SAND2006-5388C

Measuring and Communicating Progress in Predictive Capability

Martin Pilch PhD PMP

SNL/ASC/V&V Program Element Manager

Validation and Uncertainty Quantification, 1533

mpilch@sandia.gov

Presented at:

Computational Methods in Transport 2006

Tahoe City, CA

September 9-14, 2006

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.

Increasing Emphasis on Modeling and Simulation

- Roles of modeling and simulation
  - High consequence decisions
  - High consequence design
- Goals of modeling and simulation
  - *(Credible)* science-based predictive capability rather than extrapolations based on calibration and expert judgment
  - Calculating, measuring, and understanding the uncertainty in predictions

How do you measure and communicate progress in predictive capability?

Do We *Really* Want to Reveal What's Under the Hood of Our Models and Codes?

What Does it Mean "to Predict"?

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 of M&S to assert a predictive capability?

***A CS&E prediction is a M&S-based evaluation prior to or in lieu of physical measurement**

Some Attributes of Predictive Capability

You can't measure and communicate "it" unless you know what "it" is

- Representational (geometric) fidelity
- Physics and material model fidelity (predictive science)
- Code readiness for stockpile computing (SQE, code verification)
- Evidence that numerical errors are not polluting decisions i.e., solution verification
- Validated models
- Quantified margins and uncertainties with sensitivity analysis

How Much is Enough?

- **Sufficiency (or Adequacy) should be discussed in conjunction with measures of progress and this can only be discussed in an application context**
- **Graded approach based on risk tolerance can help mold customer expectations:**
  - **High risk tolerance (e.g., scoping studies)**
  - **Risk tolerance (e.g., design support)**
  - **Risk aversion (e.g., M&S-informed decisions)**
  - **High risk aversion (e.g., M&S-based decisions)**
- **Alternatively, communicate risk incurred for a given level of rigor**
  - **You get what you pay for**

Increasing Rigor Expected

This is Where We Are Going
Predictive Capability Maturity Model (PCMM)

PREDICTIVIE ATTRIBUTE	High Risk Tolerance (e.g., Scoping Studies)	Risk Tolerance (e.g., Design Support)	Risk Aversion (e.g., Qual. Support)	High Risk Aversion (e.g., Qualification)
Representation (Geometry) Fidelity	• Grossly defeatured or stylized representation based on practical considerations	• Significant defeaturing or stylization based on judgment • 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	• Unknown model form • Empirical model form speculated or calibrated to represent trends applied w significant or unknown extrapolation	• Empirical model form speculated or calibrated to represent trends applied w/o extrapolation • Physics informed models applied w significant or unknown extrapolation	• Physics informed models applied w/o significant extrapolation • Physics-based model applied w significant or unknown extrapolation	• Well accepted physics-based model applied w/o significant extrapolation
Code Readiness	• 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 against SQE standards • Sustained verification test suite w significant coverage of required F&Cs	• Code managed and assessed against SQE standards • Sustained verification test suite w significant coverage of required F&Cs and their interactions
Solution Verification	• Judgment only	• Sensitivity to discretization and algorithm parameters explored	• Numerical errors estimated	• Rigorous numerical error bounds quantified
Validation	• Judgment only • Qualitative validation w/o SET coverage or w/o IETs	• Qualitative validation with significant SET coverage and IETs	• Quantitative validation w/o assessment of var/unc in diagnostics and IC/BC *and* w/o significant SET coverage or w/o IETs	• Quantitative validation w assessment of var/unc in diagnostics and IC/BC w significant SET coverage and IETs
QMU and Sensitivities	• Judgment only	• Deterministic assessment of margins (bounding analyses) • Informal "what if" assessment of var/unc, margins, and sens	• Formal quantification var/unc, margins, and sens w/o confidence assessments	• Comprehensive quantification of var/unc, margins, and sens w explicit confidence assessments

Why PCMM?

- **Goals of the table**
  - **Measure/communicate maturity of evidence (not adequacy of results) associated with M&S in a decision context**
  - **Provide program vision so that technical and infrastructure needs can be leveraged across multiple funding lines to enhance the credibility of M&S results**
  - **Speak to the *whats*, not dictate the *hows***
- **Target audience**
  - **Decision makers and analysts who rely on CS&E**
    - **Focus on codes that solve PDEs**
  - **Program managers and academics who can make credible M&S a reality**

Measuring Progress in Representational Fidelity

Are you overlooking important effects because of judgment-based Defeaturing or Stylizations?

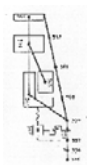
Grossly defeatured or stylized

Significant D&S based on judgment or justified lower fidelity representation

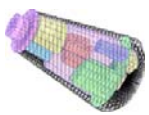
Limited D&S judged to retain the essential elements of "as built" or justified lower fidelity representation

Highest fidelity representation "as built" w/o significant D&S or justified lower fidelity representation

H Risk Tolerant | **Risk Tolerant** | **Risk Adverse** | **H Risk Adverse**



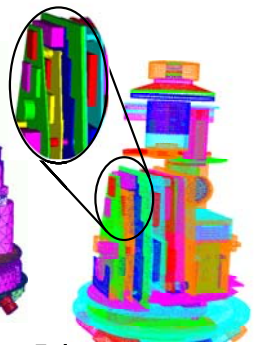
10 years ago: Shellshock 2D NASTRAN 200 dof



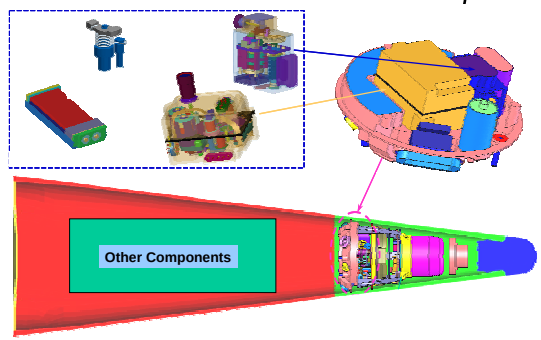
Recent Past: NASTRAN

30,000 dof

800,000 dof



Today: SALINAS MP >10M dof



Measuring Progress in Physics Fidelity

What physics is important for the application and how predictive are the models?

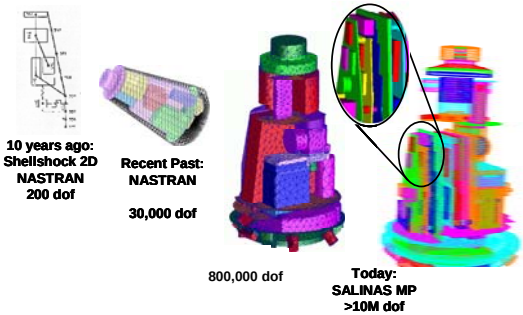
Unknown model form or empirical model form applied w sig extrap

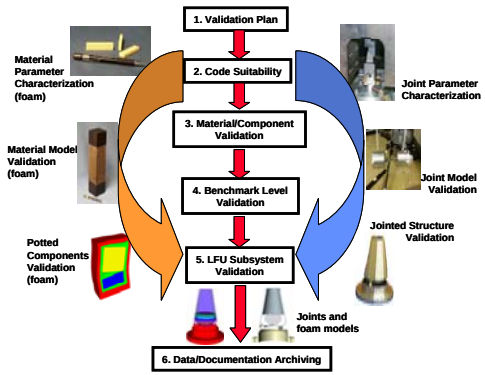
Empirical model form applied w/o sig extrap or physics informed model applied w sig/unk extrap

Physics-informed model applied w/o sig extrap or physics based model applied w sig/unk extrap

Physics-based model applied w/o sig extrap

H Risk Tolerant → **Risk Tolerant** → **Risk Adverse** → **H Risk Adverse**





Measuring Code Readiness From An Application Perspective

Are you solving the equations right?

Judgment only

Code managed to SQE standards

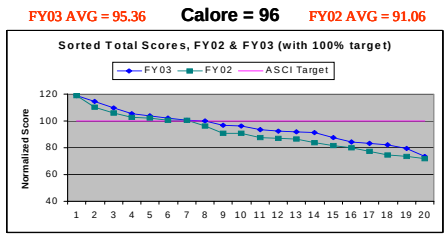
Sustained unit/regression tests w sig coverage of F&C

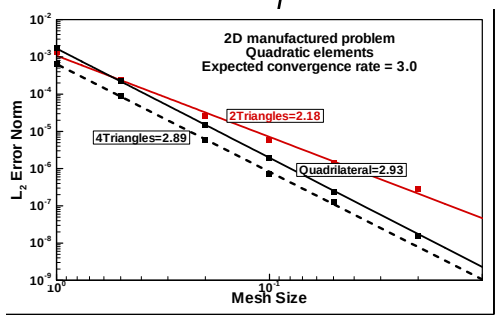
SQE +assessment + sustained VERTS w sig coverage of F&C

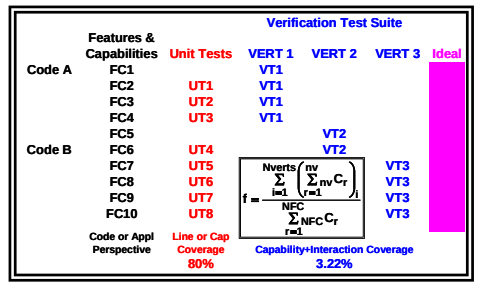
SQE(A) + VERTS w sig coverage of F&C interactions

H Risk Tolerant → **Risk Tolerant** → **Risk Adverse** → **H Risk Adverse**

Code/Code Comparisons







Measuring Progress in Solution Verification

Are numerical errors polluting decisions?

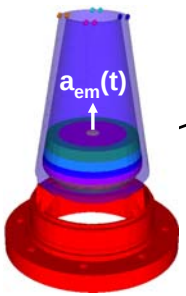
Judgment only

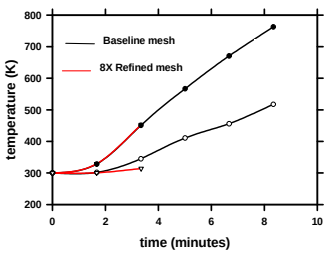
Explore sensitivity to discretization and algorithm parameters

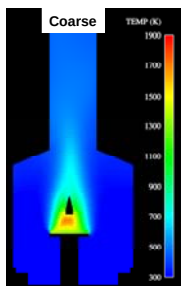
Estimate numerical errors

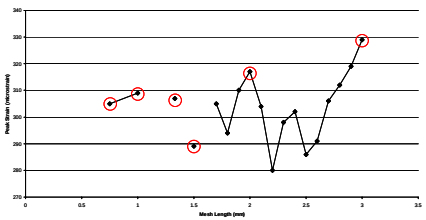
Quantify rigorous numerical error bounds

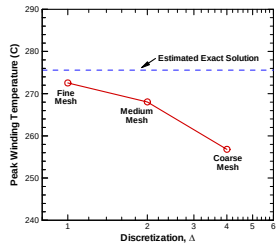
H Risk Tolerant | **Risk Tolerant** | **Risk Adverse** | **H Risk Adverse**

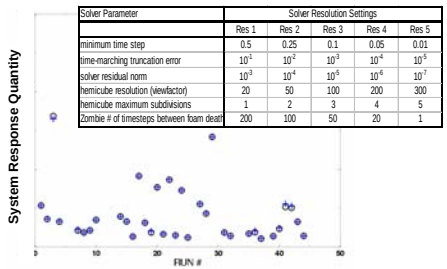


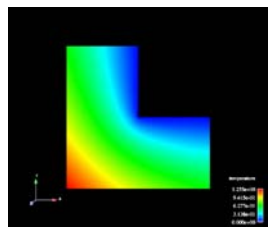












Measuring Progress in Validation

Are you solving the right equations?

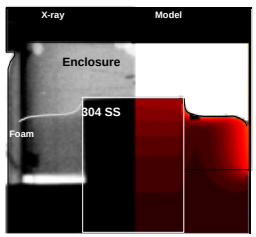
Judgment only or qual m/p comp w/o SET coverage or w/o IETs

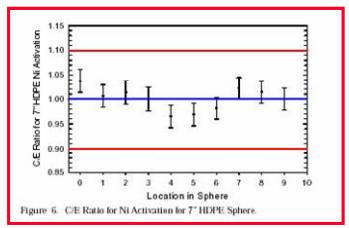
Qual m/p comps w SET coverage and IETs

Quantitative validation w/o assessment of var/unc *and* w/o SET coverage or w/o IETs

Quantitative validation w assessment of var/unc in diagnostics & IC/BC *and* SET coverage and IETs

H Risk Tolerant → **Risk Tolerant** → **Risk Adverse** → **H Risk Adverse**

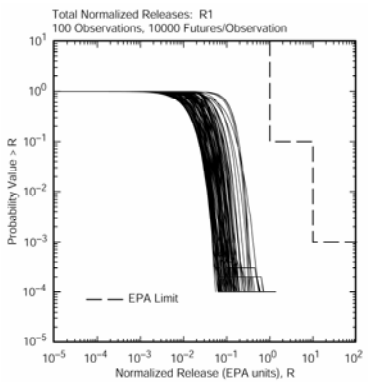


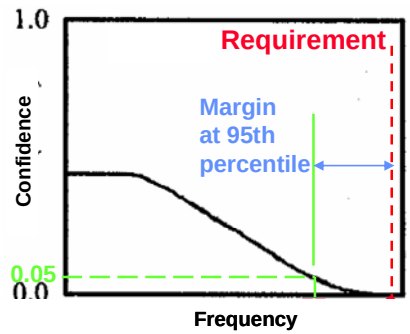


Measuring Progress in UQ/Sensitivity Analyses

What is the impact of variabilities and uncertainties in the decision context?

