

# We **TEND** to Underestimate Uncertainty in Our Predictions

SAND2010-7929P

**Martin Pilch PhD, PMP**  
**Thermal & Fluid Processes, 1514**  
**Sandia National Laboratories**  
**[mpilch@sandia.gov](mailto:mpilch@sandia.gov)**  
**505 845-3047**

**ASME International Mechanical Engineering  
Conference and Exposition  
Vancouver, British Columbia  
November 12-18, 2010**

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



**Sandia  
National  
Laboratories**

# Take the Perspective of *High Consequence* Decisions Where Predictive Modeling and Simulation Play a Key Role

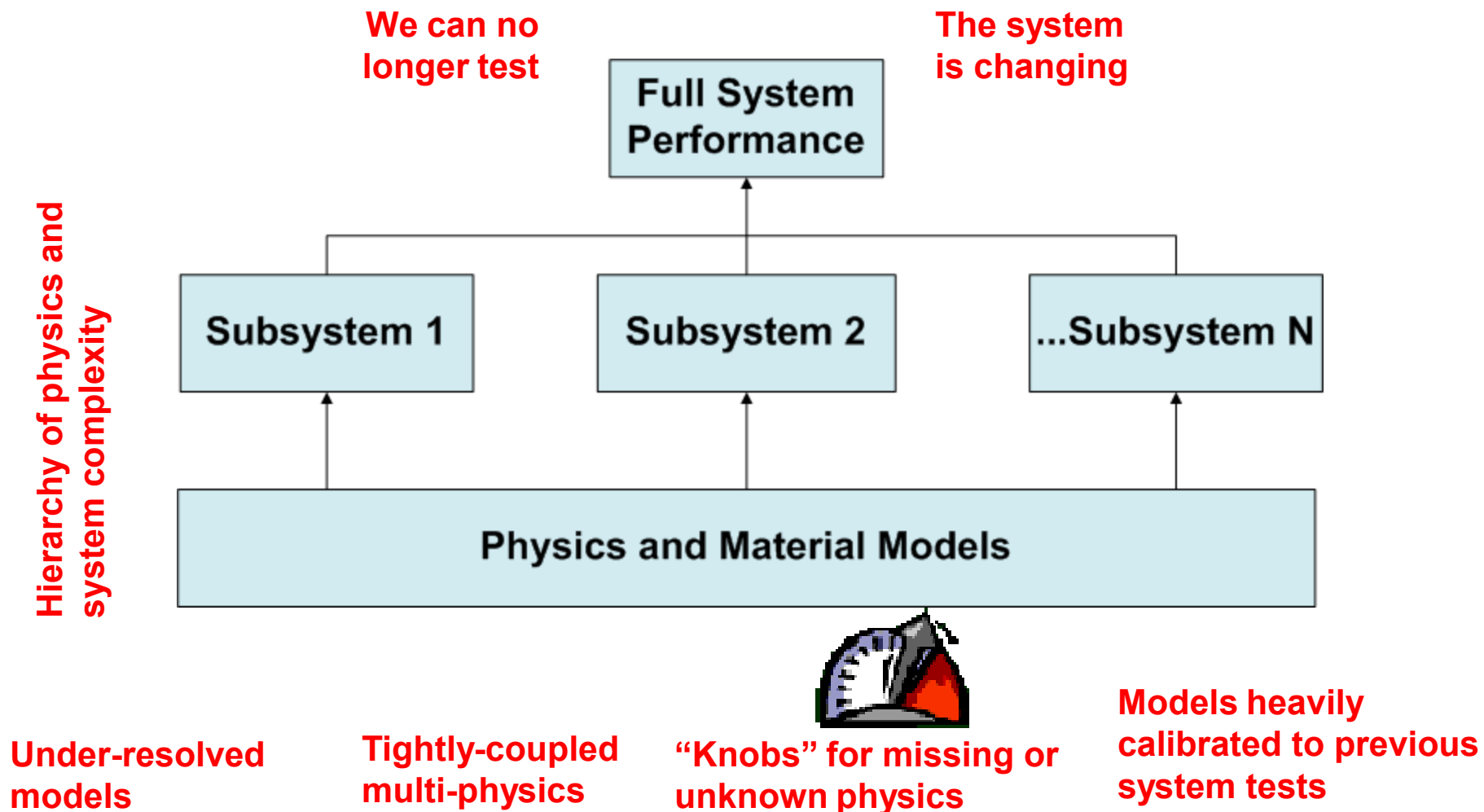


**The Safety and Security  
of Your Country**



**Your Life,  
or That of a Loved One**

# We Are Unable to Test Ultimate System Performance Until First Use



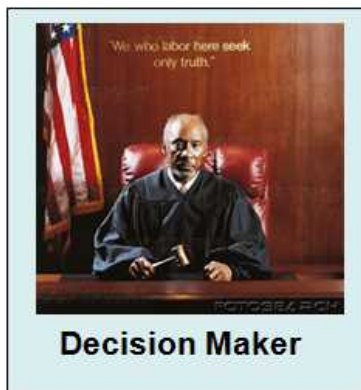
# Complex, High-Consequence, or Contentious Decisions Are Usually *Risk-Informed*

**A Lot of Other Intangibles Associated with the Decision Environment**

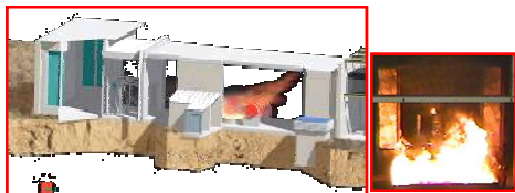


**Good Design Principles**

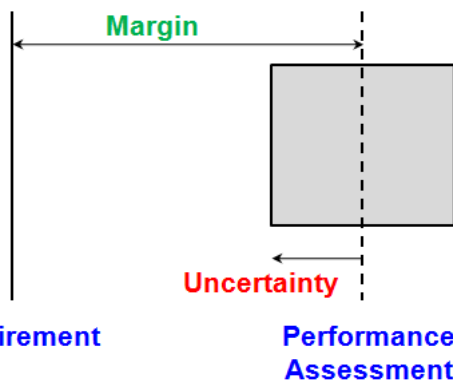
**Cost/Scheduled Constrained**



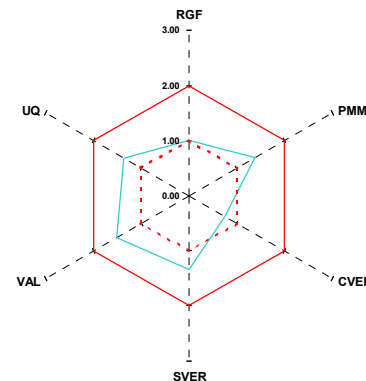
**Peer Review Early and Often**



**Testing Suitable for Model Validation and Qualification**

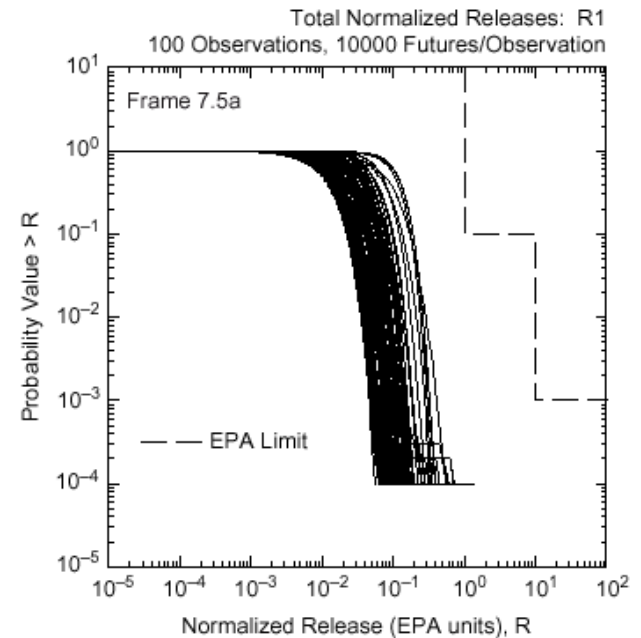
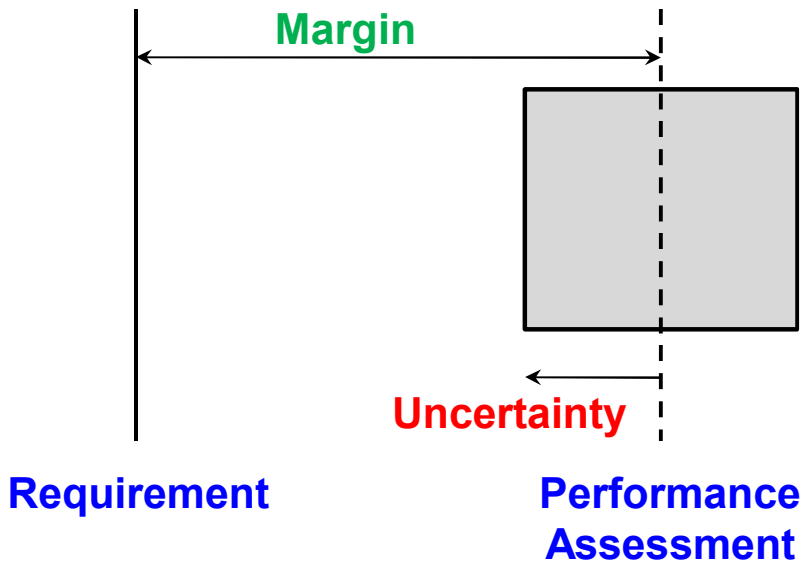


**Quantified Margins and Uncertainties**



**Credibility That is Measured and Communicated**

# Quantifying Uncertainties is Essential to Risk Informed Decision Making



SF =  
Performance/Requirement

Engineering Safety  
Factors

CF = Margin/Uncertainty

Quantified Margins and  
Uncertainties (QMU)

$p(f_{\text{Perf}} > f_{\text{Req}}) >$   
Confidence Requirement

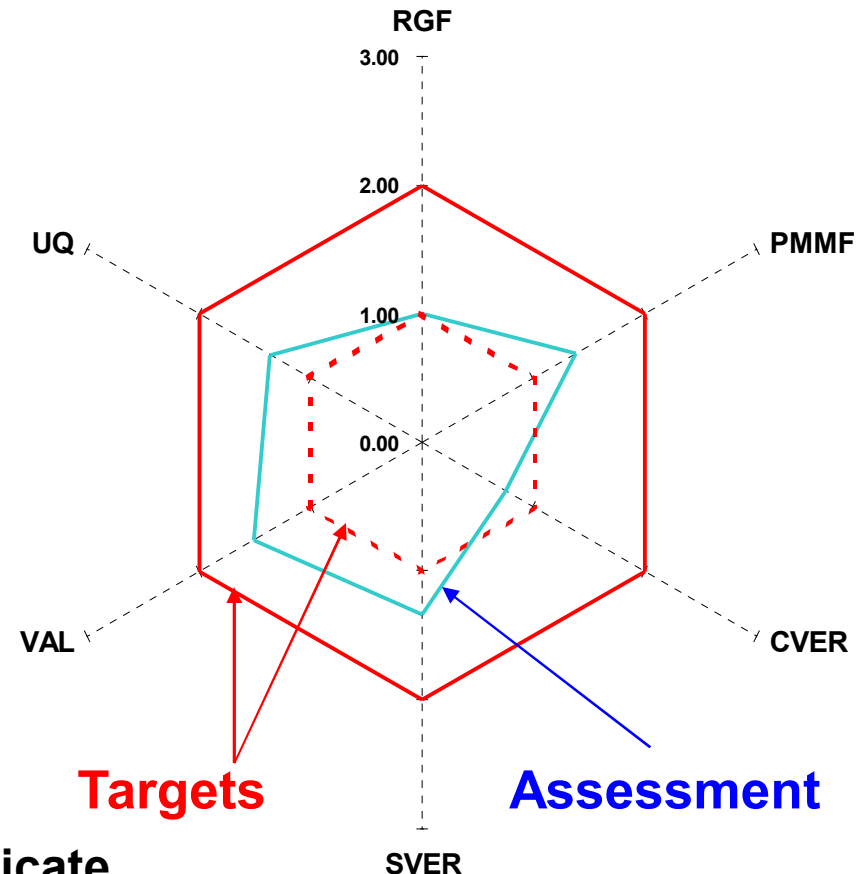
Quantitative Risk  
Assessment (QRA)

# What does it Mean to have *Predictive Capability*?

American Heritage Dictionary: *Predict* – To state, tell about, or make known in advance, especially on the basis of special knowledge

## What *special knowledge* do we demand to assert a predictive capability in M&S?

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



R&D: How do you measure and communicate progress in predictive capability?

# My Goal Today is to ....

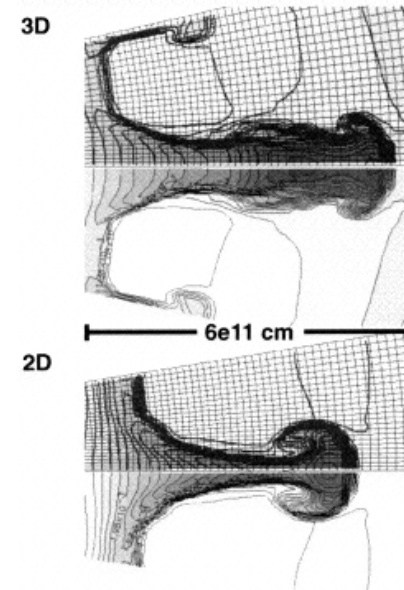
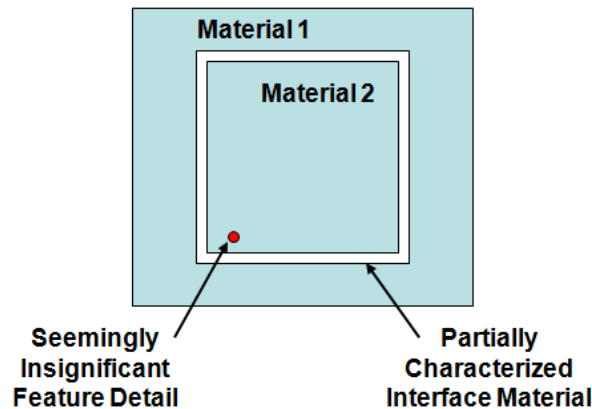
- Use the previous 6 attributes as a framework to show how bias's and uncertainties creep into our predictions
- Offer mini-tutorials (in some cases) in order to provide perspective
- Suggest some research agendas, often grand challenge in nature

**Stimulate your thinking about  
some things you might not  
have thought about**



# Representation or Geometric Fidelity

Are representation errors corrupting simulation results?



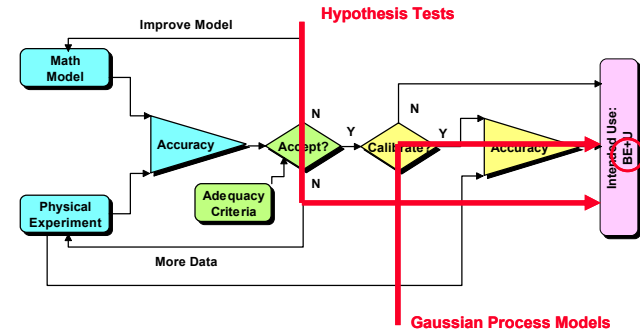
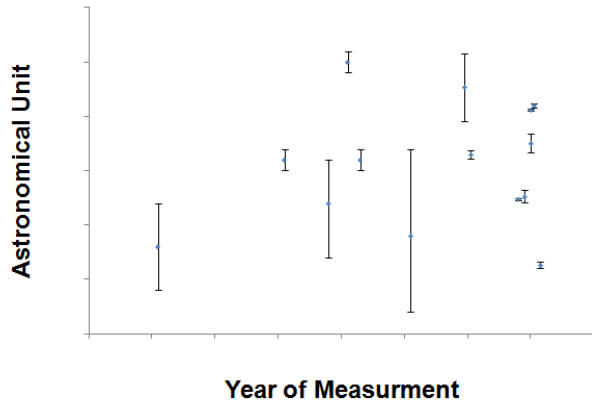
**R&D: What *evidence* is required to assert that you have adequate geometric fidelity in your models?**

- How do you quantify the impact of imperfect geometry?

**R&D: We need a practical means to deal with dimension variability in parts**

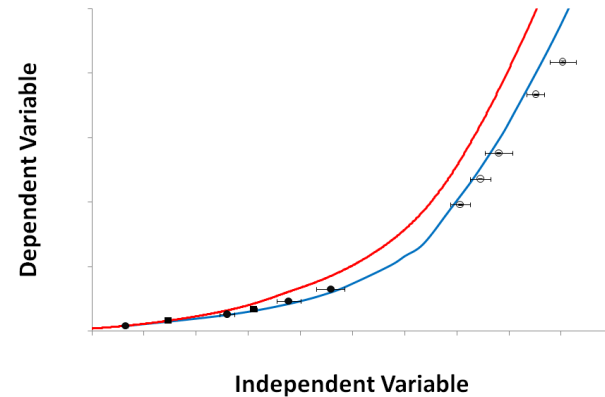
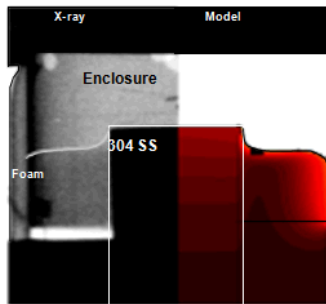
# Physics and Material Model Fidelity

## How accurate are the physics and material models?



**We Tend to Overlook Systematic Errors in Measurements**

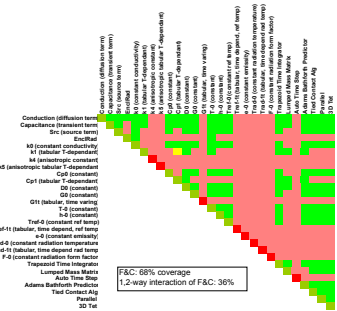
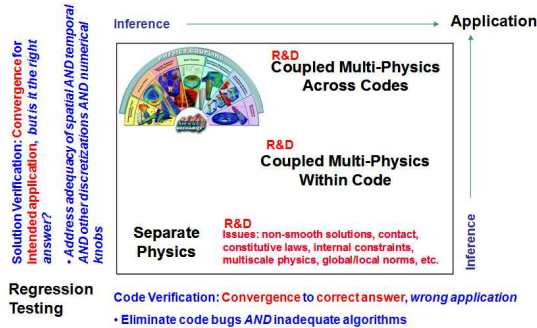
**Validation is Not Calibration**



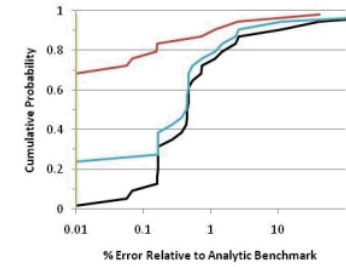
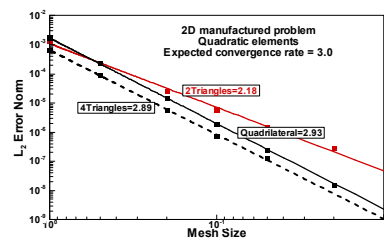
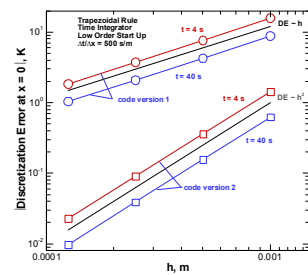
**Vugraph Norms and Curve Overlays  
Fall Short of Quantifying Model Accuracy**

# Code Verification

Are software errors or algorithm deficiencies corrupting simulation results?



