Verification and Validation

Timothy Trucano
Optimization and Uncertainty Estimation Department
William Oberkampf and Martin Pilch
Validation and Uncertainty Processes

Sandia National Laboratories
Albuquerque, NM 87185

Phone: 844-8812, FAX: 844-0918
Email: tgtruca@sandia.gov

Numerical PDEs in the 21st century:
can you teach new tricks to an old dog?

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
SAND2006-2723C

V&V - Helping puppies survive …

AKA "Taking the fun out of computational science?"

A Modest Proposal:

The purpose of computing is not insight. ☺

NNSA Advanced Simulation and Computing Program:

The purpose of computing is to provide

"high-performance, full-system, high-fidelity-physics <u>predictive codes</u> to support weapon assessments, renewal process analyses, accident analyses, and certification."

(DOE/DP-99-000010592)

Sandia National Laboratories

ASC is required to deliver high-consequence predictive computational science.

DO PREDICTIVE M&S!

What does "predictive" mean?

Large-scale computational simulations supplement or replace physical experiments and tests for stockpile stewardship.

Use the science and experience of high-consequence system design/performance assessment in rigorous decision environments.

"Risk"-Informed Decision Making

Reliability Dimension

1. What can happen?
2. How likely is it?
3. What are the likely consequences if it does happen?

Confidence Dimension

4. What is your confidence in predicting the answers to the three questions?

Foundation = V&V

SAND2006-2723C

Decisions – what do we mean by "predictive"?

"Predictability" versus "Predictive Science" versus "Predictive Capability"

- Predictability – A technical concept, conventionally arising in the consideration of complex systems. I.e. as in "predict the stability of the solar system" or "predict the evolution of a chaotic system."
- Predictive Science – might just as well be a philosophical hope in the progress of the human condition. How do you measure it?
- Predictive Capability – in particular a computational capability with some (rigorous?) basis for credible interpolation or extrapolation of current knowledge, for example experimental data.

We (ASC) believe that "predictive capability" can be measured, although such capability is always relative to the intended application.

SAND2006-2723C

Can you trust the calculation?

Three reasons you may not wish to bet your life on a calculation:

1. Wrong physics (validation)
2. Wrong numerics (verification)
   - Wrong math, algs, software
   - Lousy numerical accuracy
3. Wrong use of the results* (decisions)

(* Especially scary!)

SAND2006-2723C

Verification and Validation (V&V) Definitions

Verification – Are the equations solved correctly? (Math)
Validation – Are the equations correct? (Physics)

ASC:

- **Verification: The process of confirming that a computer code correctly implements the algorithms that were intended.**
- **Validation: The process of confirming that the predictions of a code adequately represent measured physical phenomena.**

AIAA:

- **Verification: The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model.**
- **Validation: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.**

IEEE:

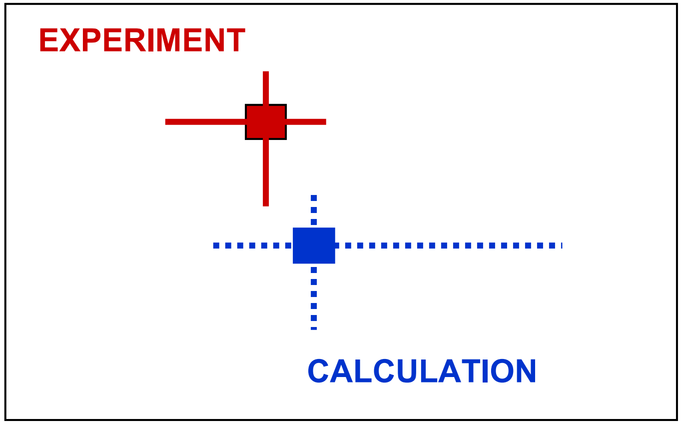
- **Verification: (1) The process of evaluating a [software] system or component to determine whether the products of a given development phase satisfy the conditions imposed at the start of that phase. (2) Formal proof of program correctness.**
- **Validation: The process of evaluating a [software] system or component during or at the end of the development process to determine whether it satisfies specified requirements.**

Verification and Validation (V&V) definitions cont.

ASME: (as of ~ April 2006):

- **Verification**: The process of determining that a computational model accurately represents the underlying mathematical model and its solution.
- **Validation**: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.

Validation lies at the heart of "**predictive codes**" and validation is not easy …



In principle, a <u>simple strategy</u>:

Converge the calculation.

Put in enough physics to insure "agreement" of calculation and experiment.

This is the 1995 charter of ASCI.

- **Experimental uncertainty (variability, bias, diagnostic fidelity) is remarkably hard to quantify.**
- **<u>Quantitative</u> expt-calc differences are uncertain quantities**
- **We aren't converging calculations yet (10 years later).**
- **What ARE the calculation error bars? Why would anybody believe the reported value? (VERIFICATION IS CRITICAL)**
- **How much physics do you need?**
- **How much agreement is good enough?**