Judgment only	Deterministic margins, informal "what if" assessment of var/unc and sens	Formal quantification of, var/unc, margins, and sens w/o conf assessment	Formal quantification of margins, var/unc, and sens w conf assessments
H Risk Tolerant	Risk Tolerant	Risk Adverse	H Risk Adverse





Predictive Capability Maturity Model (PCMM)

PREDICTIVIE ATTRIBUTE	High Risk Tolerance (e.g., Scoping Studies)	Risk Tolerance (e.g., Design Support)	Risk Aversion (e.g., Qual. Support)	High Risk Aversion (e.g., Qualification)
Representation (Geometry) Fidelity	• Grossly defeatured or stylized representation based on practical considerations	• Significant defeaturing or stylization based on judgment • 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	• Unknown model form • Empirical model form speculated or calibrated to represent trends applied w significant or unknown extrapolation	• Empirical model form speculated or calibrated to represent trends applied w/o extrapolation • Physics informed models applied w significant or unknown extrapolation	• Physics informed models applied w/o significant extrapolation • Physics-based model applied w significant or unknown extrapolation	• Well accepted physics-based model applied w/o significant extrapolation
Code Readiness	• 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 against SQE standards • Sustained verification test suite w significant coverage of required F&Cs	• Code managed and assessed against SQE standards • Sustained verification test suite w significant coverage of required F&Cs and their interactions
Solution Verification	• Judgment only	• Sensitivity to discretization and algorithm parameters explored	• Numerical errors estimated	• Rigorous numerical error bounds quantified
Validation	• Judgment only • Qualitative validation w/o SET coverage or w/o IETs	• Qualitative validation with significant SET coverage and IETs	• Quantitative validation w/o assessment of var/unc in diagnostics and IC/BC *and* w/o significant SET coverage or w/o IETs	• Quantitative validation w assessment of var/unc in diagnostics and IC/BC w significant SET coverage and IETs
QMU and Sensitivities	• Judgment only	• Deterministic assessment of margins (bounding analyses) • Informal "what if" assessment of var/unc, margins, and sens	• Formal quantification var/unc, margins, and sens w/o confidence assessments	• Comprehensive quantification of var/unc, margins, and sens w explicit confidence assessments

Measured Credibility, on Demand, for Diverse Applications

- **Decision makers need to understand predictive capability in order to make informed decisions and to efficiently leverage and make use of research dollars**
- **Progress in predictive capability needs to be measured in each individual decision context**
  - **Predictive capability is more than geometric fidelity or even physics fidelity**
  - **There is a need to define sufficiency (or adequacy) in each attribute of predicative capability based on risk tolerance**
- **The Predictive Capability Maturity Model provides a graded approach to assessing and measuring predictive capability for specific applications**

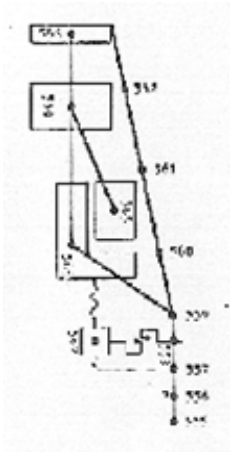
The Credibility of M&S is Critical

“Due diligence means asking the questions, even if you don’t think you’ll like the answers.”

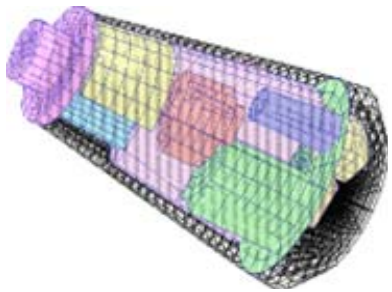
Representational (Geometric) Fidelity

Hyperlinks

Progress in Representational Fidelity in Structural Dynamics

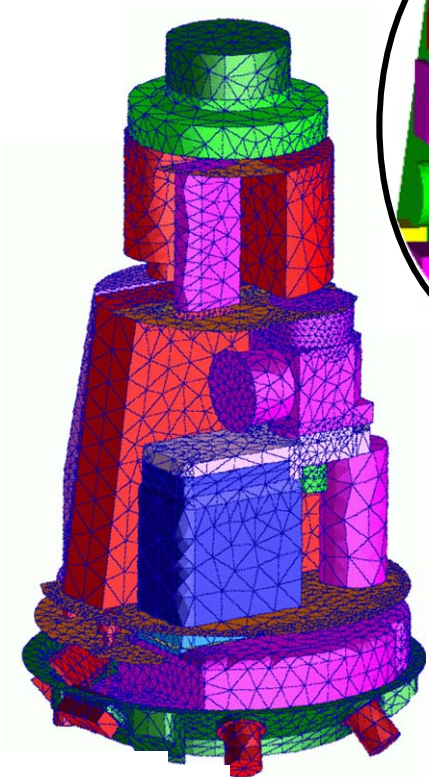


10 years ago:
Shellshock 2D
NASTRAN
200 dof

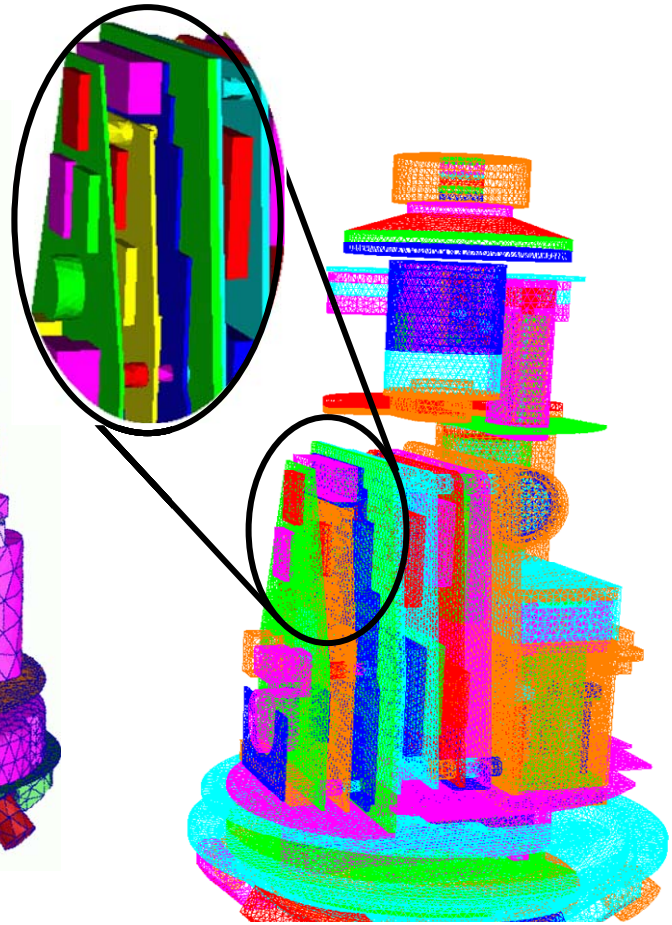


Recent Past:
NASTRAN

30,000 dof

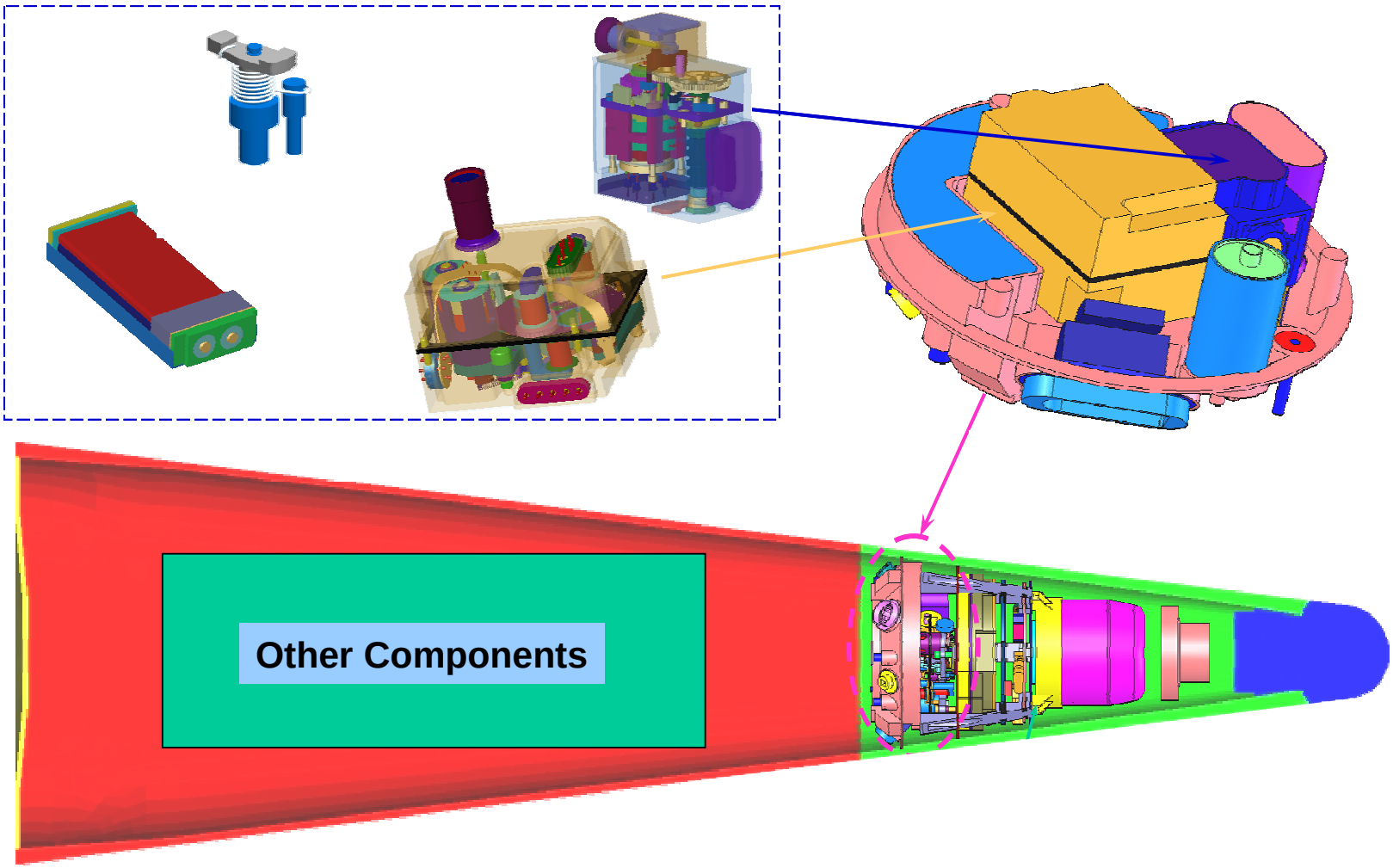


800,000 dof



Today:
SALINAS MP
>10M dof

Progress in Representational Fidelity Thermal Modeling



Physics Fidelity

Hyperlinks

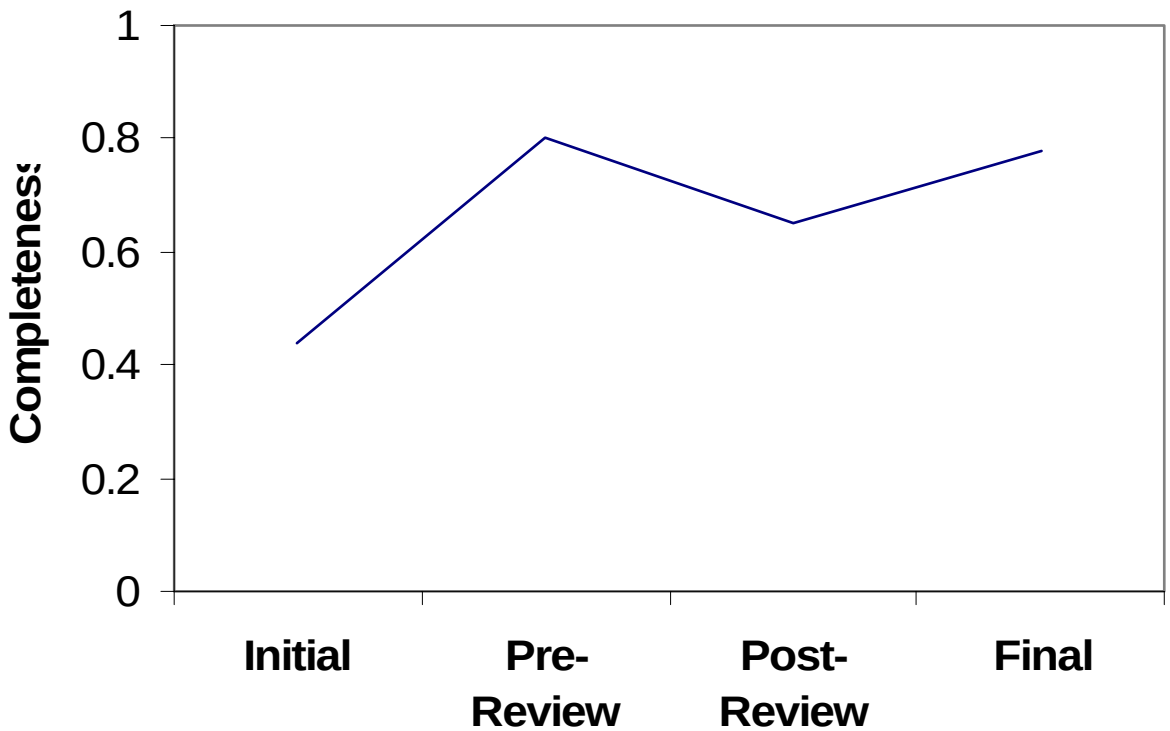
Phenomena Identification and Ranking Tables (PIRT)

Establish efficiency and sufficiency of activities

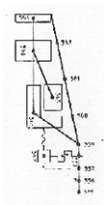
Phenomena	Importance	Adequacy: Model	Adequacy: Code	Adequacy: Validation
P1	H	H	M	L
P2	M	M	L	L
P3	L	L	L	L

