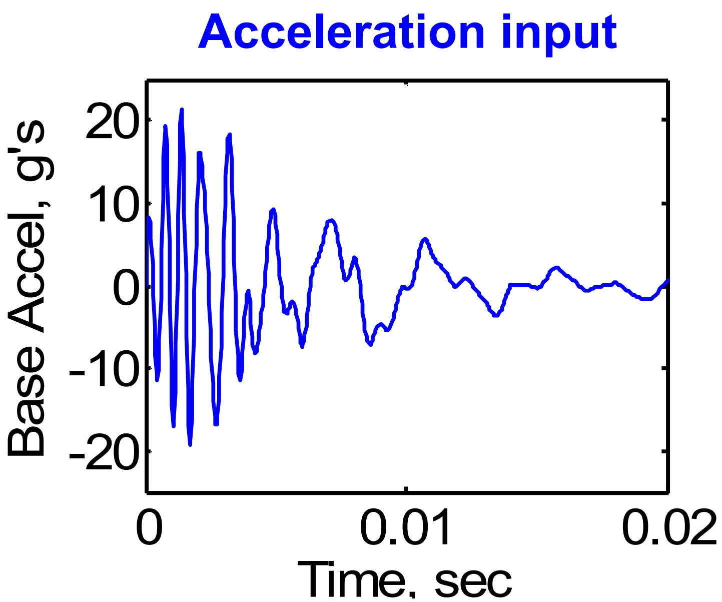
Challenge: Limited test data (1-2) and/or cannot test in the application domain

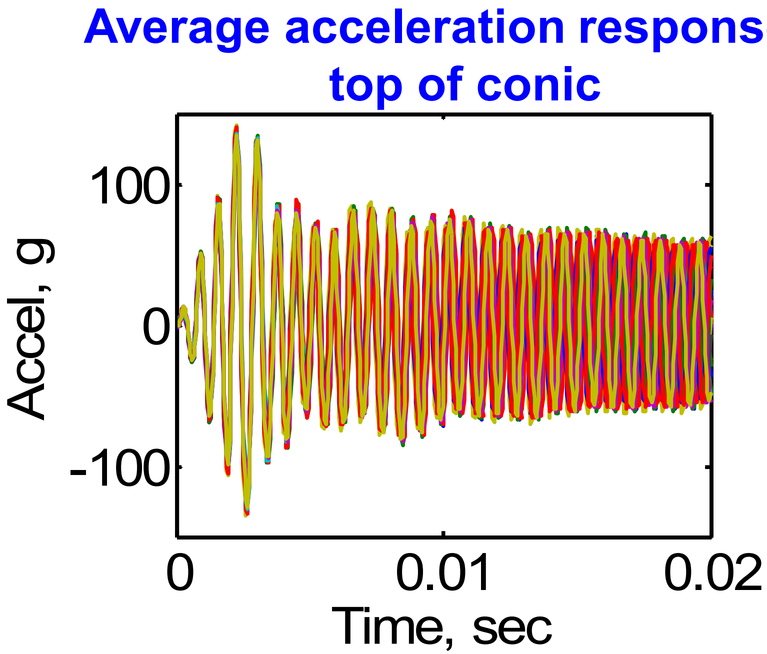


The Example (experimental side)

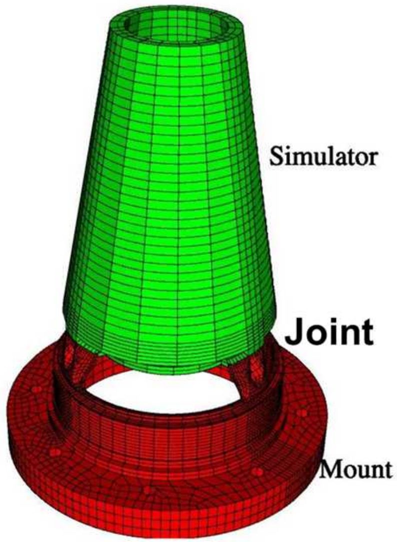


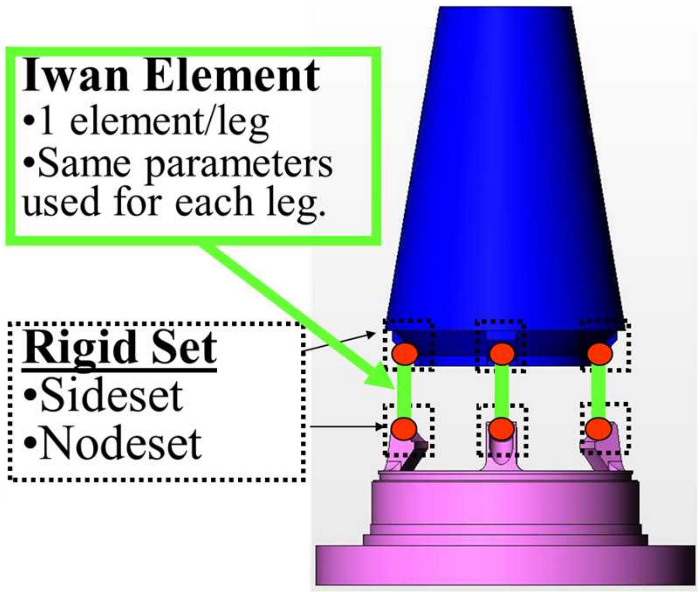
- Hardware consists of 3 top conic sections and 3 bottom sections
- 9 total combinations of top/bottoms
- 3 test repetitions
- 27 total tests

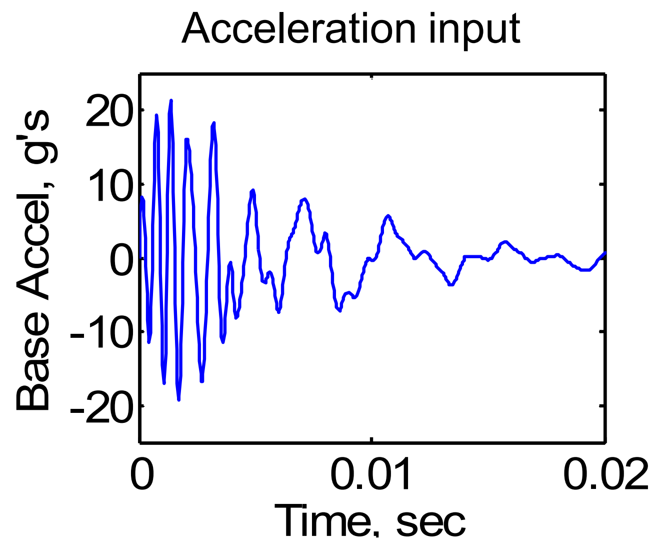


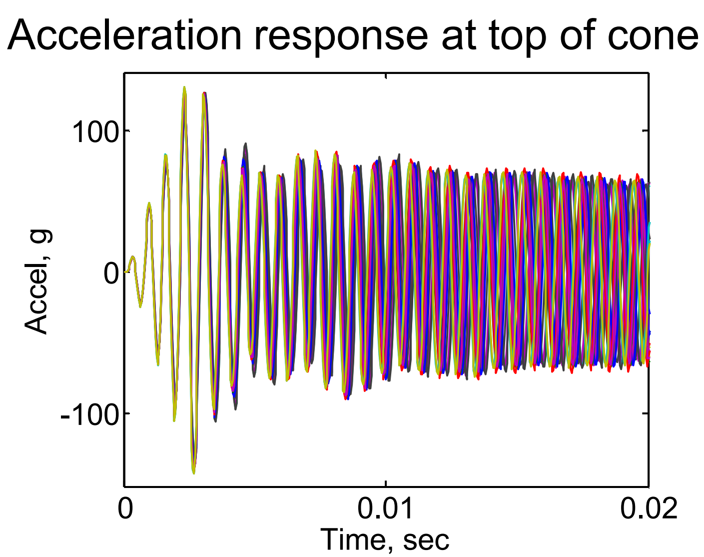


The Example (CompSim side)