~~Code1:Code2 Comparisons~~

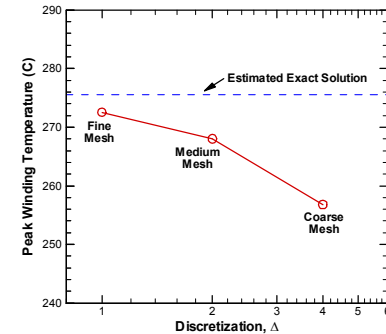
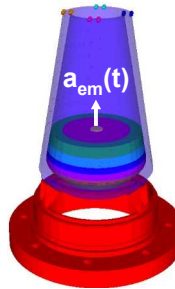
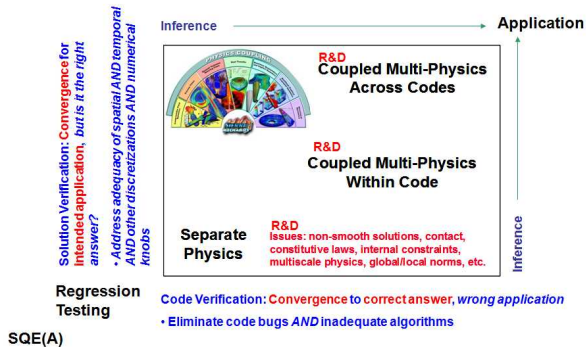


R&D: How can you estimate the impact of undetected code or algorithm deficiencies?

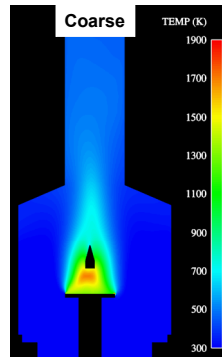
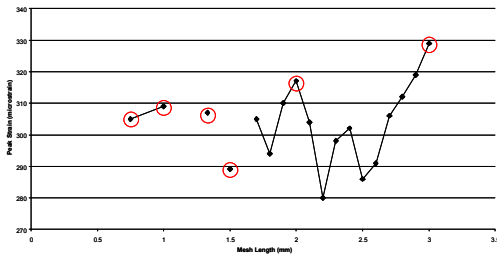
# Solution Verification

Are human procedural errors

or numerical solution errors corrupting simulation results?



SQE(A)

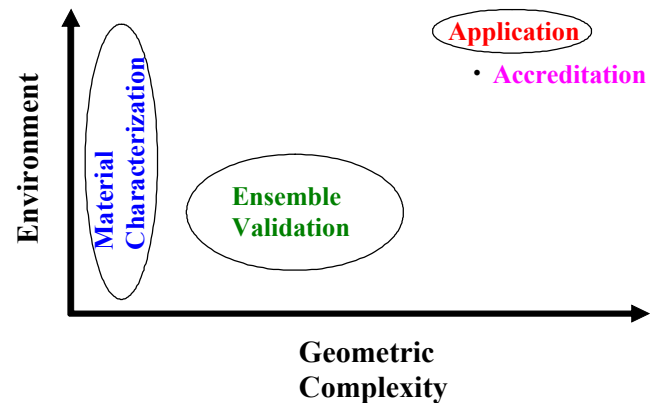
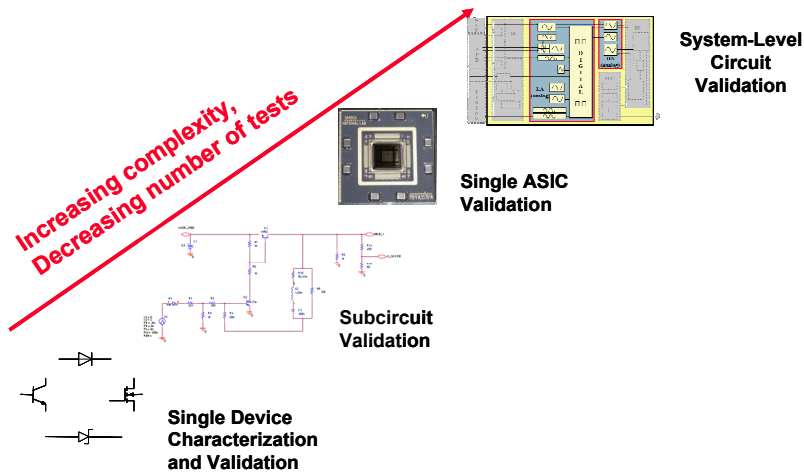


R&D: How do you make unbiased predictions with under-resolved models?

R&D: How do you quantify and propagate errors associated with non-convergent solutions?

# Validation

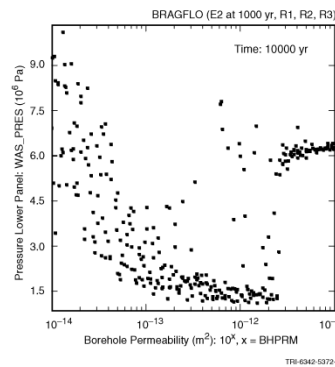
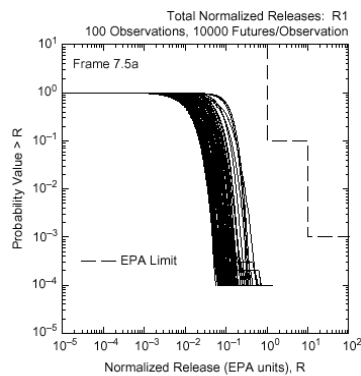
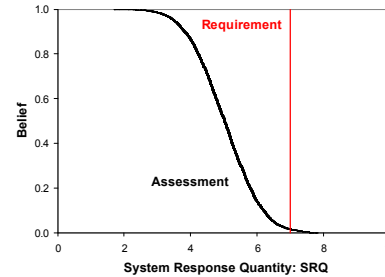
How accurate are the integrated physics and material models?



R&D: How do you use biased models to make unbiased predictions with uncertainty at the next level in the hierarchy

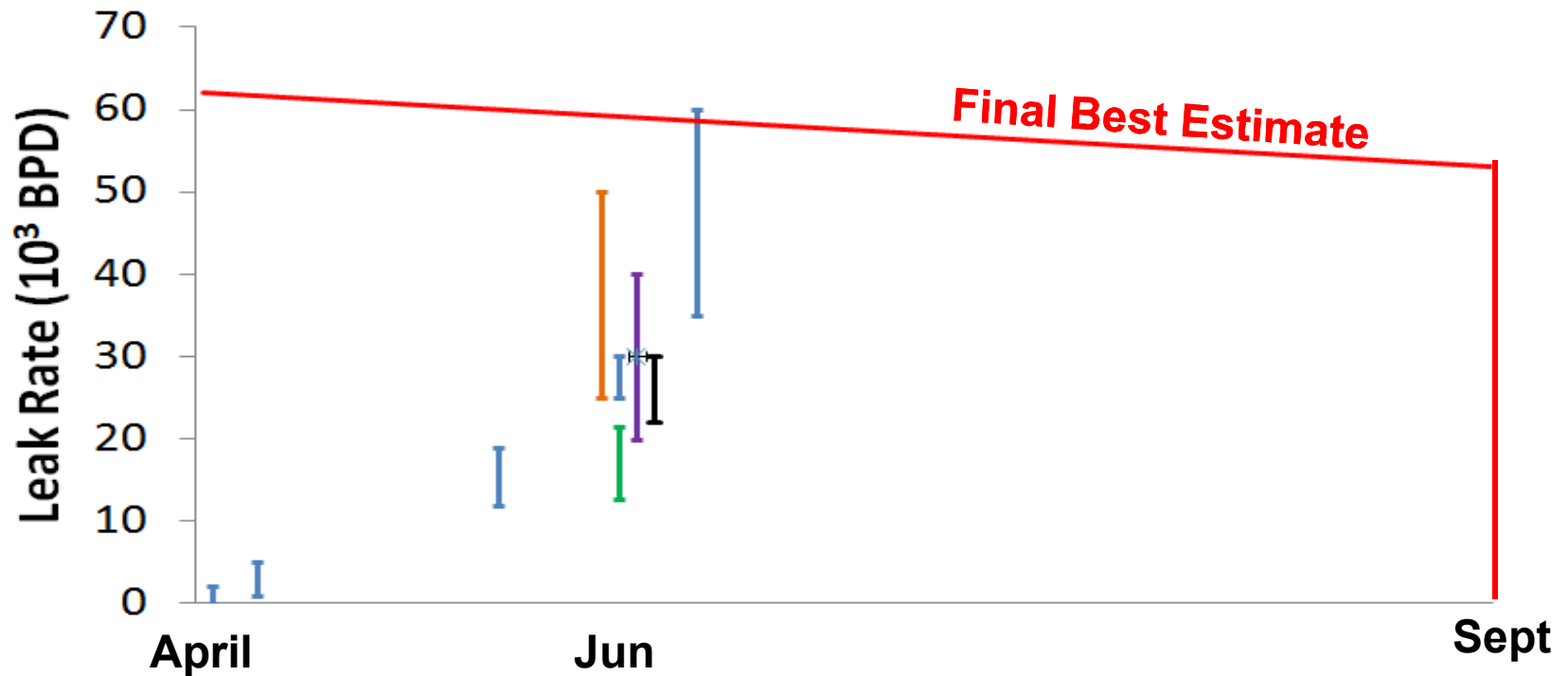
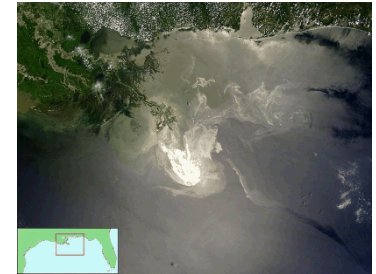
# Uncertainty Quantification

What is the impact of variabilities and uncertainties on system performance and margins?



# Measurements of Well Leak Rate Were Biased and Uncertainty Was Underestimated

Deepwater Horizon: April 22 – Sept 19, 2010



# In Decision Support, You Have to Maintain a Balanced Perspective

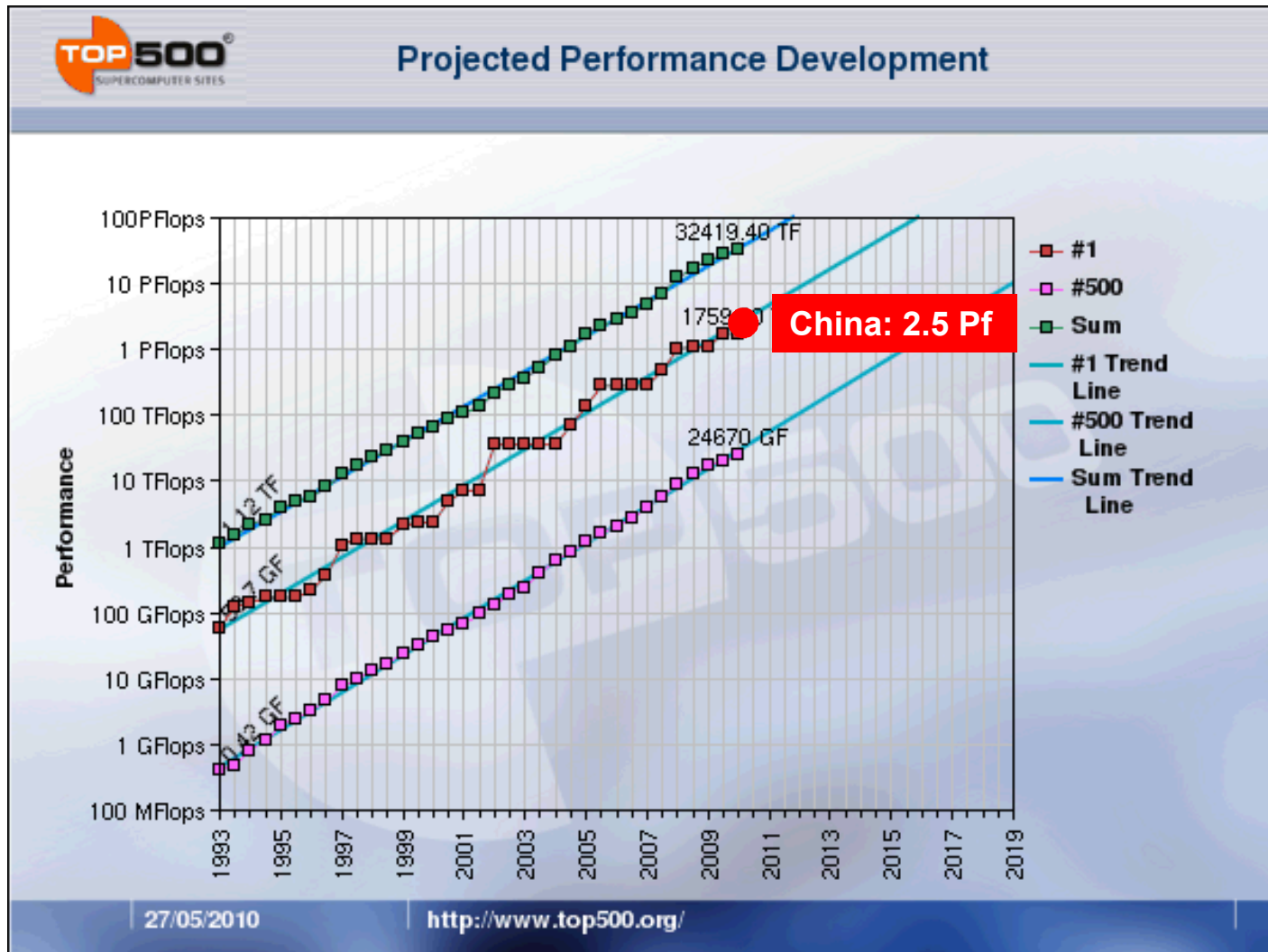
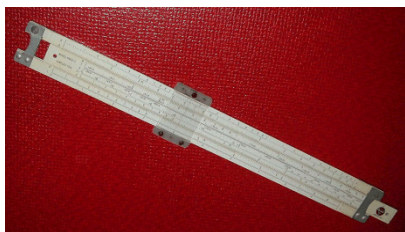


- **Uncertainties arising from:**
  - **Geometric fidelity**
  - **Physics and material model fidelity**
  - **Code verification**
  - **Solution verification**
  - **Validation**
  - **Uncertainty quantification**
- **Systematic differences and errors**
  - **Institution/institution**
  - **Analyst/analyst**
  - **Methodology/methodology**
- **Unknown/unknowns**

**R&D: How Can You Routinely Deal With the Full Spectrum of Bias' and Uncertainties in a Consistent Conceptual Framework?**

# HPC is Foundational to Quantifying Many of the Bias' and Uncertainties

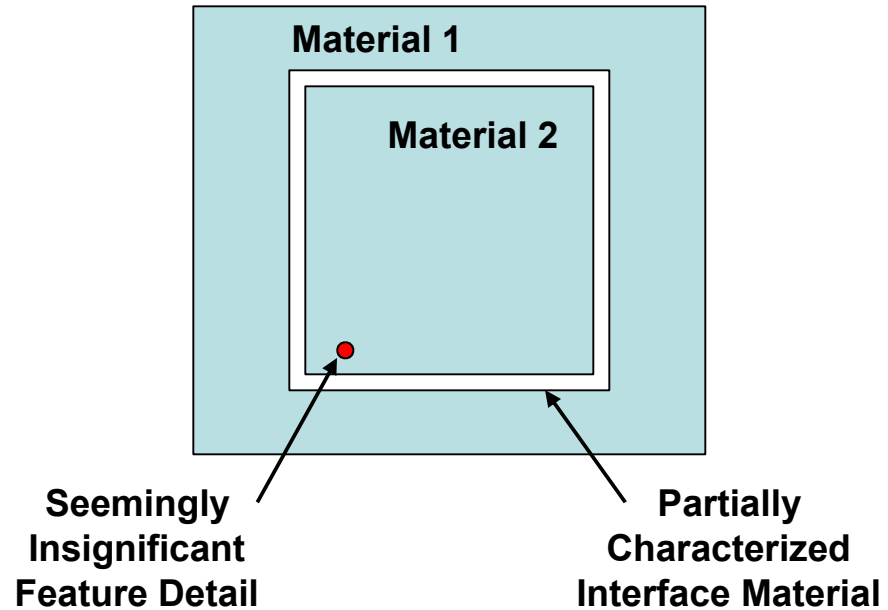
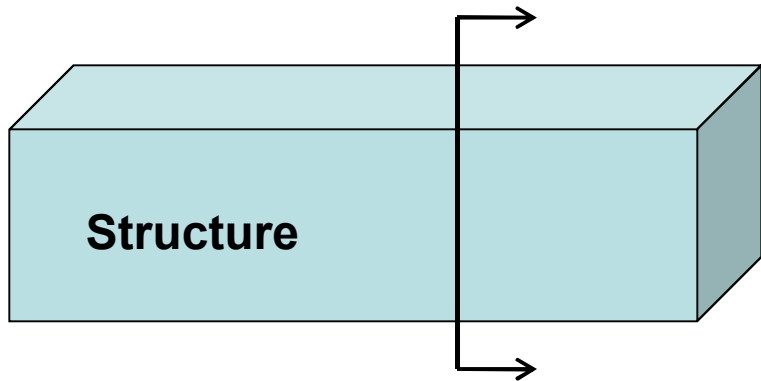
HPC (~ 1 flop/s)  
1620 - 1974



# Hyperlinked Pages

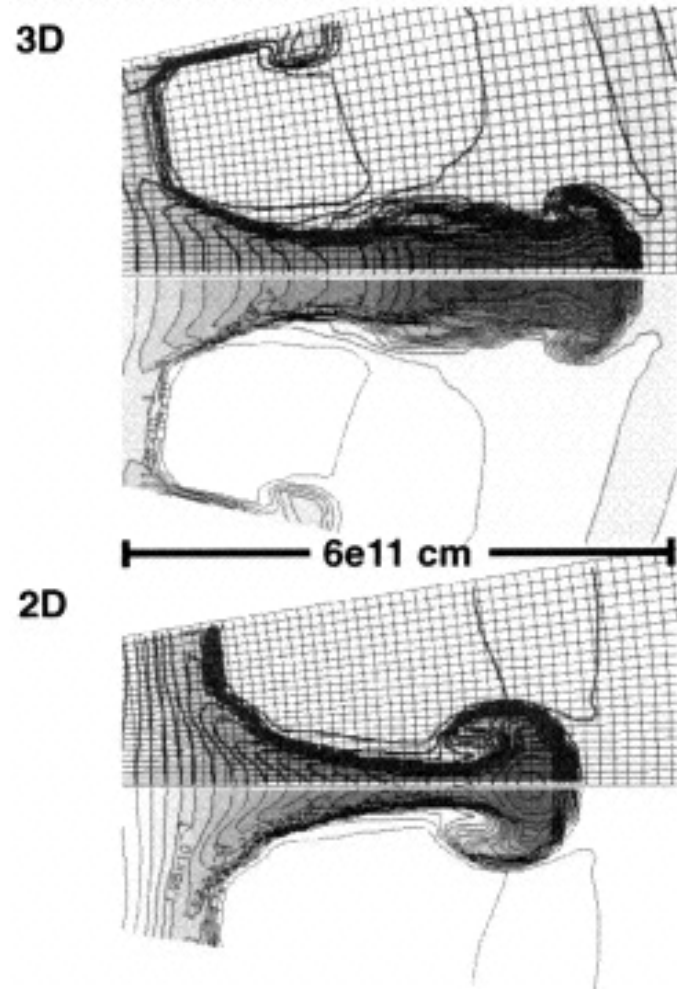
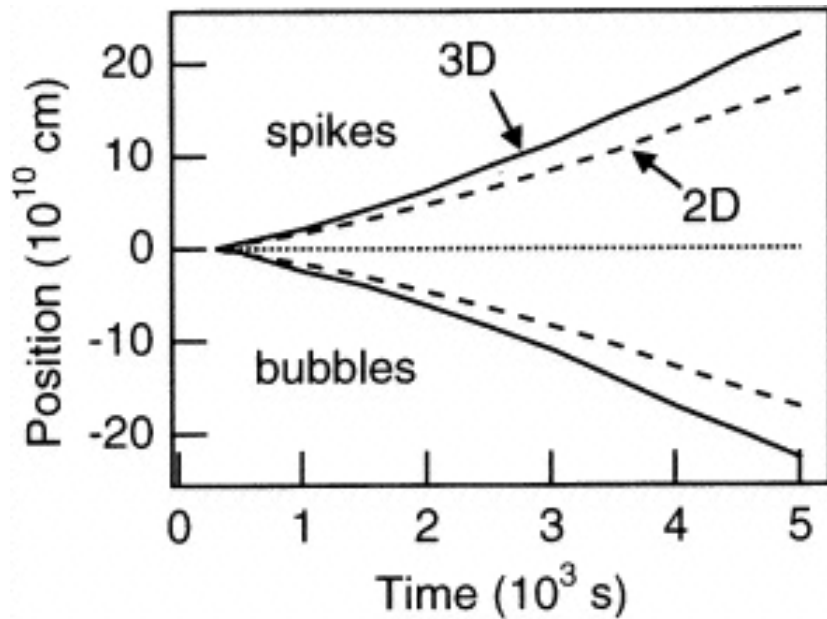
- **Listed in order as they appear in main body of talk**

