MPilchConsulting

Cautionary Tale When Using Computational Simulation to Inform Regulatory Decisions

- Recommendations to Mitigate the Risk -

Martin Pilch PhD
26 Mustang Mesa Trail
Tijeras NM 87059
USA
MPilchConsulting@gmail.com
(505) 270-2248

Workshop on Dynamic Modeling and Simulation
Cologne Germany, August 29-30, 2019

This talk will address …

- **Why caution is necessary when regulatory decisions are informed by computational simulation**
- **Recommendations to manage the risk**
- **Difference between verification, validation, and calibration**
- **What responsible calibration looks like**
- **What would it take to certify a new baseline design with simulation alone without a hardware specific certification test**

1981
PhD NE

My career in a nutshell

Then

Now

Sandia National Laboratories

19 years

Reactor Safety

Resolution of regulatory issues through integration of testing and modeling

- Direct containment heating
- Lower head failure
- NRC, industry, national labs, academia, foreign partners

16 years

Nuclear Weapon Performance and Safety

VV&UQ capability development/deployment

- Advanced Simulation and Computing (ASC) Program

Integration of CompSim with testing for

- Component/system design, environments specification, test support, qualification
- Communicating "credibility"

Quantification of margins and uncertainties (QMU)

- Methods/deployment

MPilchConsulting

VV&UQ
3 years

Sandia Laboratories: Mentoring, knowledge transfer, and peer review of CompSim activities supporting major LEPs

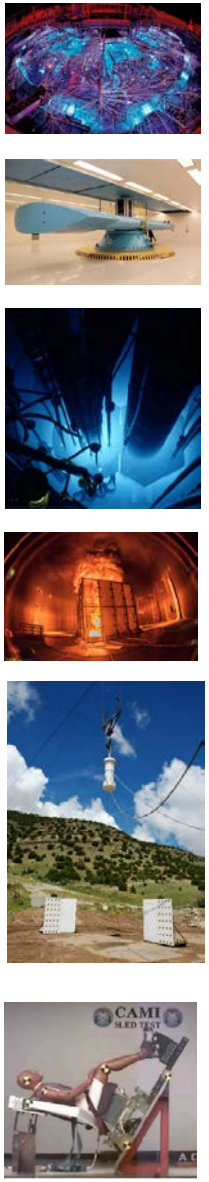
ORNL – CASL: V&V for new hifi modeling capabilities supporting nuclear power

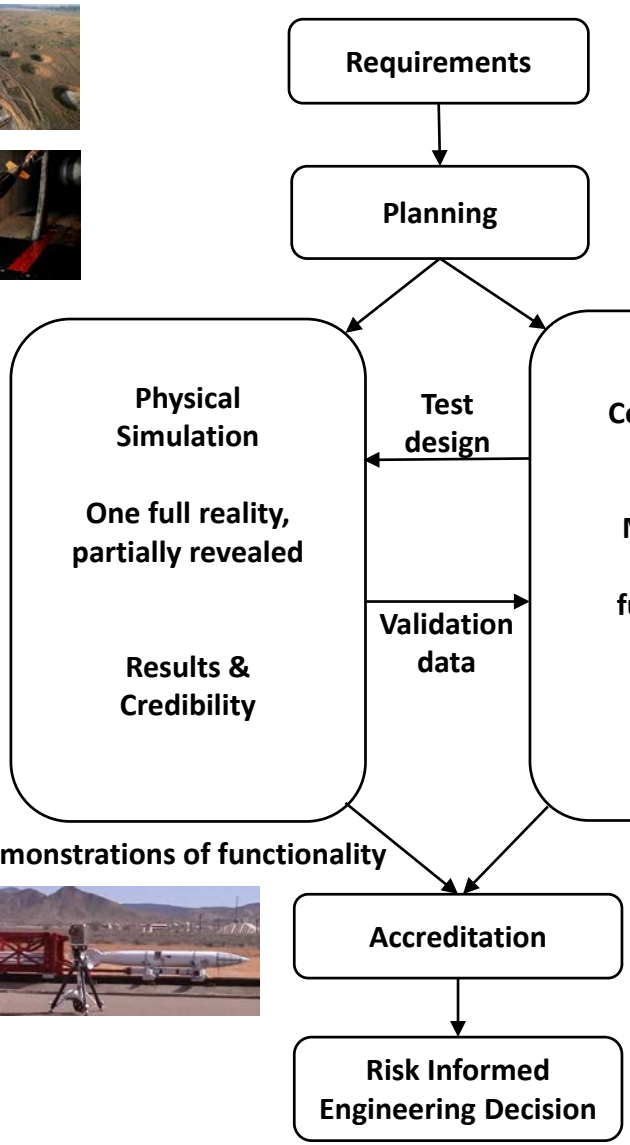
NAFEMS Booklet: *Simulation Verification and Validation for Managers*

McFarland Technologies, FAA, and HOA

One late night in my hotel room …
I thought I was hallucinating

Engineering decisions increasingly rely on computational simulation





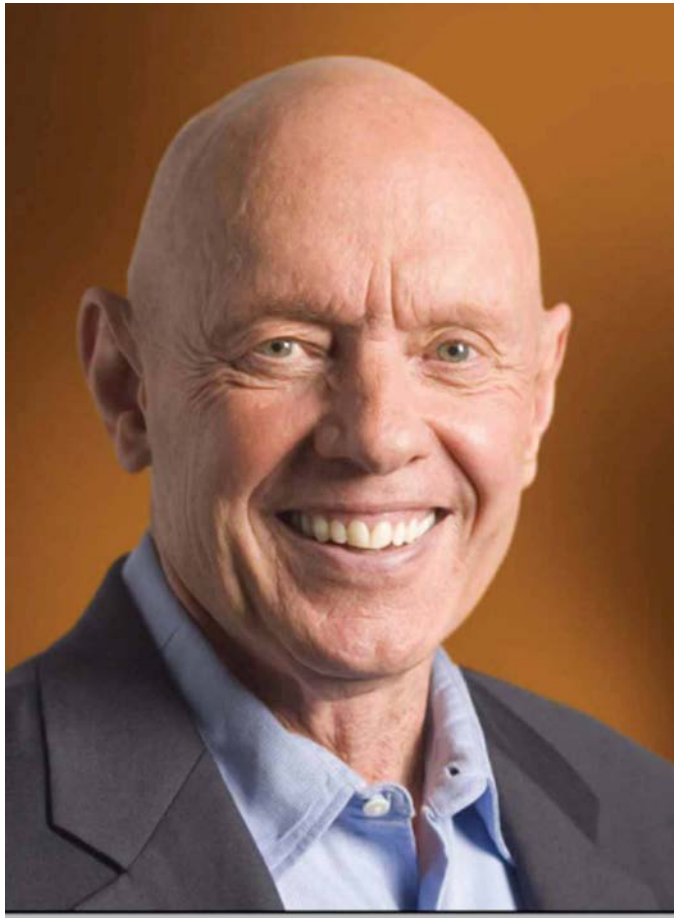
Value Proposition of Computational Simulation

- Optimize system & component design for performance, safety, and manufacturability
- Derive component environments
- Design test facilities to replicate component environments on the ground
- Margin assessment accounting for uncertainties
- Reduced system testing
- Accelerated qualification or certification

Some combination of faster, better, cheaper

Discovery!