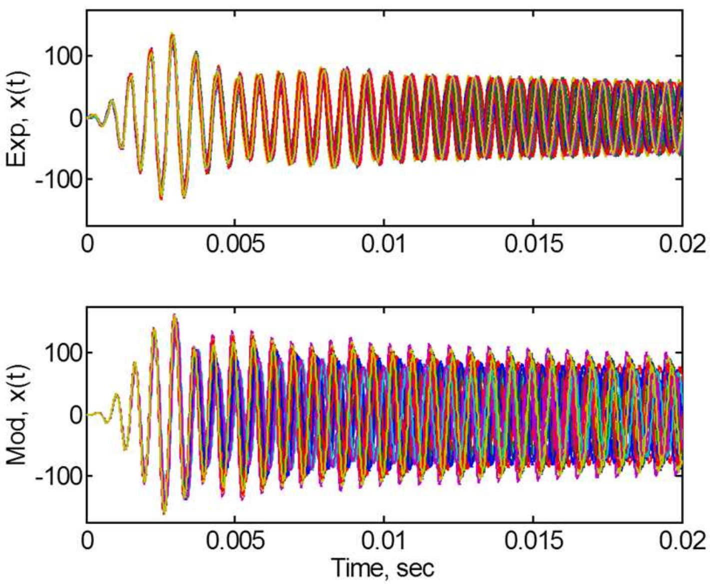


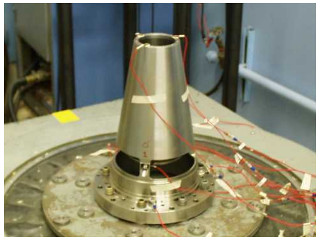


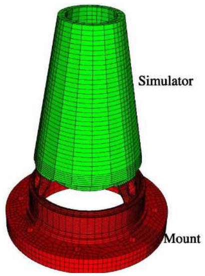


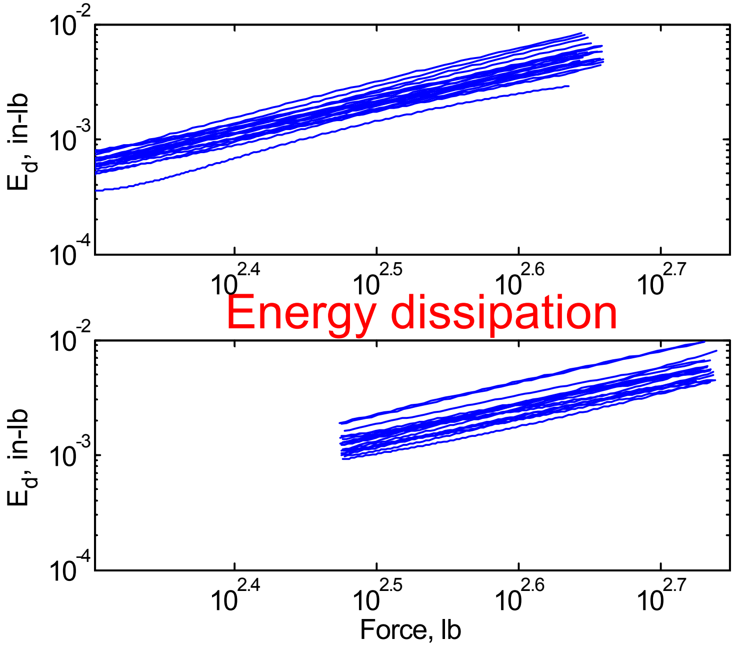
- 3D finite element model representing 3 leg hardware was created
- Bolted joints (J) are modeled using an Iwan element
- Non linear transient analysis was performed using Sierra-SD (structural dynamics)

Quantity of interest – Energy Dissipation

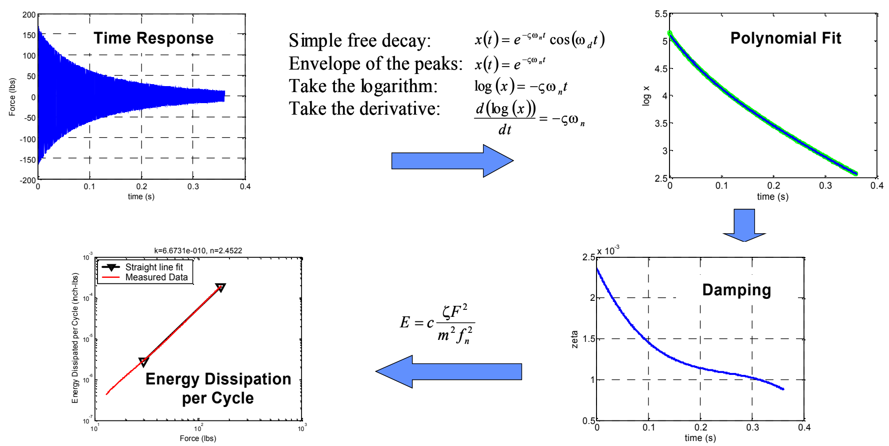






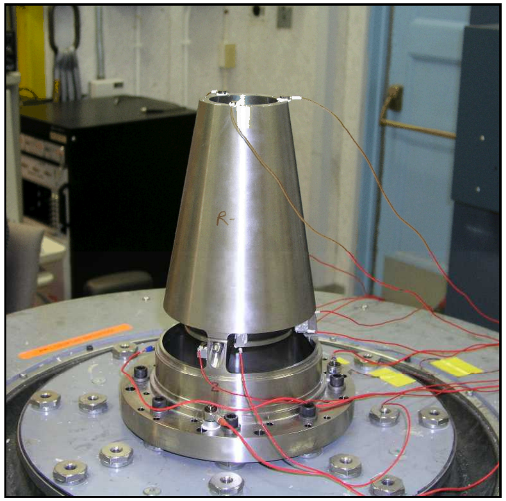


Energy dissipation per cycle from transient responses

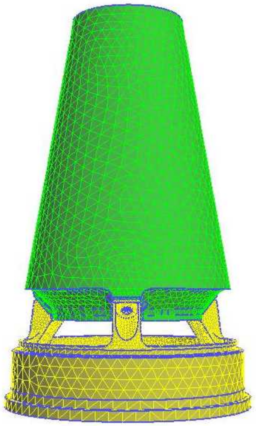


The Example (Credibility)

Response of physical system	$\overset{?}{\approx}$	Response of model system



$\overset{?}{\approx}$



Key questions:

- Are model predictions "good enough" to be used in lieu of the real thing?
- How do we establish the credibility of these predictions?

Are we ready for a challenge?

- If called upon, how do we defend our predictions?
- What evidence is necessary to support our claims?
- Short answer: It depends on what questions will be answered by the decision maker using the simulation results AND the associated risk with the decision being wrong

So, how do we do this?
Starts with an understanding of what "credible" means?

Class exercise

What is credibility and how do we establish it ???

Class exercise

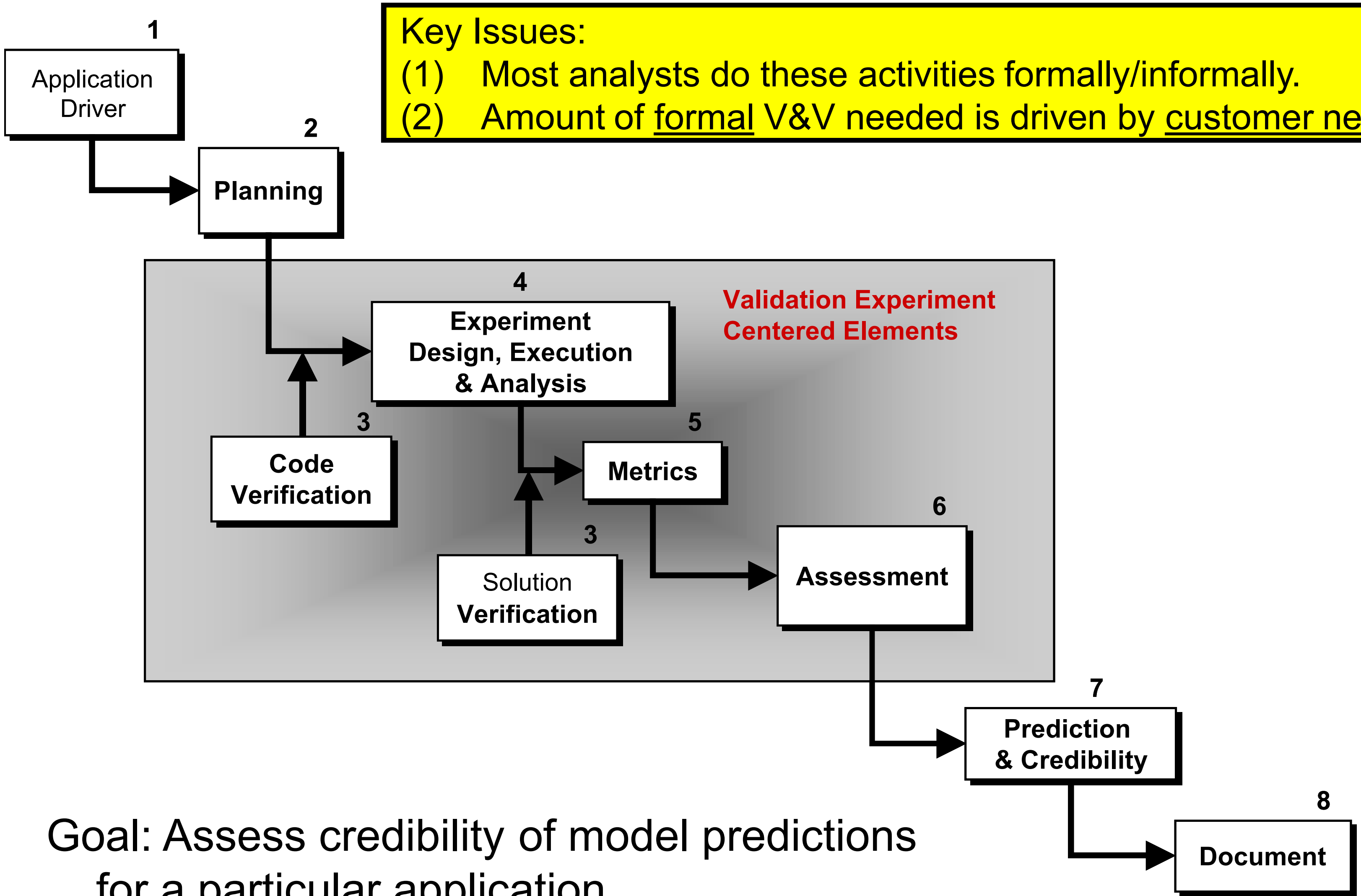
Definition of Credible: Offering reasonable grounds for being believed; (i.e. a credible account of an accident; credible witnesses)