Gap = 5

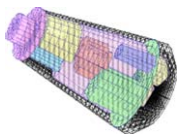
Completeness = 0.44



Low Physics Fidelity

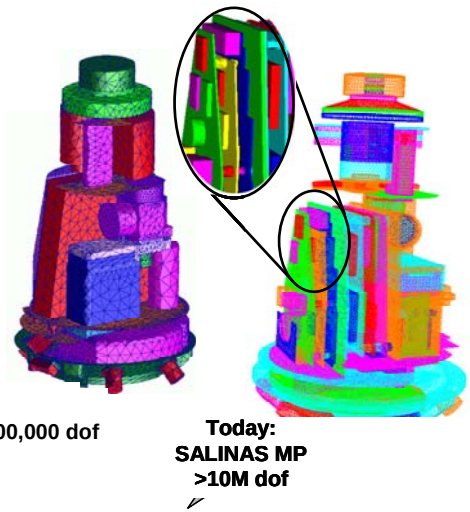


10 years ago:
Shellshock 2D
NASTRAN
200 dof



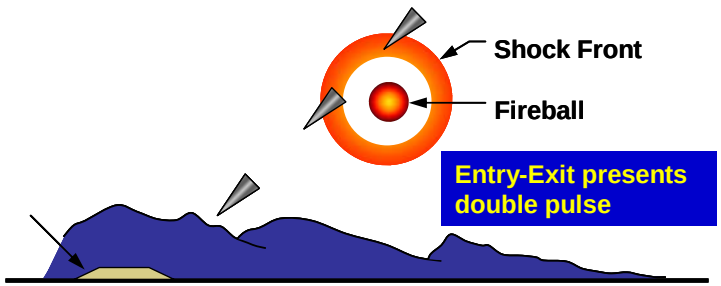
Recent Past:
NASTRAN

30,000 dof



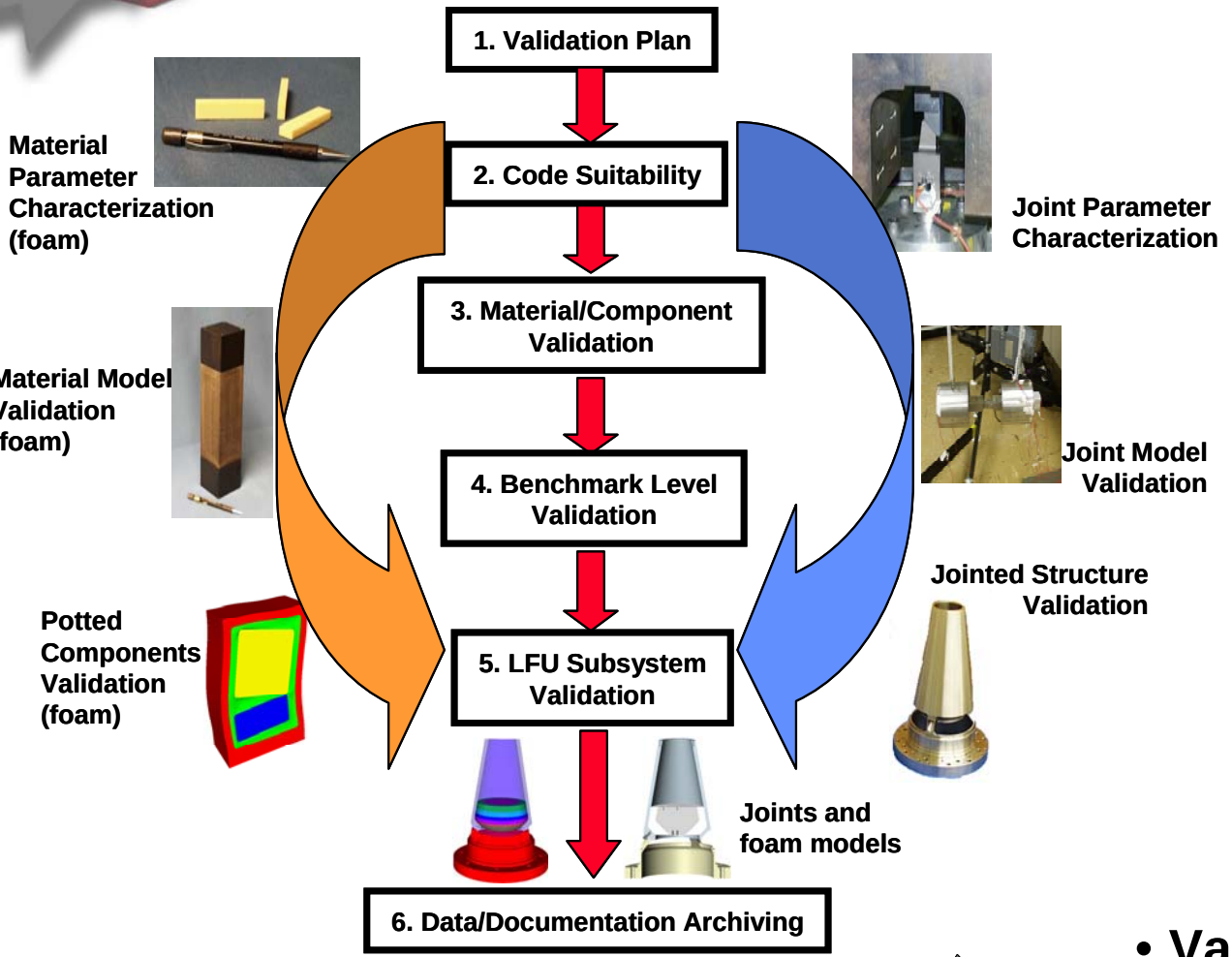
800,000 dof

Today:
SALINAS MP
>10M dof

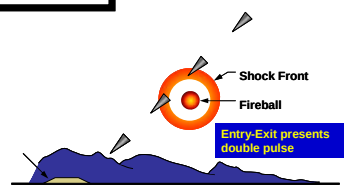


- Conduct blast test
- Calibrate model to blast test using global stiffness and damping parameters: knobs that act as surrogates for missing or unknown physics
- Use calibrated model to make prediction in tactical environments

Improving Physics Fidelity



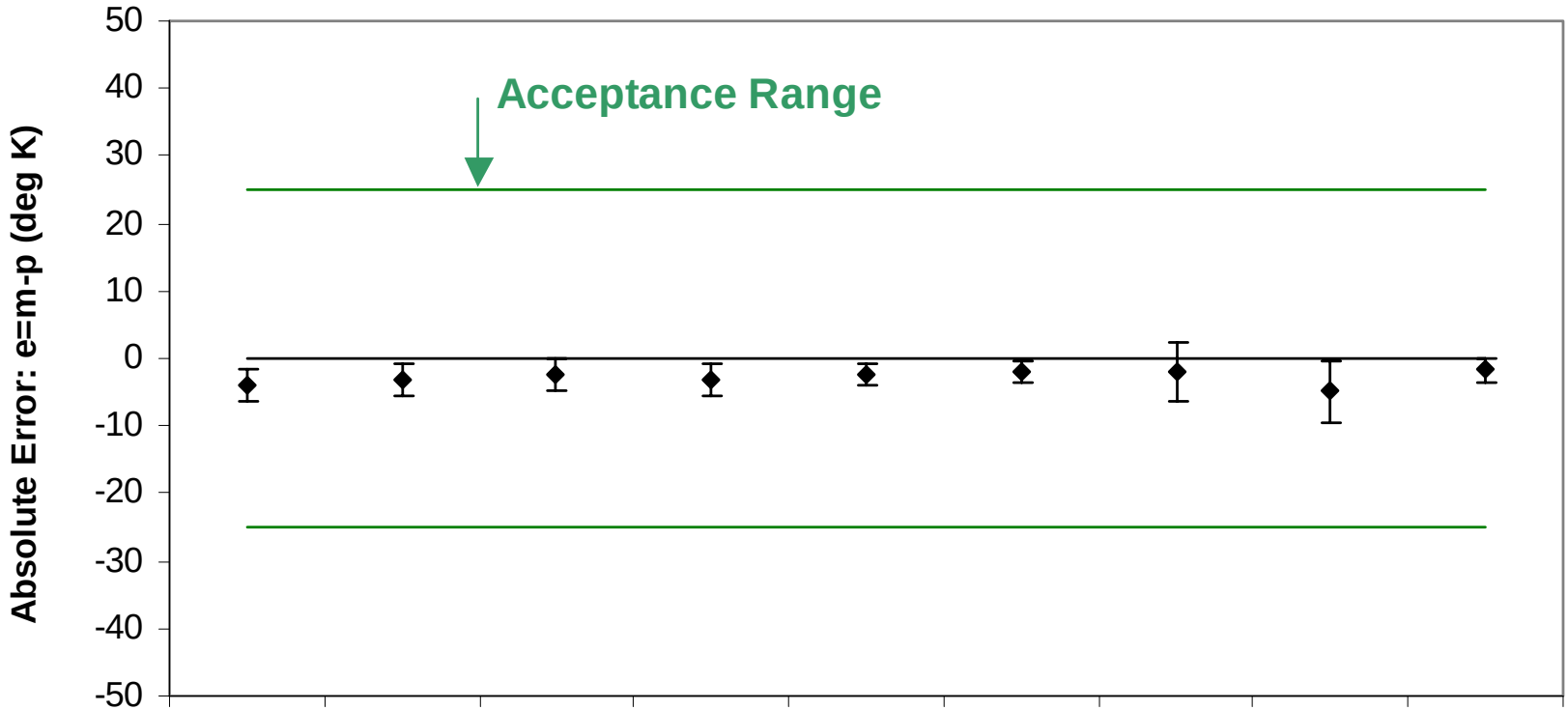
- **Physics-informed models validated against separate effects tests**



- **Validate against blast test and make prediction in tactical environments**

Well Established Physics Fidelity

e~2K for conduction



Various Locations in Test Object

Code Readiness

Hyperlinks

Attributes of Verification

Demonstrating Convergence to 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

Coupled Multi-Physics Across Codes

Coupled Multi-Physics Within Code

Separate Physics

Issues: non-smooth solutions, contact, constitutive laws, internal constraints, multiscale physics, global/local norms, etc.

Inference

Regression Testing

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

- Eliminate code bugs *AND* inadequate algorithms

SQE(A)

Code to Code Comparisons Are a Poor Substitute for Formal Verification

Code Comparison Principle (CCP)

Code 1 = assessed code Code 2 = benchmark code

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

$$\therefore \|\text{Code 1} - \text{Code 2}\|.$$

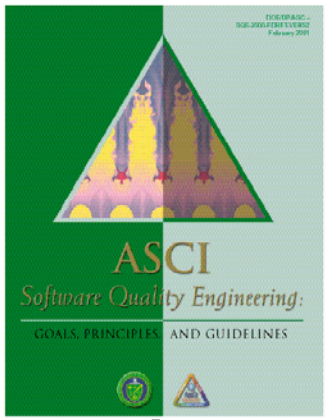
What if this term is not negligible?

- **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?

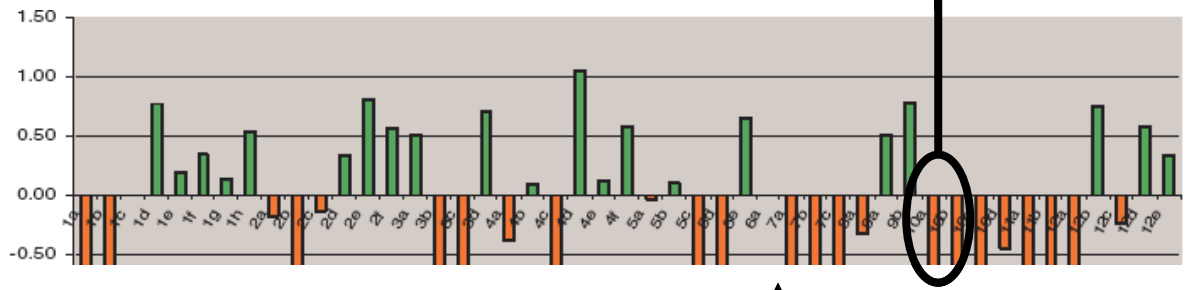
SQE(A): Demonstrated Due Diligence in the Stewardship of Codes