# Calibration Hides Errors Resulting From Inadequate Geometric Fidelity



	Relative Error (%) = (Pred - Meas) / Meas		
	Defeatured	Unphysical Calibration	Full Feature
Mode 1	34.9	-4.5	-3.5
Mode 2	38.0	-5.1	-3.8
Mode 3	6.4	-2.6	-0.7
Mode 4	3.2	-4.9	-2.3

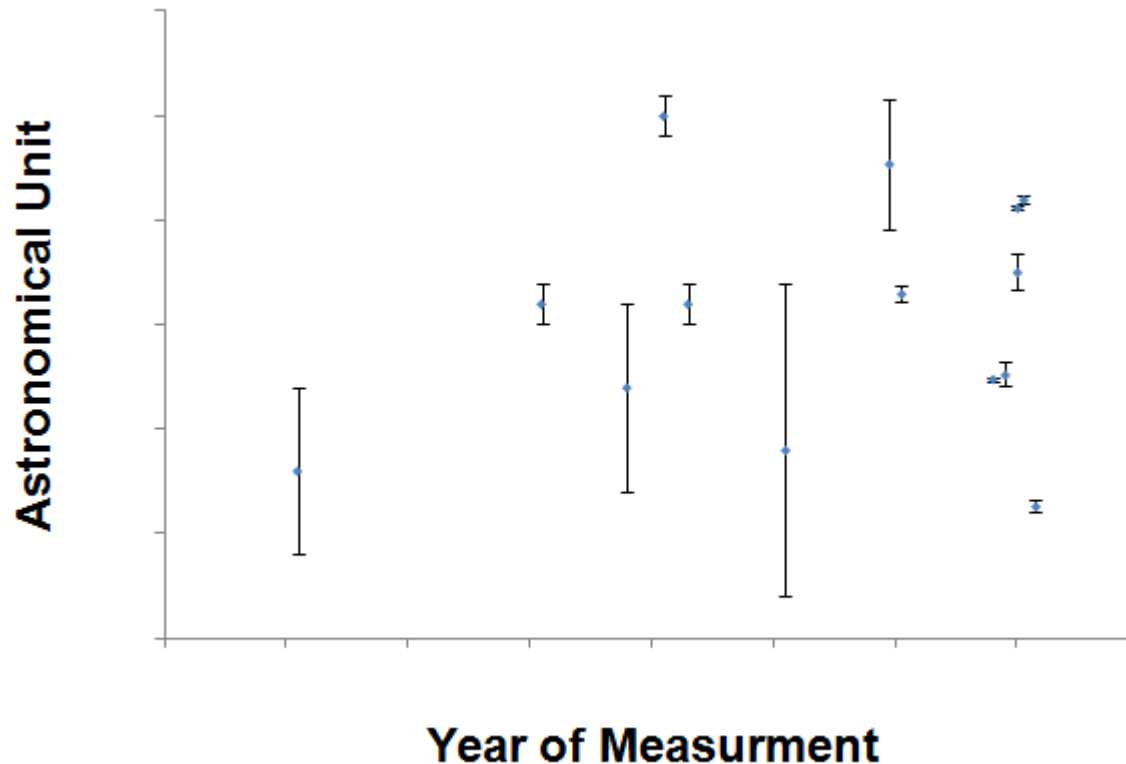
# Instabilities Behave Differently in Different Dimensions



L. Kane et. al., *Two-Dimensional vs Three-Dimensional Supernova Hydrodynamic Instability Growth*, The Astrophysical Journal, 528:989-994

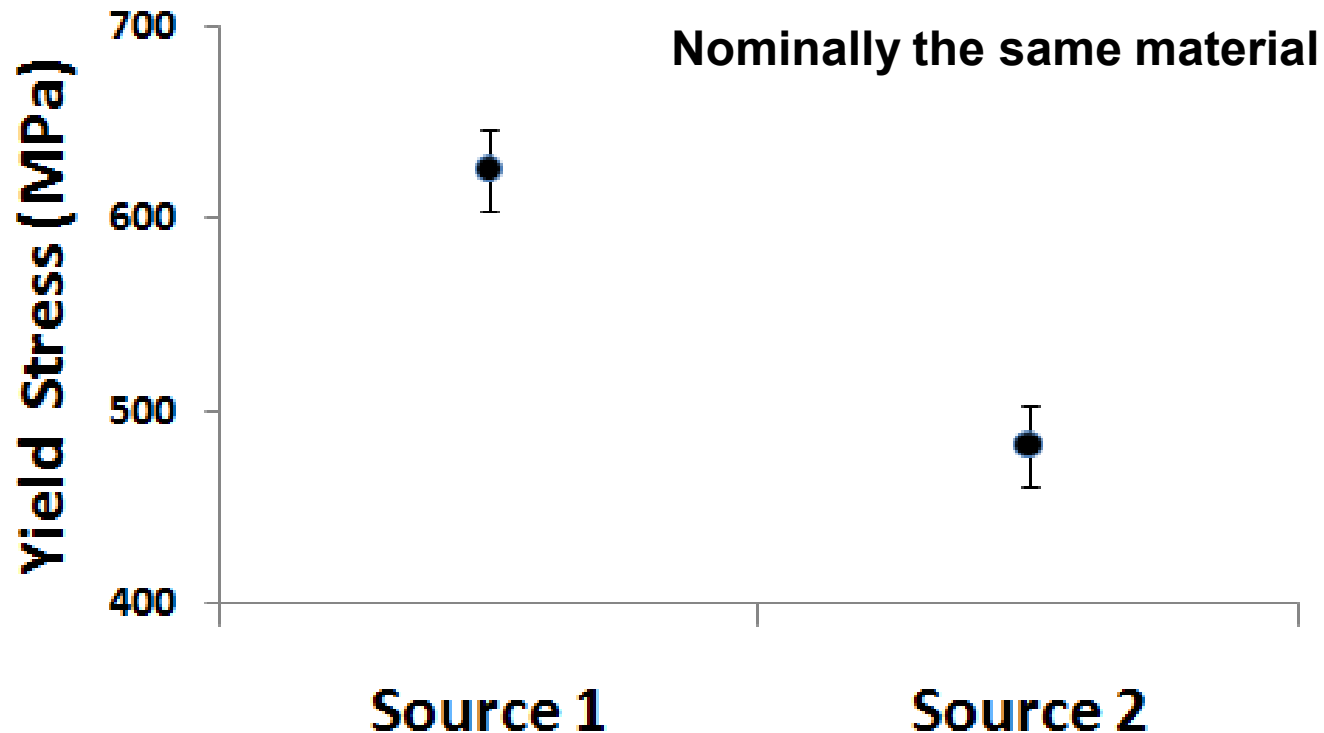
0D or 1D: What Instability?

# Systematic Errors Permeate Measurements of Even Our Physical Constants



- We might consider sample/sample variability (SE ~ 5-10%)
- **But we typically overlook systematic errors associated with vendor/ vendor, batch/batch, testing-lab/testing-lab, or processing/processing differences**

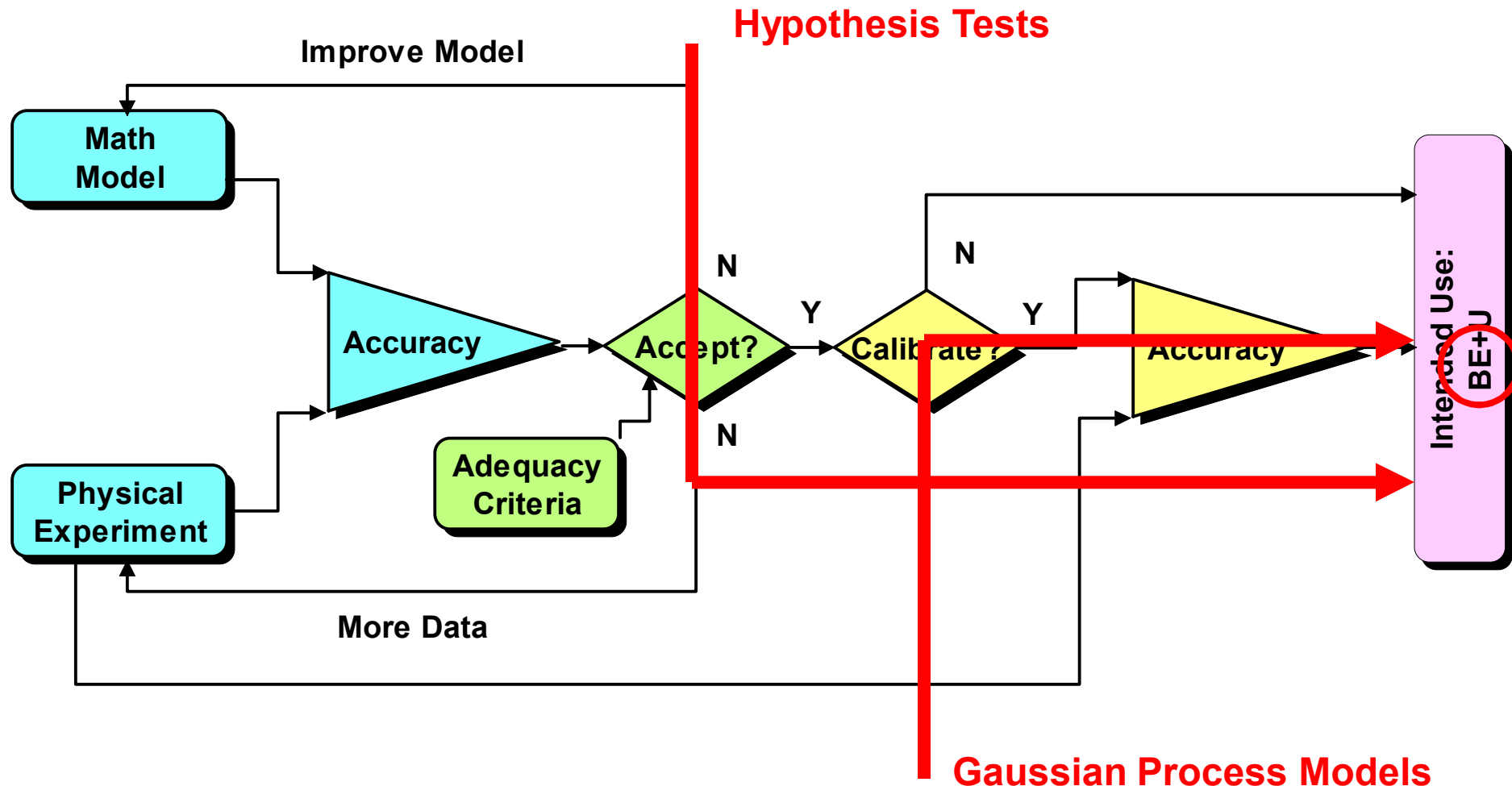
# We Tend to Overlook Sources of Systematic Error in Our Material Property Data



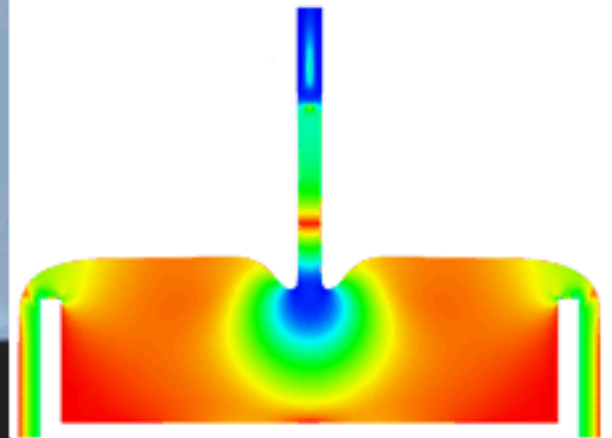
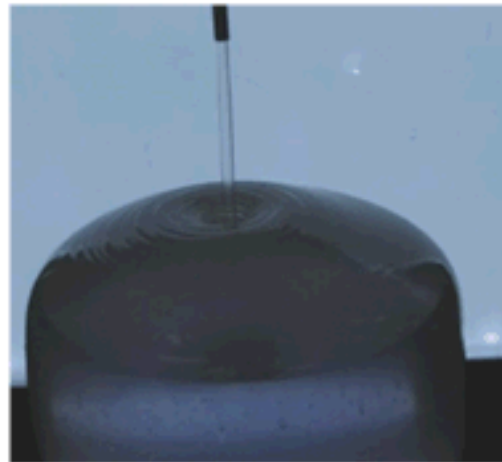
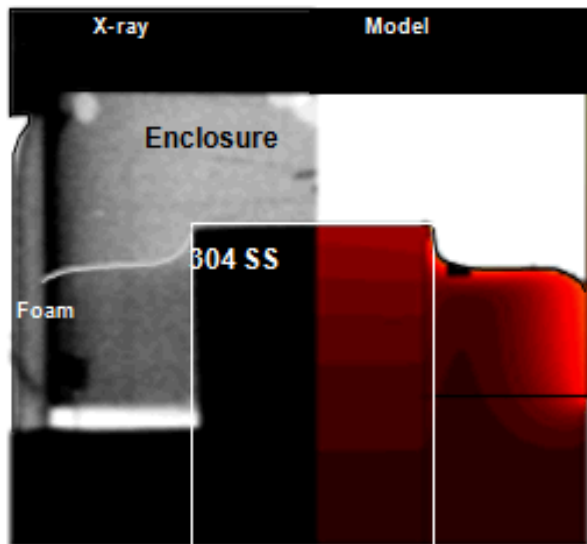
- We might consider sample/sample variability (SE ~ 5-10%)
- But we typically overlook systematic errors associated with vendor/ vendor, batch/batch, testing-lab/testing-lab, or processing/processing differences (~20-50%)

# Validation is Assessment Supporting BE=U

## Calibration is not Validation

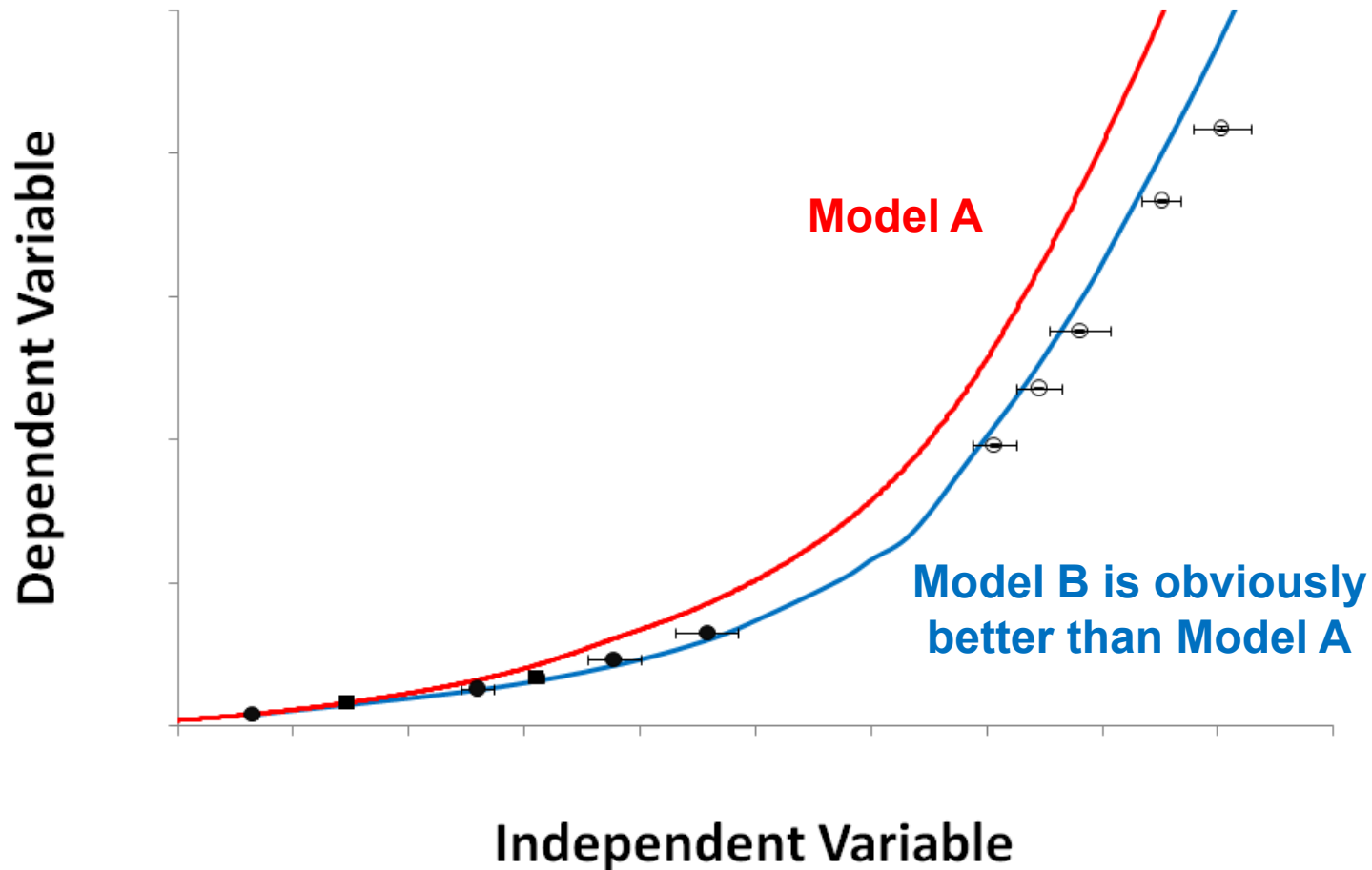


# Vugraph Norms Say Nothing About the Accuracy of Our Models



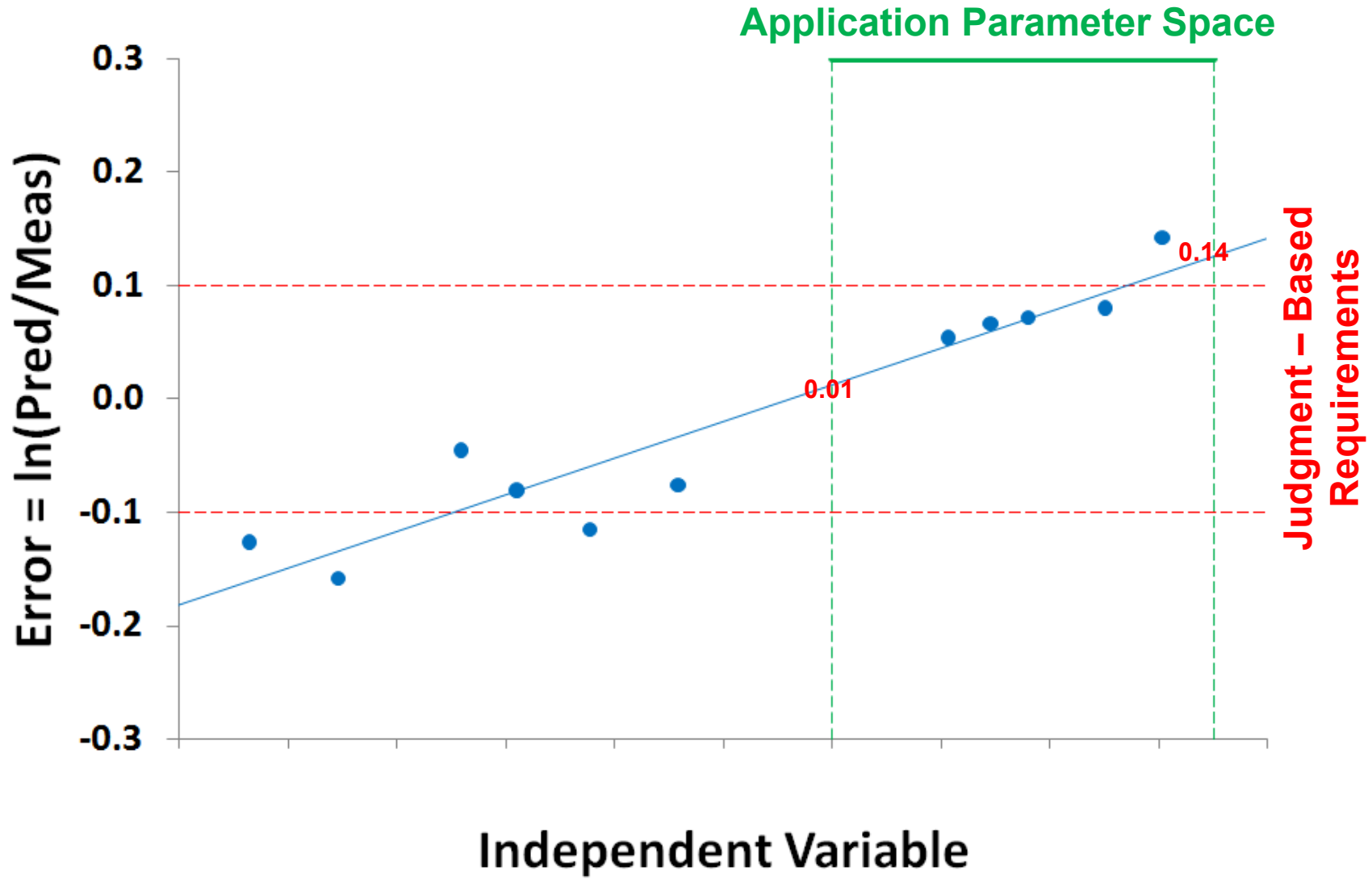
**Acknowledge that vugraph norms can be useful for  
discovery or when the ho-ha test is good enough**