With you as the end customer, try to answer the following (at high level):

- What do “credible results” look like to you?
- What evidence would you think is necessary to be able to support the claim that results are “credible”?

The V&V Process

Overview of the Sandia V&V Process



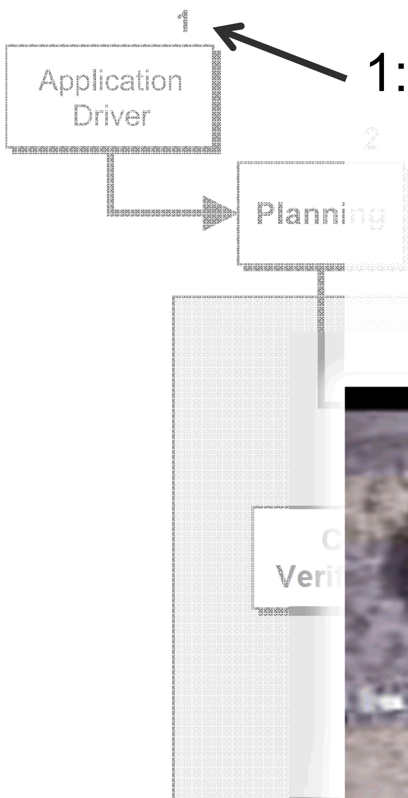
Key Issues:

(1) Most analysts do these activities formally/informally.

(2) Amount of <u>formal</u> V&V needed is driven by <u>customer needs</u>.

Goal: Assess credibility of model predictions for a particular application

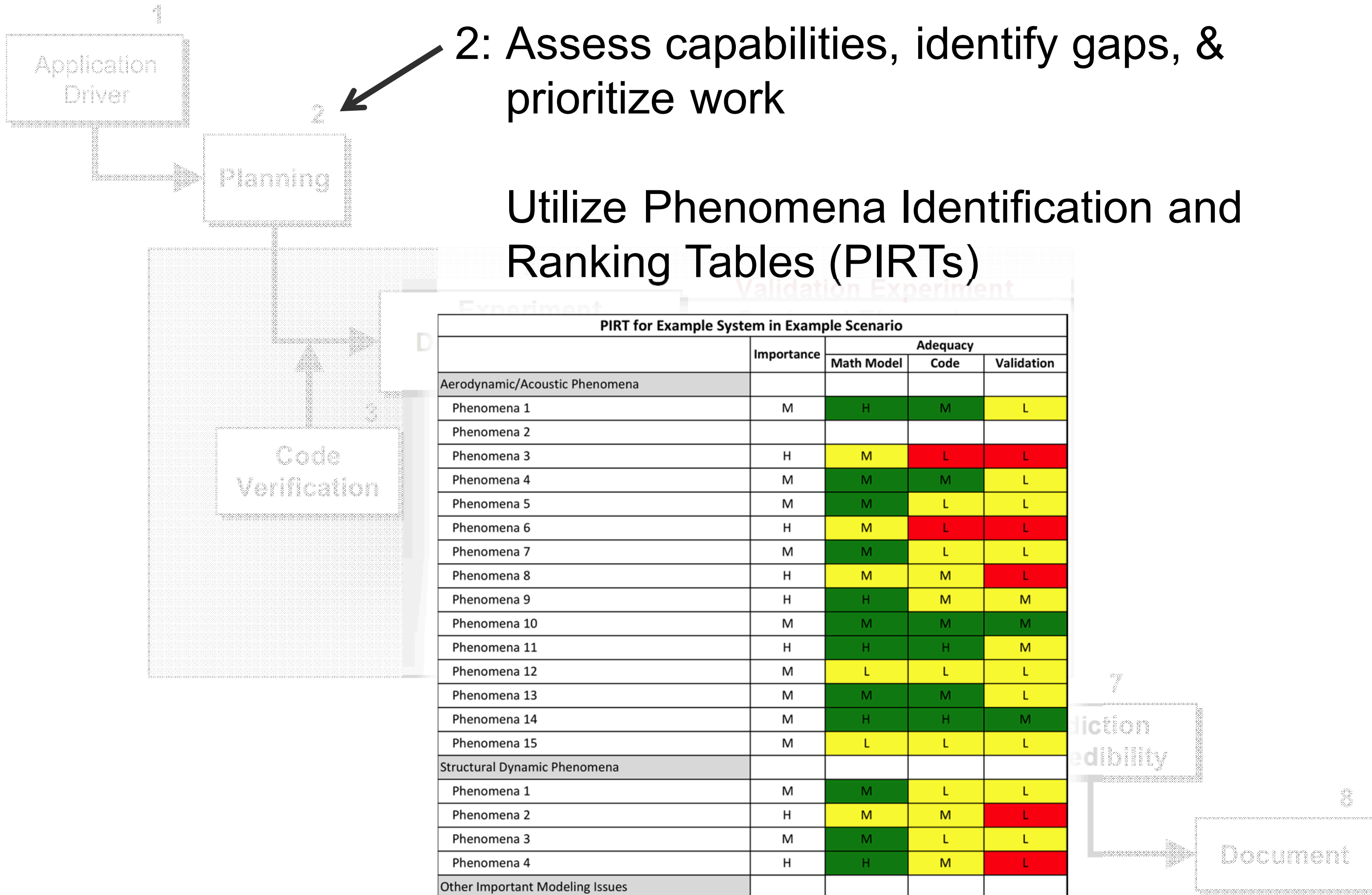
Overview of the Sandia V&V Process



1: Understand the application and requirements

What is the intended use of the model?
How ‘good’ is good enough?