Requirements



SourceForge: Issue Tracking

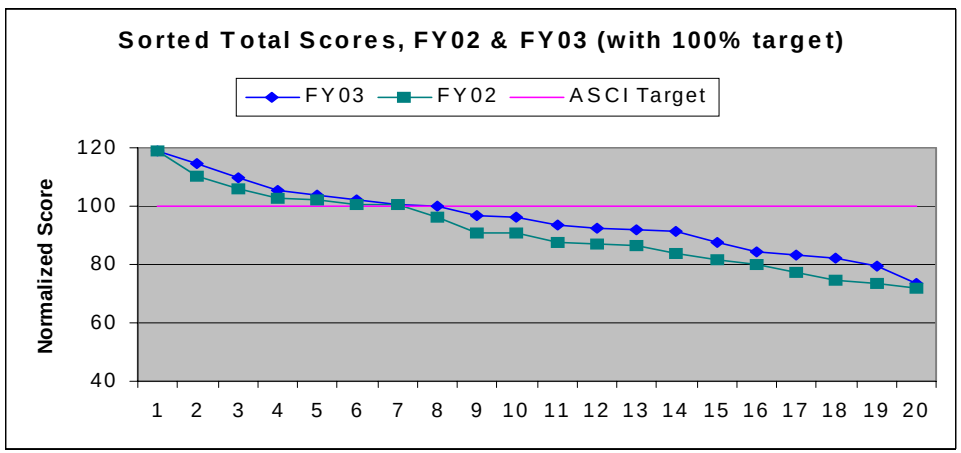
Improvement



FY03 AVG = 95.36

Calore = 96

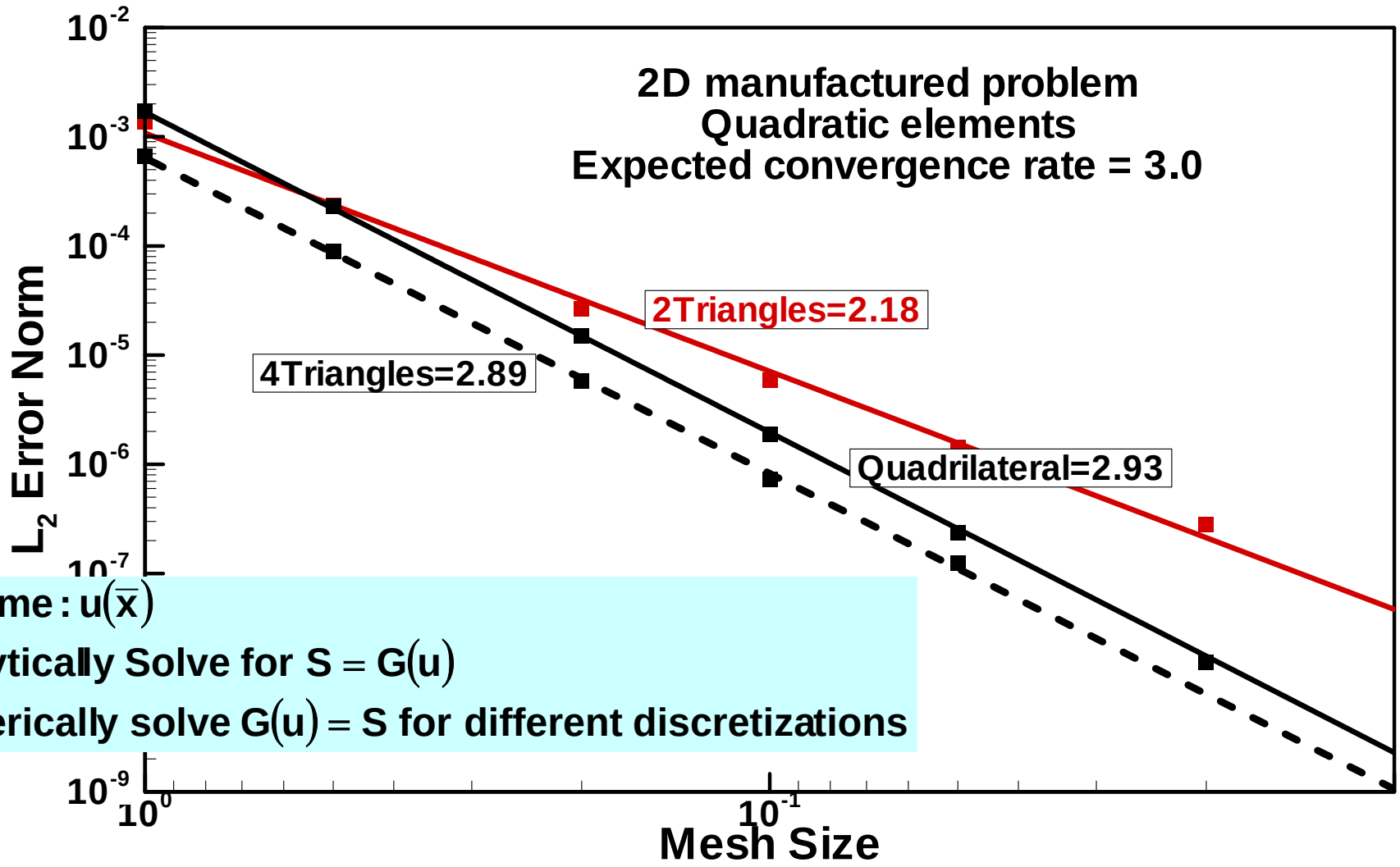
FY02 AVG = 91.06



Assessments

Verification with Manufactured Solution

CEPTRE: Radiation Transport



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

Analytically Solve for $S = G(u)$

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

Measuring Progress in Code Verification Coverage and Interactions

	Features & Capabilities	Unit Tests	Verification Test Suite: VERT 1	VERT 2	VERT 3	Ideal
Code A	FC1		VT1			
	FC2	UT1	VT1			
	FC3	UT2	VT1			
	FC4	UT3	VT1			
	FC5			VT2		
Code B	FC6	UT4		VT2		
	FC7	UT5			VT3	
	FC8	UT6			VT3	
	FC9	UT7			VT3	
	FC10	UT8			VT3	
	Code or Appl Perspective	Line or Cap Coverage	Capability+Interaction Coverage			
		80%	3.22%			

$$f = \frac{\sum_{i=1}^{Nverts} \left(\sum_{r=1}^{nv} C_r \right)_i}{\sum_{r=1}^{NFC} C_r}$$

Solution Verification

Hyperlinks

Attributes of Verification

Demonstrating Convergence to 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

Coupled Multi-Physics Across Codes

Coupled Multi-Physics Within Code

Separate Physics

Issues: non-smooth solutions, contact, constitutive laws, internal constraints, multiscale physics, global/local norms, etc.

Inference

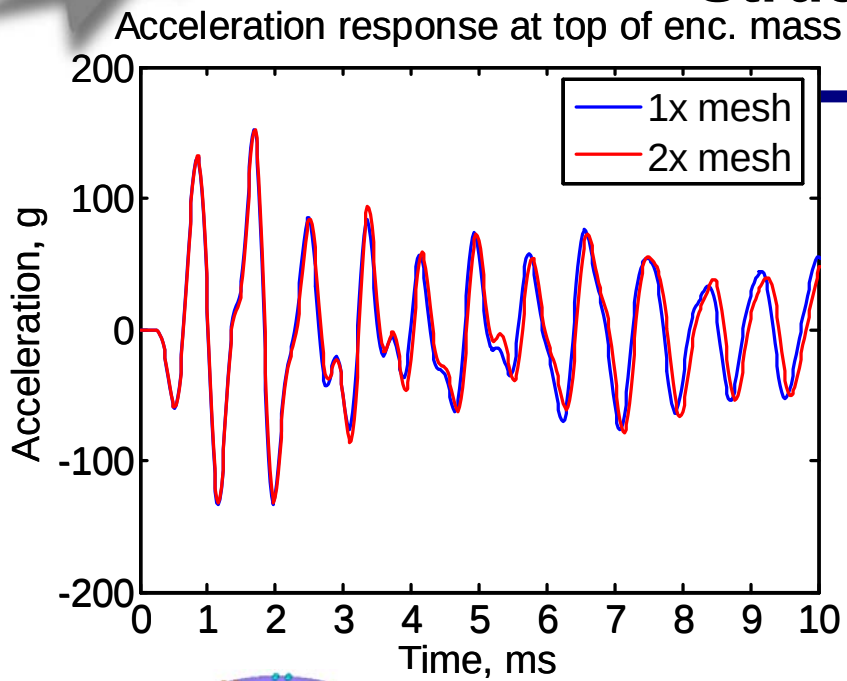
Regression Testing

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

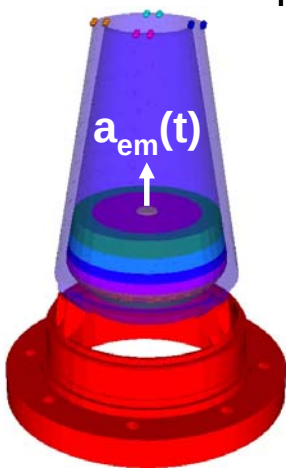
- Eliminate code bugs *AND* inadequate algorithms

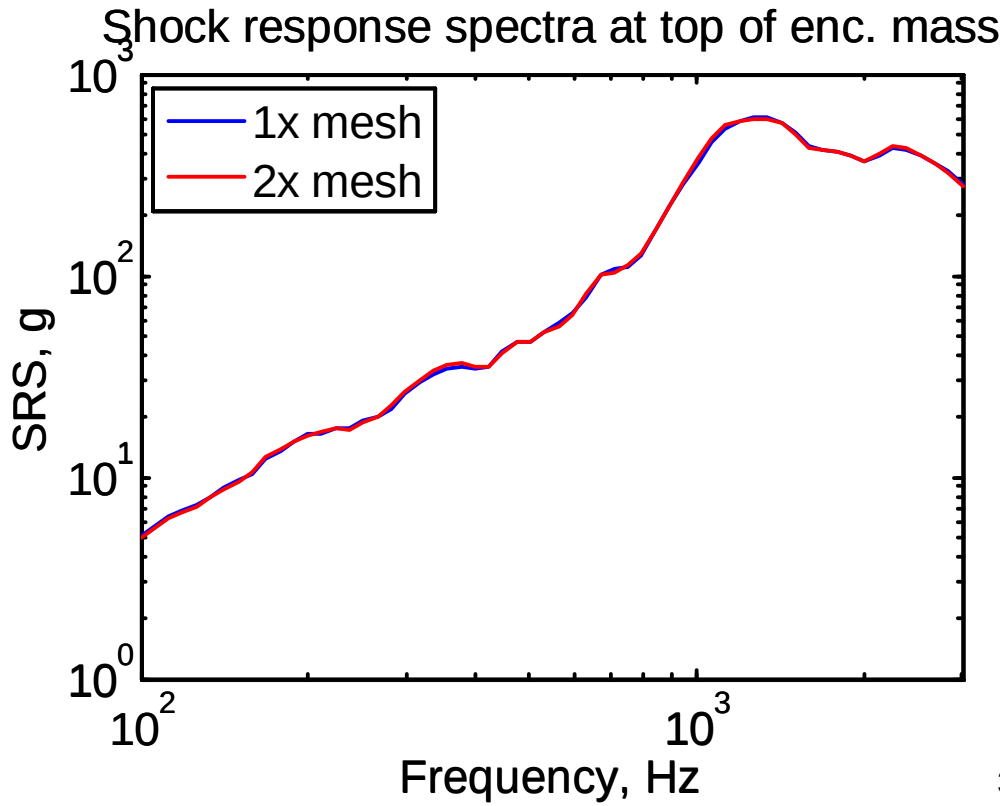
SQE(A)