Technically speaking, validation is -

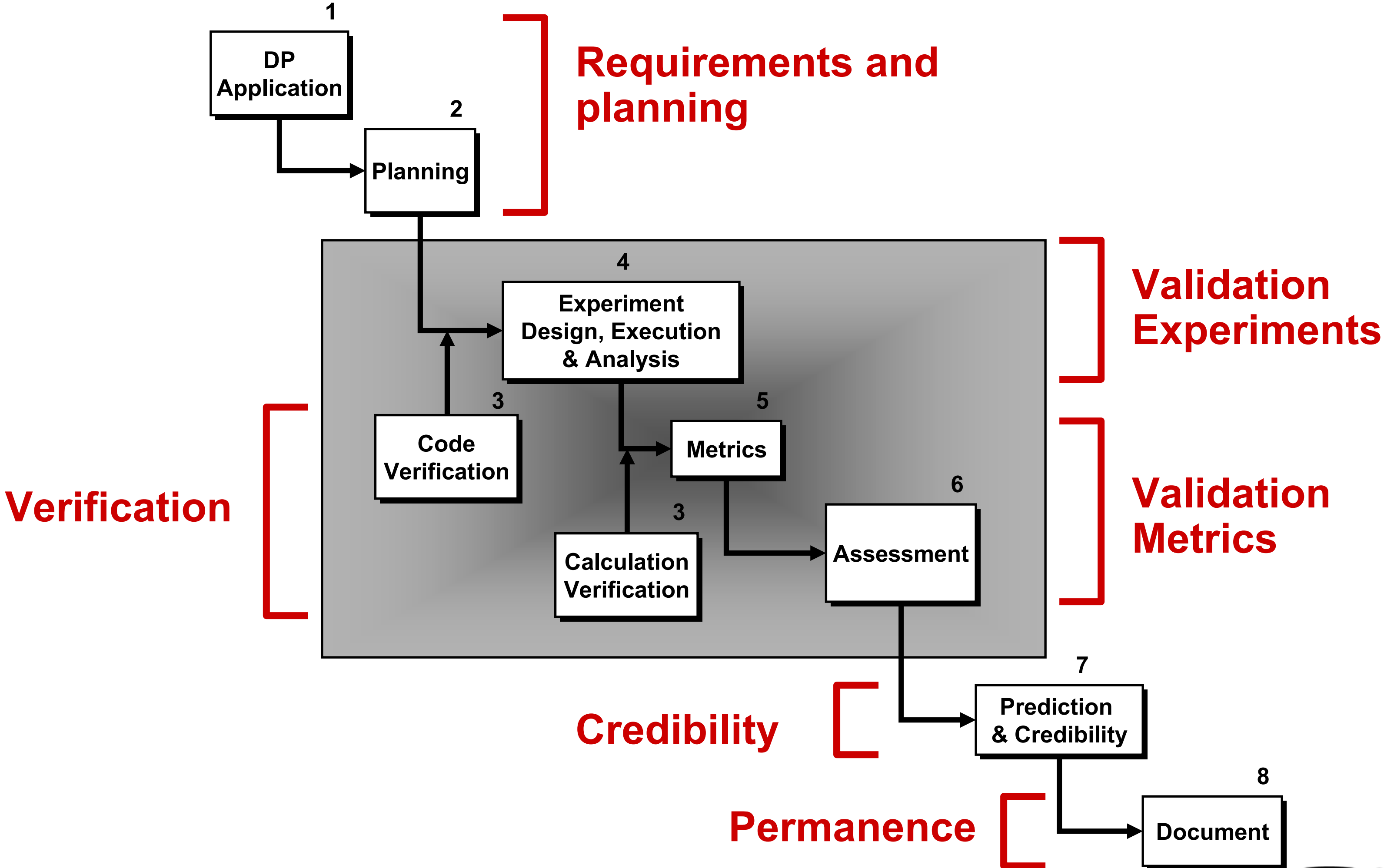
Characterization of the high-dimensional random field

Diff = "Nature – Calculation"

- **Given relatively sparse information**
- **For the purpose of making a reliability statement about "*Calculation*"**
- For example, see Jolliffe and Stephenson (2003), Forecast Verification; Wilks (1995), Statistical Methods in the Atmospheric Sciences
- See Kennedy – O'Hagan (2001), "Bayesian calibration of computer models," J R Stat Soc, Volume 63, 425–464
- Trucano et al (2006), "Calibration, validation, and sensitivity analysis: What's what," to appear in Reliability Engineering and System Safety (SAMO 2004 proceedings)

SAND2006-2723C

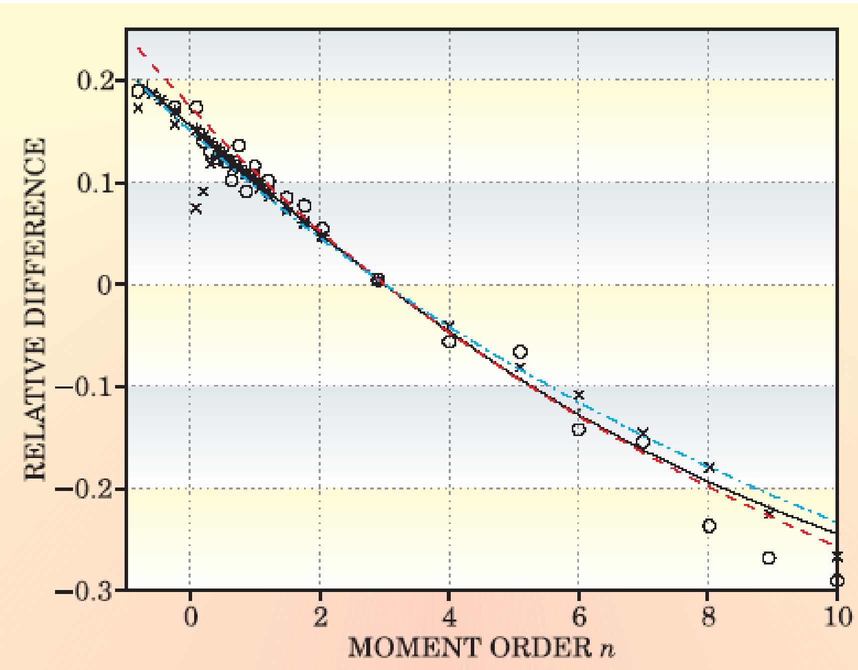
V&V is a methodology.