Overview of the Sandia V&V Process



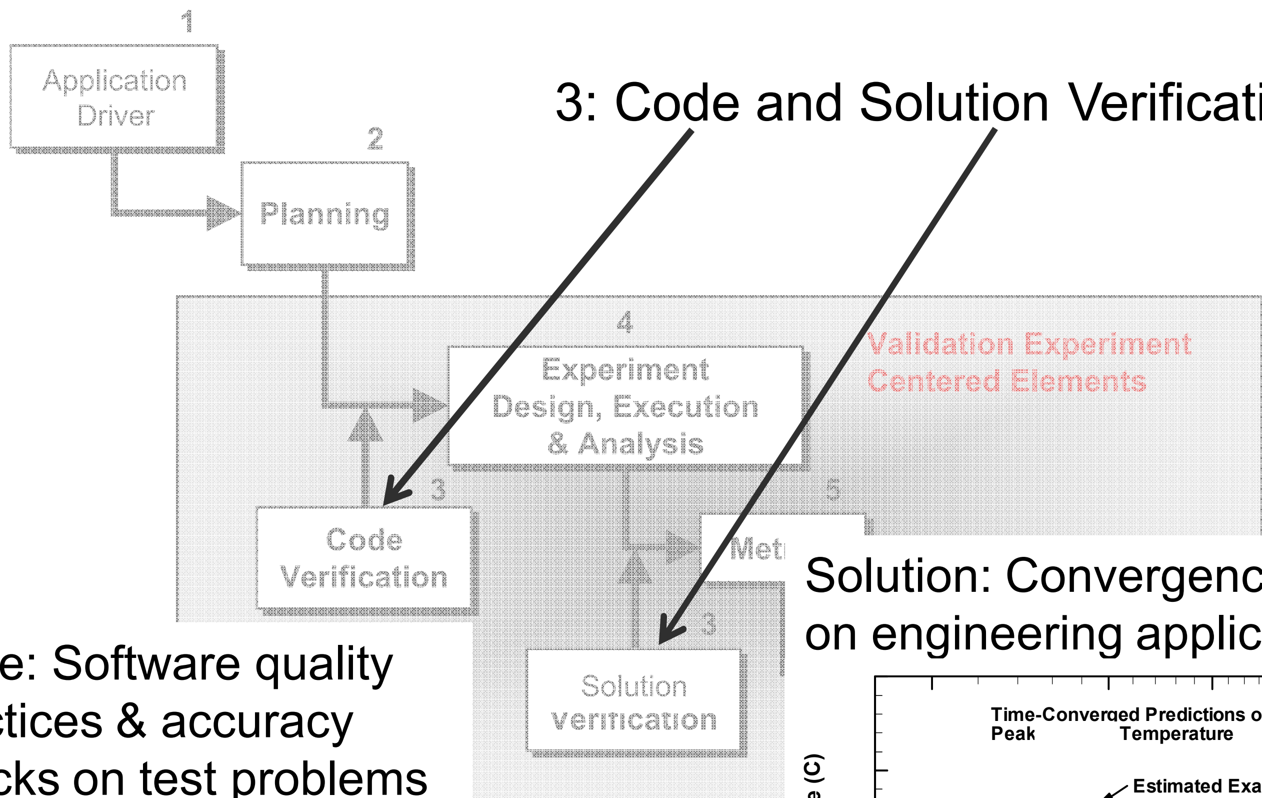
2: Assess capabilities, identify gaps, & prioritize work

Utilize Phenomena Identification and Ranking Tables (PIRTs)

PIRT for Example System in Example Scenario

	Importance	Adequacy		
		Math Model	Code	Validation
Aerodynamic/Acoustic Phenomena				
Phenomena 1	M	H	M	L
Phenomena 2				
Phenomena 3	H	M	L	L
Phenomena 4	M	M	M	L
Phenomena 5	M	M	L	L
Phenomena 6	H	M	L	L
Phenomena 7	M	M	L	L
Phenomena 8	H	M	M	L
Phenomena 9	H	H	M	M
Phenomena 10	M	M	M	M
Phenomena 11	H	H	H	M
Phenomena 12	M	L	L	L
Phenomena 13	M	M	M	L
Phenomena 14	M	H	H	M
Phenomena 15	M	L	L	L
Structural Dynamic Phenomena				
Phenomena 1	M	M	L	L
Phenomena 2	H	M	M	L
Phenomena 3	M	M	L	L
Phenomena 4	H	H	M	L
Other Important Modeling Issues				
Phenomena 1	H	L	L	L
Phenomena 2	M	N/A	N/A	L
Phenomena 3	H	N/A	N/A	M

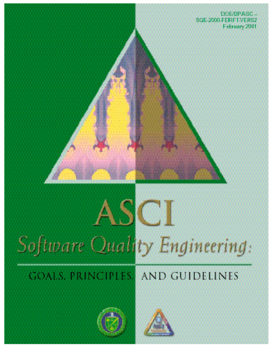
Overview of the Sandia V&V Process

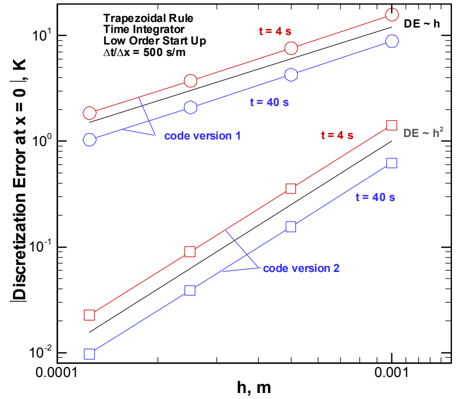


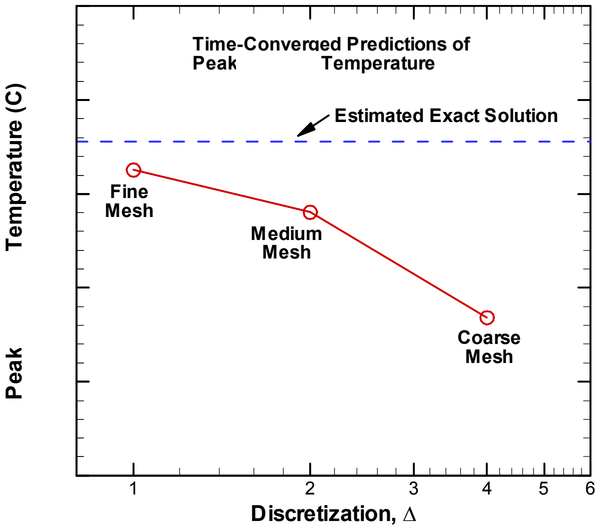
3: Code and Solution Verification

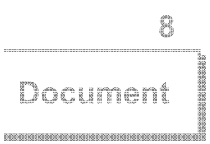
Solution: Convergence checks on engineering application

Code: Software quality practices & accuracy checks on test problems









Verification

- "Are we solving the equations correctly?"

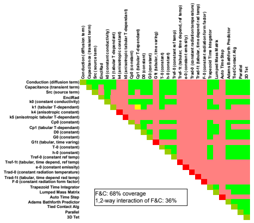
- Two aspects of this:
  - Code verification: Correctness of implemented mathematical algorithms (is your code bug-free?).
  - Solution verification: Convergence to the correct answer → numerical error estimation

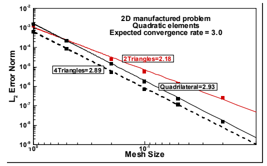
Code Verification

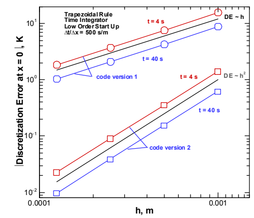
Are software errors or algorithm deficiencies corrupting simulation results?

Code1:Code2 Comparisons

- Apply good SQE processes
  - Do you have a mature code development process?
- Assess SQE processes
  - Verify that codes are developed with an appropriate level SQE maturity?
- Provide adequate test coverage
  - Can the user be confident that the code is adequately tested for the intended application?
- Quantify computation errors
  - What is the impact of undetected code or algorithm deficiencies on simulation results?

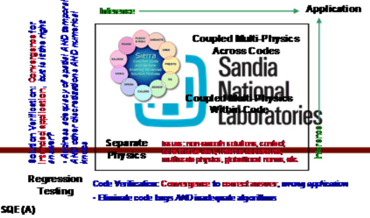






Feature Coverage Tool (FCT)

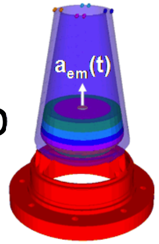
- An **objective** way to know if a verification test exists that **uses the features you are interested in**.
- A metric to communicate gaps to development team.
- Evidence for the CompSim credibility (PCMM table for simulations).

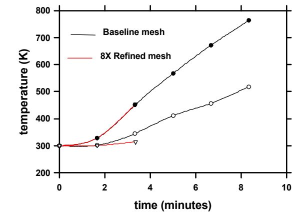


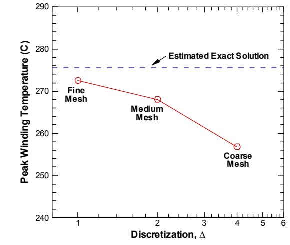
Solution Verification

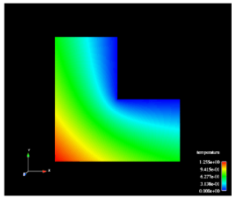
Are human procedural errors or numerical solution errors corrupting simulation results?