# Curve Overlays are State of the Art for Validation Comparisons in Journals

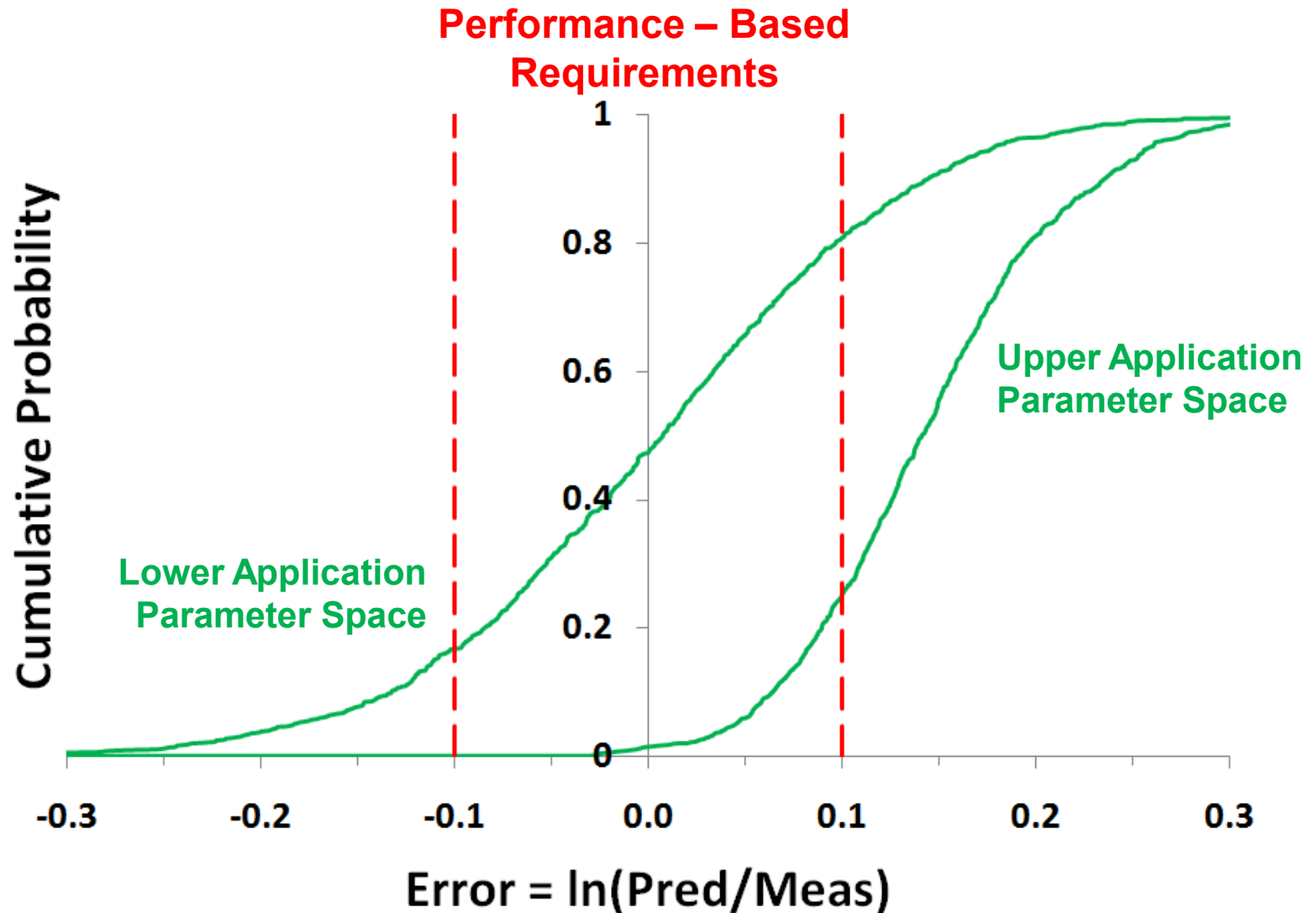


How accurate is the Model B?

# It is the Norm That Our Physics Models Are Biased and do not Fully Reproduce Trend Behaviors



# We Have to be More Rigorous When Dealing With Performance-Based Requirements



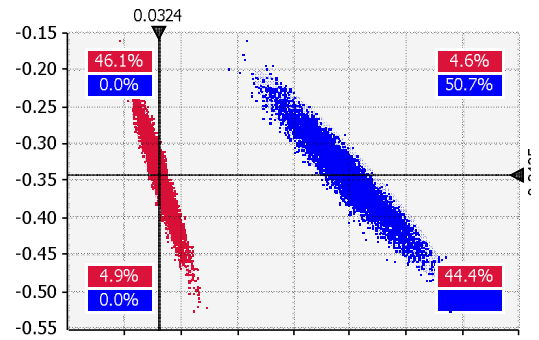
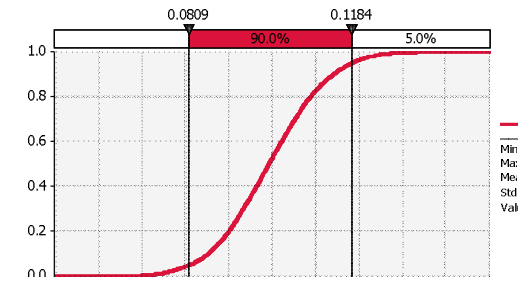
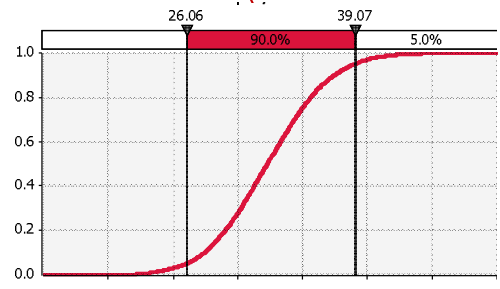
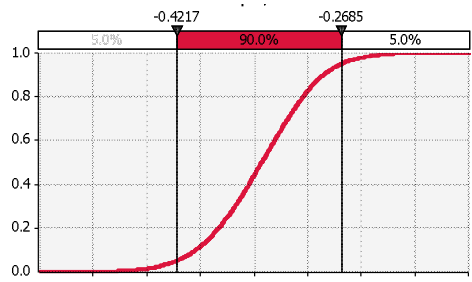
# R&D: How do you use a Biased Model to Make Unbiased Predictions with Uncertainty?

Reality = Model( $x_{nom}$ ) \* Exp(-Error)

**Bias and Trend Correction**

**Test/Test Variability**

Error = N( **a** + **bx** ;  $\sigma$  )



# Code and Solution Verification

Demonstrates **Convergence** to the **Correct Answer** for the **Intended Application**

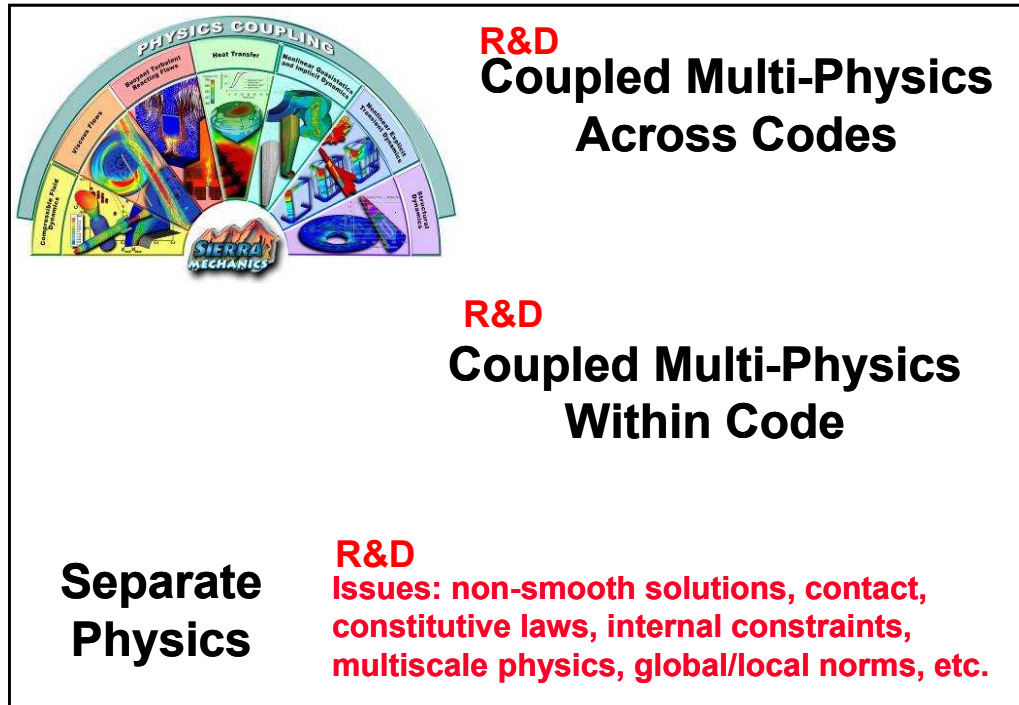
**Solution Verification: Convergence for intended application, but is it the right answer?**

**• Address adequacy of spatial AND temporal AND other discretizations AND numerical knobs**

Inference



Application



Inference

Regression Testing

**Code Verification: Convergence to correct answer, wrong application**

**• Eliminate code bugs AND inadequate algorithms**

SQE(A)

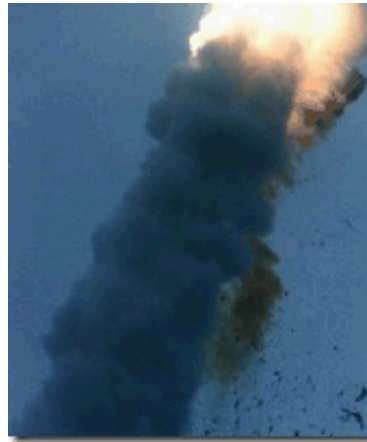
**R&D: Is it possible to demonstrate convergence to the correct answer for the intended application without inference?**



# Software Errors Are Costly: Both in Dollars and Reputation



**Hartford Coliseum  
(\$70M)**



**Ariane 5  
(\$500M)**



**Mars Climate  
Orbiter (\$125M)**

<http://www.devtopics.com/20-famous-software-disasters/>

# Code to Code Comparisons

## Are a Poor Substitute for Formal Verification

Trucano et. al., On the Role of Code Comparisons in Verification and Validation

### Code Comparison Principle (CCP)

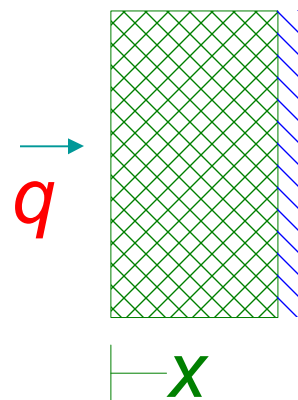
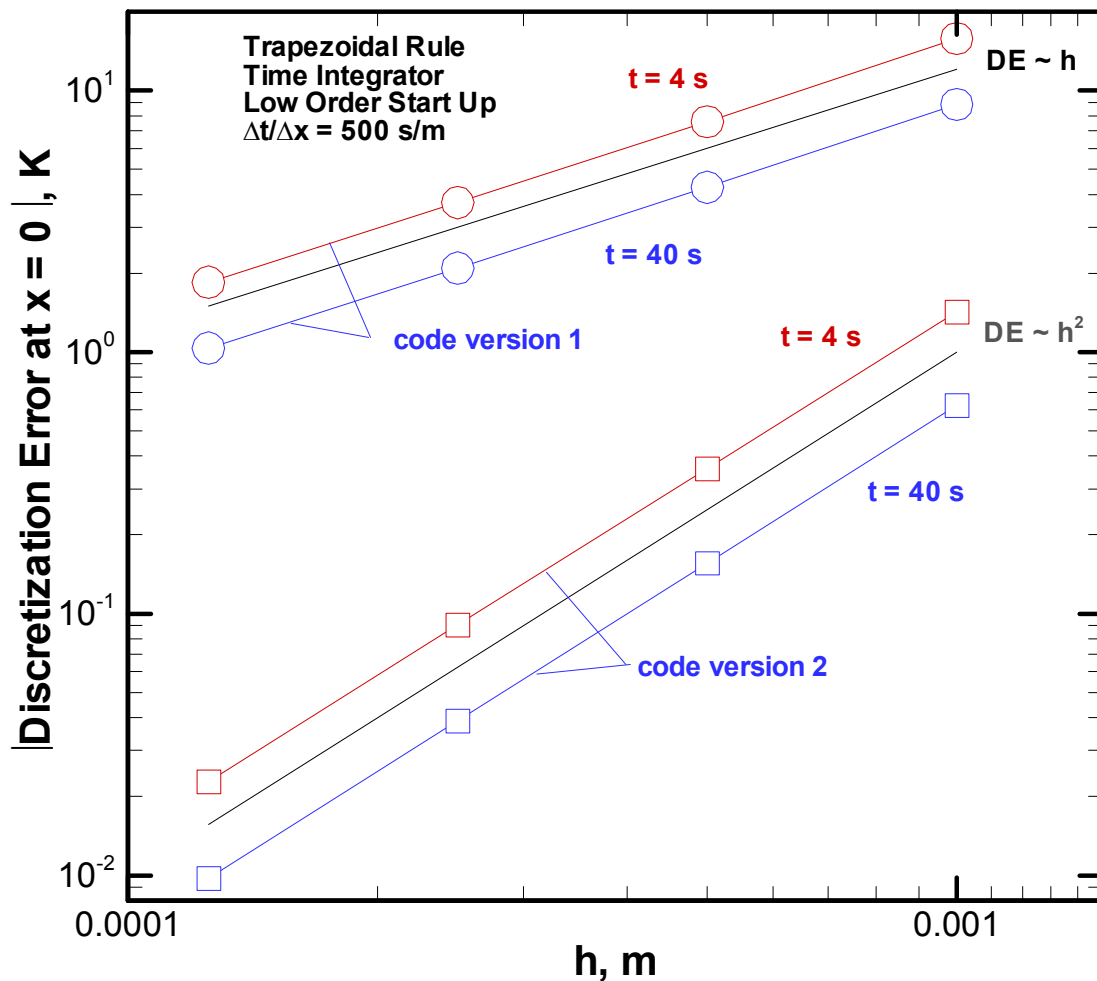
**Code 1 = assessed code      Code 2 = benchmark code**

$$\|\text{Code 1} - \text{Truth}\| \leq \|\text{Code 1} - \text{Code 2}\| + \|\text{Code 2} - \text{Truth}\|$$

- $\|\text{Code 1} - \text{Code 2}\|$  · What if this term is not negligible?**
- **Could be that Code 1 models are different from Code 2 models**
  - **Could be a bug in Code 1 or Code 2**
  - **Could be an algorithm flaw in Code 1 or Code 2**
  - **Could be that Code 1 or Code 2 model is not converged**

**Points to path for better code-to-code comparisons; but if Code 2 is formally verified, why not verify Code 1 to the same verification test suite? And if not, why bother with the code-to-code comparison?**

# “Order of Convergence” is a Sensitive Metric for Detecting Algorithm Deficiencies



- Transient response of planar 1-D slab to constant flux with analytic solution as the benchmark
- Code bug discovered and fixed based on priority and resource availability. Status tracked in code issue log, which can be accessed by analysts