Let's begin with the end in mind
…….. and cautionary tale #1

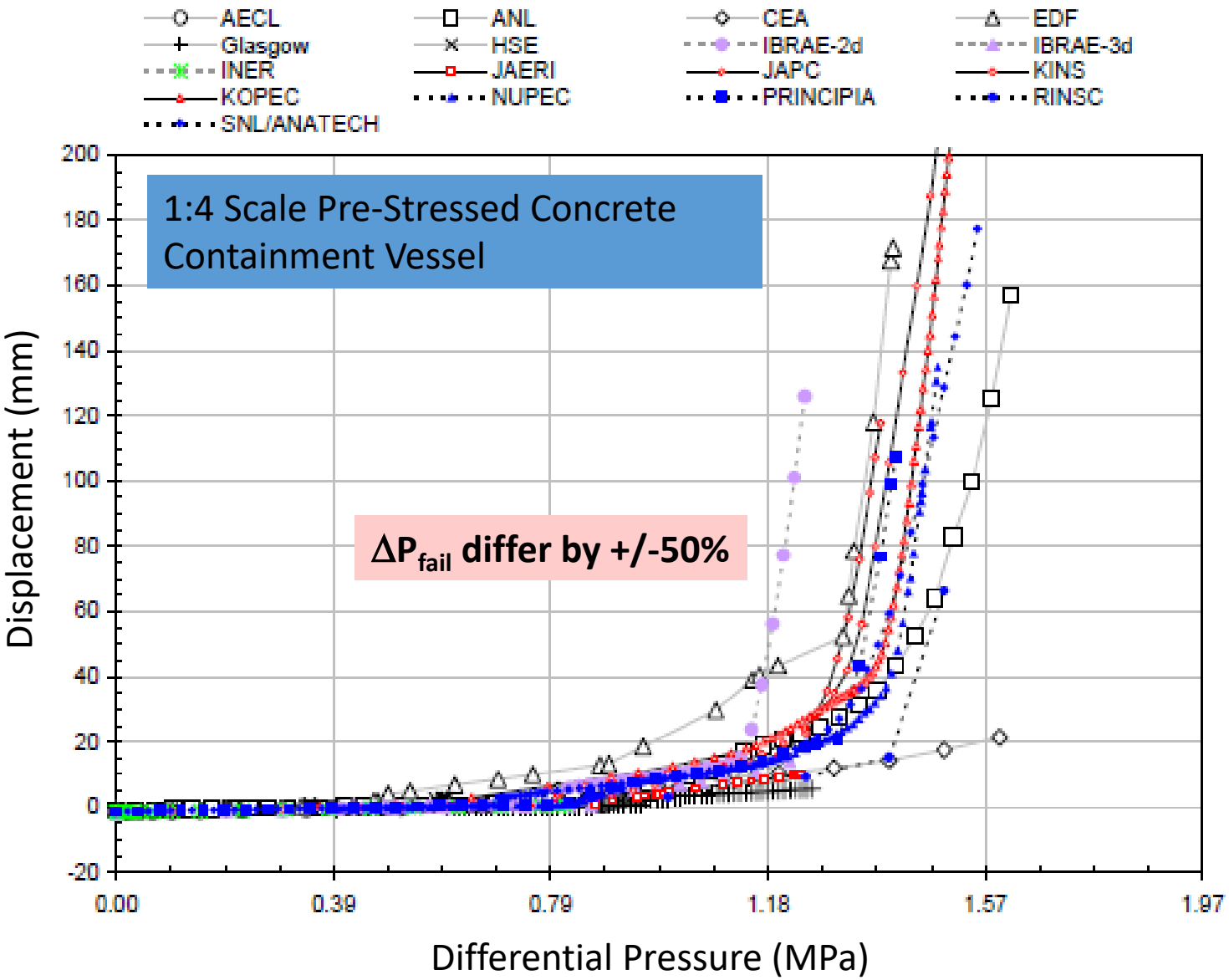


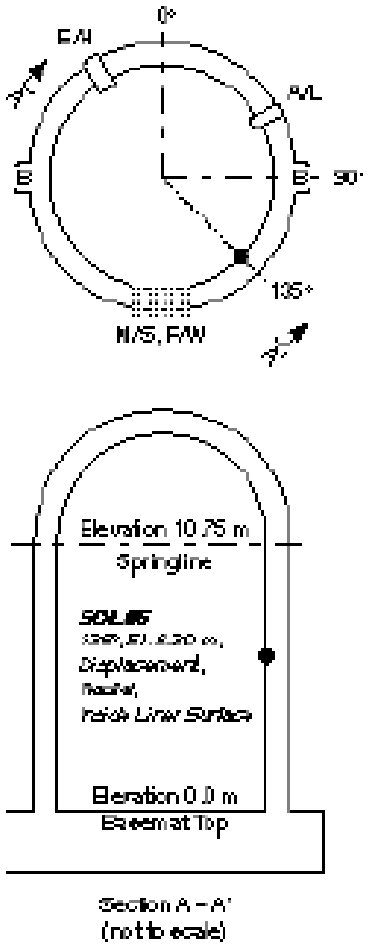
Steven Covey, *Seven Habits of Highly Effective People*

- **Reputable organizations**
- **Employing qualified analysts**
- **Using accepted codes and methodologies**
- **Will produce results that differ wildly from each other when assessing a common problem**

***Simulation-based* regulatory decisions are risky!**

Institutional modeling differences were revealed through blind predictions





….and many more!

- **Time limited on how many examples to present**
- **Others, I can't speak to because of nondisclosure agreements**
- **I'm approached with anecdotal stories every time I give this message**

Accept it!

Analyst-to-Analyst (A2A) variability is the norm, not the exception

What can you do to minimize the risk?

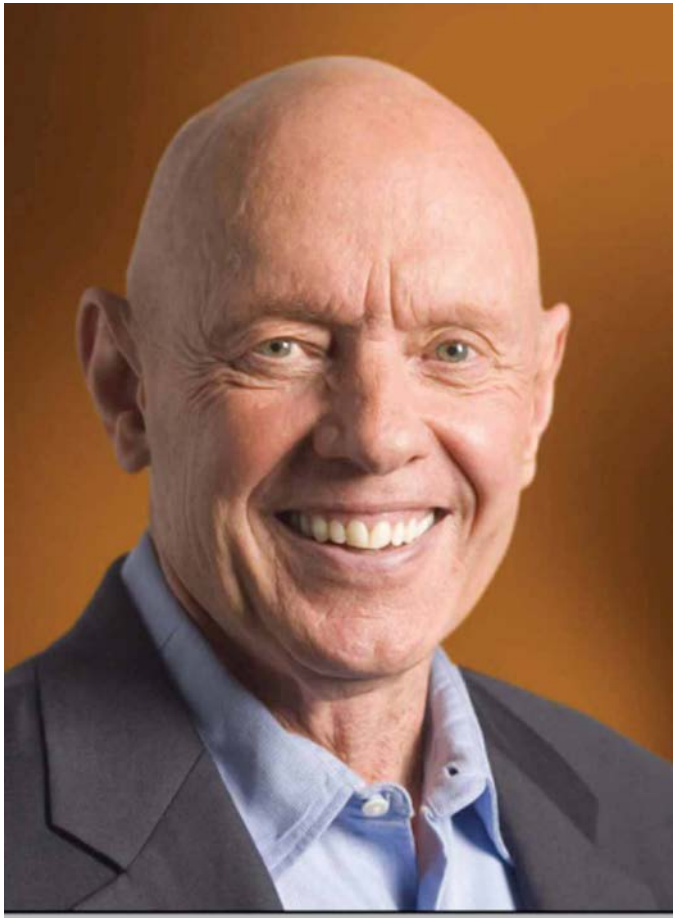
Develop a culture of predictiveness as a necessary element in managing the risk

Simulation Results

Credible predictions provide evidence of completeness and correctness, communicated in a forthright and understandable manner

- What are the physics and material models that are needed and what are your capabilities?
- What is the geometric fidelity in the model and why do you think it is adequate?
- Why do you believe that code bugs or algorithm deficiencies are not corrupting simulation results?
- Why do you believe that numerical errors are not biasing simulation results?
- How accurate are the models and how applicantion relevant is the validation database?
- Do uncertainties in simulation results impact decisions and what are the dominant contributors to uncertainty?

Let's continue with the end in mind
…….. and cautionary tale #2



Steven Covey, *Seven Habits of Highly Effective People*

- **Simulation results are not as predictive as you think**

We'd like to think that our simulations will ….

Converge to the correct answer for the intended application

Code Verification:

Converge to the correct answer

Solution Verification:

Converge for the intended application

Model Sustainment:

Reproducibility of application results over time

- Mission relevant acceptance tests
- Usage-based verification

Quality Code Development:

Ensure correctness/robustness/stability during code development

- Code development managed to accepted SQE standards
- Regression (and Unit) Test Suite (RTS)

But we use commercial codes ….

Converge to the correct answer for the intended application

Code Verification:

Converge to the correct answer

- VERification Test Suite (VERTS) mapped to capability matrix and maintained as acceptance tests
- Training for new staff

Solution Verification:

Converge for the intended application

- Building blocks: CVER, component, subsystem maintained as acceptance tests
- Common strategy for dealing with non-convergent and/or calibrated features and capabilities

Model Sustainment:

Reproducibility of application results over time

- Mission relevant acceptance tests
- Usage-based verification: NQA-1, feature coverage, application-specific line coverage, RTS+VERTS mapping to application PIRT

Quality Code Development:

Ensure correctness/robustness/stability during code development

- Code development managed to accepted SQE standards – look for independent certification of SQA: NQA-1, ISO, CMMI, etc
- Regression (and Unit) Test Suite (RTS): line/function coverage and nightly diff's

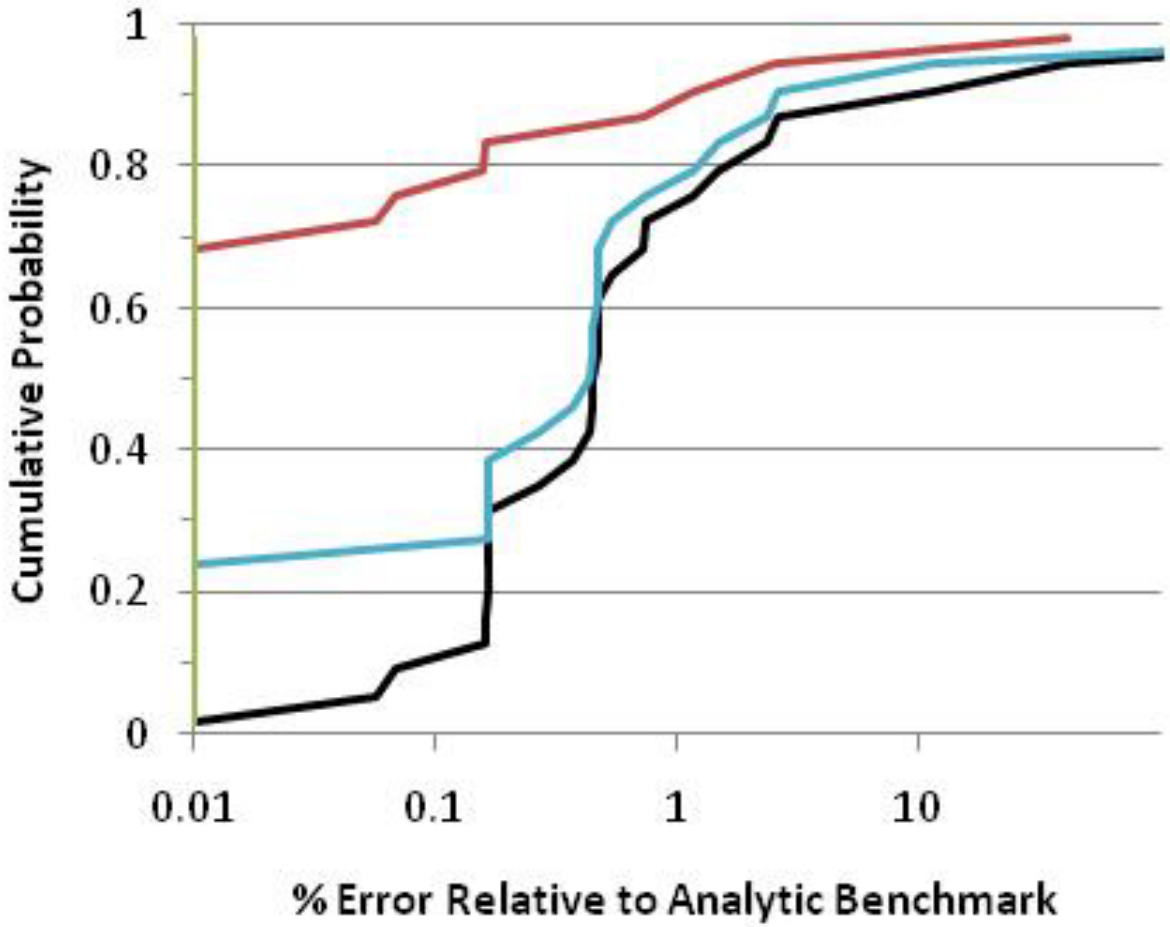
Mathematically ill-posed code
"features and capabilities" are not predictive