Sensitivity to Mesh Parameters
Structural Dynamics



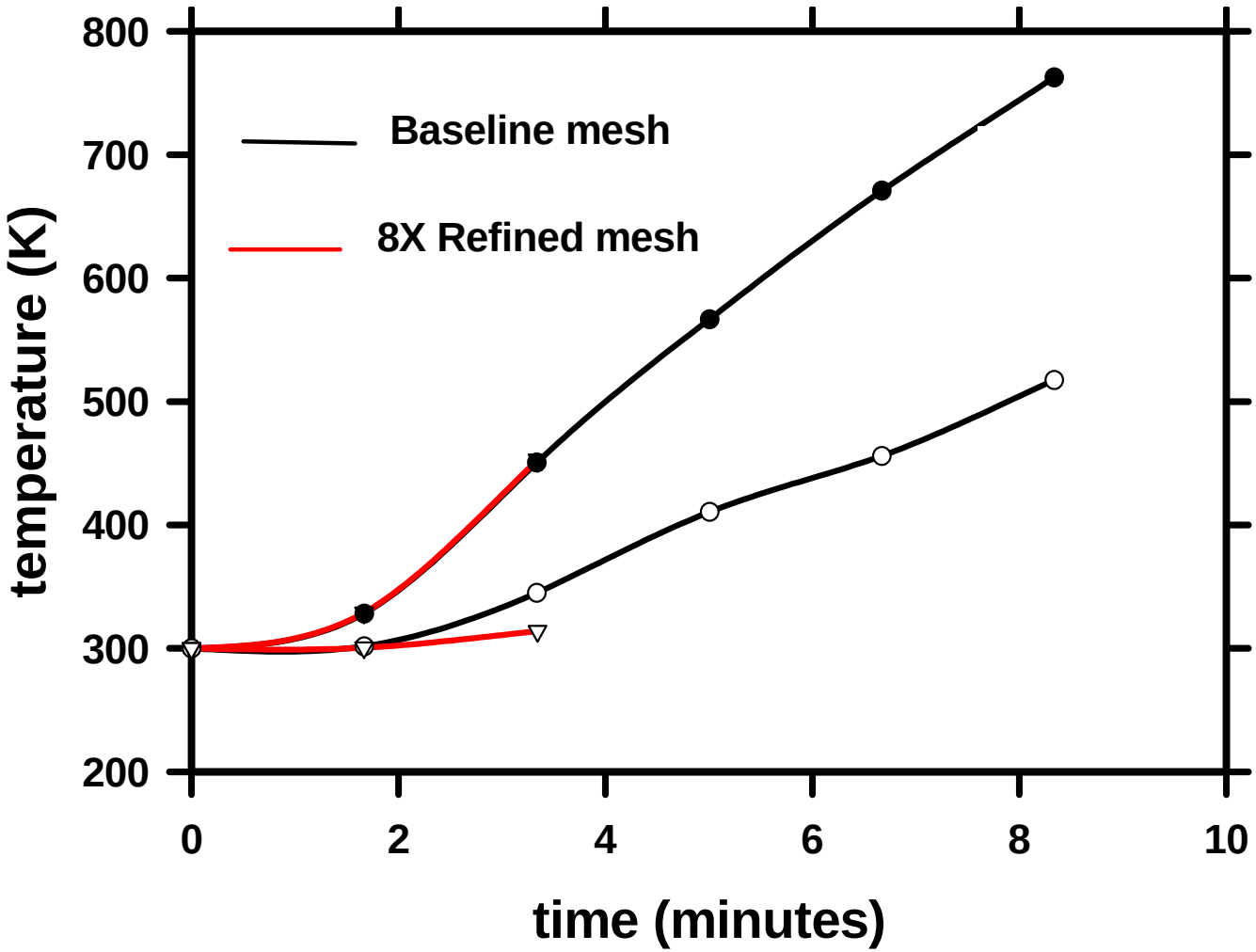
Max. relative error between SRS: +/- 5%



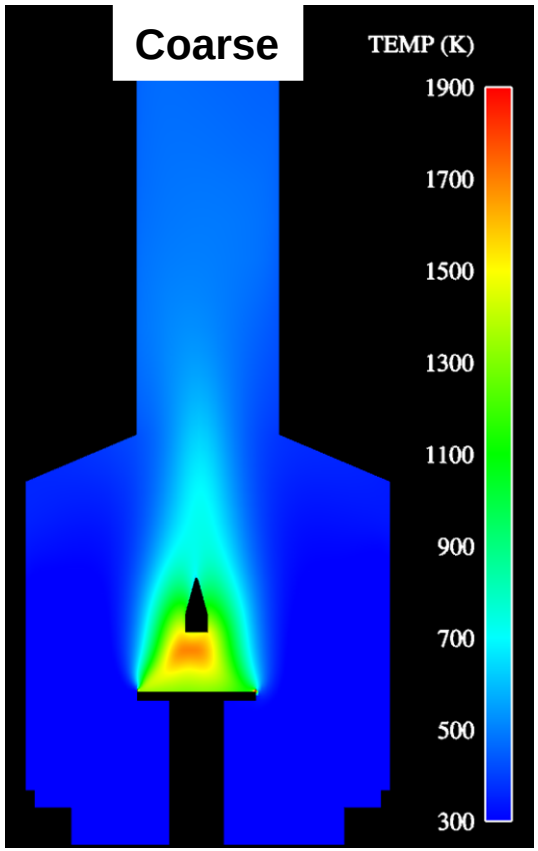


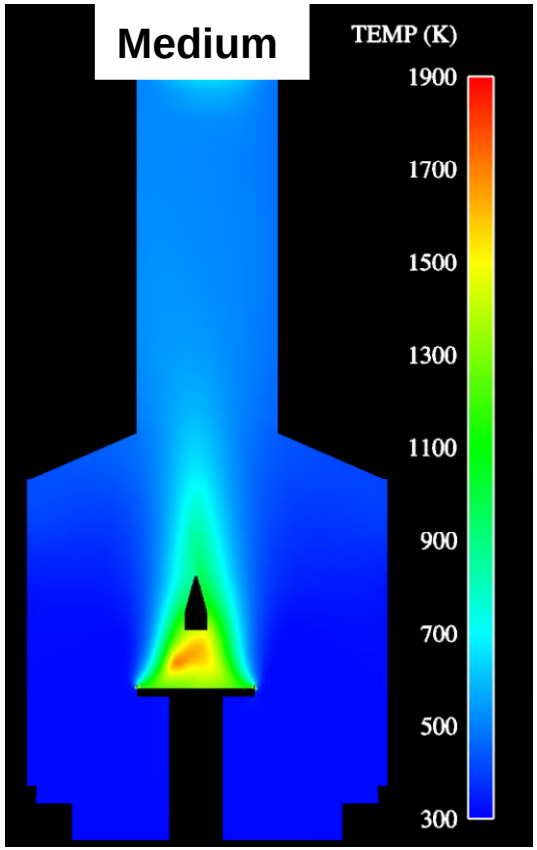
Solution Verification on High Fidelity Models is Hard

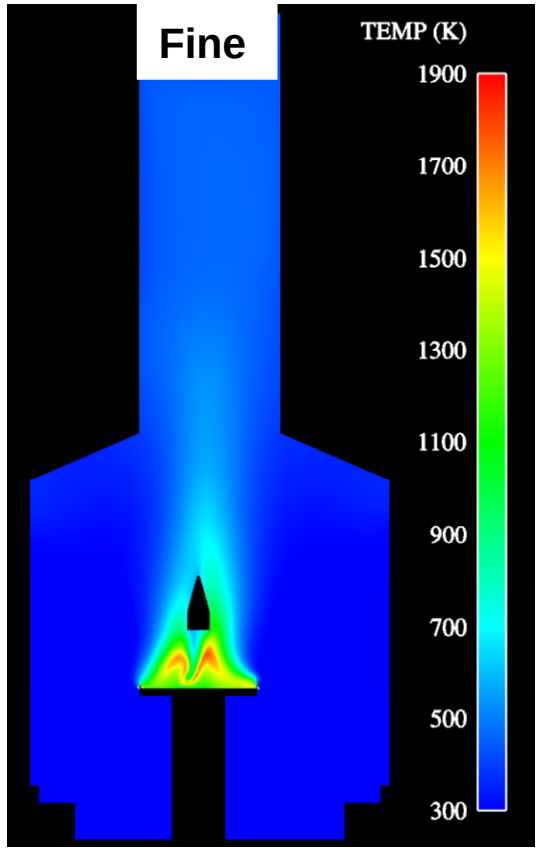
Solution Verification: Is the Discretization Adequate?