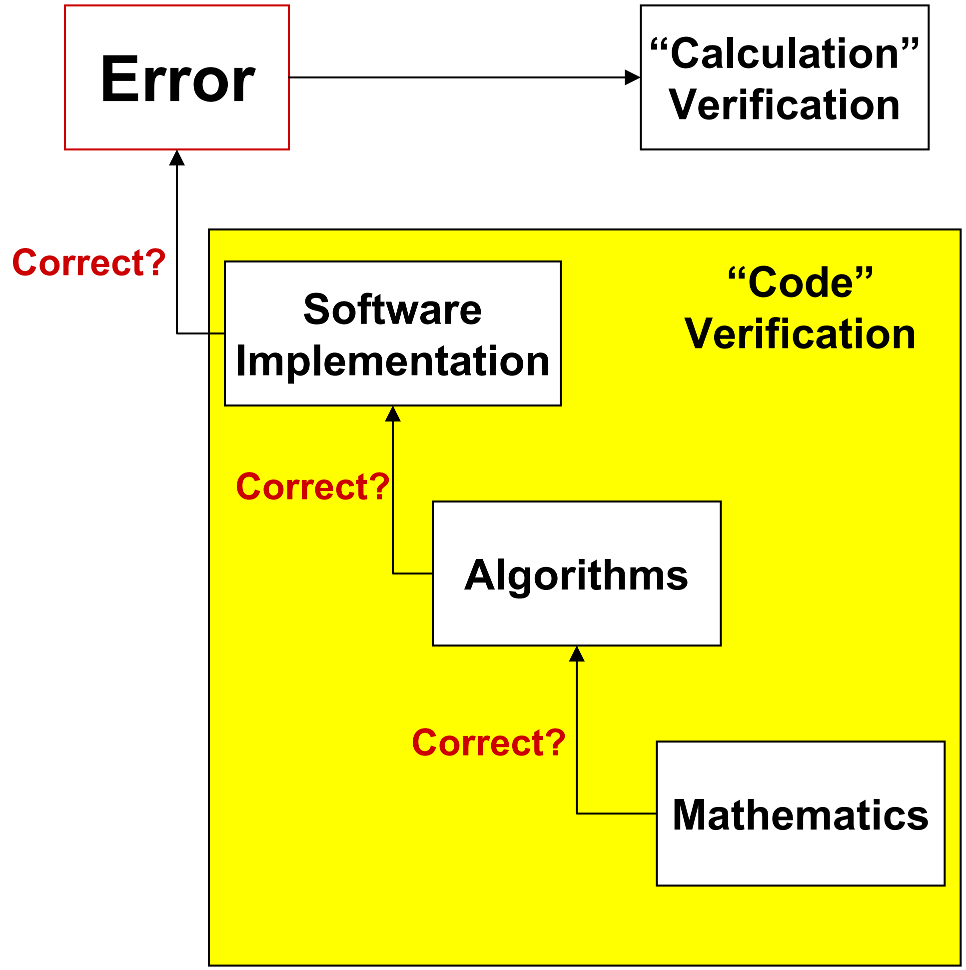
SAND2006-2723C

Consider the following comparison with data:

Where are the error bars?



- What is the computational (crosses and stars) error?
- "Good agreement" with experimental data (circles) does not imply <u>numerical</u> accuracy!
- (A DNS resolution study is present.)



To believe any error statement requires "code verification."

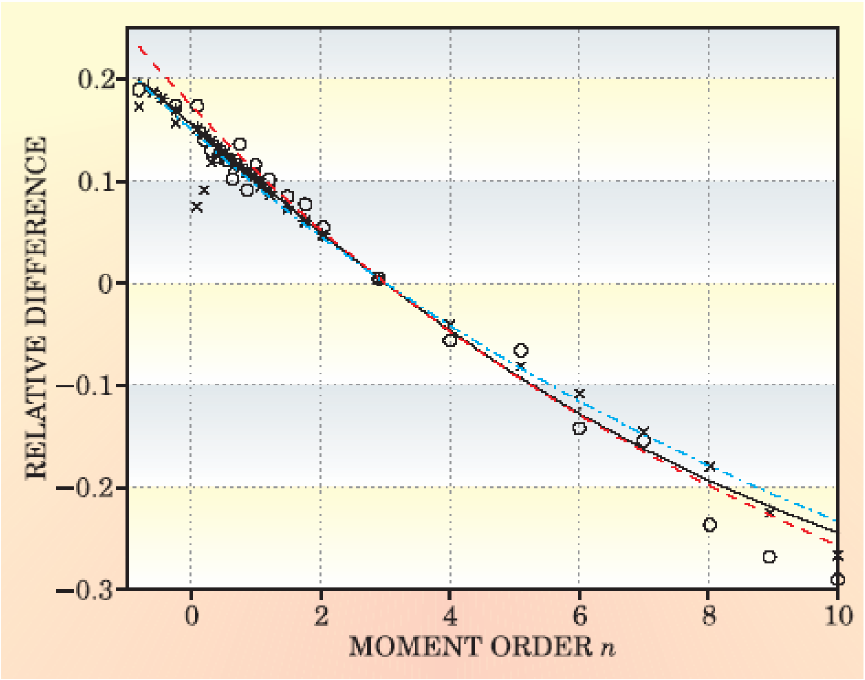
The challenge of verification is VERY HARD – and quite appropriate for this group!

- **Credible error statements rest on a significant code verification foundation:**
  - Proofs that math and algorithms are correct
  - Proofs that the software has no bugs
  - Anything less is an approximation and has epistemic (lack-of-knowledge) uncertainty attached to it
- **Error statements themselves come from a (presently) limited technology base:**
  - Convergence studies (highly empirical – can I take these to the bank?)
  - A posteriori error estimation (not in my favorite equations)
  - Error "models" with intrinsic uncertainty ("the error probably is…")
- **My experience has been that the average scientist who does ASC calculations has no idea where the asymptotic regime for the numerical methods is for the given problems. In fact, not even the experts are clear about it.**

What about validation?