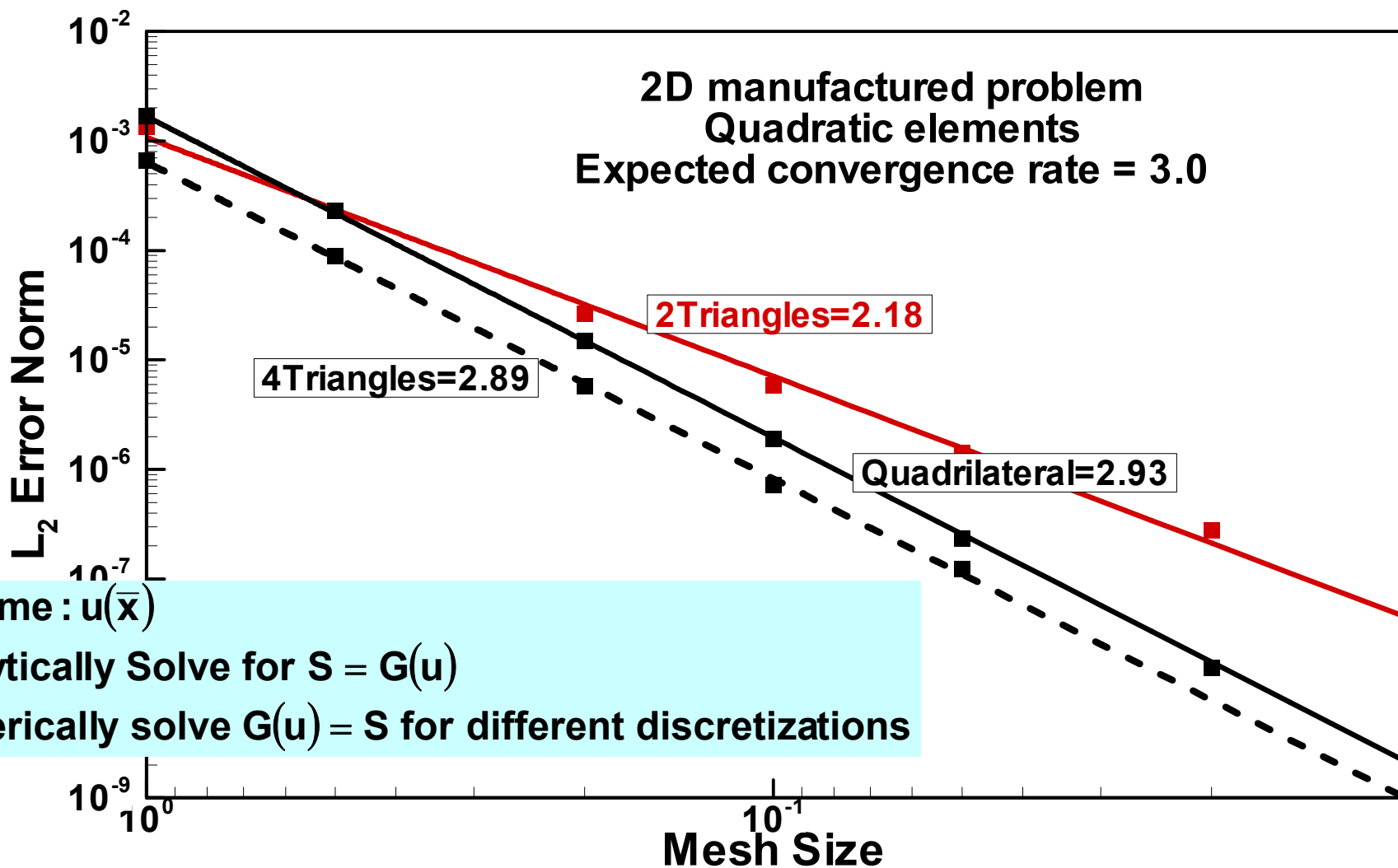
Modeled as full 3-D object



Sandia  
National  
Laboratories

# No Exact Analytic Solution?

## Verification with a Manufactured Solution

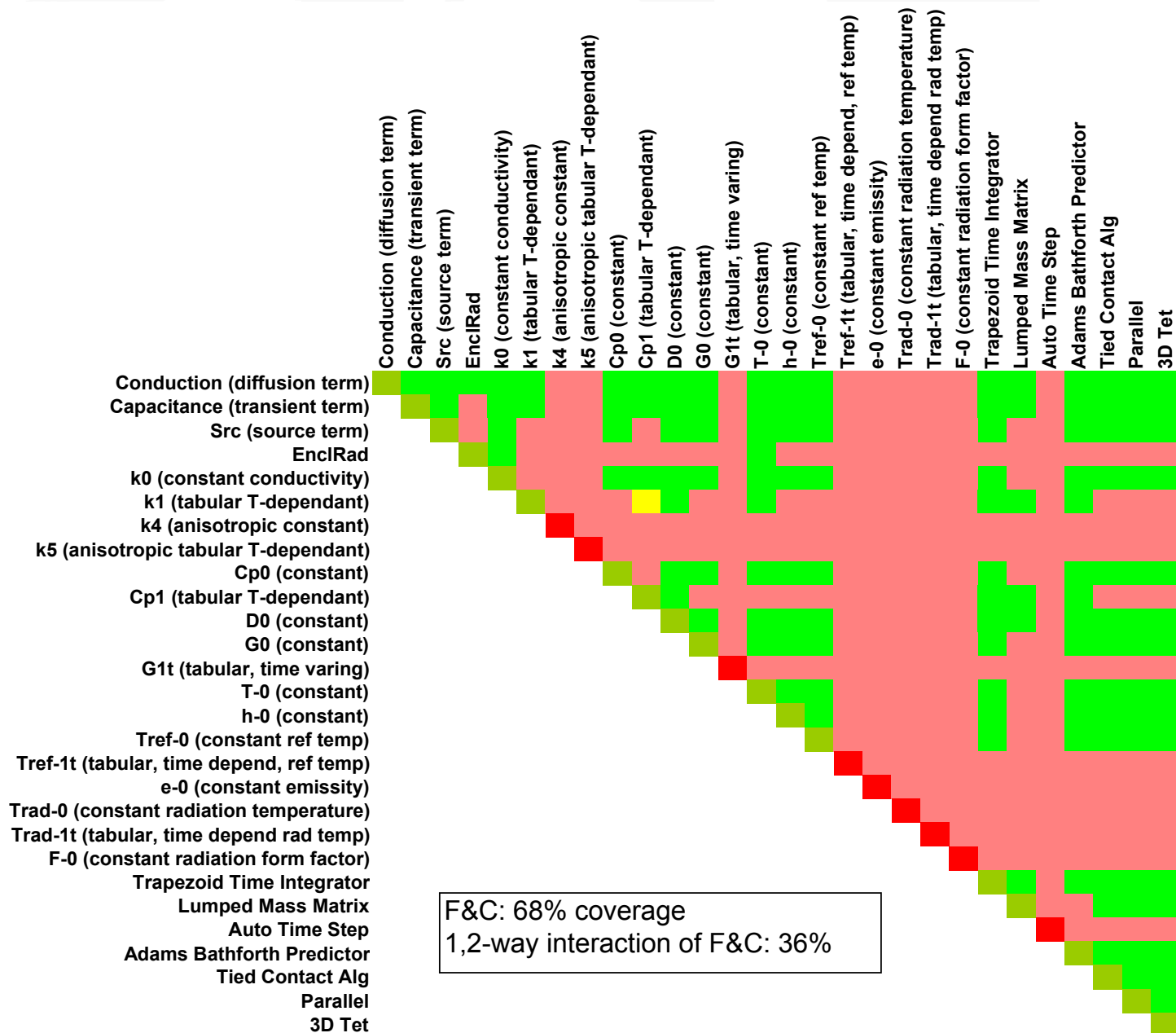


Assume :  $u(\bar{x})$

Analytically Solve for  $S = G(u)$

Numerically solve  $G(u) = S$  for different discretizations

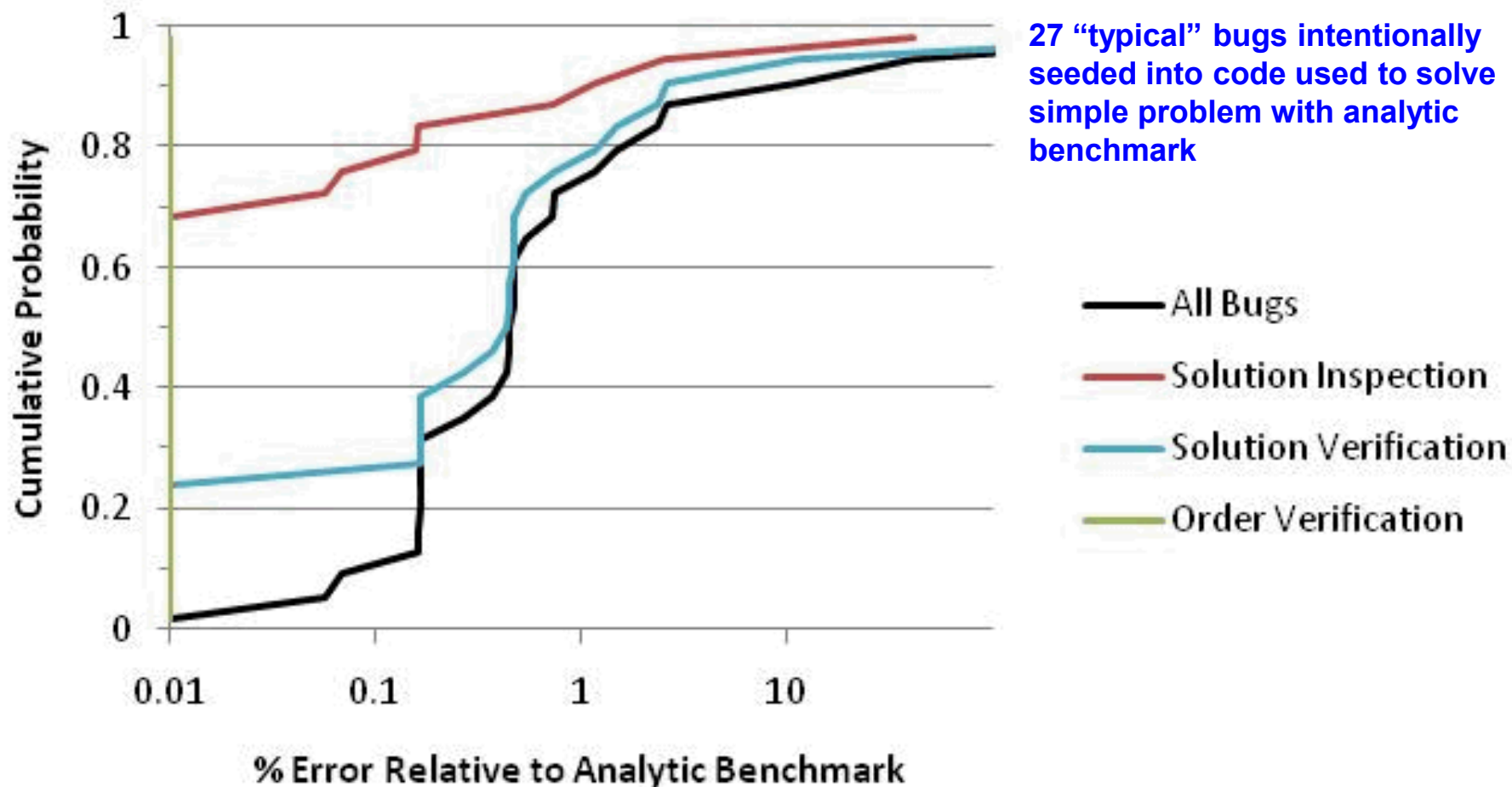
# A Better Way to Measure Test Coverage is to Focus on *Features/Capabilities and Their Interactions* for your *Specific Application*



F&C: 68% coverage  
 1,2-way interaction of F&C: 36%

# What is the Impact of Undetected Code or Algorithm Deficiencies on Simulation Results?

Knupp et al., *Impact of Coding Mistakes on Numerical Error and Uncertainty in Solutions to PDE's*, Sand2007-5341



# Code and Solution Verification

Demonstrates **Convergence** to the **Correct Answer** for the **Intended Application**

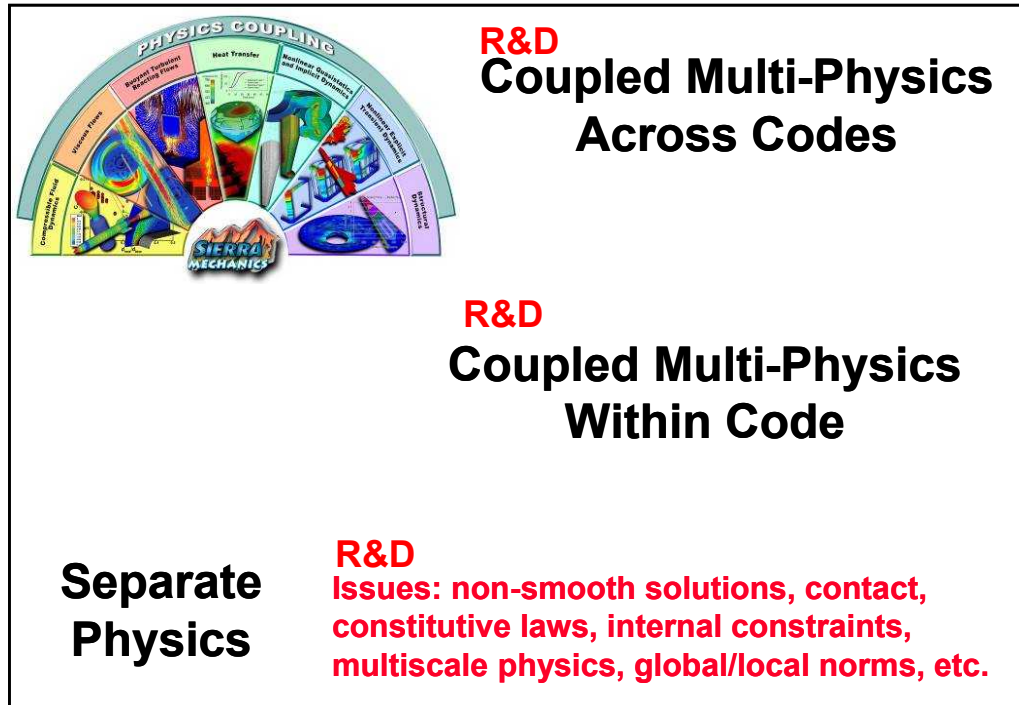
**Solution Verification: Convergence for intended application, but is it the right answer?**

**• Address adequacy of spatial AND temporal AND other discretizations AND numerical knobs**

Inference



Application



Inference

Regression Testing

**Code Verification: Convergence to correct answer, wrong application**

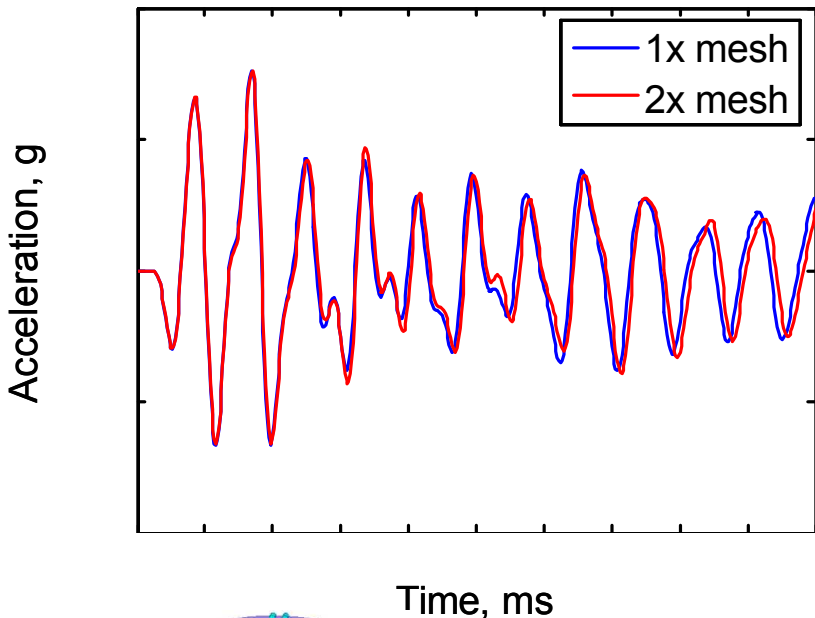
**• Eliminate code bugs AND inadequate algorithms**

SQE(A)

**R&D: Is it possible to demonstrate convergence to the correct answer for the intended application without inference?**

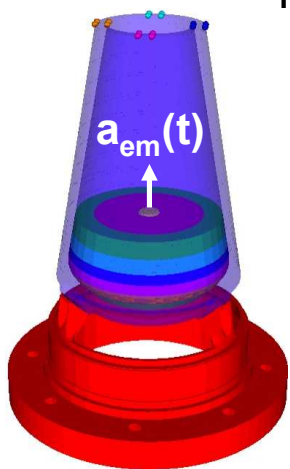
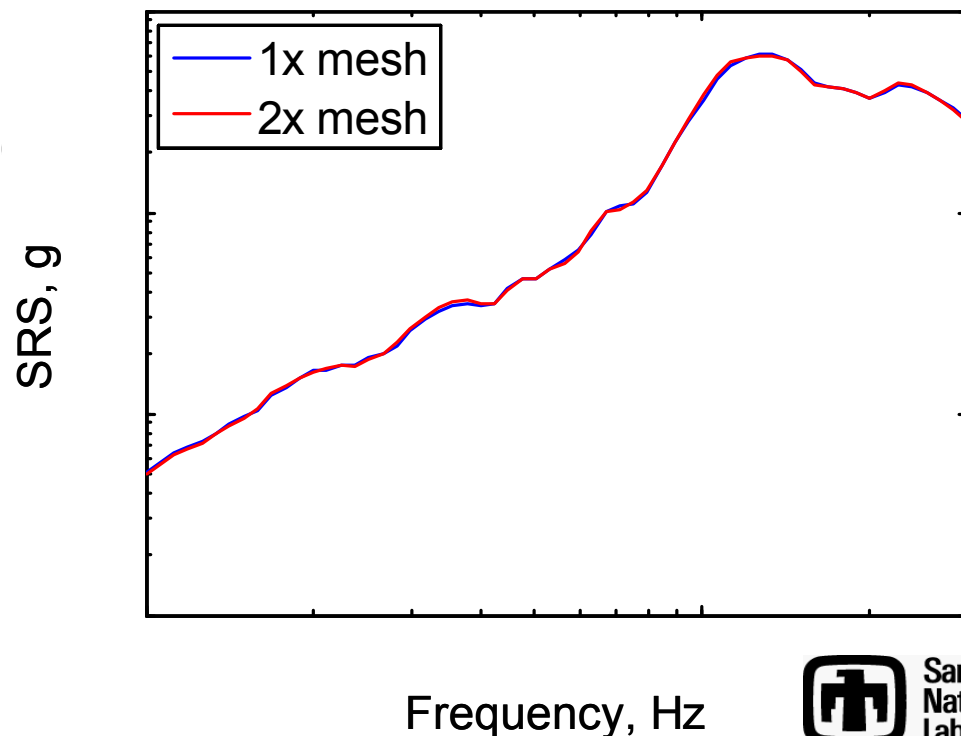
# It's Common to Explore Sensitivity to Mesh Parameters