- Element death or element merging
- Spot welds
- Mass scaling
- Switch functions
- Discontinuous flow regime maps and closure laws
- Mesh dependent sub-grid models
- Contact

Avoid if possible; and if you can't,
acknowledge and develop a common strategy to deal with them

Undetected code or algorithm deficiencies affect simulation results whether you know it or not

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



27 "typical" bugs intentionally seeded into code used to solve simple problem with analytic benchmark

All Bugs

Solution Inspection

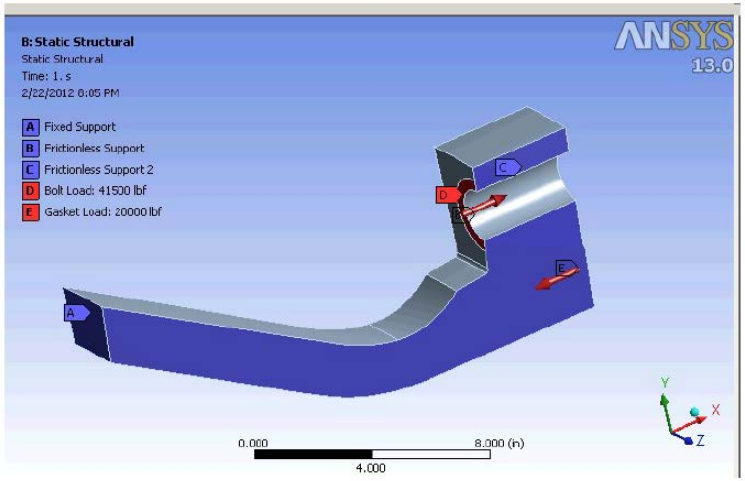
Solution Verification

Order Verification

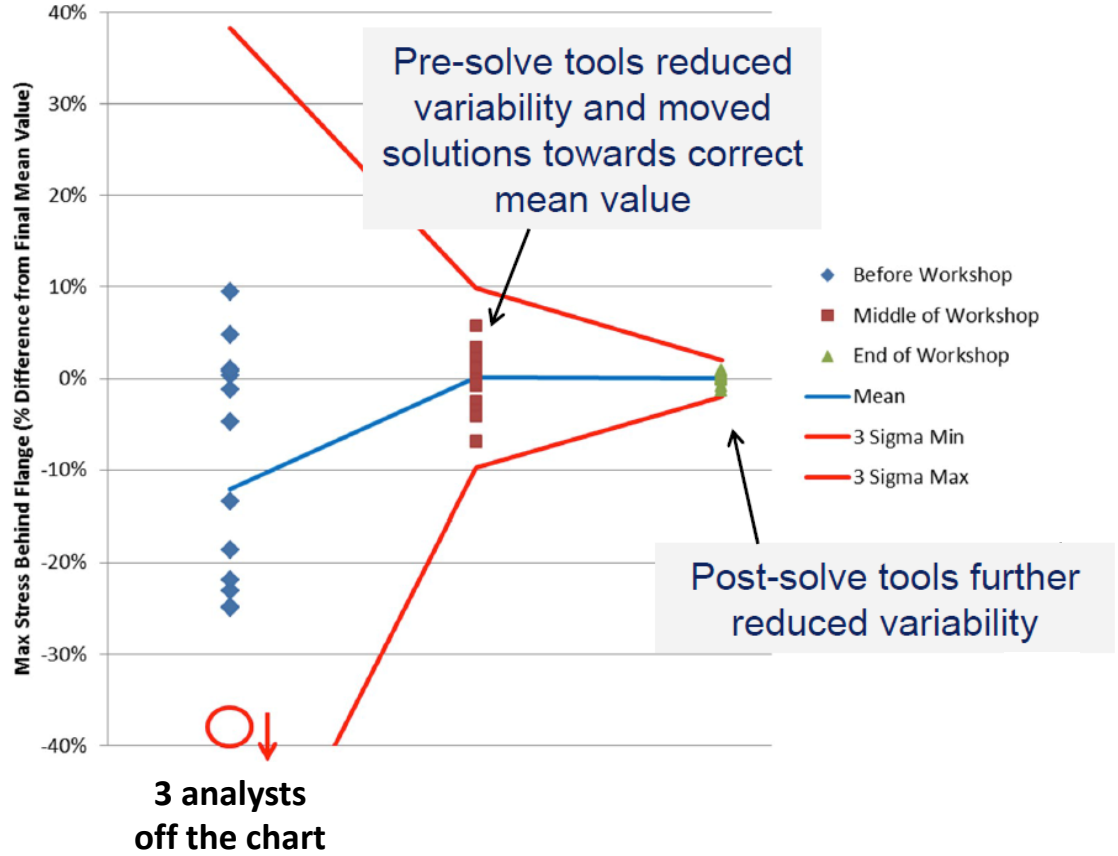
Quantifying convergence behavior is a powerful way of finding code bugs

Training helps reduce A2A variability

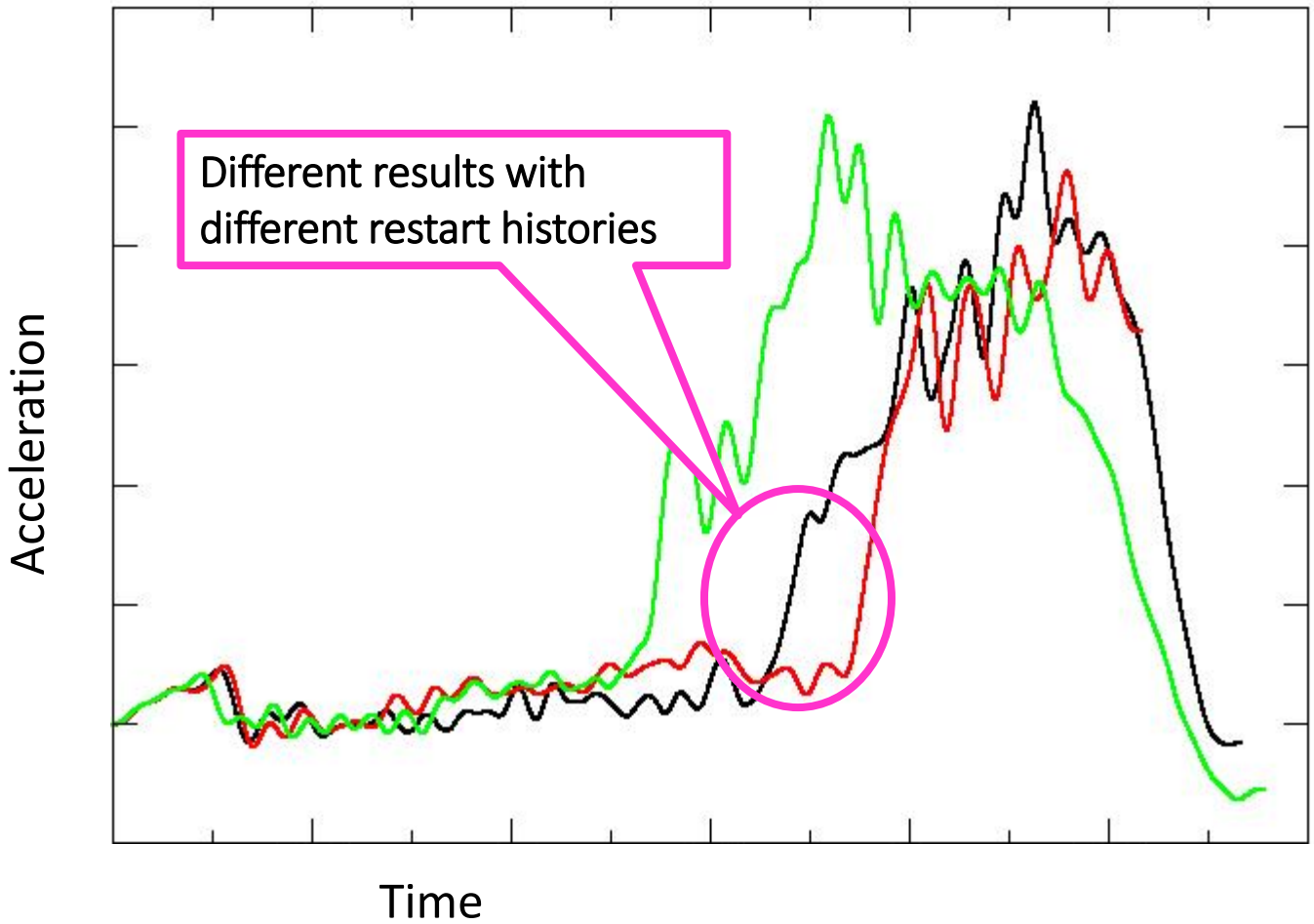
G. D. Westwater, *Variation Studies on Analyst Contribution to FEA Stress Analysis Uncertainty*, ASME V&V Symposium 2014