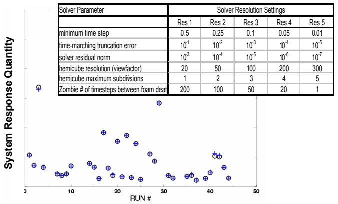
- Quantify numerical solution errors
  - What is the impact of numerical solution errors o relevant system response quantities (SRQs)
- Verify all simulation inputs and outputs
  - Have we corrupted simulation results with incorrect inputs or post processing errors?
- Perform technical review
  - Verify that the solution verification activities are relevant, adequate, and executed in a technically sound manner



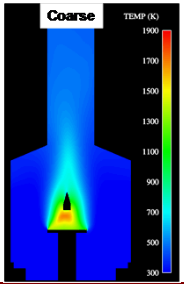


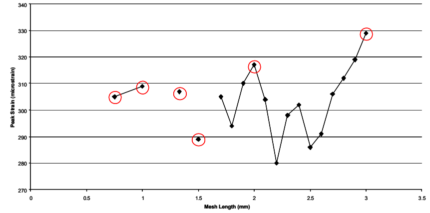






Solver Parameter	Solver Resolution Settings				
	Res 1	Res 2	Res 3	Res 4	Res 5
minimum time step	0.5	0.25	0.1	0.05	0.01
time-marching truncation error	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
solver residual norm	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
hemicube resolution (viewfactor)	20	50	100	200	300
hemicube maximum subdivisions	1	2	3	4	5
Zombie # of timesteps between foam deat	200	100	50	20	1

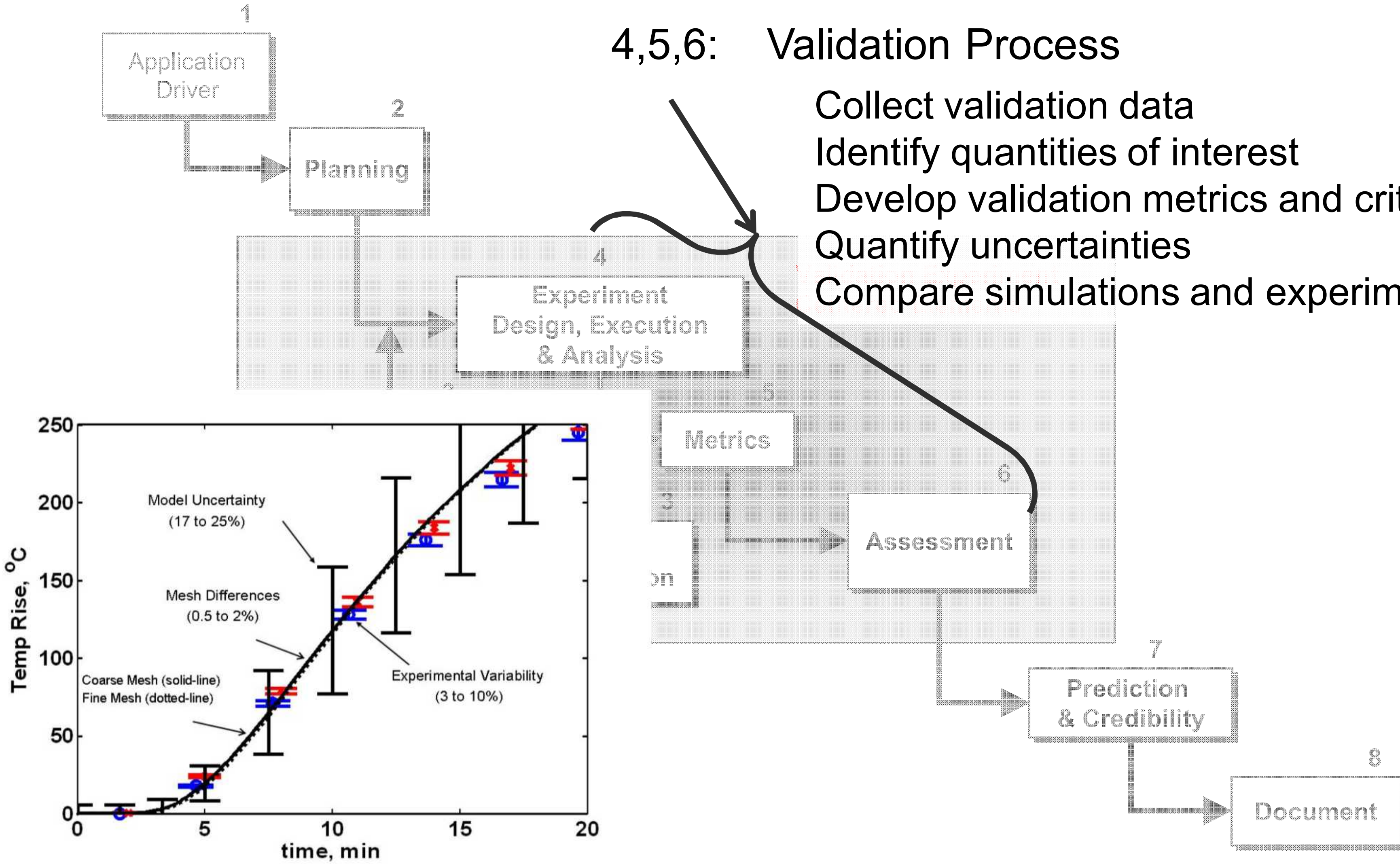




Percept

- Enable engineering analysts to easily incorporate solution verification in their workflow during modeling and simulation.
- Provide a scalable software package to help automate convergence analysis.
- Create uniformly and adaptively refined sequences of unstructured grids.
- Help manage the tradeoff of computer resources versus numerical discretization error.

Overview of the Sandia V&V Process



4,5,6: Validation Process

- Collect validation data
- Identify quantities of interest
- Develop validation metrics and criteria
- Quantify uncertainties
- Compare simulations and experiments

Uncertainty quantification

- It tries to determine how likely certain outcomes are if some aspects of the system are not exactly known.
- Epistemic (Reducible uncertainty)
  - Lack of knowledge about the appropriate value to use
  - Reduced through increased understanding or more data.
- Aleatory (Irreducible uncertainty)
  - Cannot be reduced by further data
  - Variability (due to part-to-part, test-to-test variation, etc.)
  - Usually modeled with probability distributions

Uncertainty Quantification

- What is the impact of variability & uncertainties on system performance and margins?
- What are the key uncertain parameters?
  - What do you know about them (bounds, probability distributions, other)?
- What are the key uncertain physics models?
- How do you propagate uncertainty through your simulation model?

Uncertainty Quantification

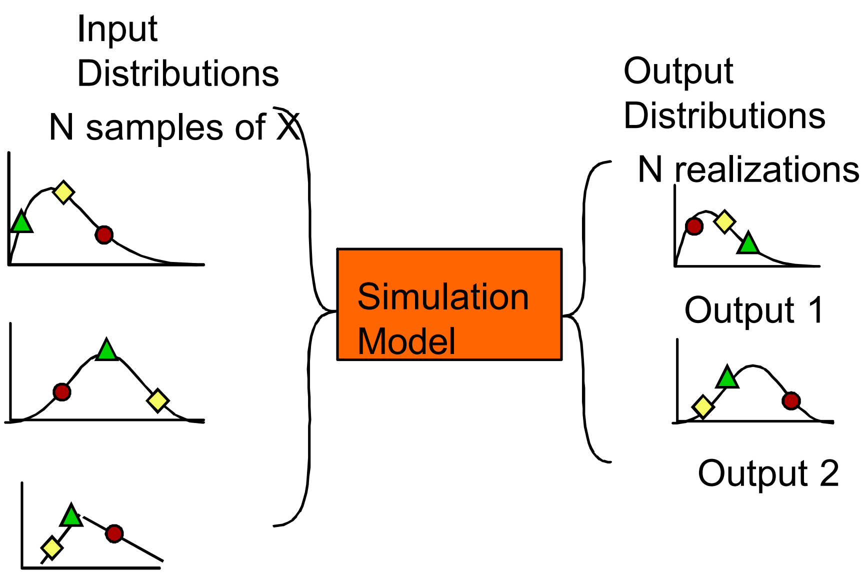
- How do we quantify uncertainty?
  - Through data from experiments on multiple pieces of hardware
    - Pros: Best way to quantify unit-to-unit variability
    - Cons: Expensive
  - Through historical data from legacy system
    - Knowledge from the past might not be relevant to the future
  - Through the use of models representing the system behavior
    - Pros: in principle, model can be run many times in a stochastic way to quantify uncertainty
    - Cons: not always an accurate representation of the real system behavior
  - A combination of experiments and models