Calorimeter Fire
BVG Solutions

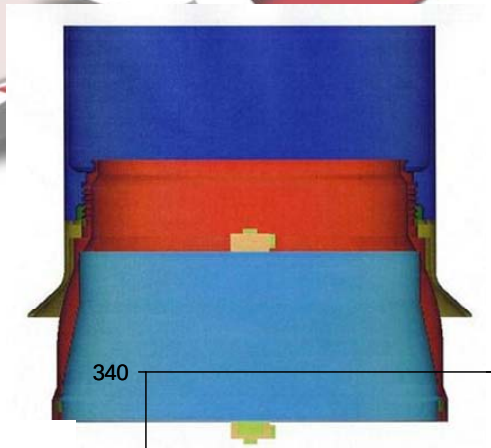


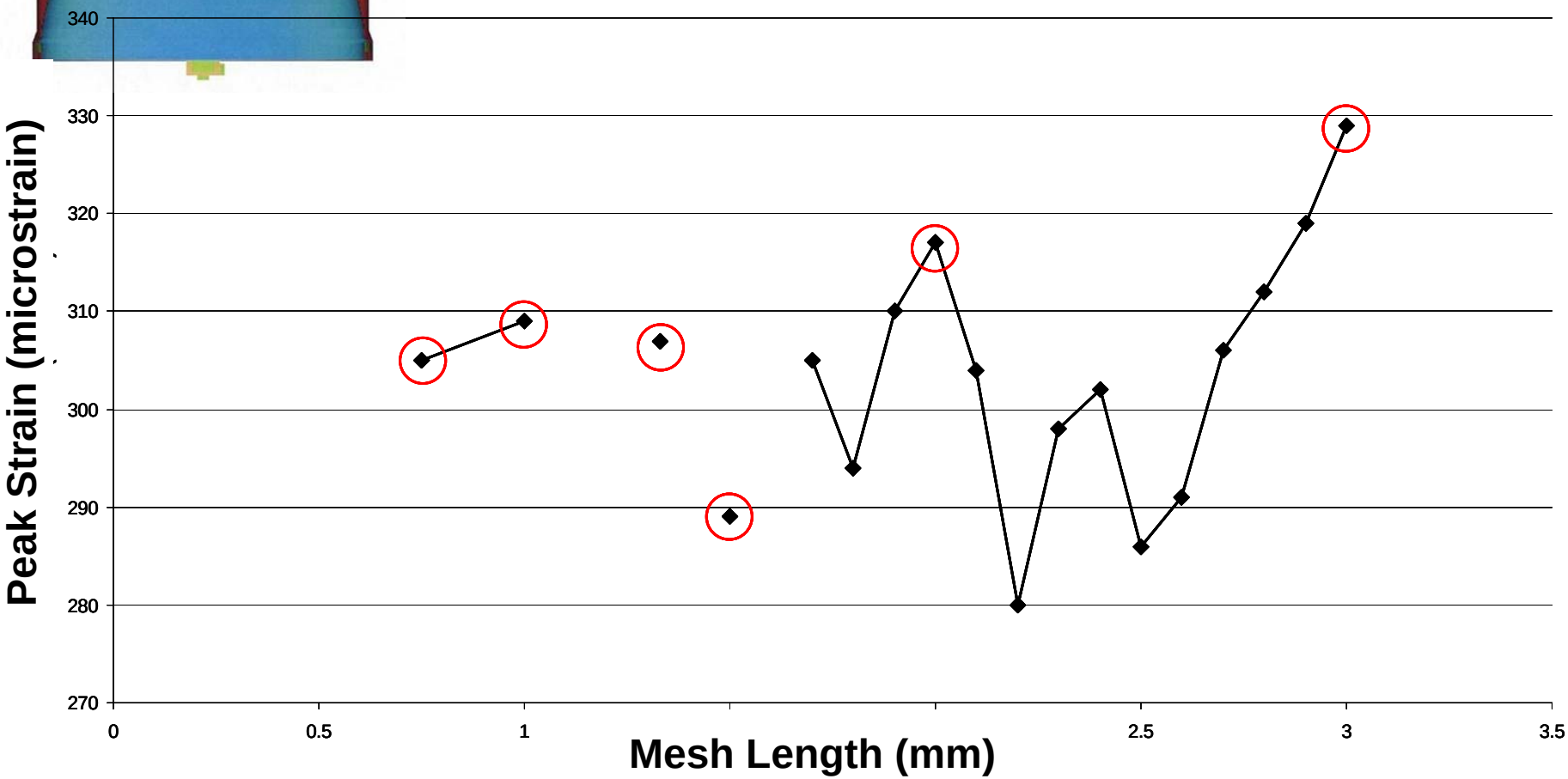




Calculation Verification for a Threaded Assembly

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

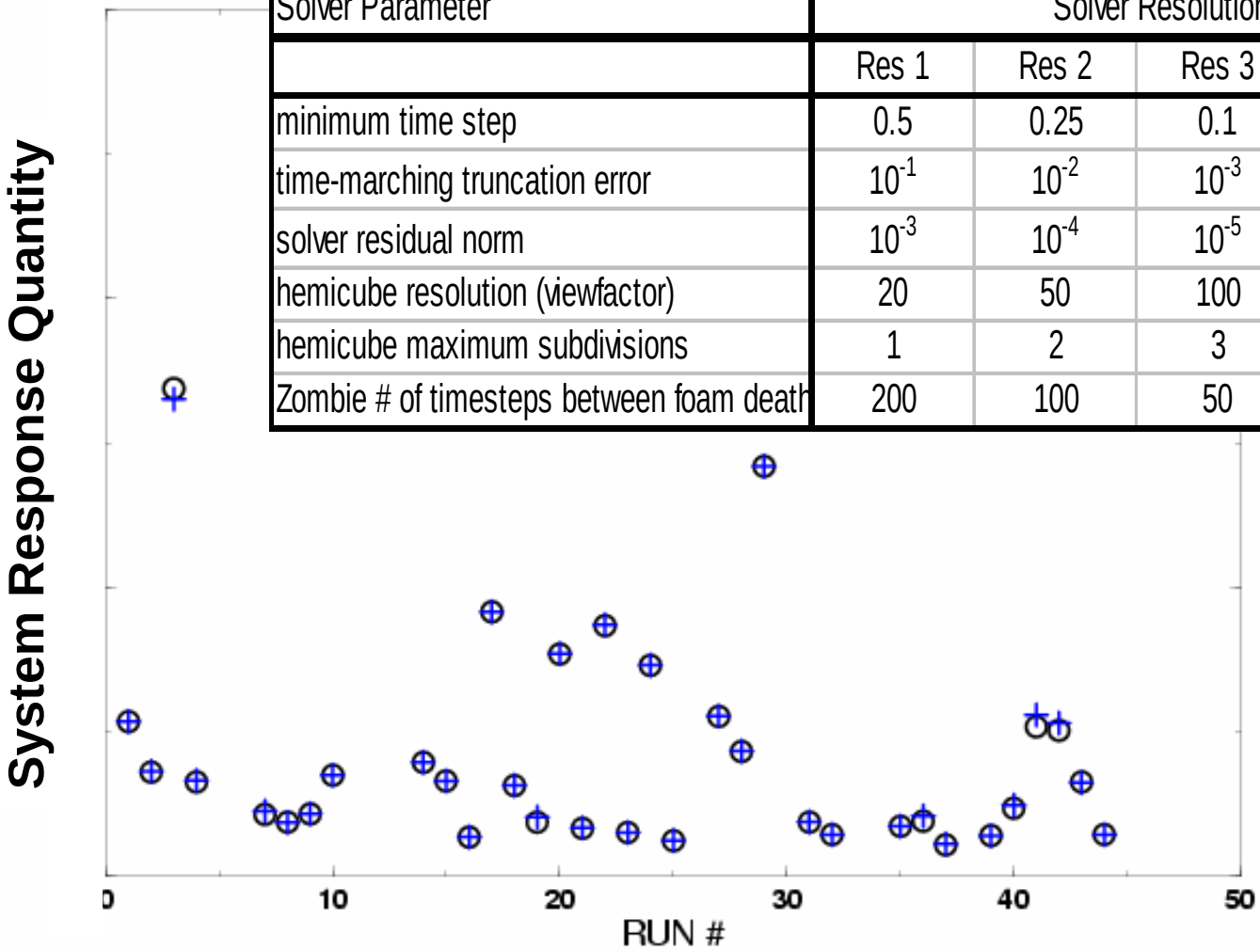




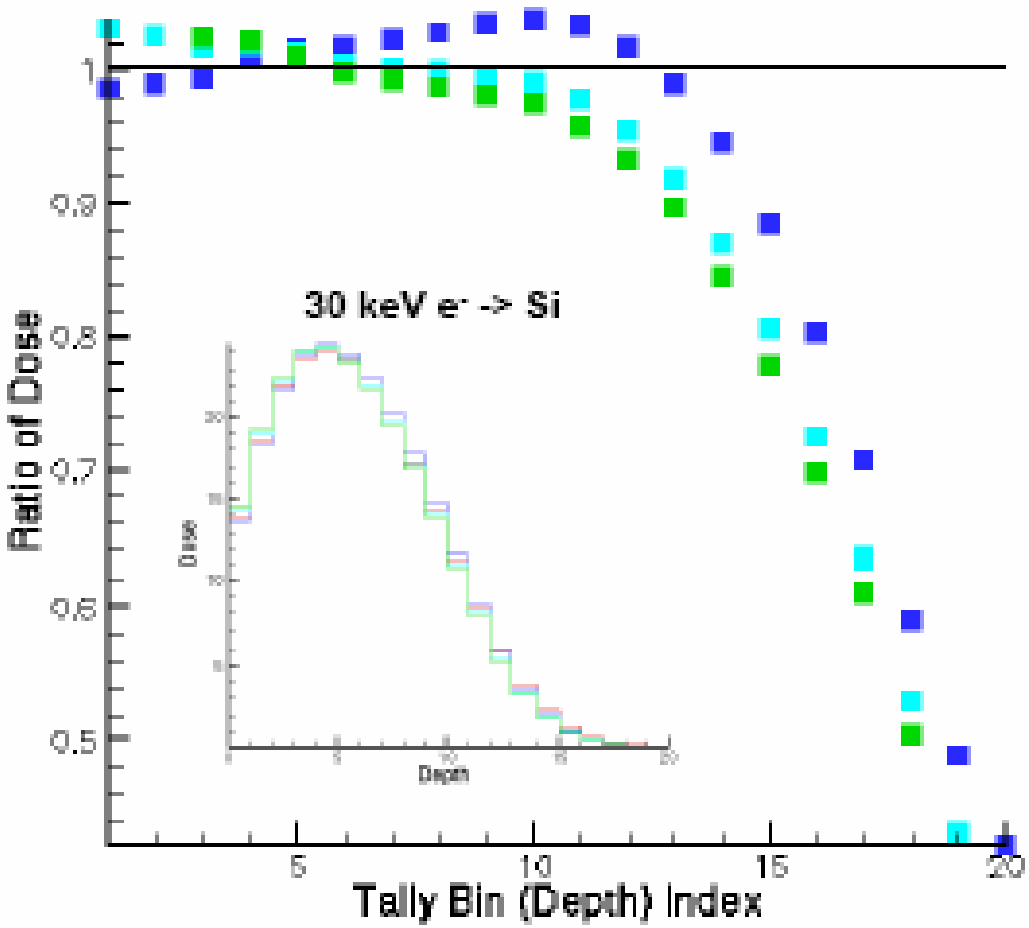
Solver Resolution Over UQ Parameter Space

Solution Verification: Are the solver settings adequate?

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 death	200	100	50	20	1

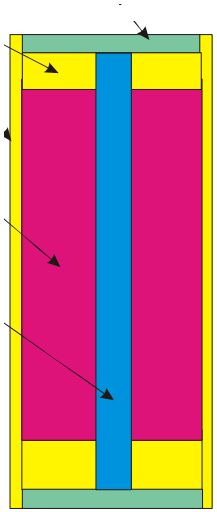


Dose Sensitivity to Electron Boundary Crossing Algorithm

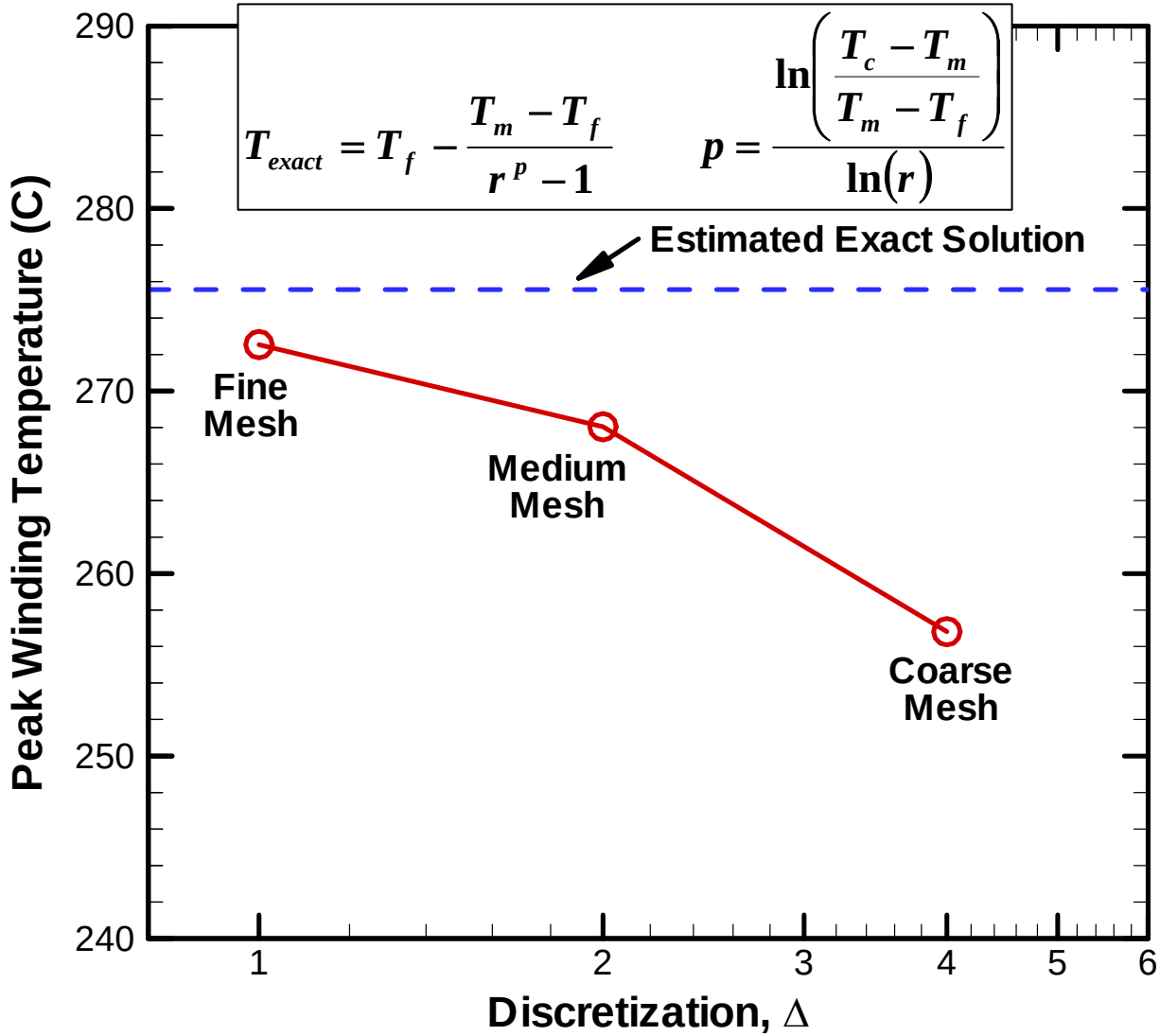


- Evaluation of ITS electron boundary-crossing error: (All with respect to no internal boundaries, default substep size.
- Blue: internal boundaries, default subtep size
- Cyan: Internal boundaries half-default substep size
- Green: Internal boundaries quarter-default substep size

Numerical Errors
Pollute Validation Assessments



$$T_{exact} = T_f - \frac{T_m - T_f}{r^p - 1} \qquad p = \frac{\ln\left(\frac{T_c - T_m}{T_m - T_f}\right)}{\ln(r)}$$



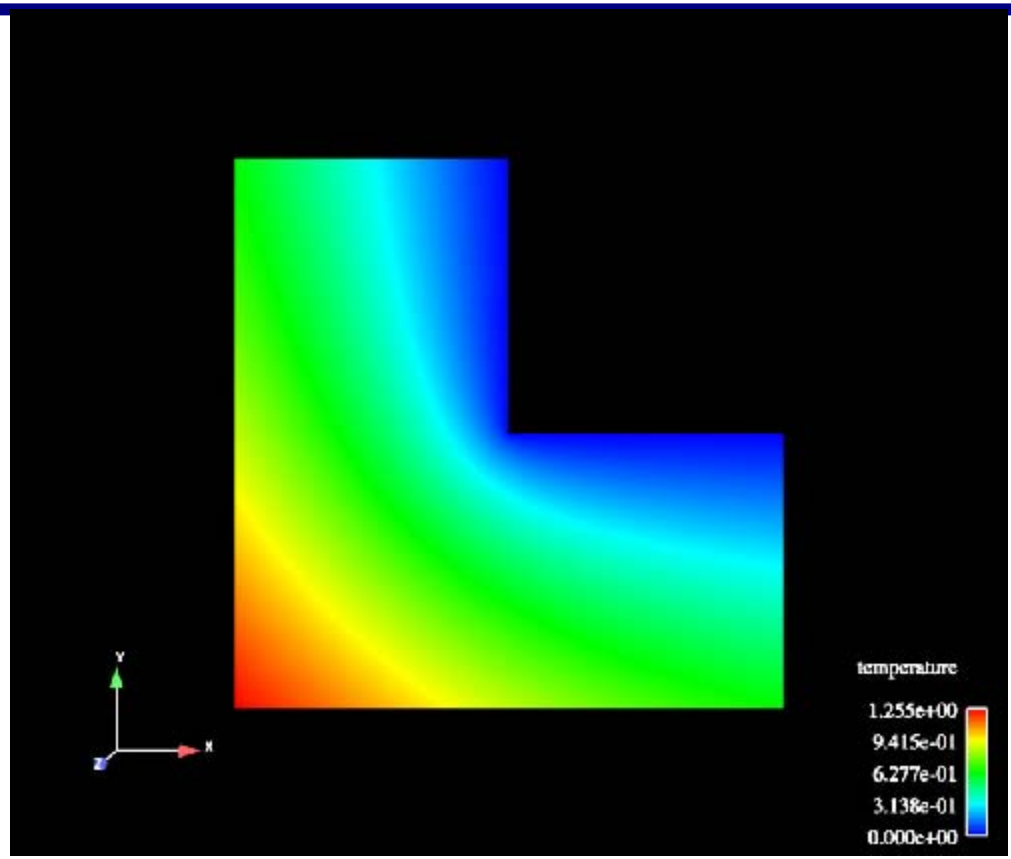
Verification of Error Estimator and Adaptive Algorithm

- 2D Exact Solution:

$$u = r^{2/3} \sin\left(\frac{2}{3}\theta\right)$$

- Linear elements
- ZZ error estimator

- Feedback adaptive algorithm:



$$\text{if } \left(||e^*||_{H_1(\omega_j)} > 0.995 \max_{1 \le i \le N_\omega} ||e^*||_{H_1(\omega_i)} \right) \text{ then refine } \omega_j$$

Validation

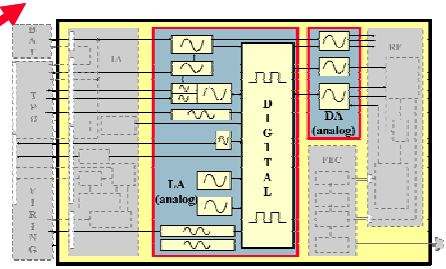
Hyperlinks

Science-Based Validation Experiments

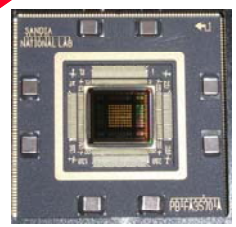
Validation: Are You Solving the Right Equations?

Hierarchal Validation: Right answer for the right reason

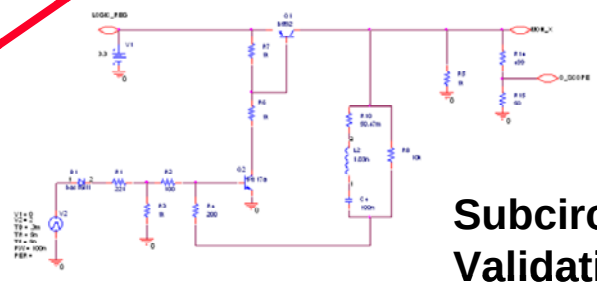
Increasing complexity,
Decreasing number of tests



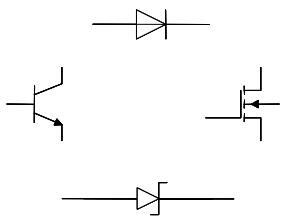
System-Level Circuit Validation



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**

Science-Based Validation Experiments

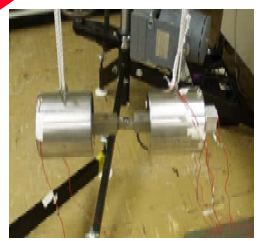
Validation: Are You Solving the Right Equations?

Hierarchal Validation: Right answer for the right reason

Full System Test

Mockup with jointed structure and foam embedded object

Jointed structure validation



Single joint validation

Joint parameter characterization

Increasing complexity, Decreasing number of tests

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

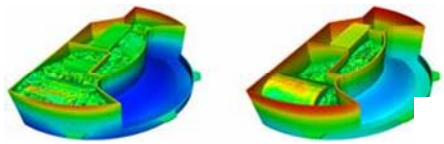
Science-Based Validation Experiments

Validation: Are You Solving the Right Equations?