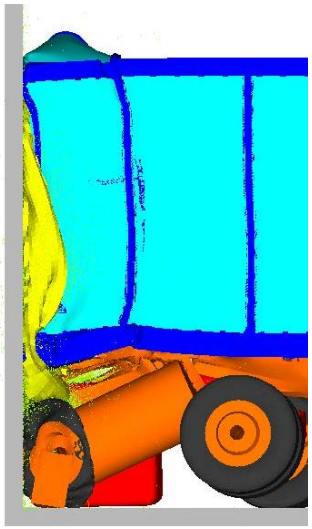


- **Variability study conducted as part of training**
  - **14 participants, 3 countries**
  - **Wide range of experience levels**
- **Before training, analysts asked to mesh and solve the model**
  - **All aspects except mesh predefined**
- **Solve again once pre-solve element quality metrics taught**
- **Solved again at class conclusion using all skills gained**



Different restart histories, processor counts, platforms, operating systems, or code releases can produce different results

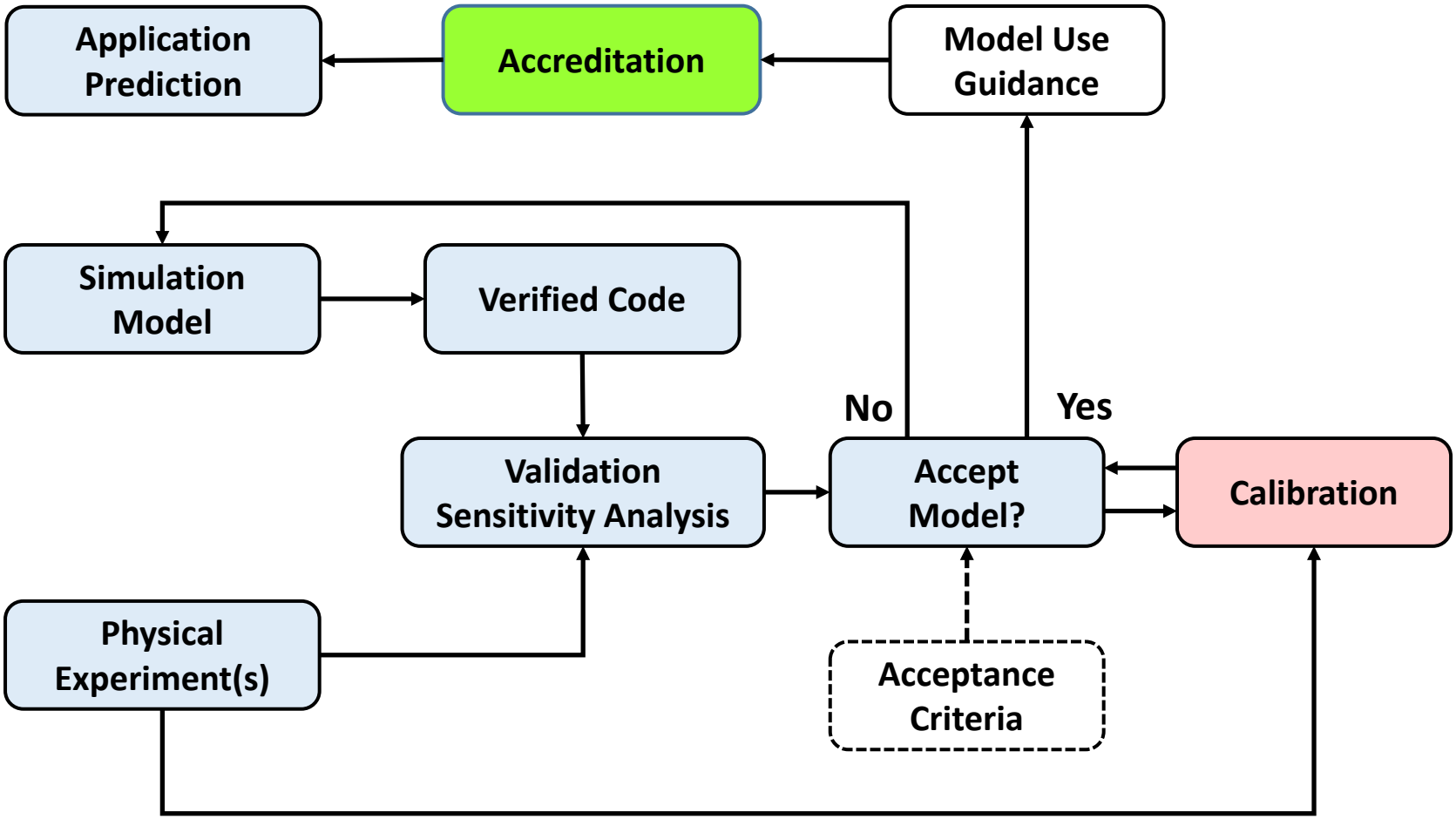




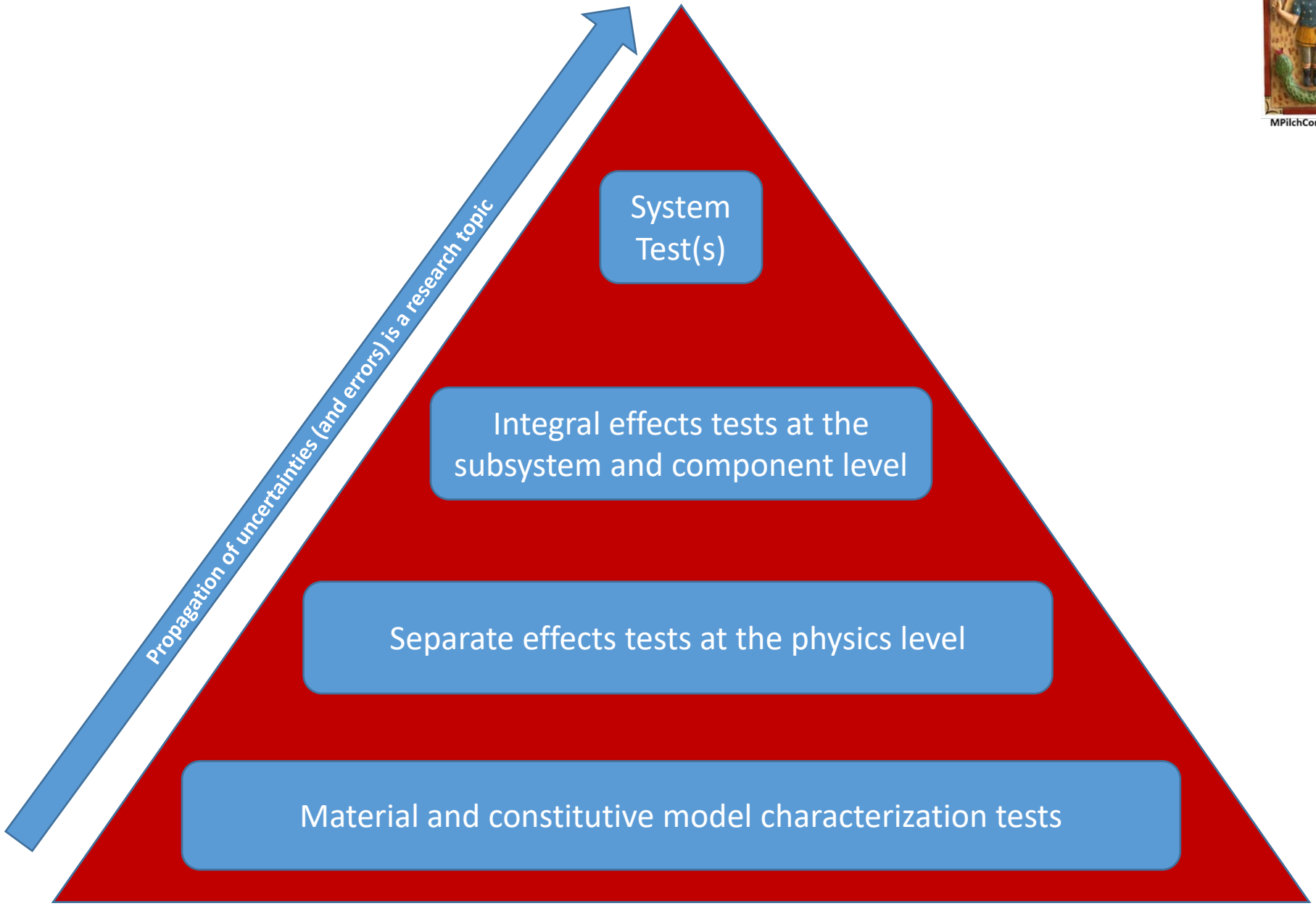
Sometimes regulatory decisions can be effected

Model, verification, validation, calibration, accreditation, and prediction: what's what