Some sources of uncertainty

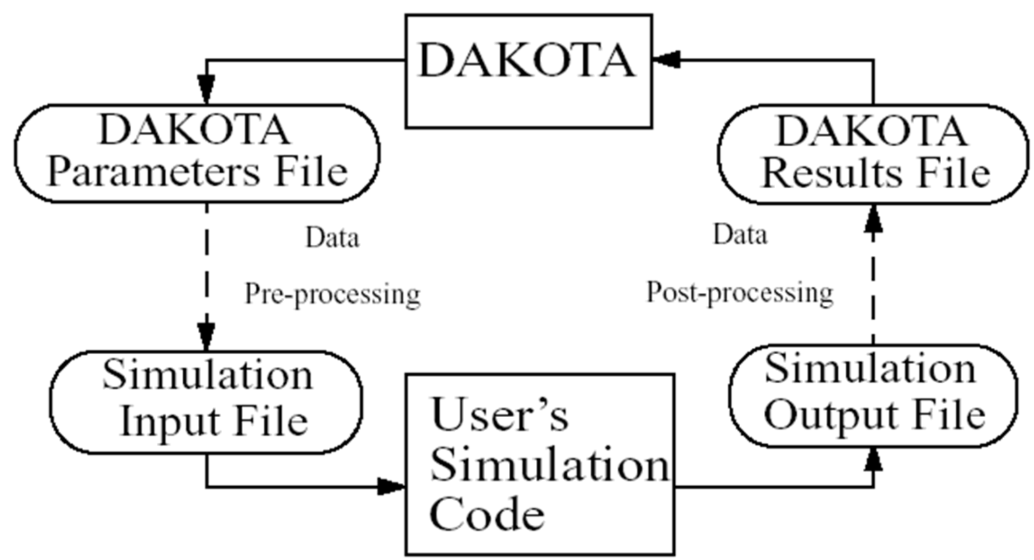
- The model structure, i.e., how accurately a mathematical model describes the true system for a real-life situation, may only be known approximately.
- The numerical approximation, i.e., how appropriately a numerical method is used in approximating the operation of the system.
- Mesh approximation
- Input and/or model parameters
  - may only be known approximately.
  - may vary between different instances of the same object for which predictions are sought.

Common UQ Method: Monte Carlo Method

- Assume distributions on the uncertain input values
- Sample from those distributions
- Run the model with the sampled values
- Repeat to build up a distribution of the outputs.



This process is enabled using DAKOTA



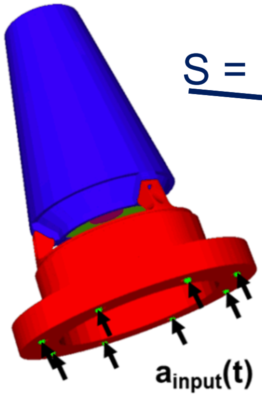
Validation

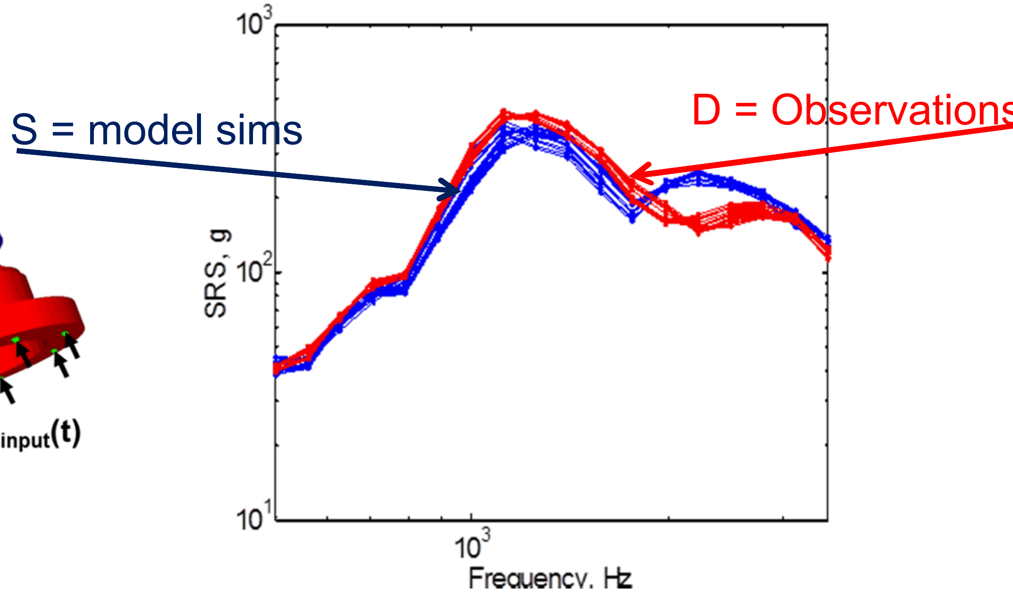
- Model Validation:
  - The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model
  - Are we solving the correct equations?

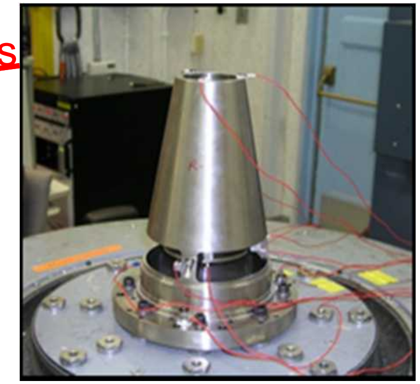
Some observations on validation

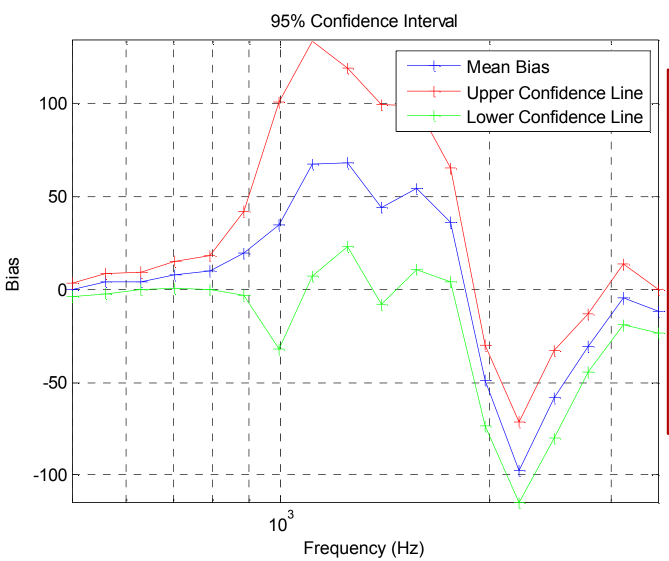
- Uncertainty in both model simulations and experimental data should be included
- A criteria for determining the required accuracy needs to be defined a priori and should be relative to the intended use of the model
- Comparing 2 lines on a plot ("vugraph norm") is way of doing validation but not very rigorous

Validation





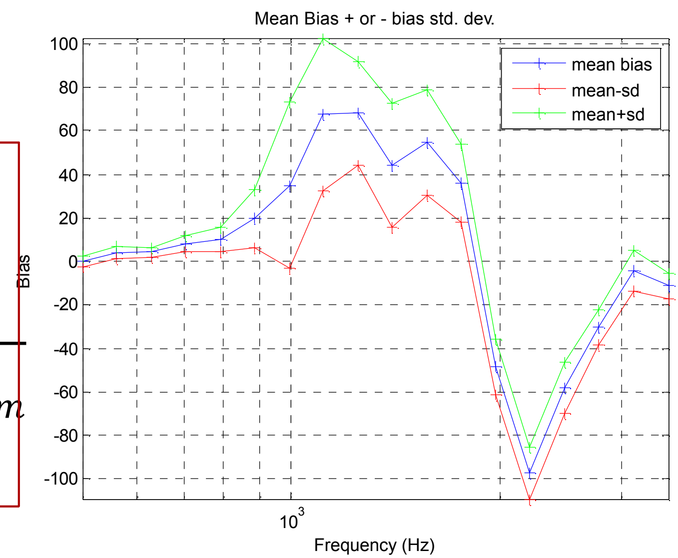




Bias and uncertainty

$$E = S - D$$

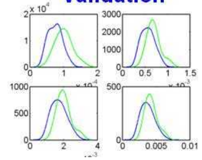
$$u_{val} = \sqrt{u_D^2 + u_{input}^2 + u_{num}^2}$$



Process to identify sources of uncertainty

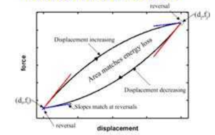
From this process, sources of uncertainty can be quantified