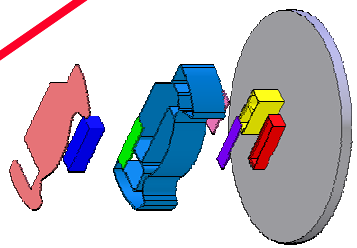
Hierarchal Validation: Right answer for the right reason

Increasing complexity,
Decreasing number of tests

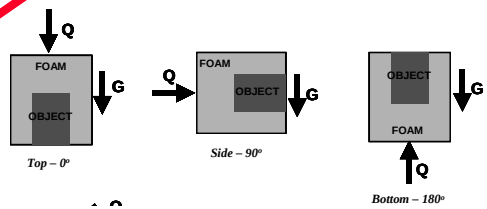
Full System Test



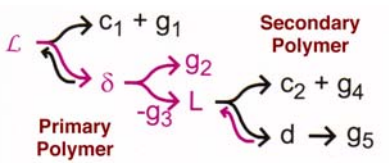
Validation Real Sub-systems



Validation with mockups



Foam recession

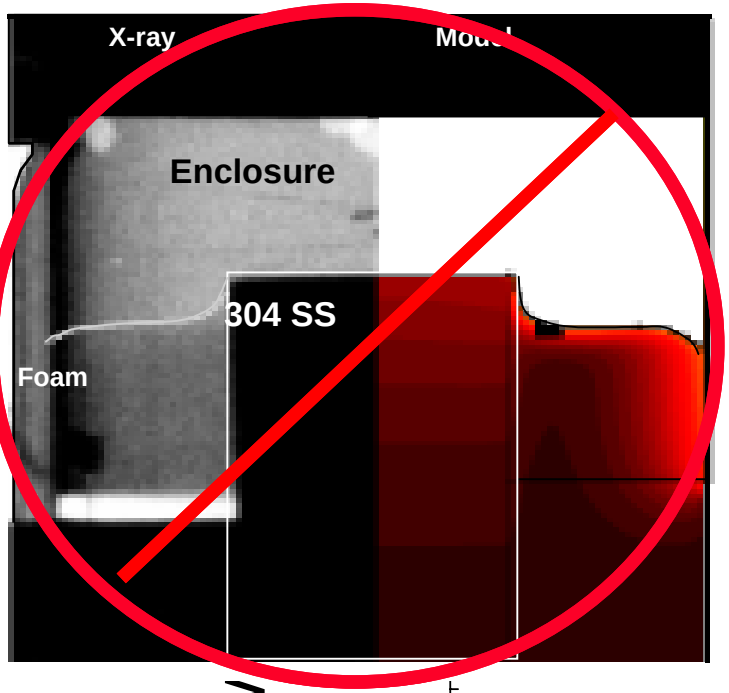


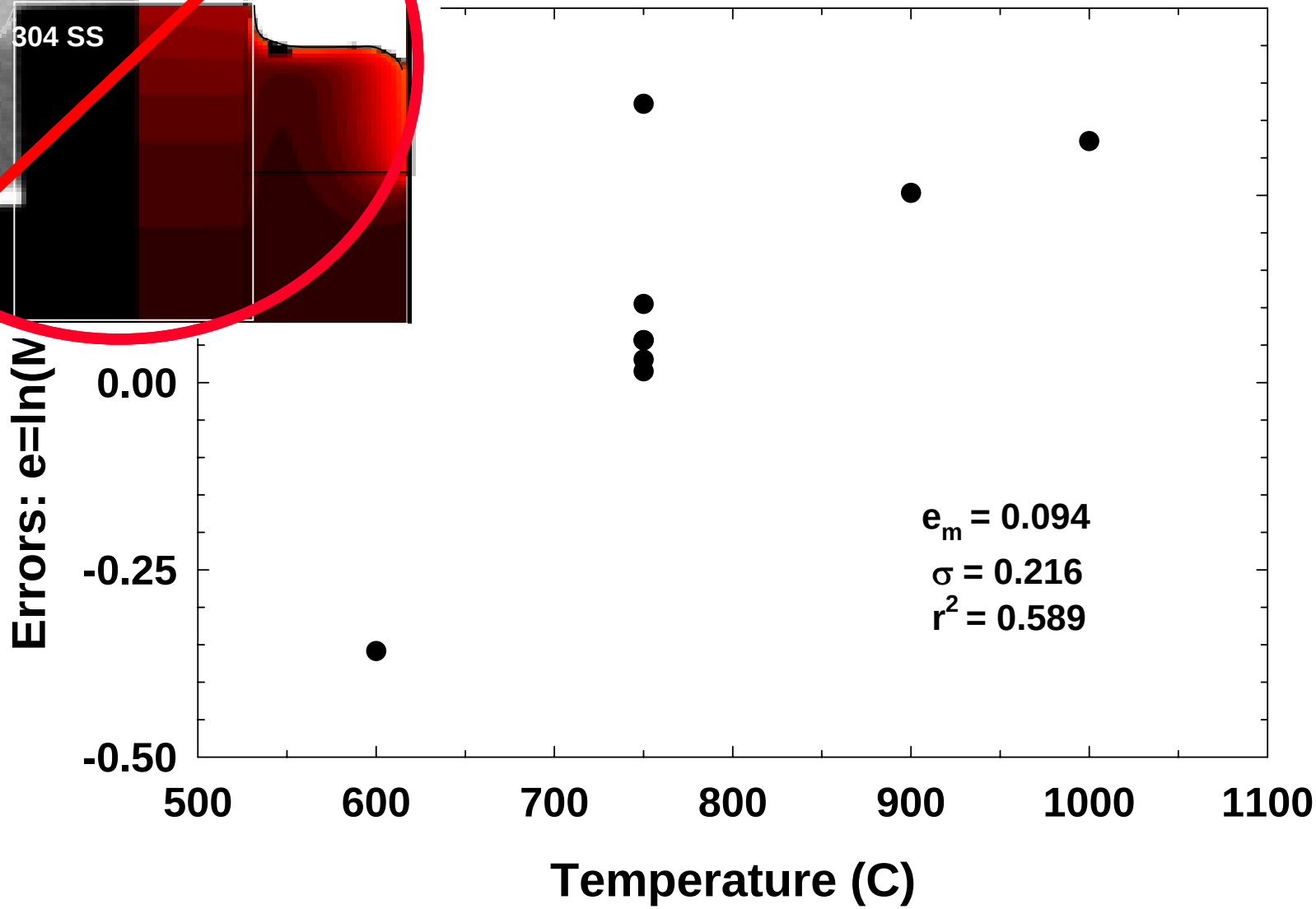
Chemistry characterization/validation

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

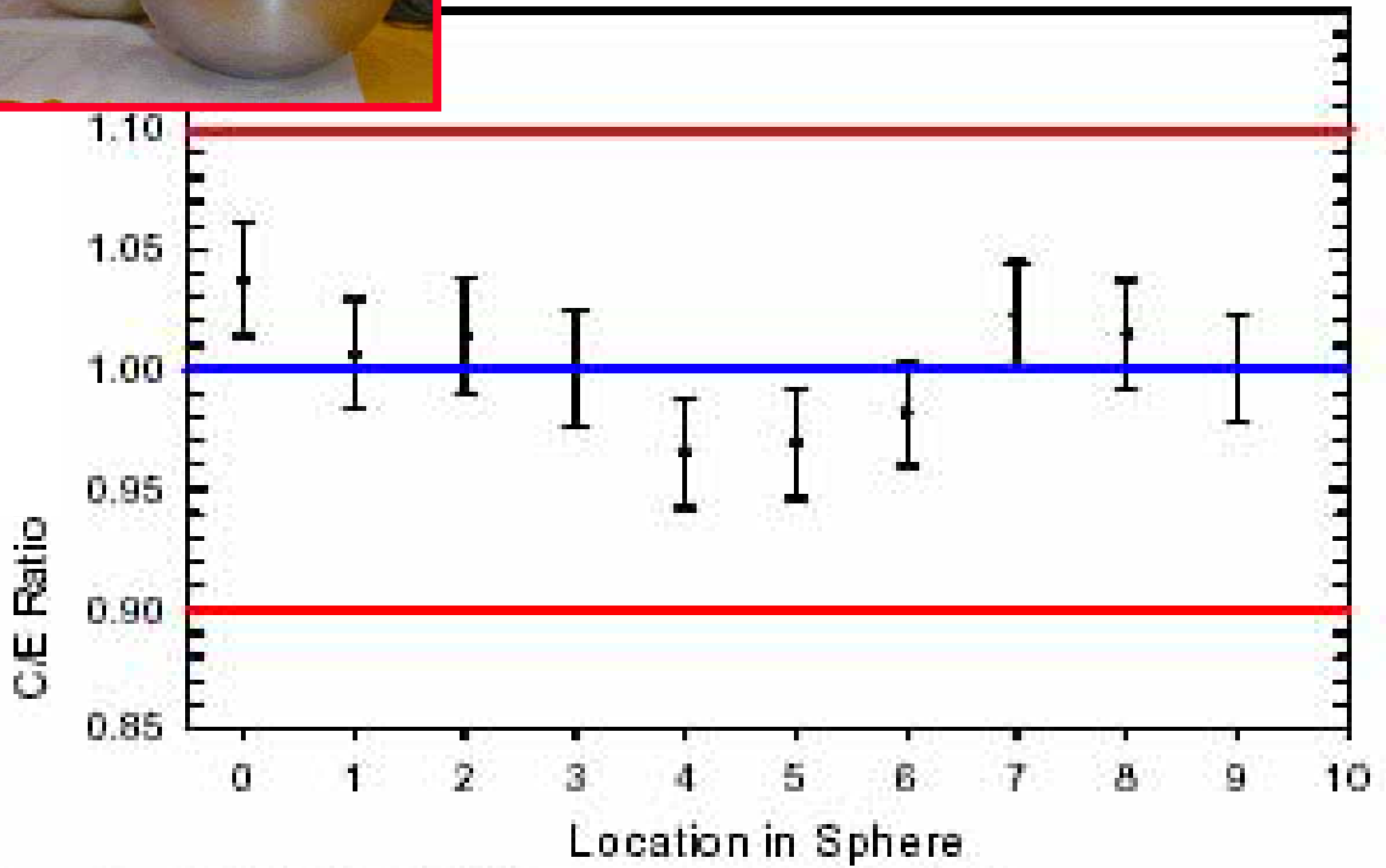
Validation is Statistical

Vugraph Norms Are Not Adequate





Neutron Attenuation in Test Objects

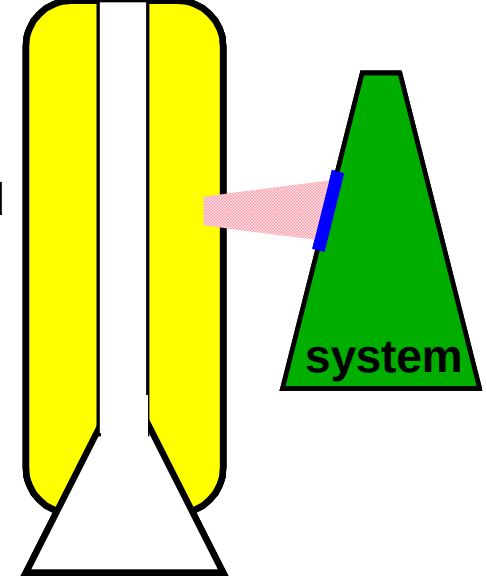


QMU and Sensitivities

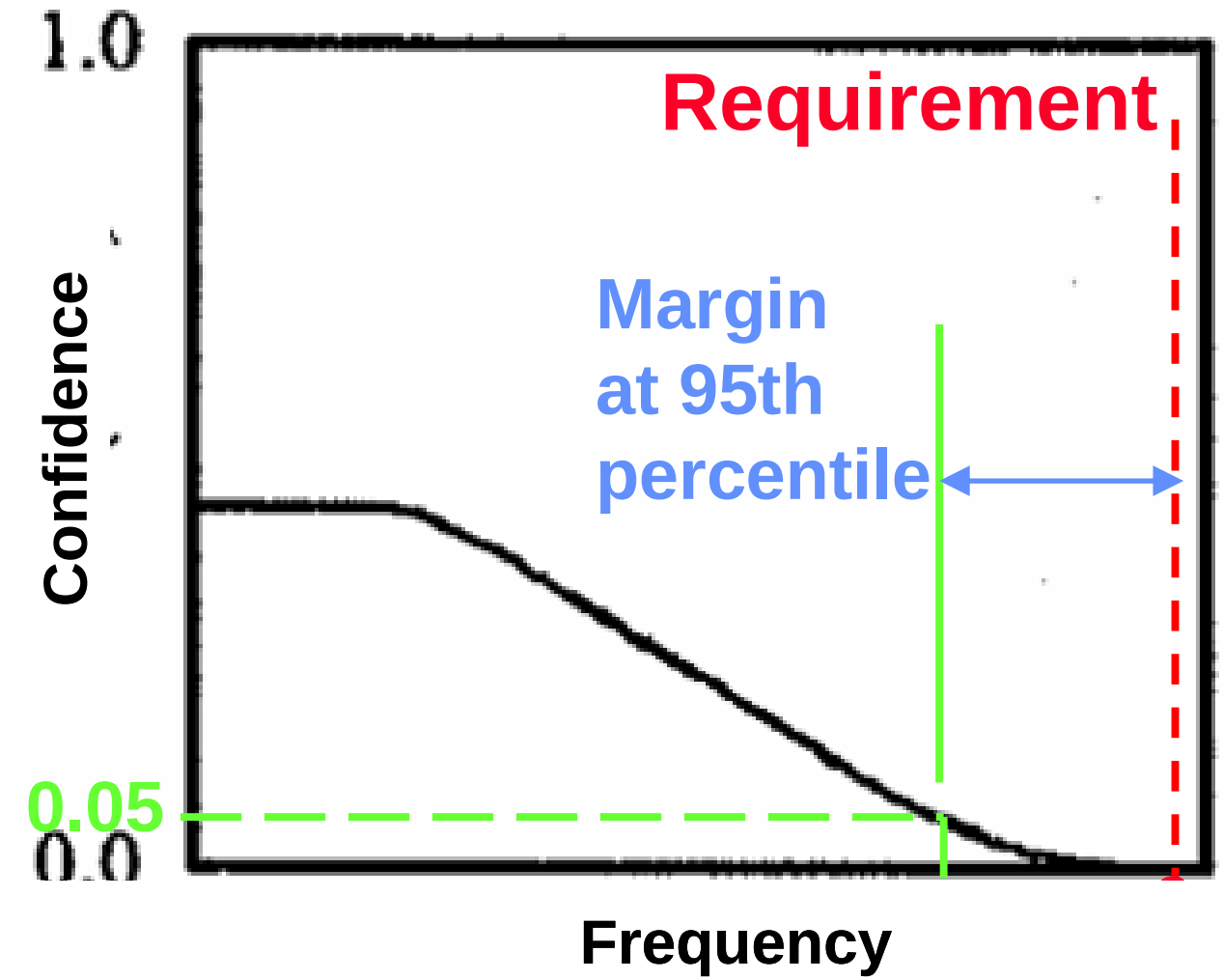
Hyperlinks

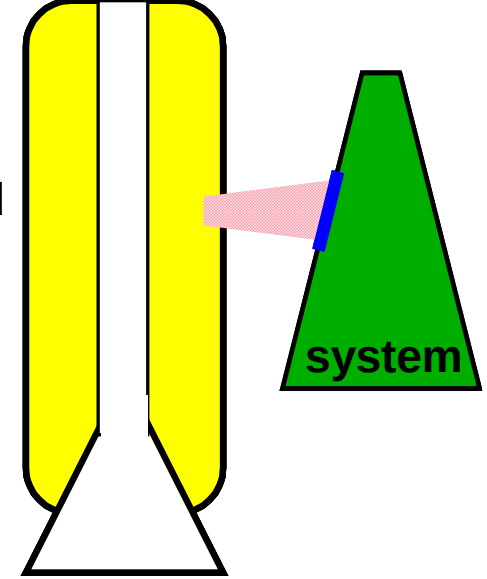
Aleatory and Epistemic Uncertainties

- **Aleatory uncertainty: Inherent randomness in behavior of system under study (frequency interpretation)**
  - **Alternatives: Variability, stochastic uncertainty, irreducible uncertainty, type A uncertainty**
  - **Examples: component failures or material properties derived from statistically significant testing under conditions relevant to intended application**