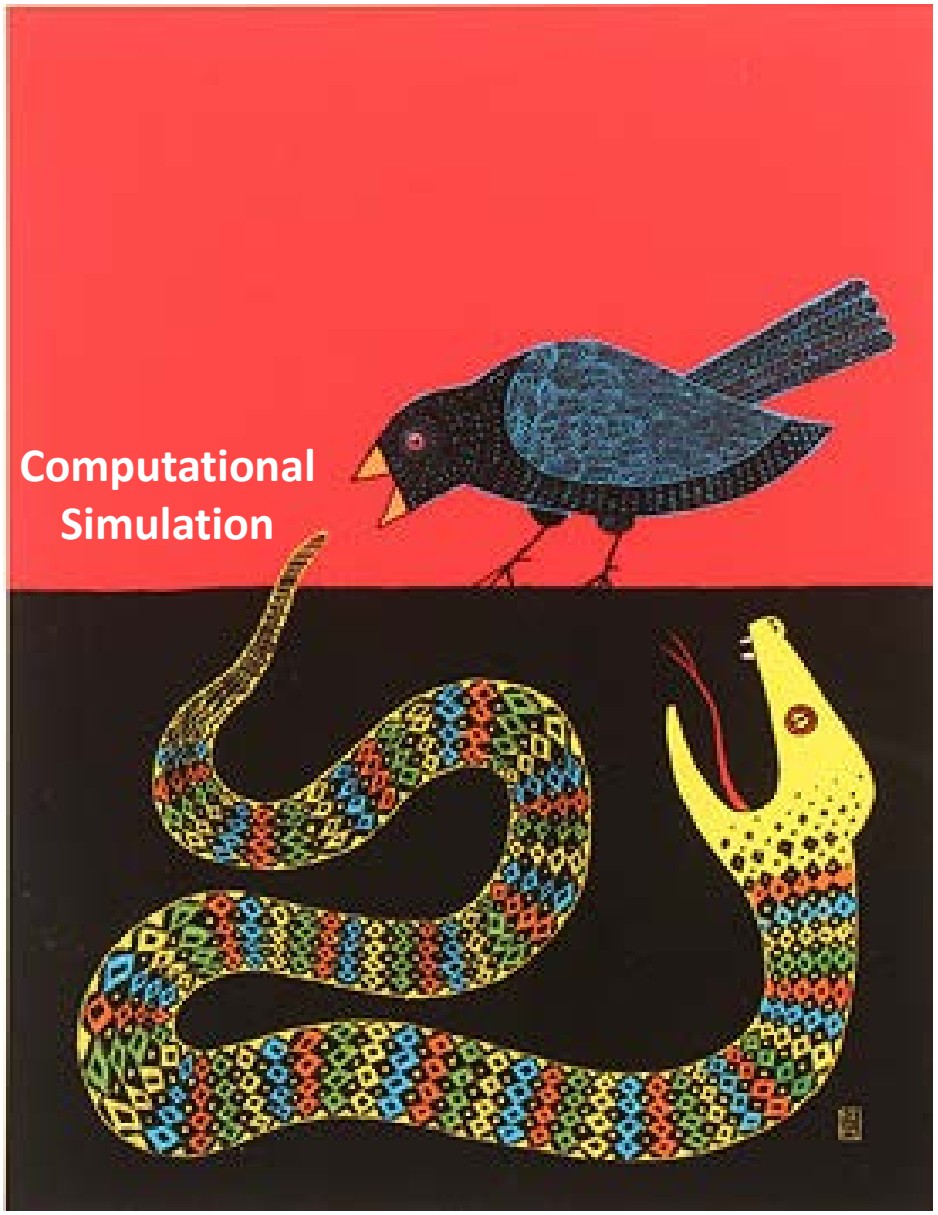
T.G. Trucano et. al., *Calibration, validation, and sensitivity analysis: what's what*, Reliability Eng & Sys Safety, Vol 91, Oct-Nov 2006



Anchor models to a hierarchy of validation tests



It's risky to eliminate all testing, especially for new or radical designs



Avoid:
Surprise + High Consequence

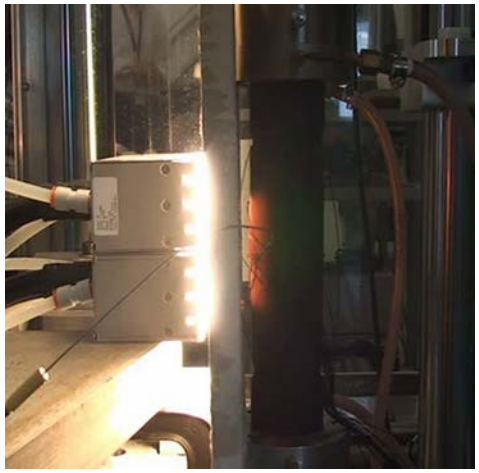
Allow Nature the opportunity to reveal its unknown-unknowns

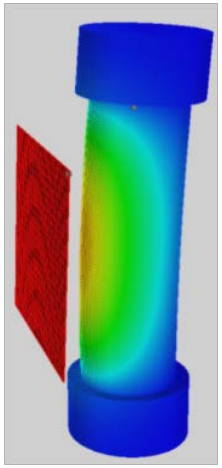
- **Think simulation-informed, not simulation-based decisions**
- **Integration of testing and simulation is often the best strategy for regulatory decisions**
- **Some unknown/unknowns are really should have been knowns – Organizational Memory and Peer Review can be very important**

Trust but verify before you validate or make any other prediction

It's a delivered capability, just use it

Pipe bomb experiments



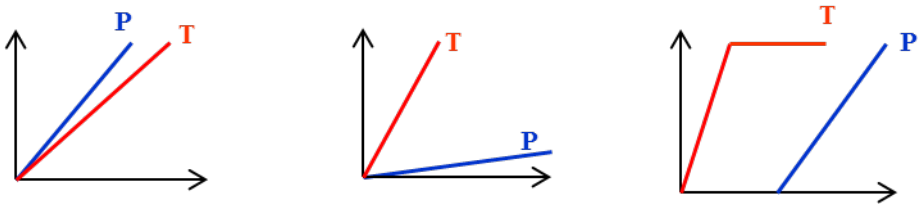


Successful delivery of coupled thermal-mechanical "capabilities"

- Demonstration calculations looked good!

Oops! We learned after delivery

- Coupled thermal-mechanical calculations were not convergent – code bug corrected
- Code bug discovered and corrected in the failure model
- Solver settings and discretization (space/time) were inadequate



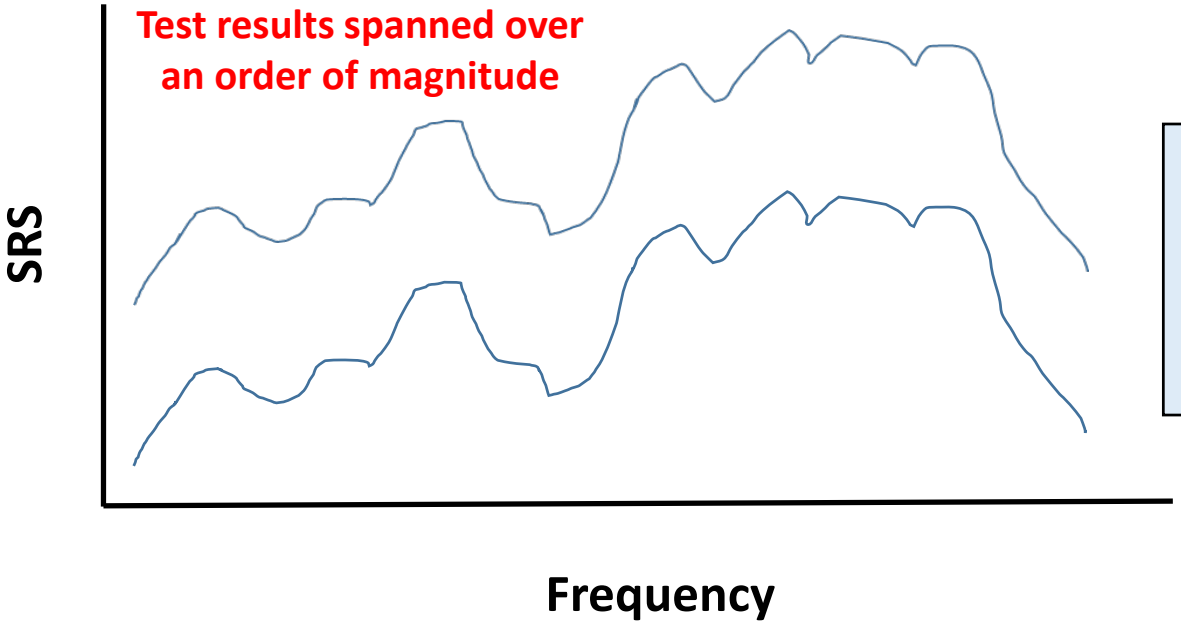
Ramp the temperature and pressure independently to failure

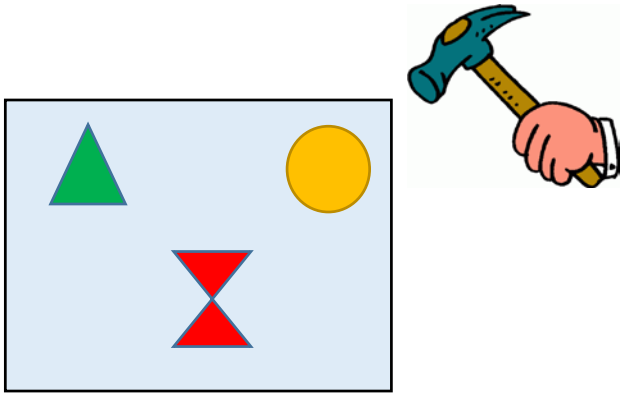
It's not a capability until it has been verified, validated, and documented AND its use in applications is robust