Acceleration response at top of enc. mass

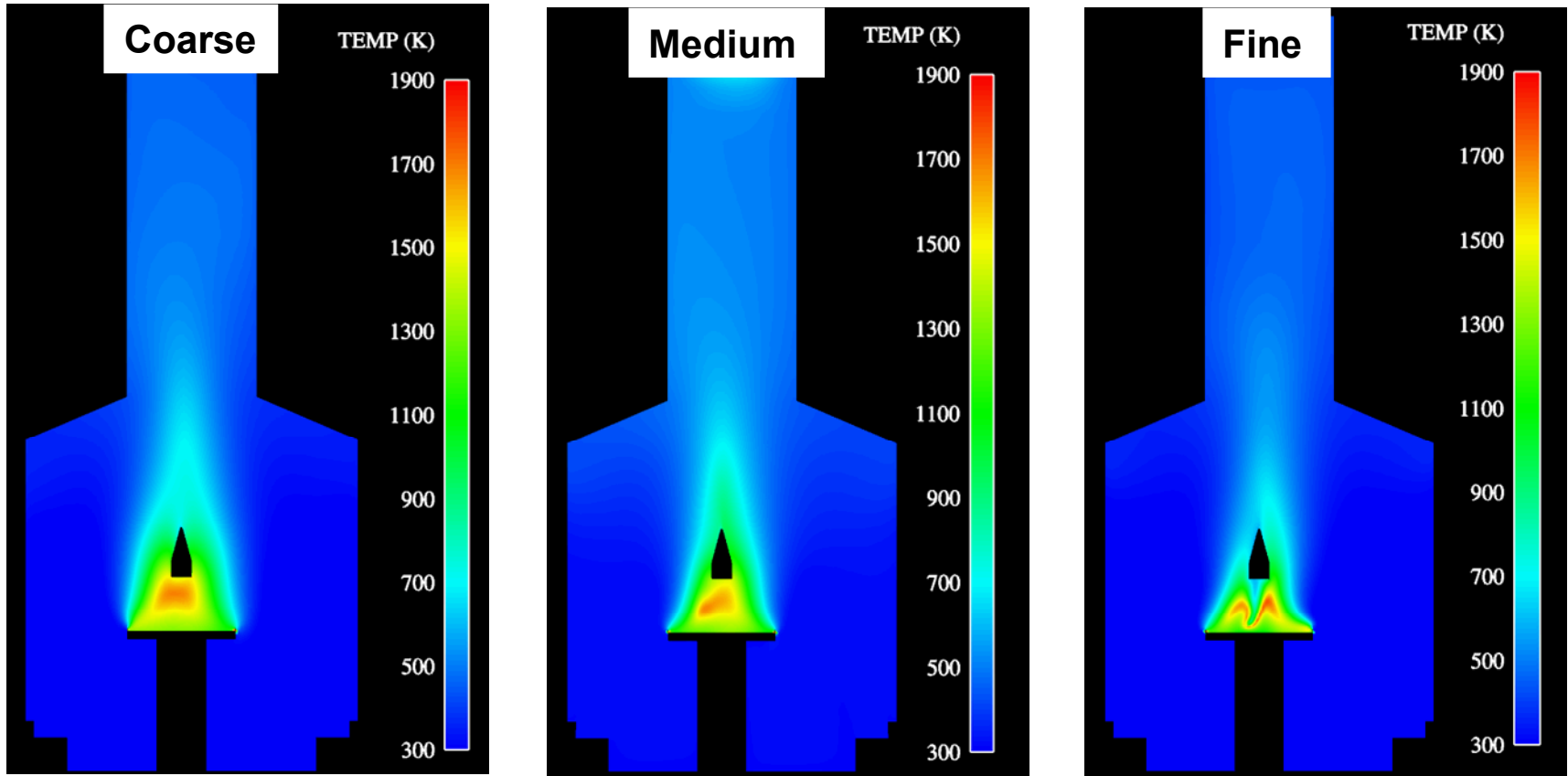


Max. relative error between SRS: +/- 5%

Shock response spectra at top of enc. mass

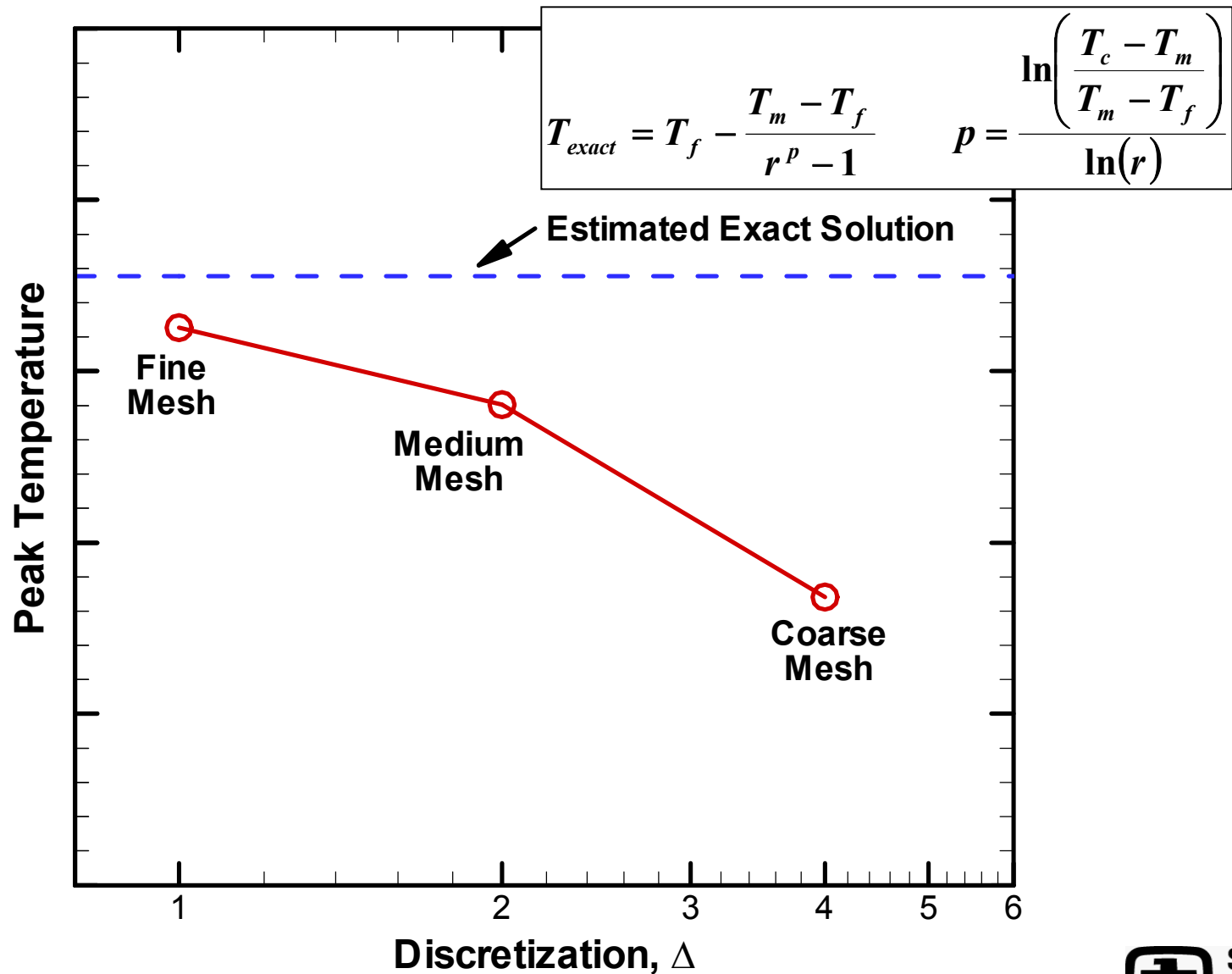


# You *Don't* Want to do a Discretization Study When Constitutive Models are Mesh Dependent and Calibrated to a Specific Mesh



Discretization leads to unphysical bifurcation of solution

# Numerical Errors Can Pollute Validation Assessments



# Solutions Don't Always Converge

Ryan Maupin, ESA-WR, LANL: IMAC-XXIV 1/31/06

Threaded assembly

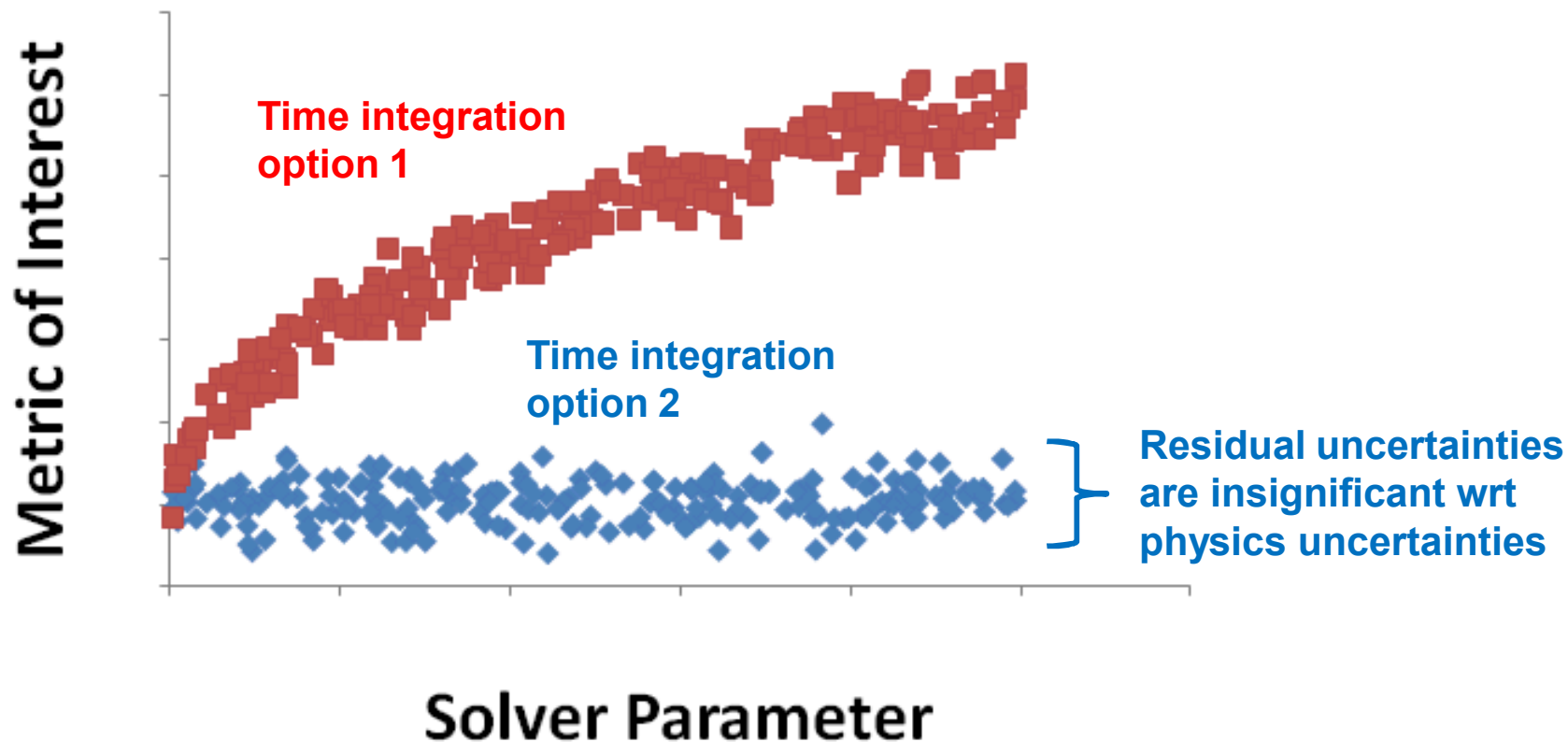
Errors = +/-20%

Peak Strain

Increasing Mesh Length →

R&D: How do you quantify and propagate errors associated with non-convergent solutions in a system analysis?

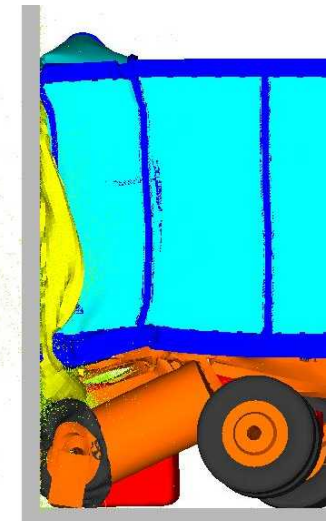
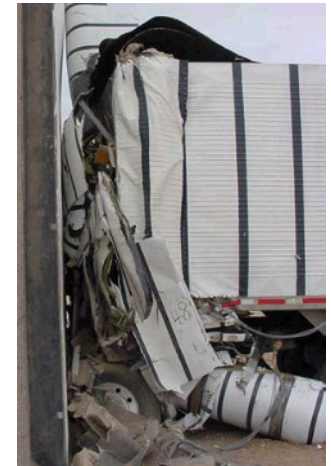
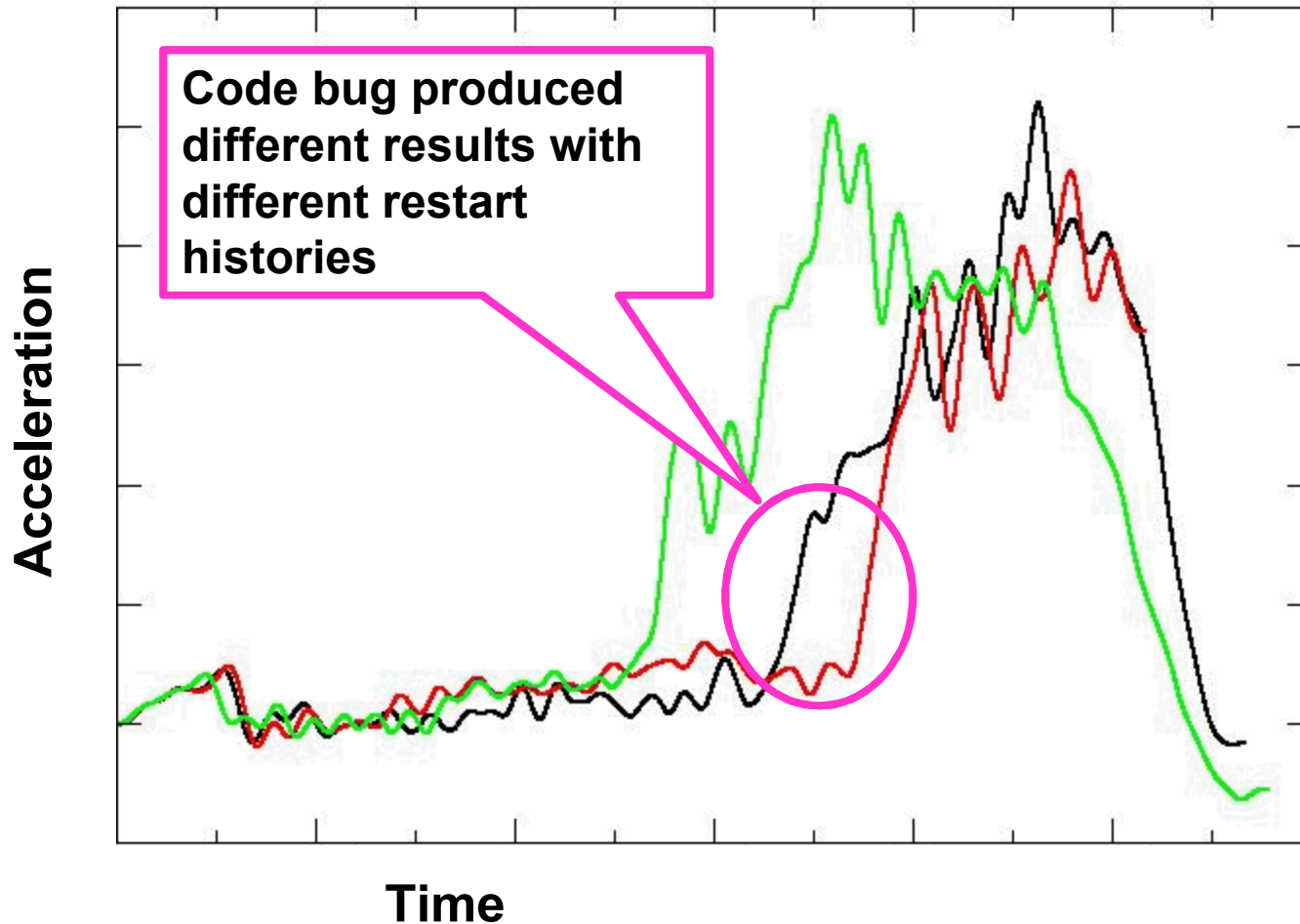
# Solution Verification Must Address Solver Settings as Well as Discretization Parameters



- 23 independent solver parameters explored with 230 LHS samples

- Sensitivity study identified 3 dominant solver parameters
- Guided appropriate selection of sensitive solver parameters

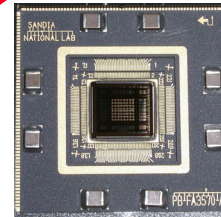
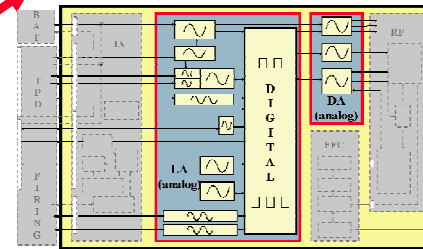
# Different Restart Histories, Processor Counts, or Platforms Sometimes Lead to Different Results



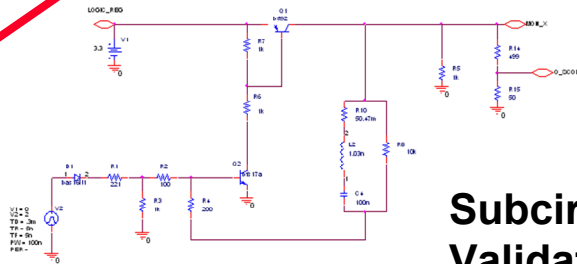
# A Hierarchy of Science-Based Validation Experiments Ensures Models Get the Right Answer for the Right Reasons

Hierarchical Validation: Right answer for the right reason

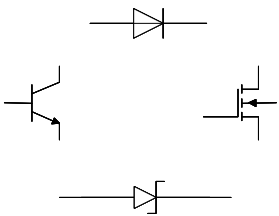
**Increasing complexity,  
Decreasing number of tests**



Single ASIC Validation



Subcircuit Validation



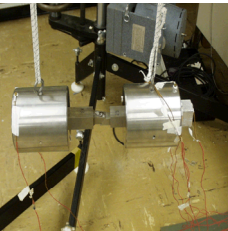
Single Device Characterization and Validation

- Application relevant parameter space
- Formal DOE and replicate tests
- Attention to diagnostic bias and precision

# A Hierarchy of Science-Based Validation Experiments Ensures Models Get the Right Answer for the Right Reasons

**Hierarchical Validation: Right answer for the right reason**

**Increasing complexity,  
Decreasing number of tests**



**Joint parameter characterization**



**Single joint validation**



**Jointed structure validation**



**Full System Test**

**Mockup with jointed structure and foam embedded object**

- **Application relevant parameter space**
- **Formal DOE and replicate tests**
- **Attention to diagnostic bias and precision**

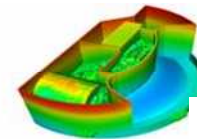
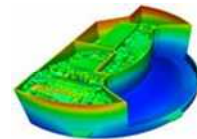
# A Hierarchy of Science-Based Validation

## Experiments Ensures Models Get the Right Answer for the Right Reasons

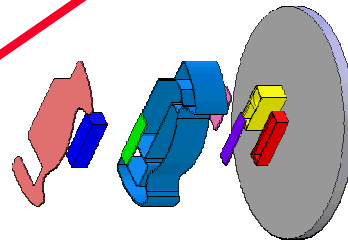
Hierarchical Validation: Right answer for the right reason



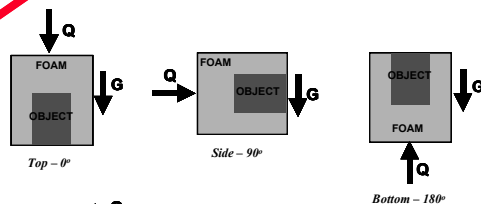
Full System Test



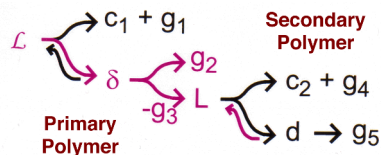
Validation Real Sub-systems



Validation with mockups



Foam recession



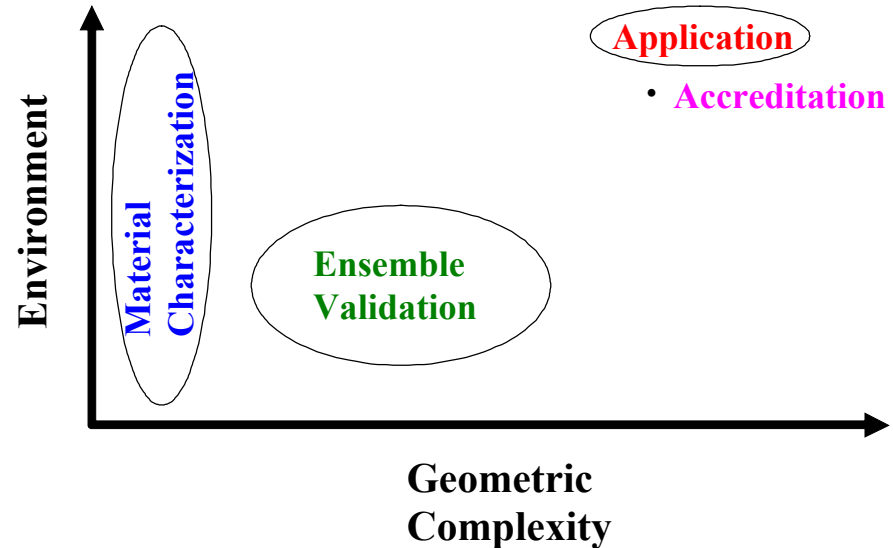
Chemistry characterization/validation

- Application relevant parameter space
- Formal DOE and replicate tests
- Attention to diagnostic bias and precision