- Validation centers on physical accuracy.
- Uncertainty is prominent in validation.
- Are there ideas that help?

Where are the error bars?

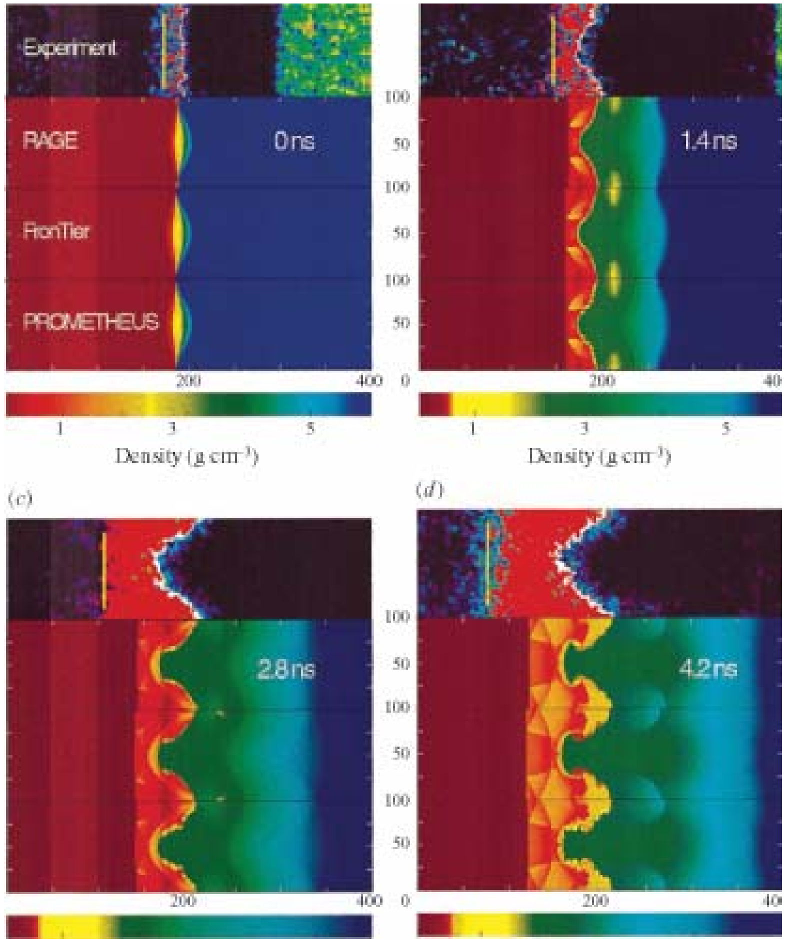


- Assume calculations are converged, say their error bar is the size of the symbol.
- What does the comparison mean?
- THERE ARE NO EXPERIMENTAL ERROR BARS (i.e. experimental uncertainty quantification).

Have we detected a trend?

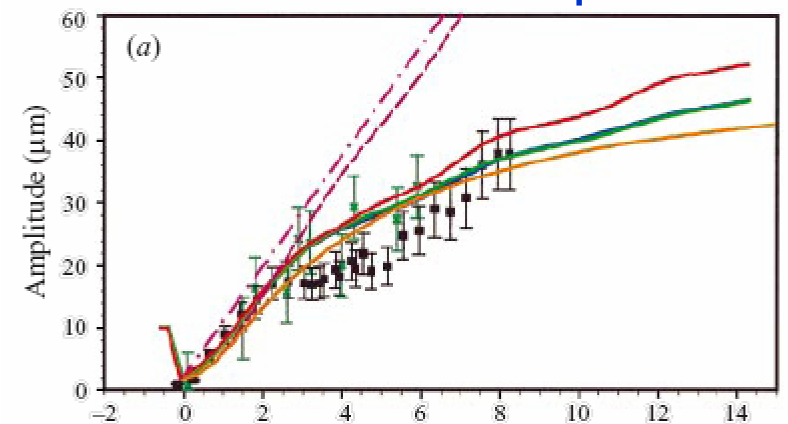
Where is the numerical accuracy estimation and experimental uncertainty quantification in these kinds of comparisons?

Mystery Calculation #1



- "…the large-scale features are measured [experimentally] and they are ***well described*** [emphasis mine] by all three simulations."
- Is it validation?
- Is it verification?

Quantitative growth of amplitude in time. No calculation errors presented.