There may be more system-to-system variability in hardware than you think

It's calibrated, just use it

- Disassembly and reassembly with same components or nominally identically replacement components





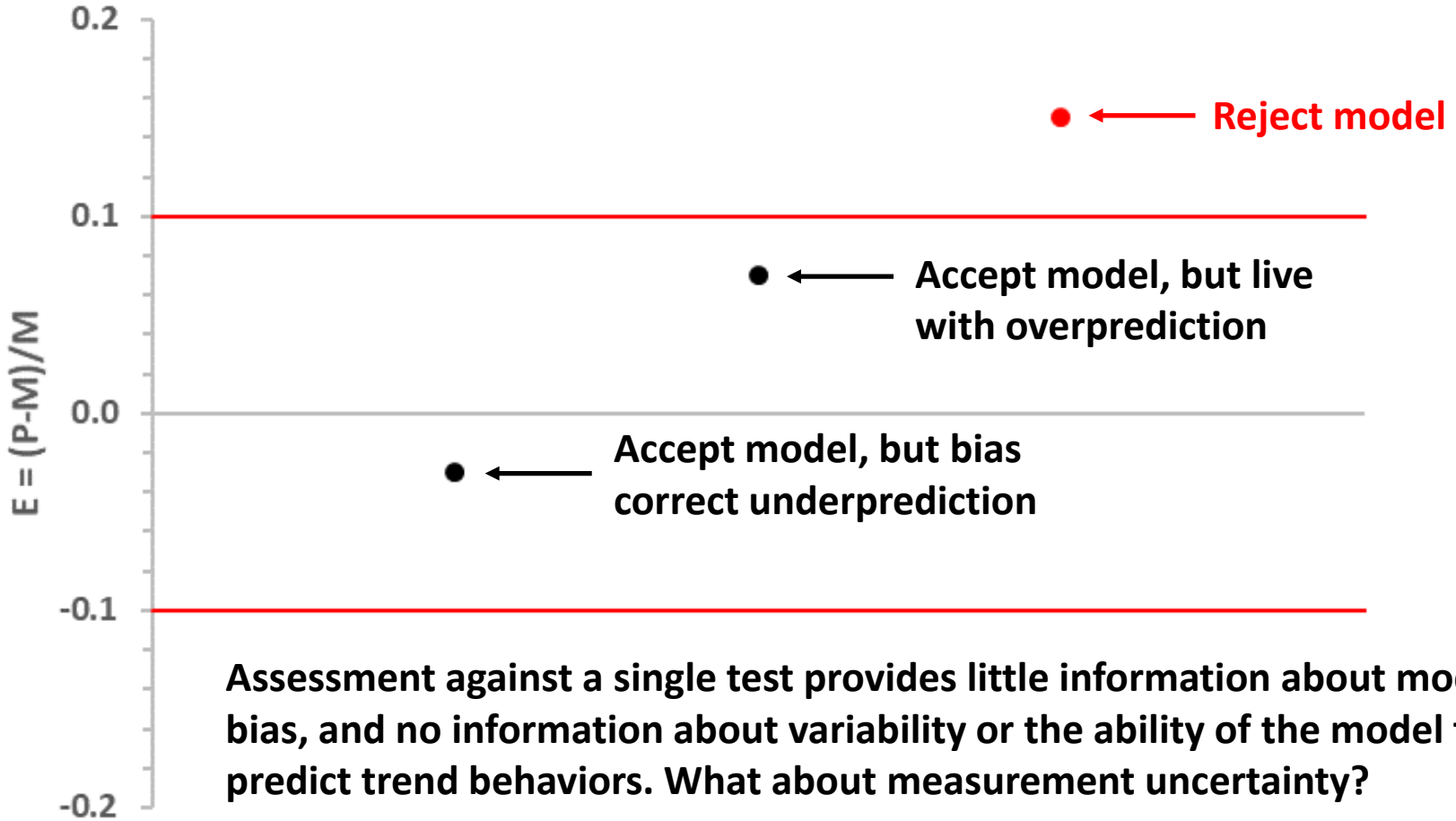
- You can't predict hardware variability – it can only be quantified with ensemble testing
- Calibration to a single test hardwires in a systematic bias of unk magnitude and direction

FAA currently allows certification of a design using simulation if ……

"replace one test with one simulation"

Its validated, just use it!

- **Model is "validated" against a nearby baseline design**
- **Modified requirement is satisfied: 1500 lbf → 1430 lbf**



Assessment against a single test provides little information about model bias, and no information about variability or the ability of the model to predict trend behaviors. What about measurement uncertainty?

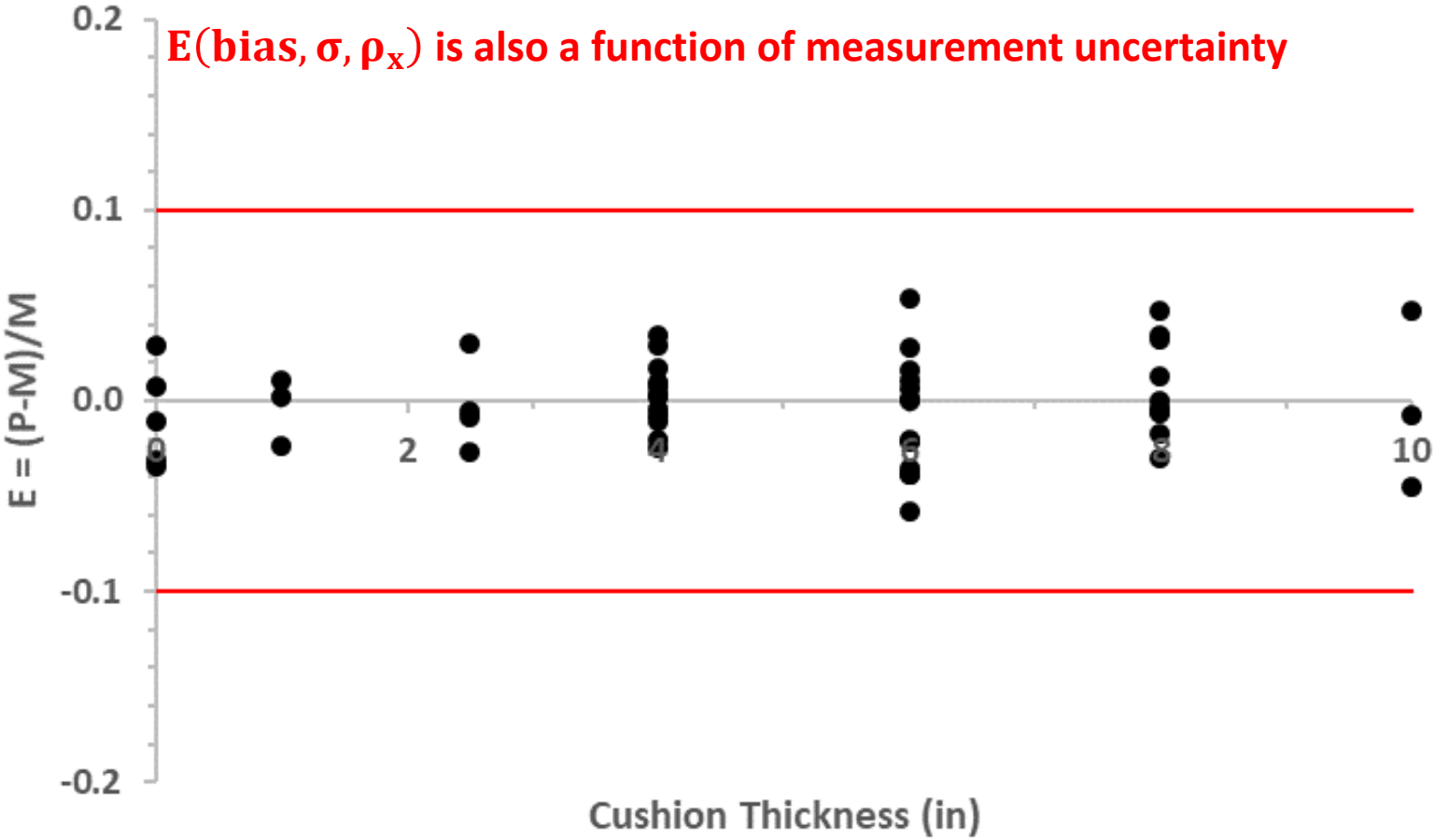
What would it take to certify a new baseline design with simulation alone?

- **Evidence that you can successfully predict a wide range of designs using a *common modeling approach***
  - **VALidation Test Suite (VALTS) under configuration control**
  - **Simulation capability is assessed/re-assessed against the VALTS when new simulation capabilities become available**
  - **Relevance of the VALTS to new design can be communicated in an understandable way**
  - **You can provide credible answers to the six questions below the surface of the iceberg**

Quantitatively assess simulation bias, ability to predict trend behaviors, and relevance of the VALTS

Where is the new design compared to the existing VALTS?

Its validated, just use it!