# Validation Challenge Workshop Tied it all Together

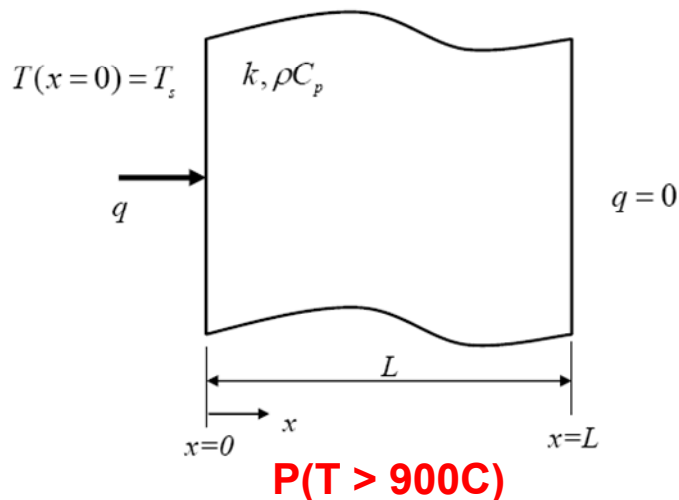
Computer Methods in Applied Mechanics and Engineering, Vol 197, Issues 29-32, May 2008

- **Regulatory Assessment (Application)**
  - Validation is best judged in the application context, which often involves a rigorous assessment against design or regulatory requirements
- **Accreditation**
  - Subsystem or full-system testing with application hardware under conditions that more closely represent the application of the model
- **Ensemble Validation**
  - Separate physics or low order interactions of important physics in stylized or de-featured geometries often for environments that are not fully representative of the application parameter space
- **Material Characterization**
  - Identification of material properties or constitutive-law parameters

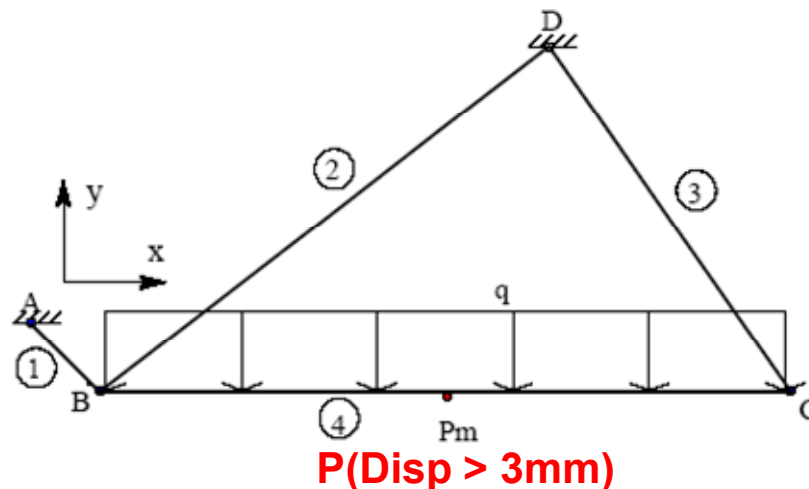


# “Regulatory Requirements” Were Evaluated for Thermal, Mechanical, and Dynamics Problems

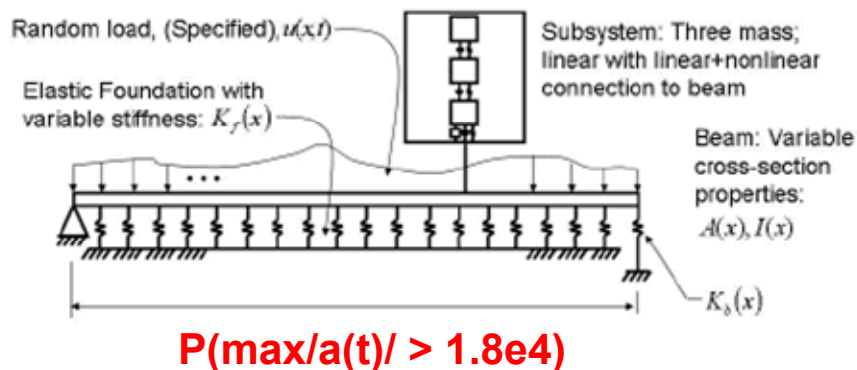
## Thermal



## Structural Static Frame

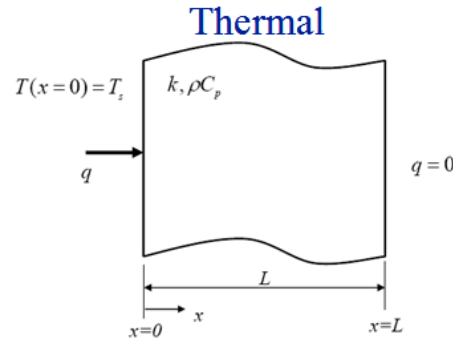
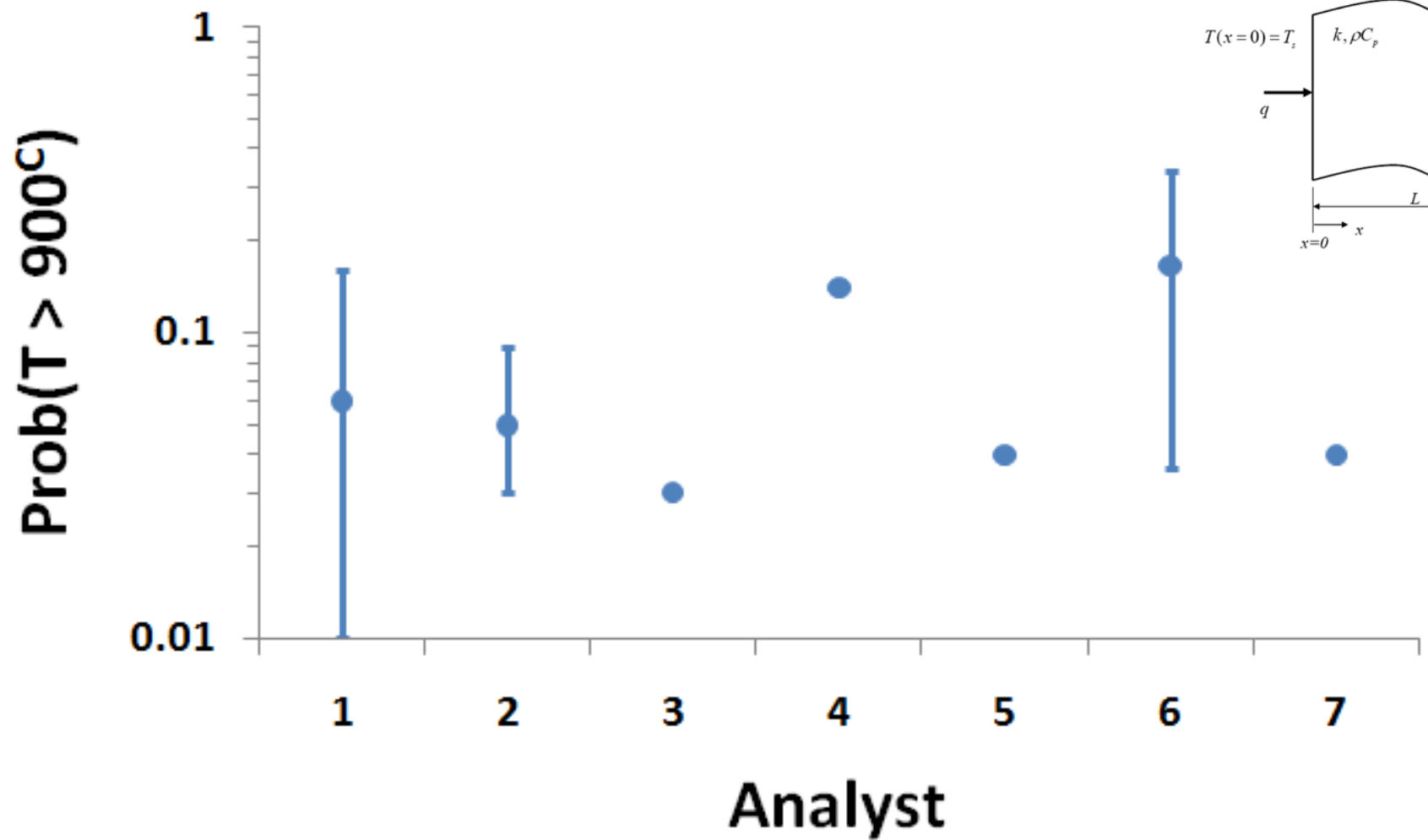


## Structural Dynamics

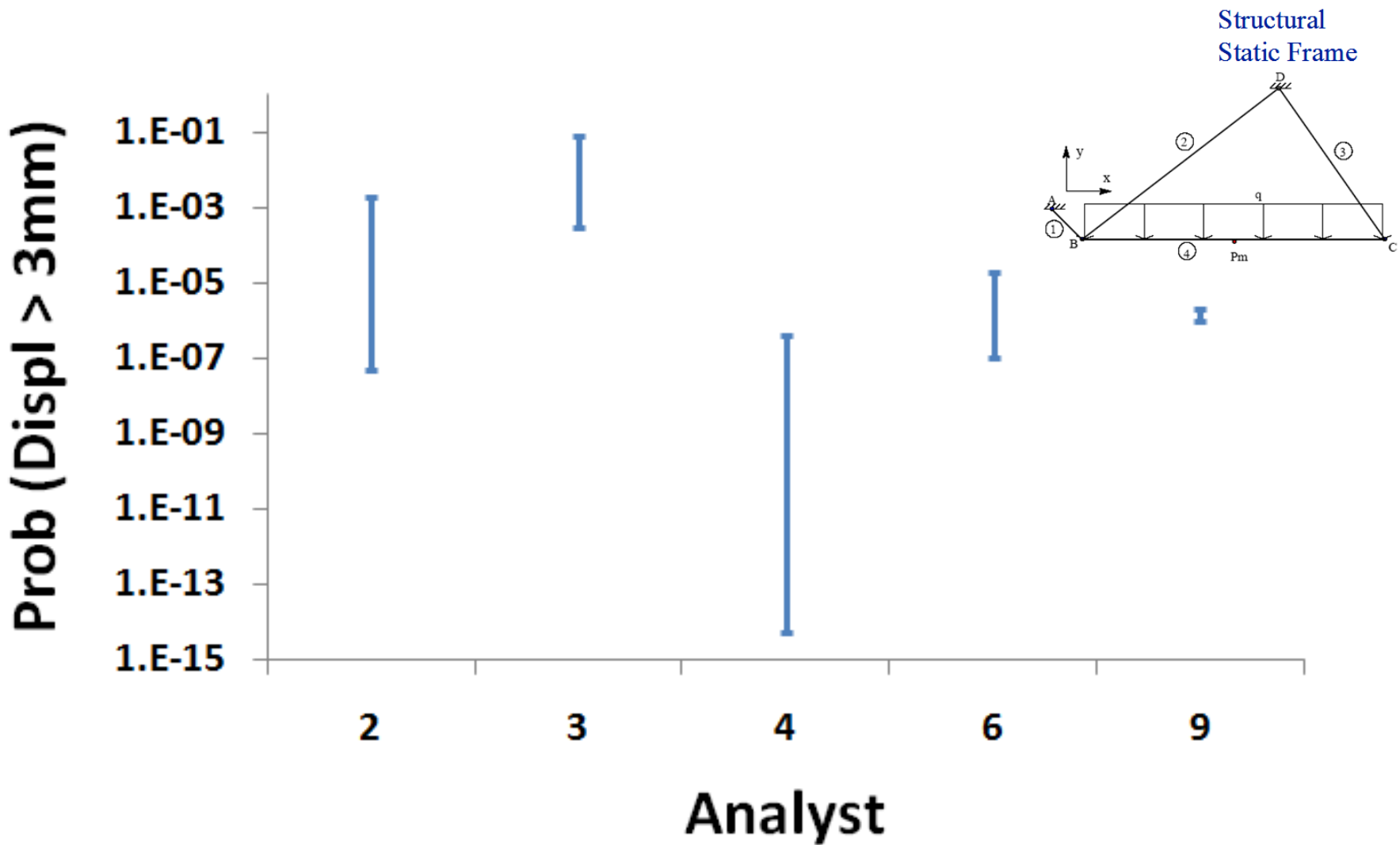


• Truth model never revealed

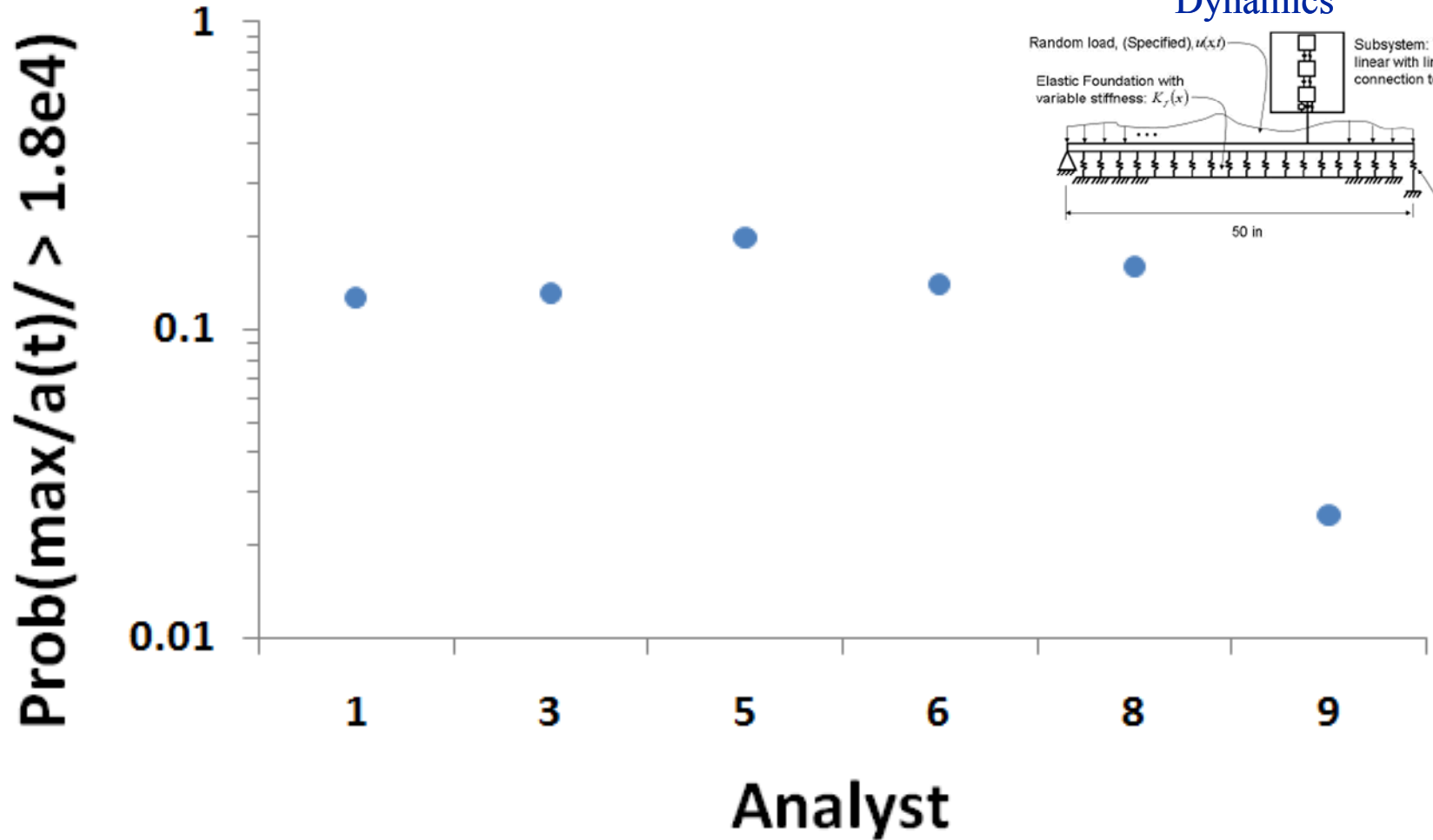
# Analyst/Methodology Differences are Comparable to Other Sources of Uncertainty



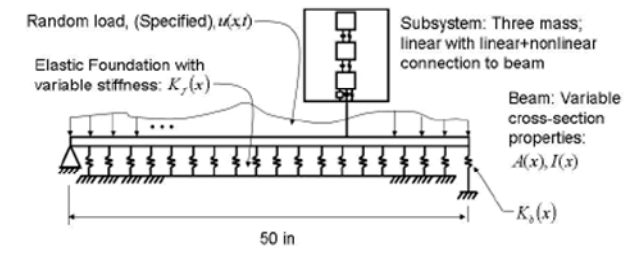
# Analyst/Methodology Differences are Comparable to All Other Sources of Uncertainty



# Analyst/Analyst Methodology/Methodology Differences Leads to a Spectrum of Results



## Structural Dynamics





# Uncertainty Has a Dual Personality

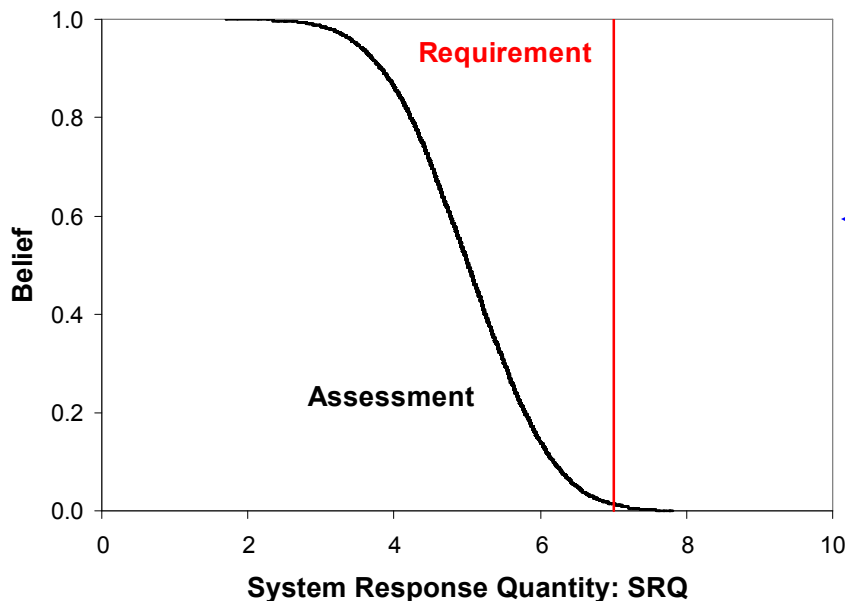
Jon C. Helton, *Conceptual and Computational Basis for the Quantification of Margins and Uncertainty*, SAND2009-3055

- **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**)