Model Validation

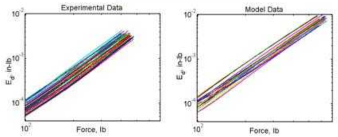


Representation and Geometric Fidelity

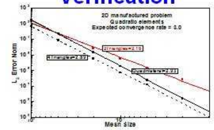
Physics and Material Model



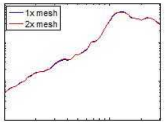
Uncertainty Quantification

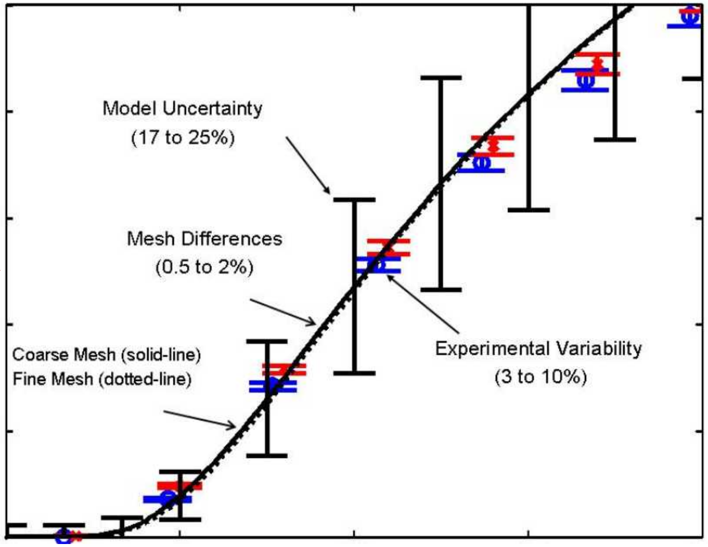


Code Verification



Solution Verification

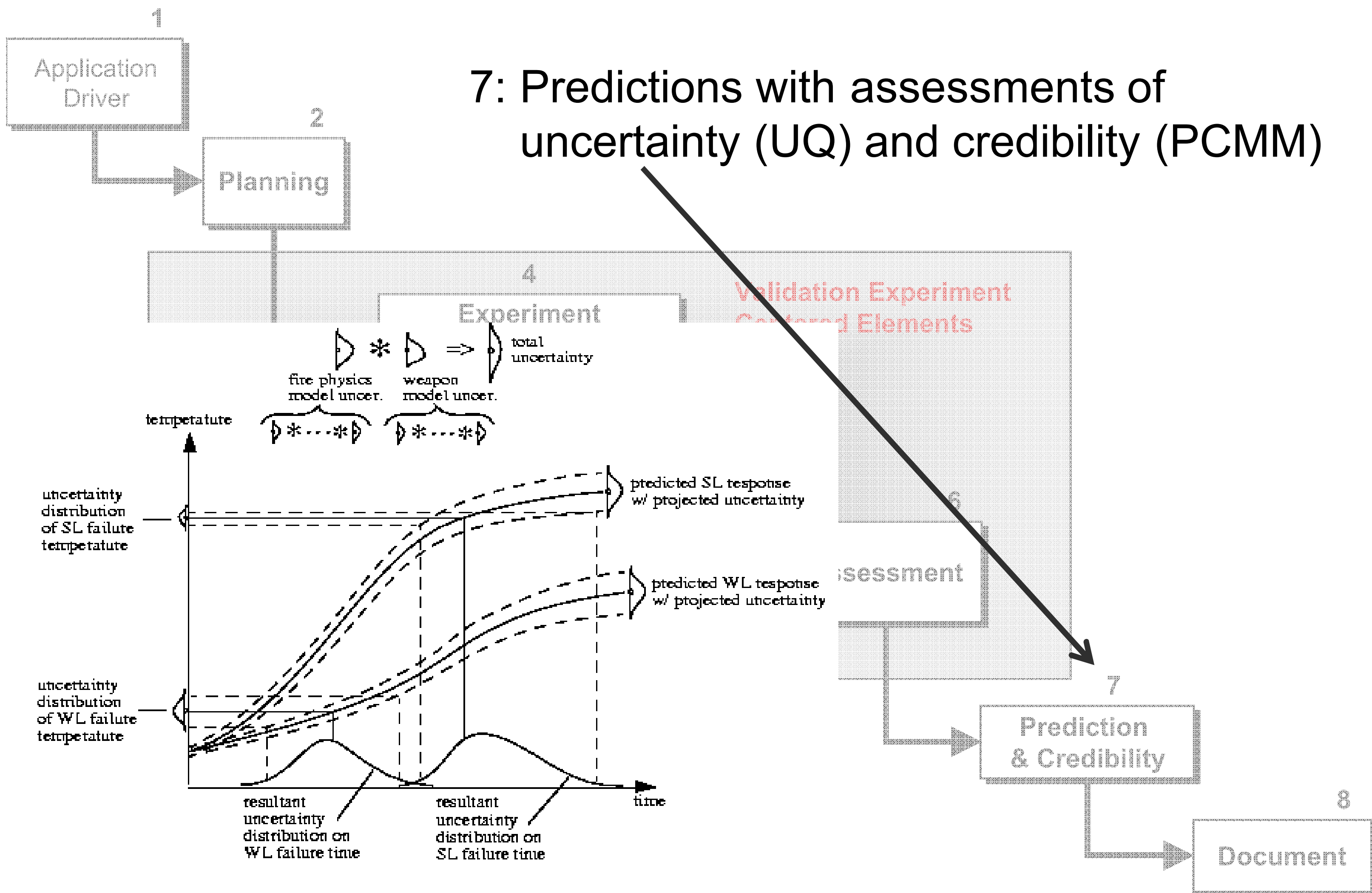




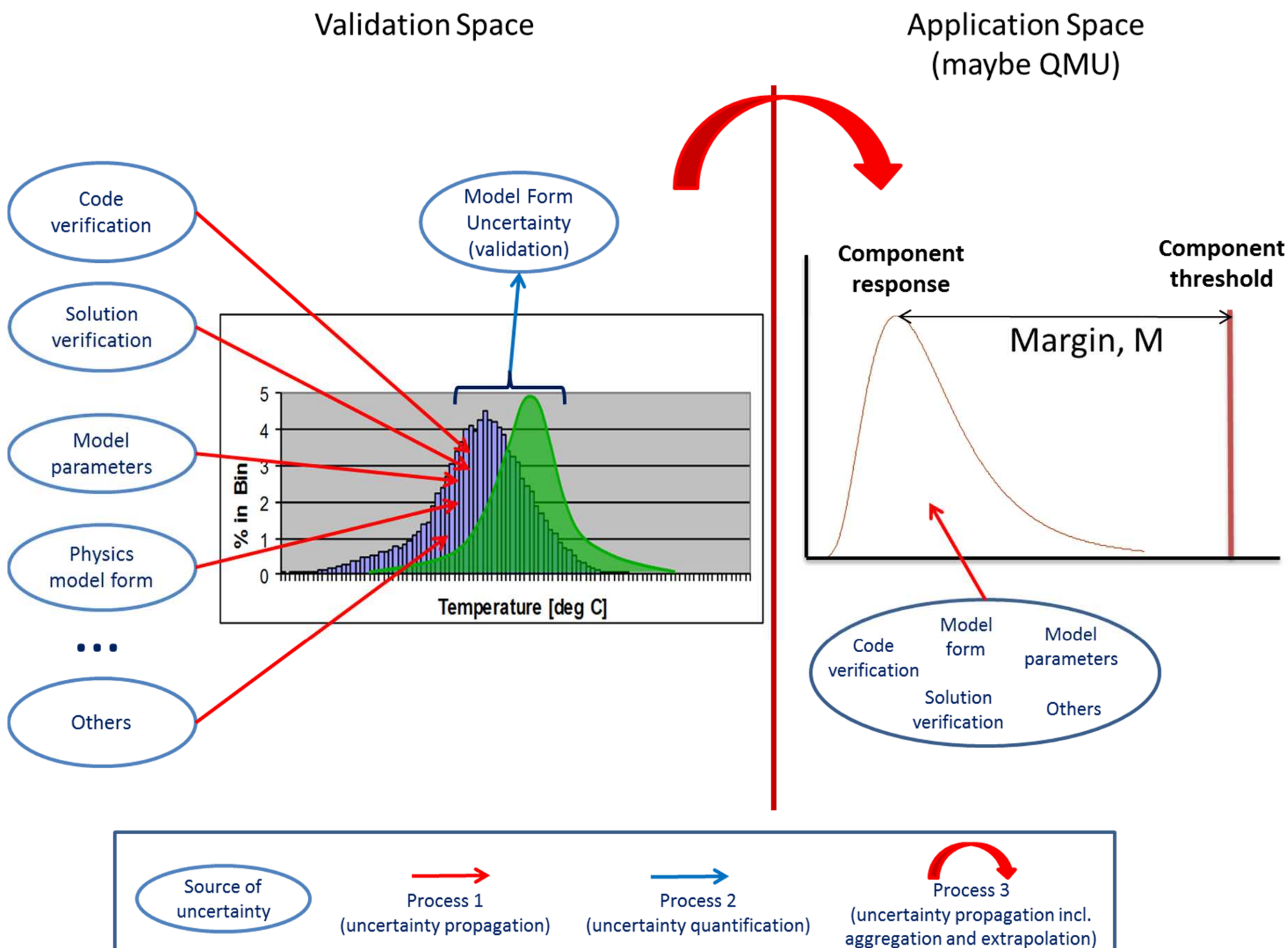
And can be aggregated to the performance quantity of interest (total uncertainty)