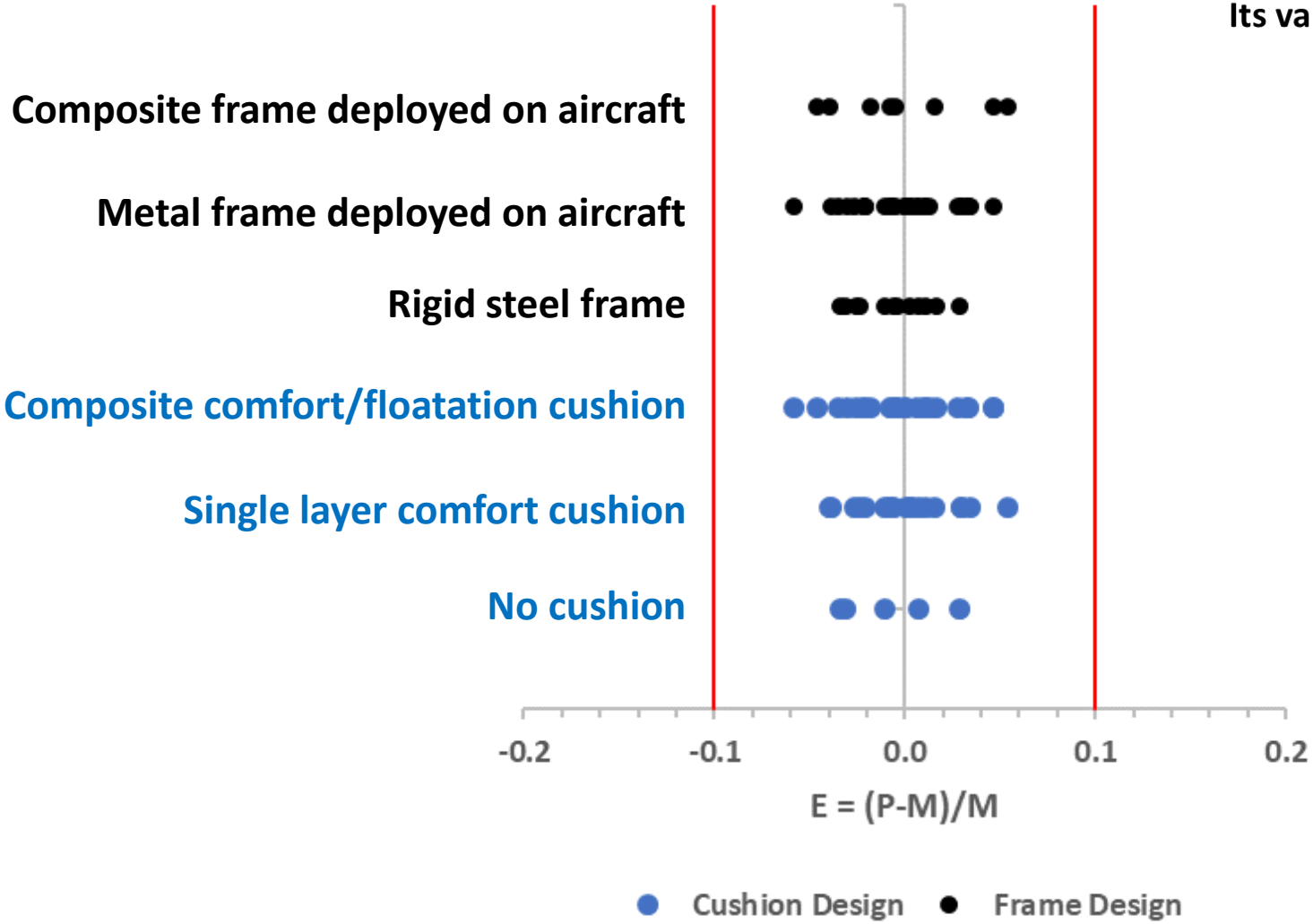


Quantitatively assess simulation bias, ability to predict trend behaviors, and relevance of the VALTS

Where is the new design compared to the existing VALTS?

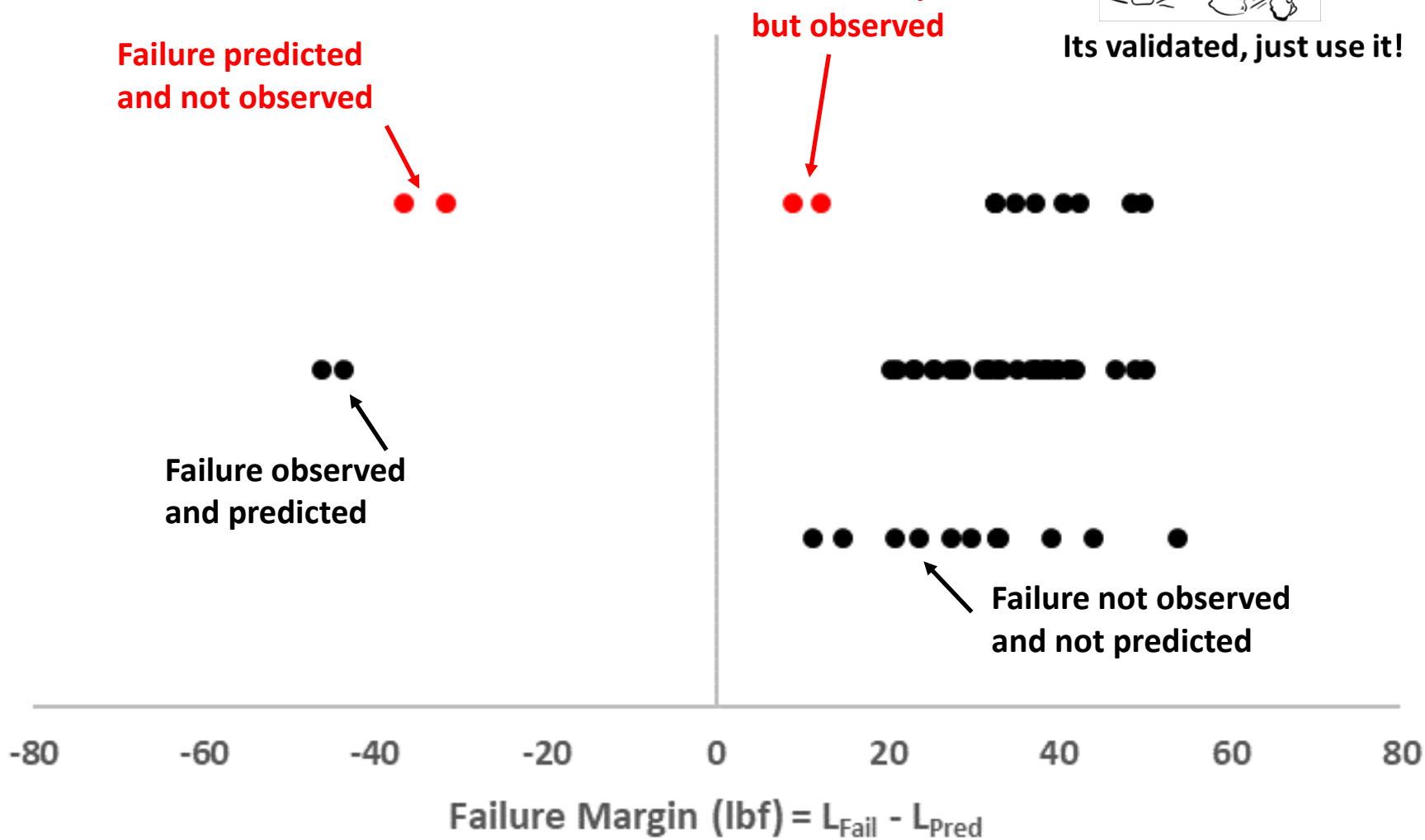
Its validated, just use it!

$E(\text{bias}, \sigma, \rho_x)$ is also a function of measurement uncertainty

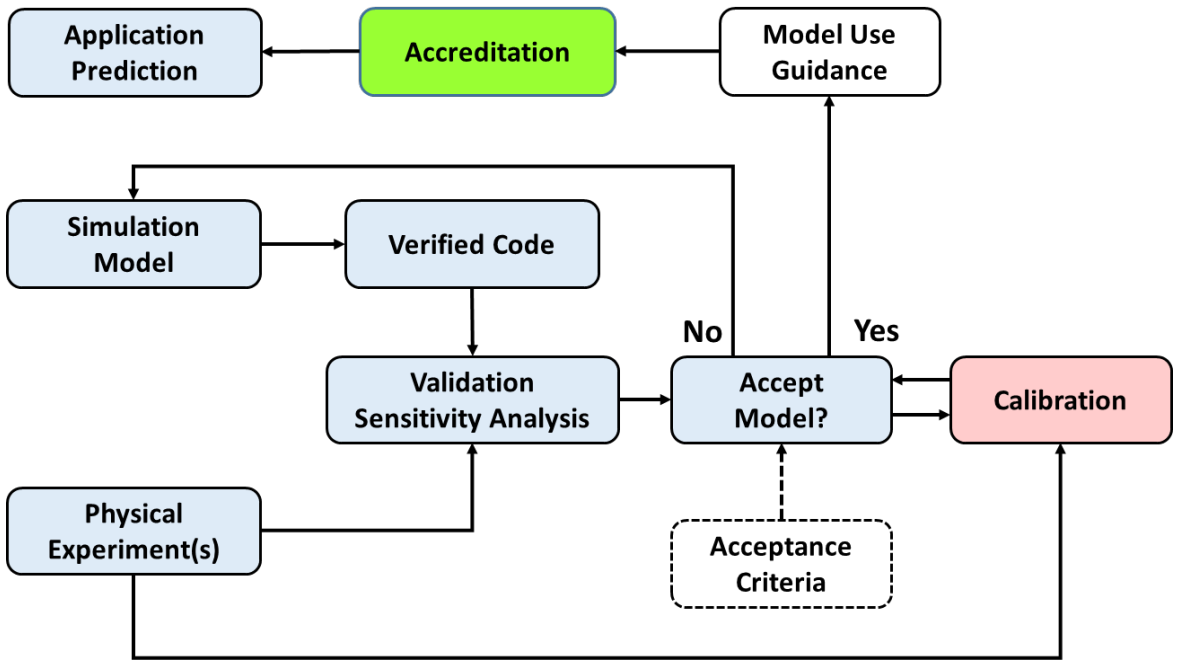


Performance aside, why do you believe that your new design will not fail?