- **Epistemic uncertainty: Lack of knowledge about appropriate value to use for a quantity that is assumed to have a fixed value in the context of a specific analysis (confidence or belief interpretation)**
  - **Alternatives: state of knowledge uncertainty, subjective uncertainty, reducible uncertainty, type B uncertainty**
  - **Examples: representative scenarios, unknown parameters in frequency distributions, parameters or models with defensible bounds but no sense of frequency**



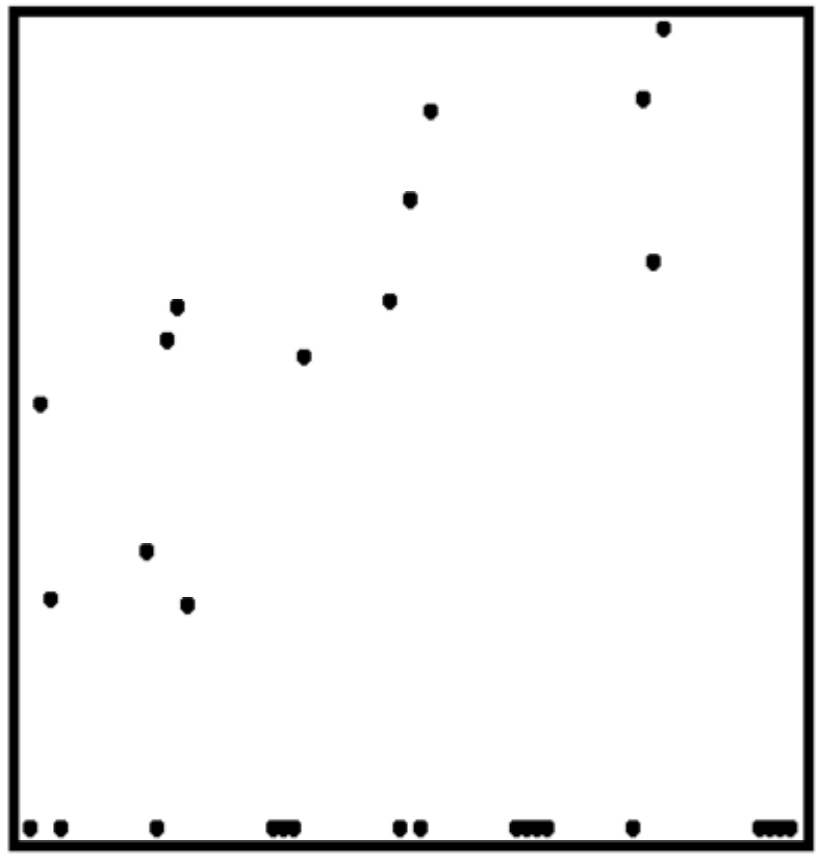
Quantified Margins and Uncertainties

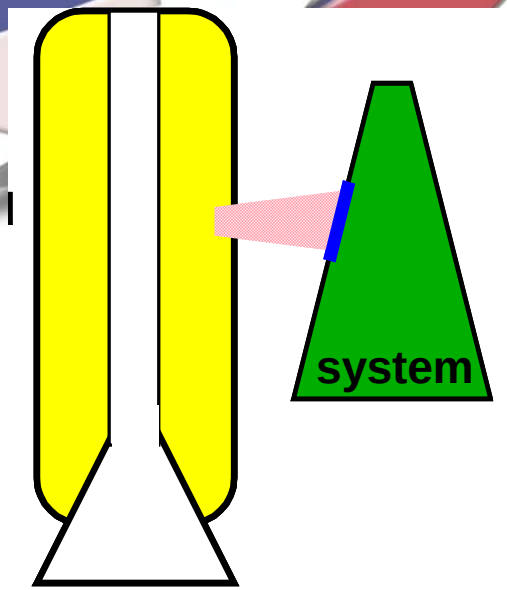




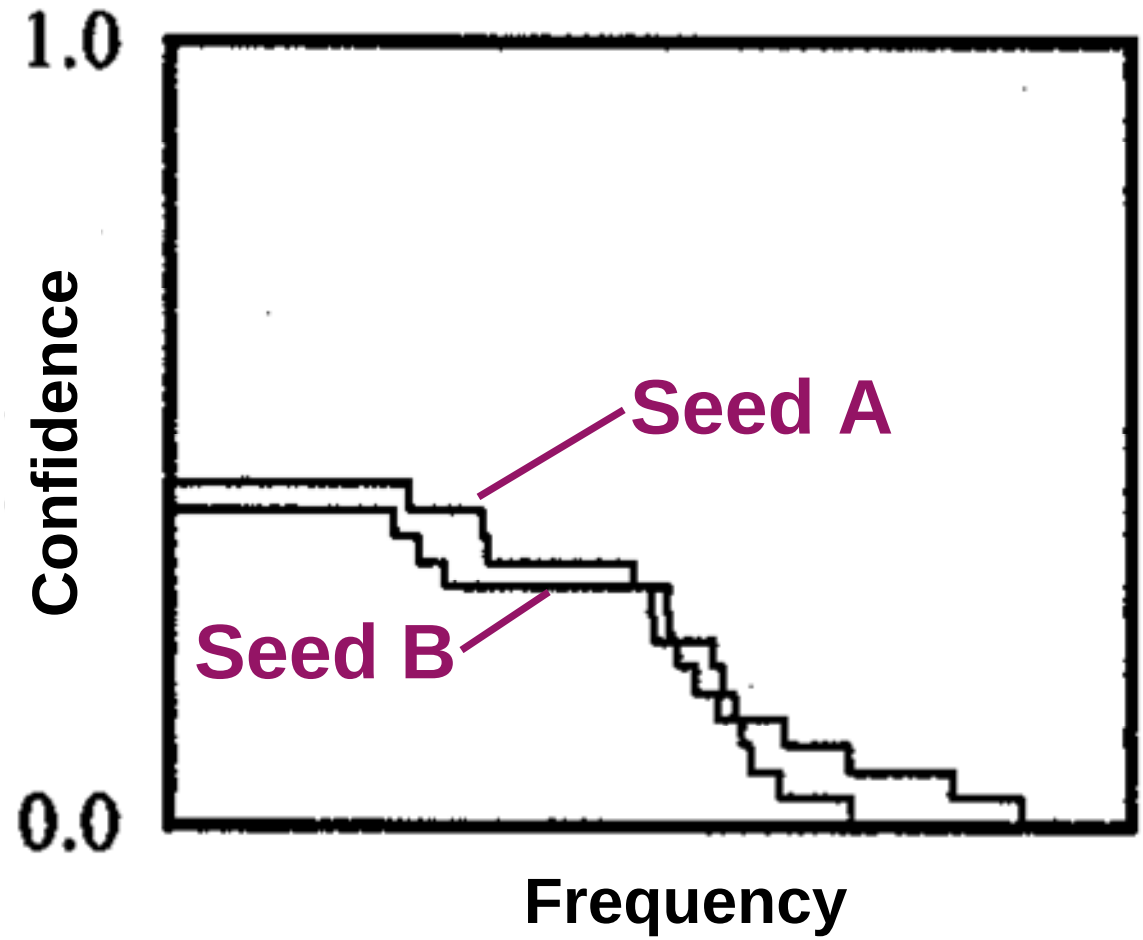
Sensitivity Analysis

$$\mathrm{SCorr} = 0.809$$





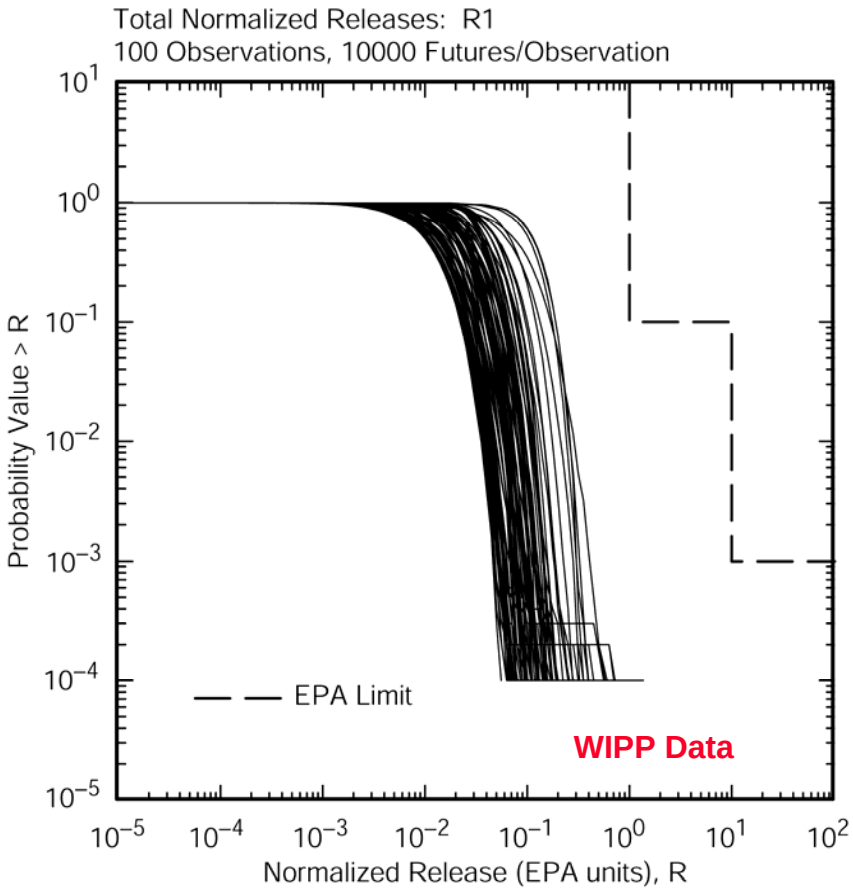
UQ Solution Verification

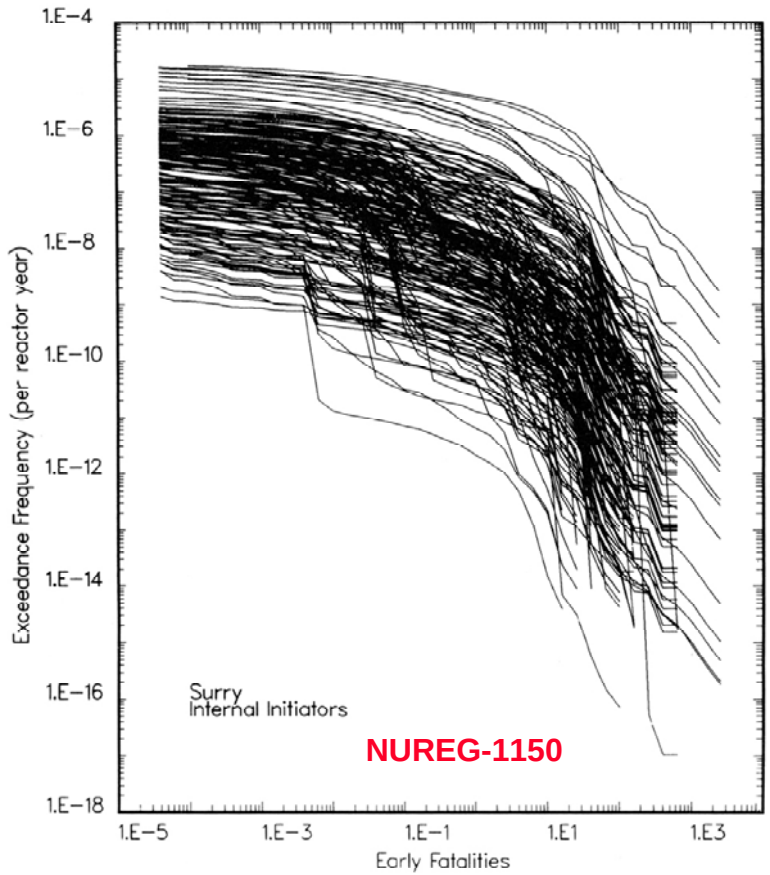


Seed Effects from limited sampling

WIPP and NUREG-1150 Precedents

High Consequence Regulatory Issues in the National Interest Addressed Primary Through Modeling and Simulation





Lessons Learned: (1) Seek BE + Uncertainty

(2) It takes more than one shot to get it right