Mystery Calculation #2

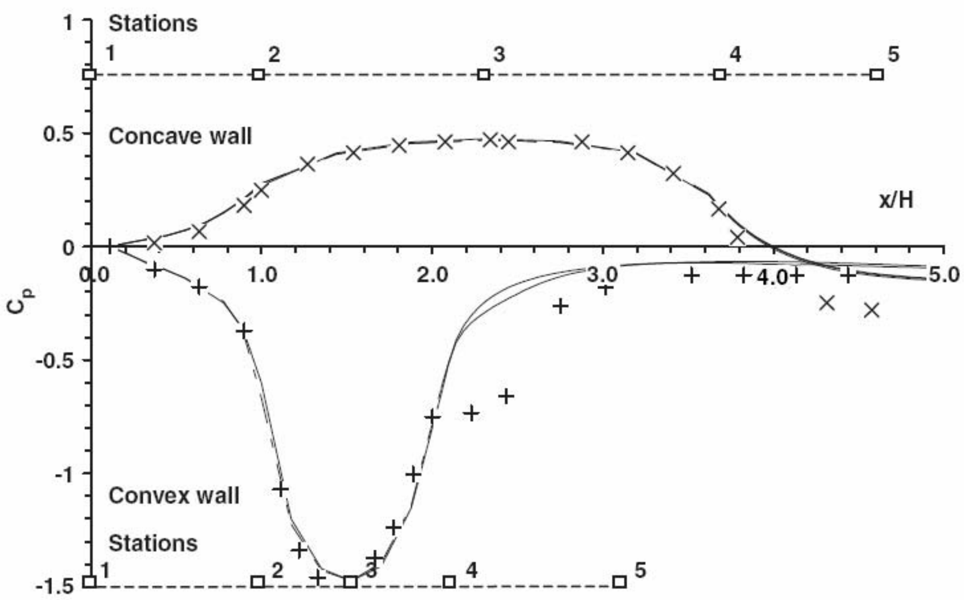


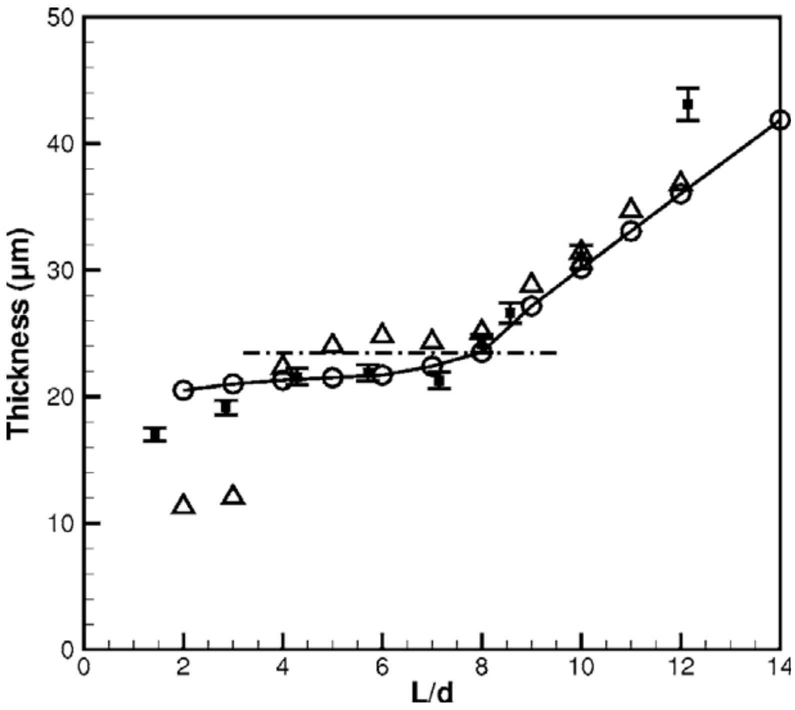
Figure 5. Pressure coefficient on the concave and convex walls of the duct x, + experiment, —RSM, —— standard k–ε, --- realizable k–ε, -·-· RNG k–ε.

- Comparison with experiment:
  - No experimental error bars.
- Grid sensitivity studies reported but not quantified.
- No numerical error quantification.
- Commercial code used.

SAND2006-2723C

Sandia National Laboratories

Mystery Calculation #3

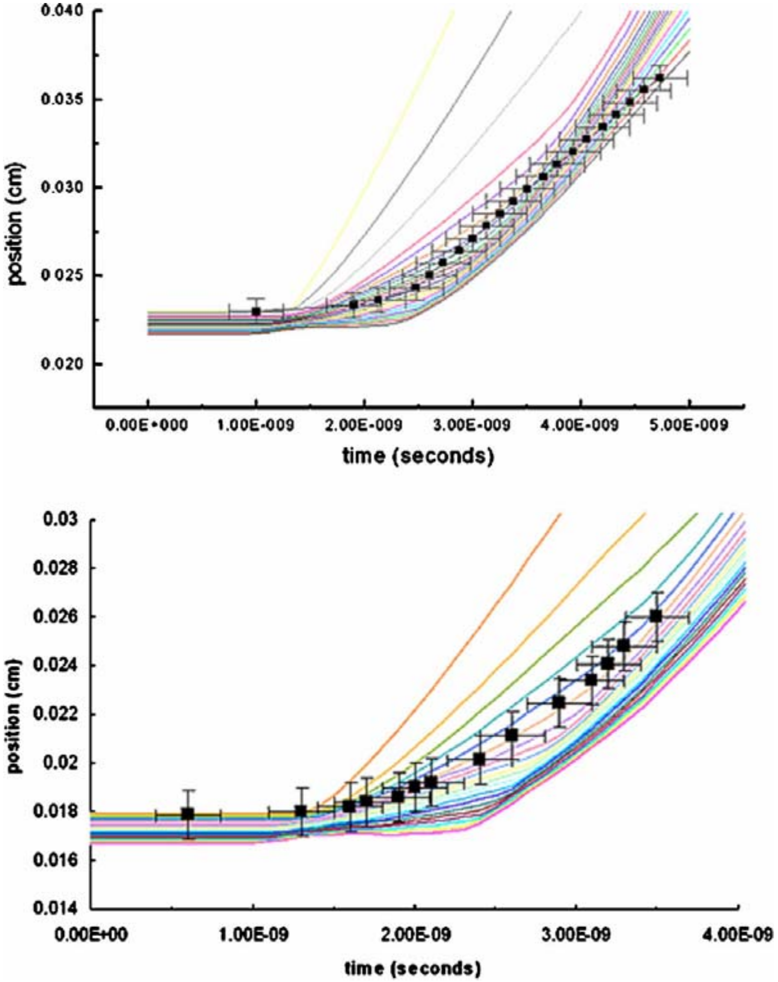


- **Comparison with experiment:**
  - **Experimental error bars.**
- **No reported convergence studies.**
- **No numerical error quantification.**

"These numerical methodologies have already been validated … [references]. The accuracy of the numerical results has been demonstrated thanks to qualitative and quantitative comparisons between numerical simulations and experimental or analytical references."