Its validated, just use it!



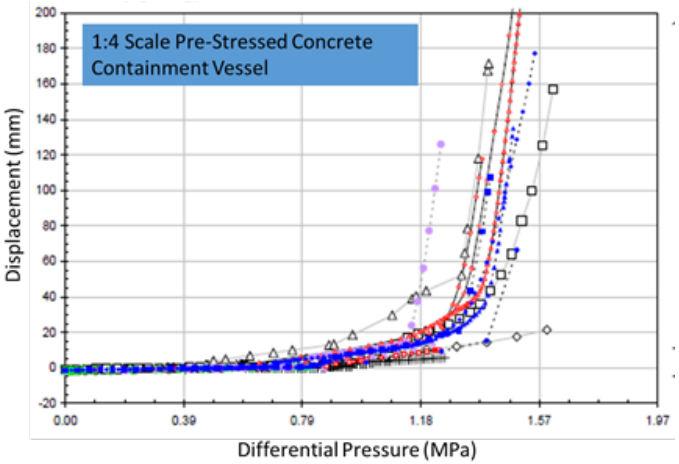
…But what if the validation assessment shows the model is biased or fails to predict trend behaviors?



- **Improve the model or validation data**
  - **Better constitutive models**
  - **Make sure numerical errors are not polluting results**
  - **Make sure there is no bias or errors in measurements**
- **Calibrate to VALTS or full system tests – not just a single test**
  - **Bias correct model predictions**
  - **Provide model use guidance e.g., FAA rules for bias correction**
  - **Calibrate parameters to correct trend behaviors**

Is calibration the antidote that leads to more accurate extrapolations?

Ok, then calibrate it!



Calibrate to pre-stressed concrete containment

Extrapolate to reinforced concrete containments?

Extrapolate to steel containments?

It certainly removes A2A variability where there is data, but what about when extrapolating?

Expect 17 different groupings of calibration parameters such that extrapolated results will all behave differently

Calibration is often a necessary evil that sweeps *all* problems under the rug even if you don't know they exist

We calibrate to empirically compensate for ……

- Unknown model parameters not accessible through SETs,
- Temporary surrogate for missing or unknown physics and material models,
- Inadequate representation of environments,
- Inadequate geometric fidelity
- Code bugs and algorithm deficiencies,
- Numerical errors,
- User errors,
- Hardware has large variability and you don't know it,
- Test measurements are biased, and you don't know it,
- We get different answers with different restart histories, processor counts, platforms, operating systems, or code releases
- We just don't understand why the test or model behaved the way it did

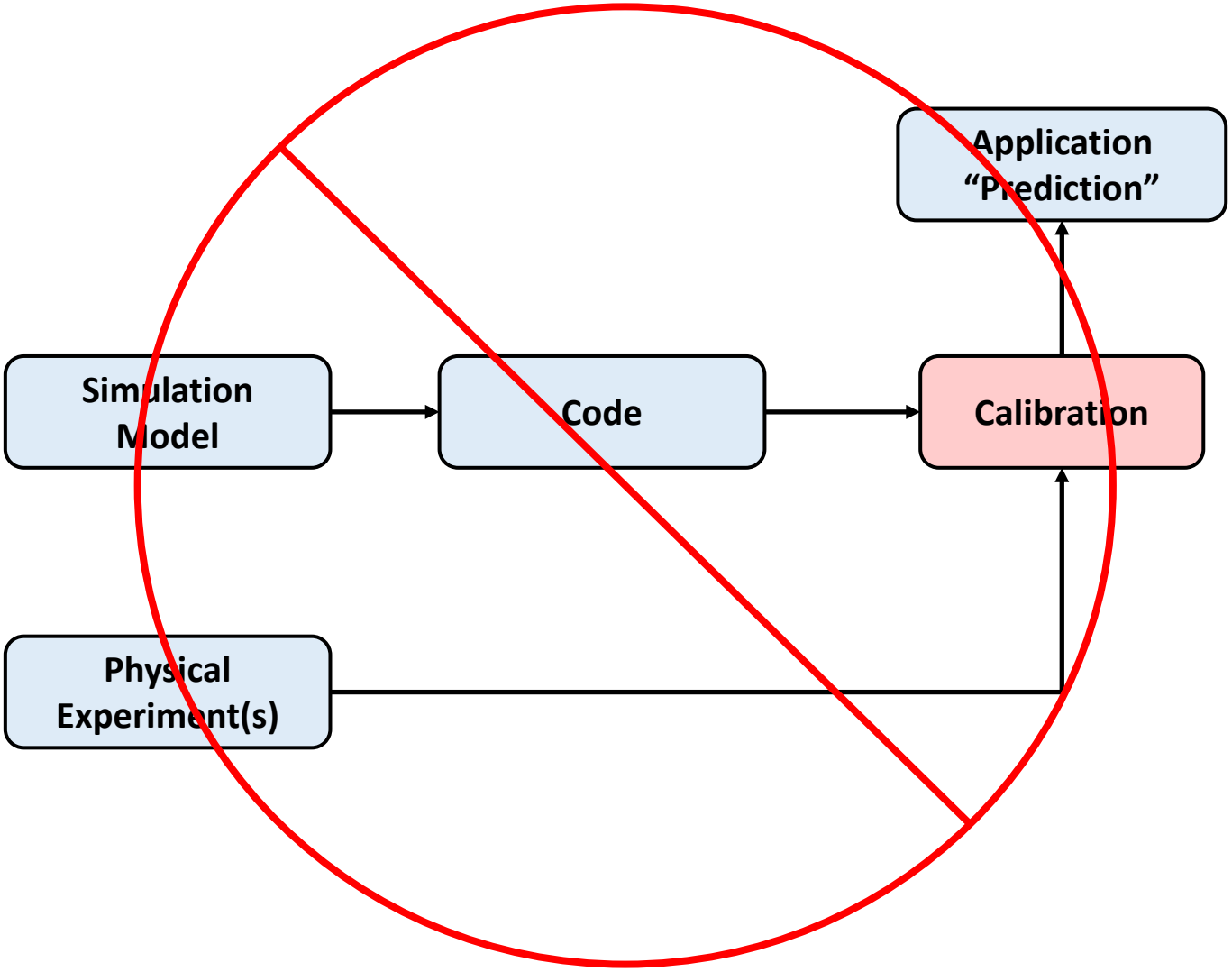
Caution!

- We usually calibrate because models are demonstrably wrong!
- We are walking on thin ice when we calibrate parameters in models that have nothing to do with why the discrepancy exists
- …… and then we extrapolate expecting more accurate "predictions" !?!?!

Responsible calibration will …

- **Acknowledge the model is calibrated**
- **Minimize the number of calibration parameters**
- **State what parameters are calibrated and why**
- **Minimize as many sources of error as possible and acknowledge the rest**
- **Be anchored to an ensemble of tests, rather than a single test, using common calibration values**
- **Assess only nearby conditions holding all calibration parameters constant**

Calibration cannot be viewed as an adequate substitute for verification and/or validation



Regulatory decisions should be *informed* by margins and uncertainties

Pilch, Trucano, and Helton, *Ideas Underlying Quantification of Margins and Uncertainty (QMU)*, SAND2006-5001, September 2006

Supporting evidence of credibility

+

$$\text{Predicted Lumbar Load} = \frac{\text{Model}}{[1-\mathrm{E}(\text{bias},\sigma,\rho_X)]} < \text{Requirement?}$$

- Distribution
- Guidance for bias correcting model
- Includes hardware variability and model form uncertainty
- Representative of VALTS, not a specific design
- Policy or guidance for failures to predict trend behaviors

What is your regulatory decision policy?

- Regulate to the mean, informed by uncertainty
  - USNRC has this policy
  - FAA test-based policy for hardware variability
- Regulate to prescribed confidence level
  - Current FAA policy for simulation-based certification

Summary and recommendations

- **Value proposition of Computational Simulation is compelling**
  - **Some combination of faster, better, cheaper, and discovery**
  - **Simulation-informed certification of new baseline seat design w/o hardware specific certification test**
- **Cautionary notes**
  - **Analyst-to-analyst variability is the norm, not the exception**
  - **Simulation results are not as predictive as you think**
- **Manage the risk by developing a culture of predictiveness**
  - **Seek the right answer for the right reason – V&V**
  - **Provide evidence that you can predict performance and failure potential for a wide range of designs – VALTS**
  - **Calibrate responsibly, transparently, and only when necessary**
  - **Provide training in the use of codes AND core V&V practices**
  - **Develop simulation governance that reflects these principles**
- **Think simulation-informed decisions not simulation-based decisions**
  - ***Simulation-based* certification of new radical design is risky – testing is a hedge against potential unknown/unknowns**

Sword: Courage to make risk-informed decisions in the face of Uncertainty

Scales: Weigh the evidence of Verification and Validation

Snake: Metaphor for high consequence decisions that computational simulation informs