# You Have to Pick a Framework for Representing and Propagating Epistemic Uncertainties

$$\text{SRQ} = \sum_{i=1}^{10} X_i \quad X_i = [0,1]$$

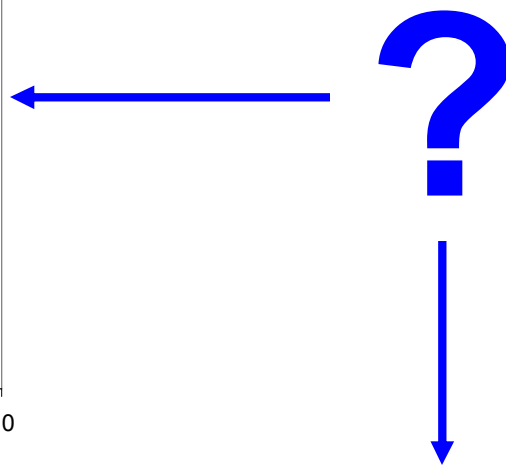
Requirement :  $\text{SRQ} \leq 7$



$$\text{Belief}(\text{SRQ} > 7) = 0.014$$

**Probabilistic framework**

- Bayesian methods
- SOA for risk assessment community

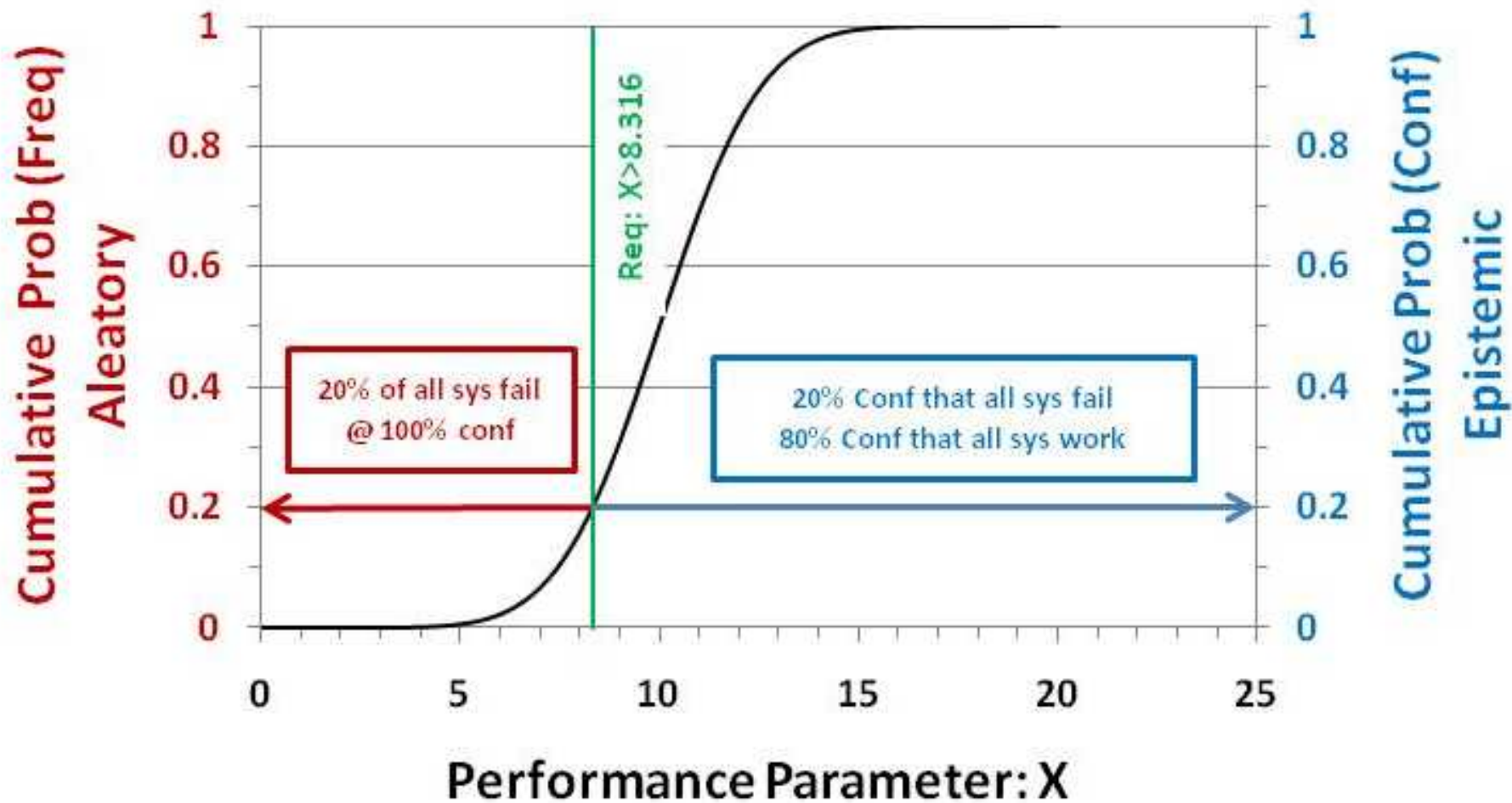


$$\text{SRQ} \equiv [0,10]$$

**Non-probabilistic framework**

- Interval analysis
- Dempster/Shافر (evidence) theory

# The Distinction Between Aleatory and Epistemic Uncertainties Matters



# Realistically, How Many Uncertainties do you Have?

Consider only parameter uncertainties to keep it simple

This is representative of something we recently did

Source	Realistically	What We Considered
IC/BC	4	0
Geometry 200 parts x 3 dimensions	600	0
Material Props k, $\rho$ , $C_p$ , $\epsilon$ for 50 materials	200	60
Constitutive Models	15	15
Failure Modes	17	17
<b>Total</b>	<b>836</b>	<b>92</b>

You'd like the analysis to tell you what's important, but we typically assume the dimension down to a manageable set

# Curse the Curse of Dimensionality!

Two Level Factorial Design:  $2^{100} = O(10^{30})$

Different requirements leads to different approaches

## R&D: efficient/accurate probability estimation

- Isolated local artifacts may be of secondary importance because probability is an integrated metric
- Most approaches exploit assumptions of smoothness and structure
  - LHS
  - Eliminate or under-resolve secondary dimensions
  - Reduce dimensions through SA and screening
  - Adaptive UQ methods

## R&D: find the “needle in the haystack”

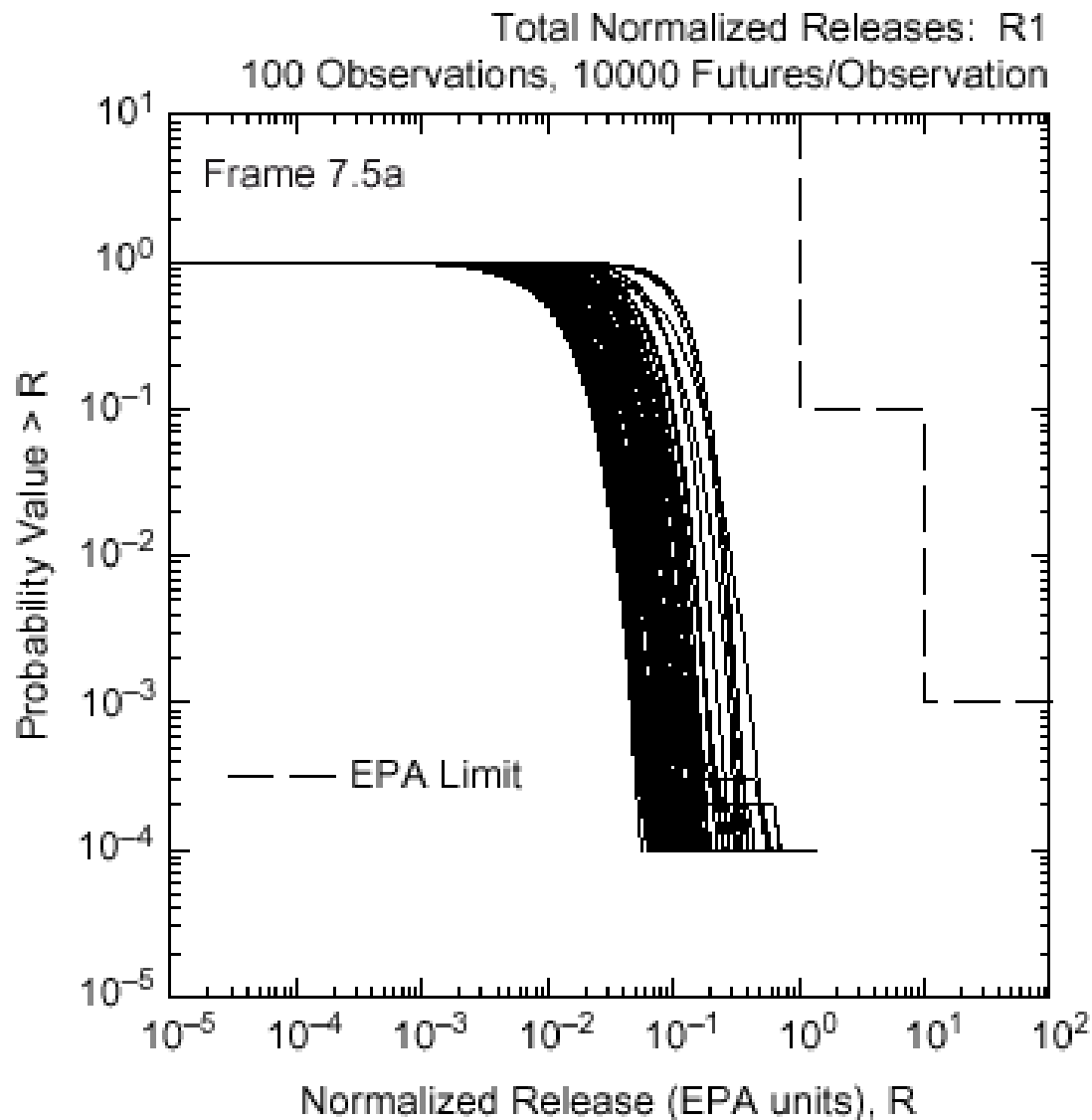
- Identification and resolution of isolated local occurrences without assumed existence of global trends
- Most approaches make a few simplifying assumptions about smoothness
  - Directed search for outliers i.e., optimization
  - Trade-off between algorithmic efficiency and assumed smoothness

R&D: How do you define, measure, and communicate meaningful “coverage” metrics based on your function evaluation budget?

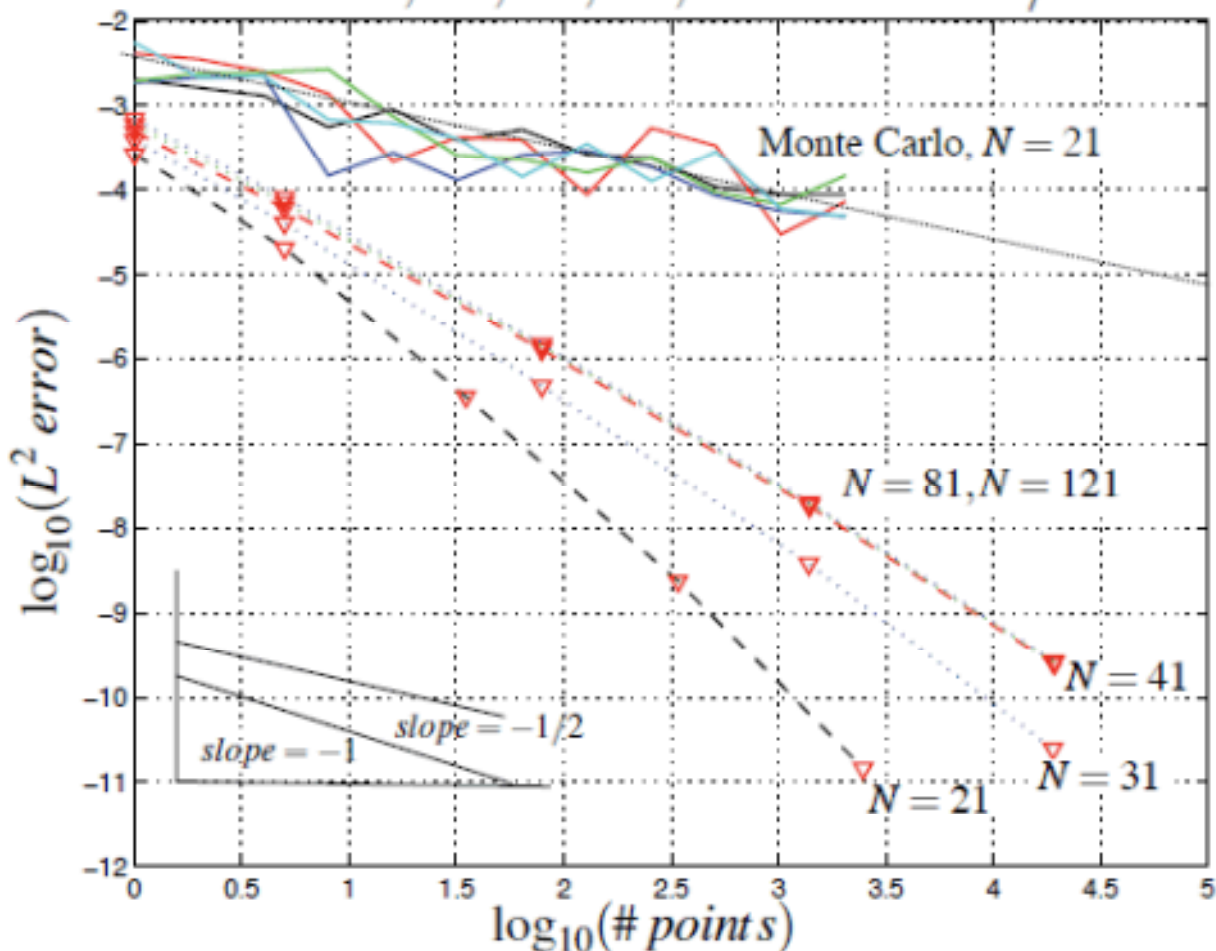


# The "Curse" Has a Big Brother

2<sup>nd</sup> order probability is the SOA for high profile risk assessments

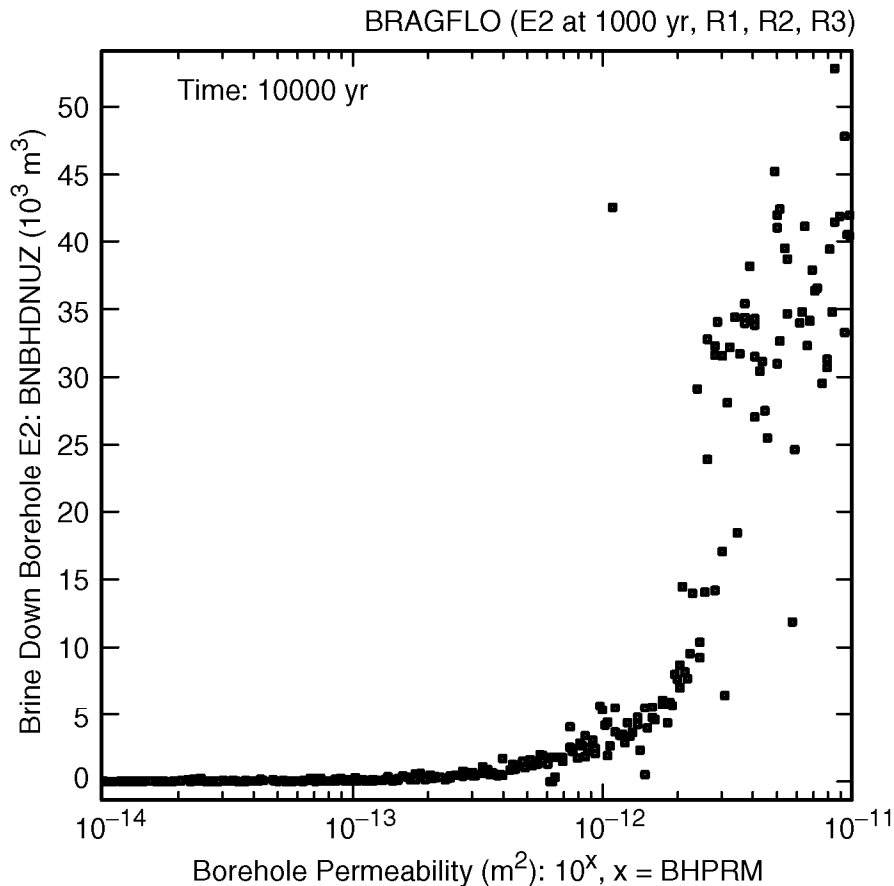


# R&D: Efficient UQ Methods Has to Be a Research Priority

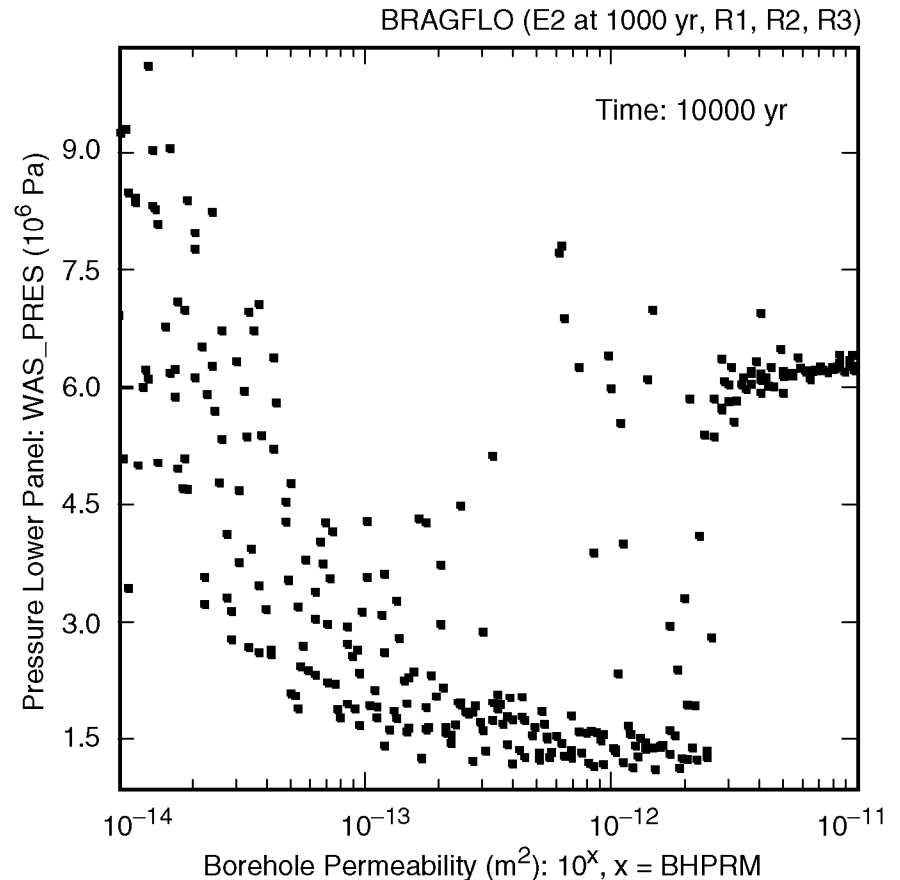


R&D: What is the error in my statistic of interest (mean, 99 percentile, failure probability, etc) when you have a fixed function evaluation budget and don't have an analytic benchmark?

# Assumptions of Linearity Could Miss-Characterize Important Features



TRI-6342-5371-0



TRI-6342-5372-0

**Sensitivity study form WIPP performance assessment**

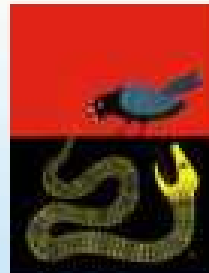
# FAQ: What About the Unknown/Unknowns?



Be Proactive: Begin With Unkn/Unkns in Mind



Analysis Paralysis: Extreme Risk Aversion



Unknown/Unknowns = Surprise + High Consequence



Ignore It: What I Don't Know Can't Hurt Me



Hubris: Nothing Can Go Wrong

# We Have to Proactively Manage Risk With Unknown/Unknowns in Mind