SAND2006-2723C

Mystery Calculation #4

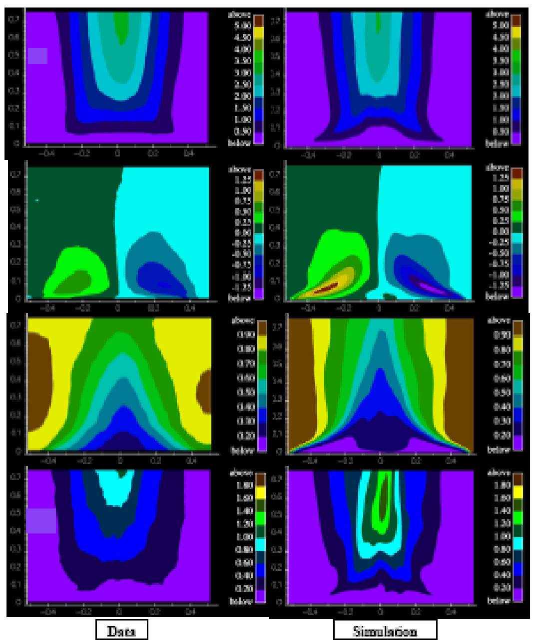


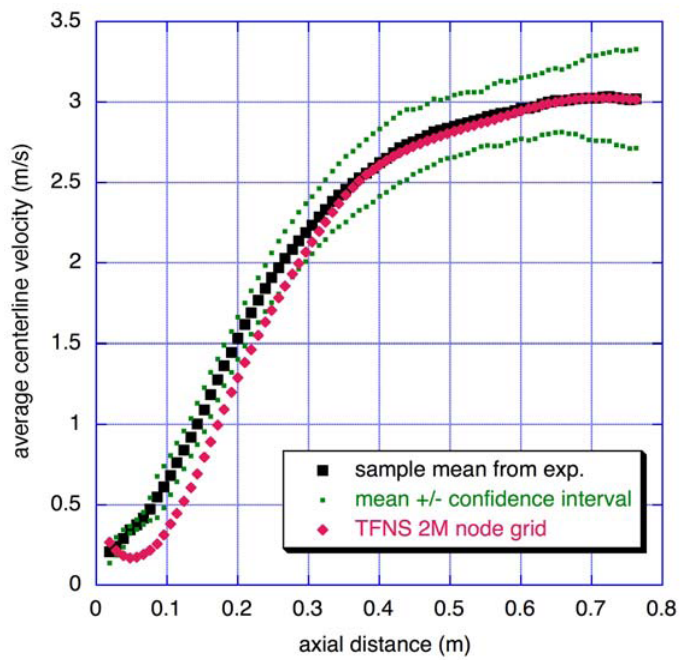
- Comparison with experiment:
  - Experimental error bars.
- No reported convergence studies.
- No numerical error quantification (the different curves are different zones in the calculation).
- The actual purpose of the study was diagnostic development.

Sandia National Laboratories

Viewgraph norm to quantitative differences

Tieszen, et al (2005), "Validation of a Simple Turbulence Model Suitable for Closure of Temporally-Filtered Navier-Stokes Equations Using a Helium Plume," SAND2005-3210.





SAND2006-2723C

Sandia National Laboratories

The wisdom of about ten years of my work briefly summarized:

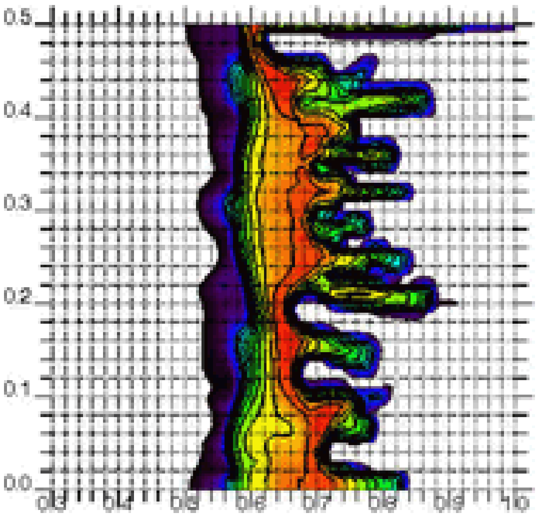
1. Don’t compare with experimental data to assess NUMERICAL ACCURACY.

2. Don’t compare with experimental data that have no error bars (diagnostic resolution, variability and bias).

SAND2006-2723C

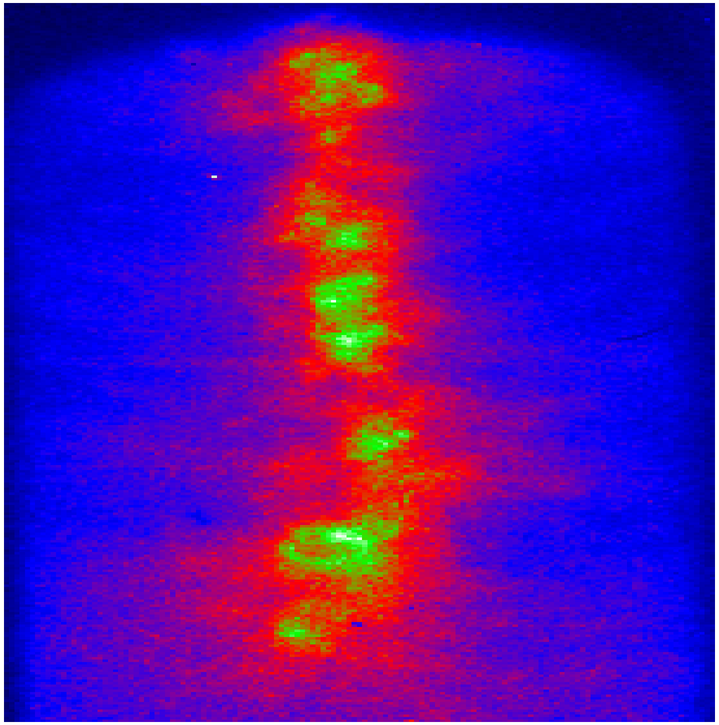
This wisdom may not be much help…

How do we compare this…



2-D r-z shell implosion calculation.