Overview of the Sandia V&V Process

7: Predictions with assessments of uncertainty (UQ) and credibility (PCMM)

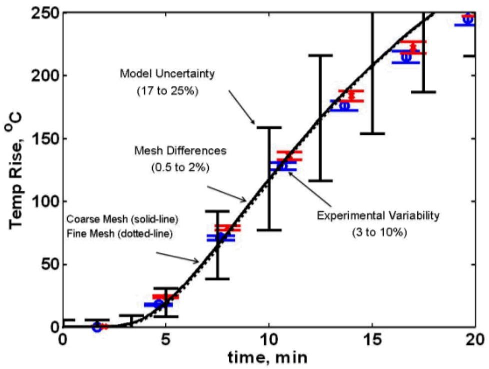


From validation to application space



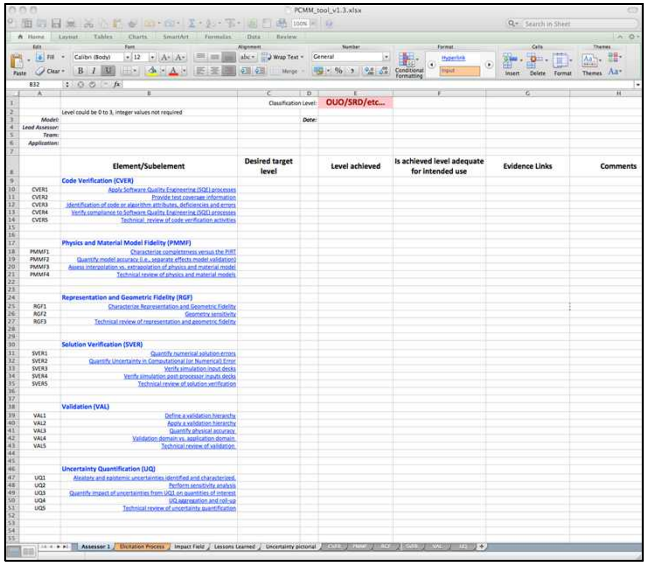
Note that QMU ≠ UQ …

UQ{ U_model(Validation)
+ E_numerical(Verification)
+ U_aleatoric(Variabilities) }



and

PCMM(credibility)



QMU entails both the numerical aspects (UQ) and the concept of how credible are the my simulations (i.e. what is their “pedigree)

What Gives CompSim Results Credibility?

Sandia National Laboratories

- PCMM is an assessment of credibility

Seven components

1. RGF: Representation and geometric fidelity
2. PMMF: Physics and material model fidelity
3. CVER: Code verification
4. SVER: Solution verification
5. VAL: Validation
6. UQ: Uncertainty quantification
7. Documentation and archiving

M&S (items 1–2)

V&V (items 3–5)

Focus here

UQ (item 6)

Predictive Capability Maturity Model
PCMM

Credibility Assessment

- The Predictive Capability Maturity Model (PCMM) is a communication tool for informing stakeholders of the level of maturity of an application-specific simulation capability
  - It is a multidimensional, qualitative metric
    - Determine readiness for stockpile issues
    - Identify gaps in credibility of application
    - Measure progress of integrated simulation effort
  - 6 Dimensions of the model:
    - Geometric fidelity
    - Physics fidelity
    - Code Verification (inc. SQE)
    - Solution Verification
    - Model Validation
    - Uncertainty Quantification

PCMM allows to qualitatively measure our CompSim "due diligence"

PCMM is intended to be a communication and a planning tool

It is not intended to be a report card

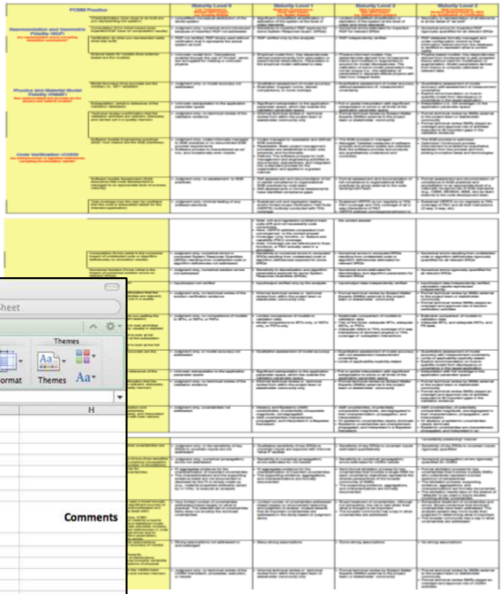
Evolution PCMM Tool

Sandia National Laboratories

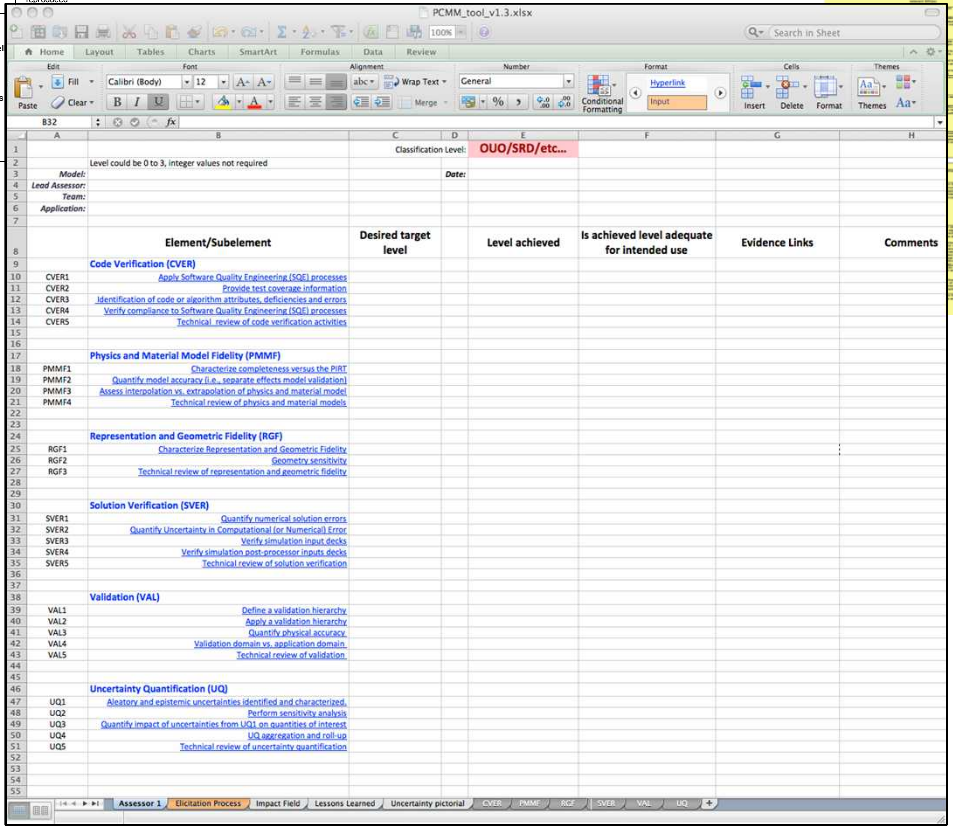
2007

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 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 stylization of 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&C) 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 an 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
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 experimental 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	
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	

~ 2009



2012-13



V&V Resources

- ESP 700 – Introduction to Verification, Validation and Uncertainty Quantification
  - POC: Josh Mullins (jmullin@sandia.gov)
  - Topics: Code and Solution verification, Basics of UQ, Dakota overview, model validation.
  - Videos are available
- Dept. 1544 - Validation and Uncertainty Quantification Processes
- Dept. 1441 - Optimization and Uncertainty Estimation

Summary

- The basic terminology relating to V&V/UQ was presented
- One of the main reasons for having a V&V/UQ process is to increase the confidence in CompSim results
  - PCMM is a way to communicate this confidence
- For more information, please contact
  - Josh Mullins, jmullin@sandia.gov; (505) 284-9169
  - Walt Witkowski (manager), wrwitko@sandia.gov; (505) 844-3869