…with THIS?



Experimental spectroscopic image of Z-pinch liner stagnation.

…and claim we know what we are doing?

SAND2006-2723C

Back to decisions – what do we mean by "predictive"?

- M&S typically informs decision making under uncertainty.
- ASC World: "Quantitative Margins and Uncertainty" (QMU)
  - Technical performance margins for engineered systems
  - Uncertainty in the underlying information and characterization of margins
  - Decisions required that reflect this uncertainty
- Many complex factors enter into using M&S in a complex technical endeavor, like Stockpile Stewardship (or climate warming policy).
- Our bottom line: Produce, communicate, and use M&S in the form of:

Best Estimate Plus Uncertainty

Example of BE+U: "Rivers of Blood"

- **Inflation projections from the Bank of England (February 2005 Inflation Report)**
- **Hendry: "Surprisingly, reporting of forecasts alone was the norm for the Bank, even until relatively recently; and it is still the norm among many forecasters." [Hendry and Ericsson, <u>Understanding Economic Forecasts</u>, MIT, 2001]**

Inflation Report: February 2005

Chart 6.3
Current CPI inflation projection based on market interest rate expectations

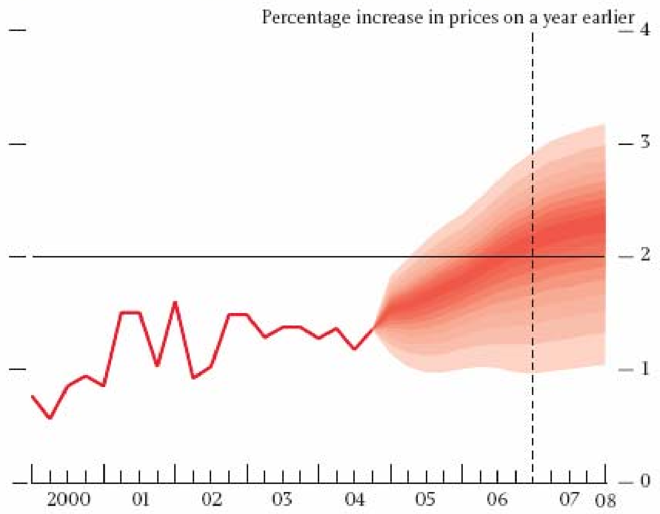
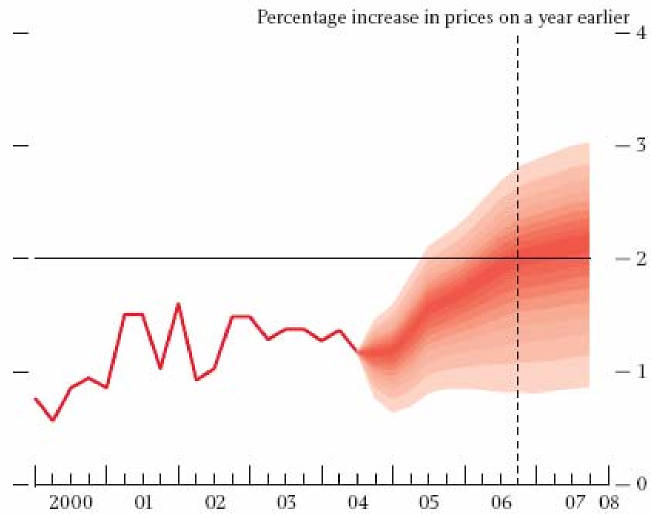


Chart 6.4
CPI inflation projection in November based on market interest rate expectations

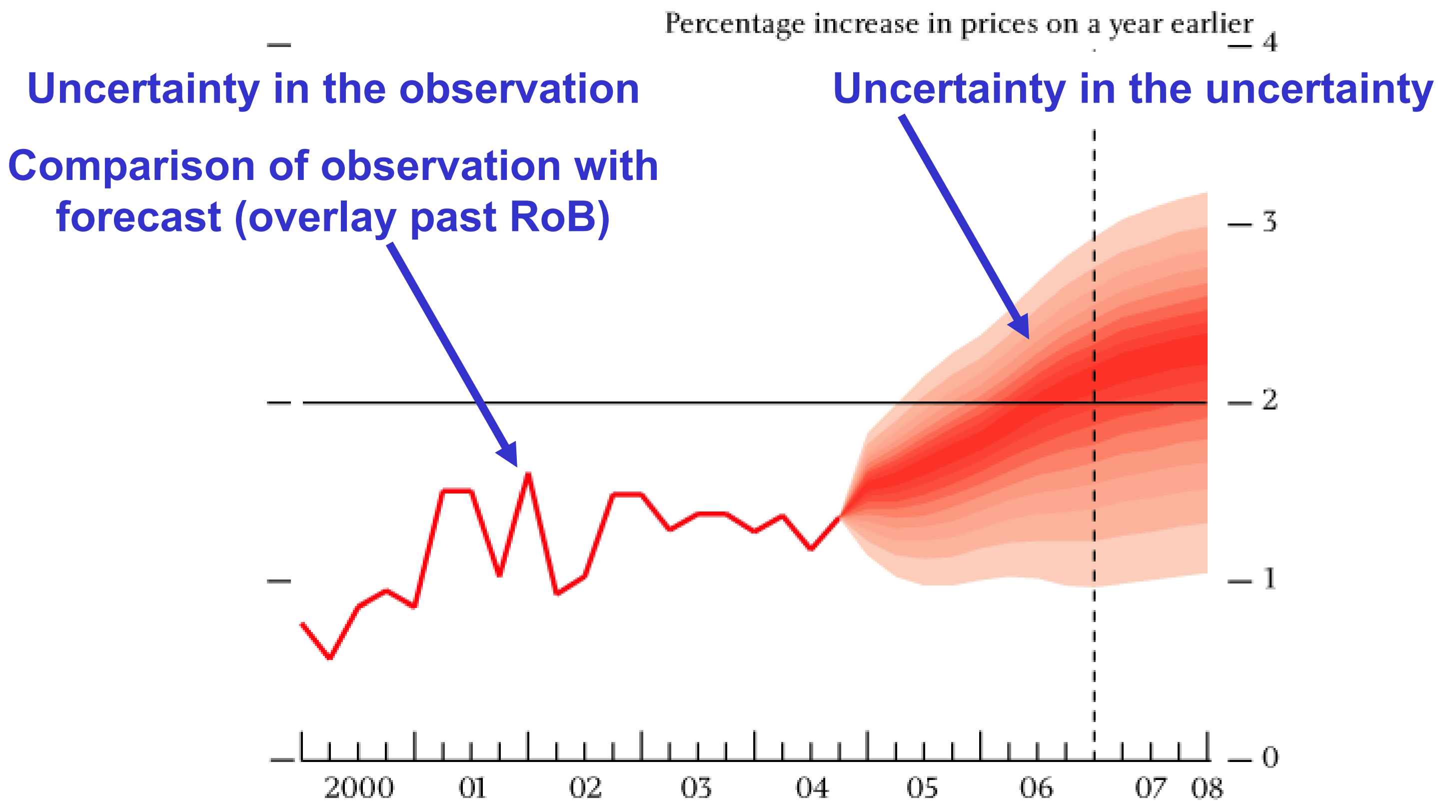


SAND2006-2723C

Sandia National Laboratories

"Rivers of Blood" – note missing elements

- **Where is the comparison of observation with prediction?**

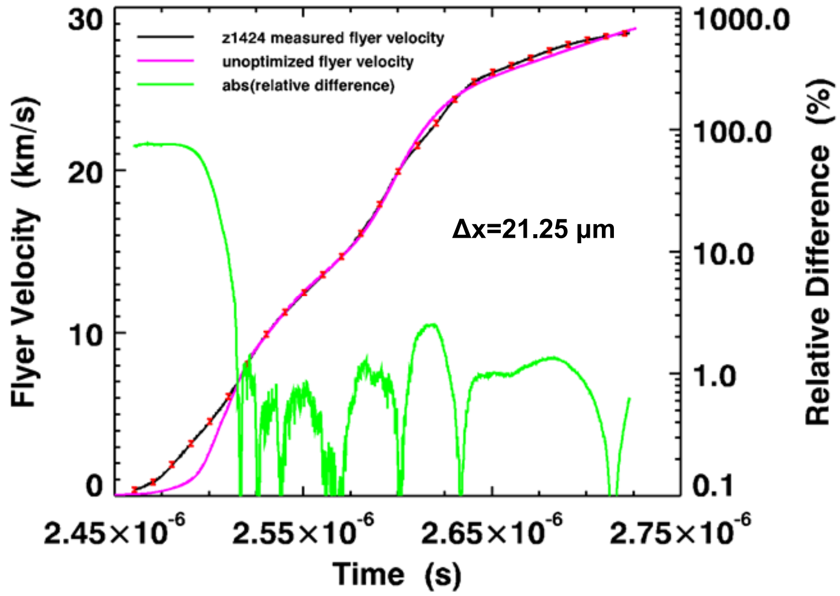


SAND2006-2723C

Lemke example: 1-D MHD driven "flyers"

- ASC V&V milestones have allowed us to put together significant parts of this logic in productive ways (but still not completely!):

z1424 Measured & Unoptimized Velocity

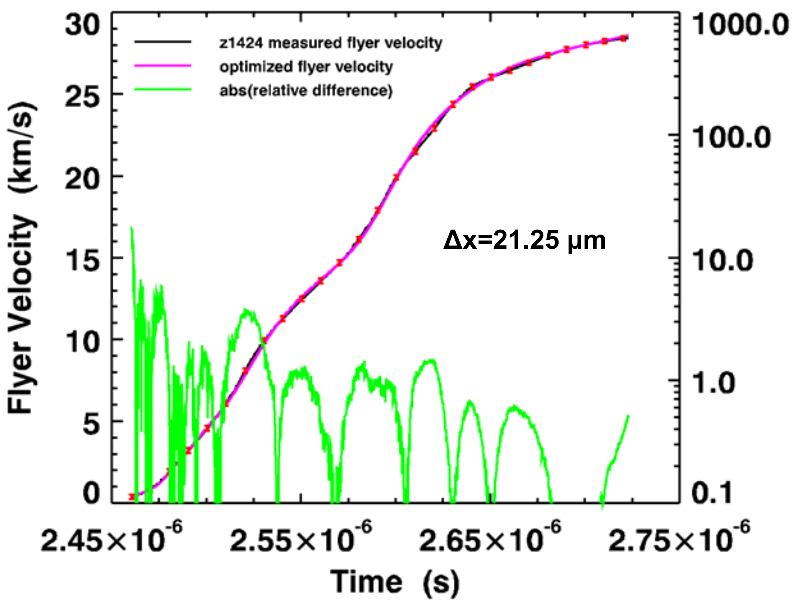


- A 1-D validation study
- Experimental data has error bars.
- Numerical error in the calculation is quantified (not represented here)
- The plot is more than a spaghetti plot.

Lemke example: calibration = BE

- The DAKOTA optimization toolkit can be unleashed once you know what you are doing.

z1424 Measured & Optimized Velocity

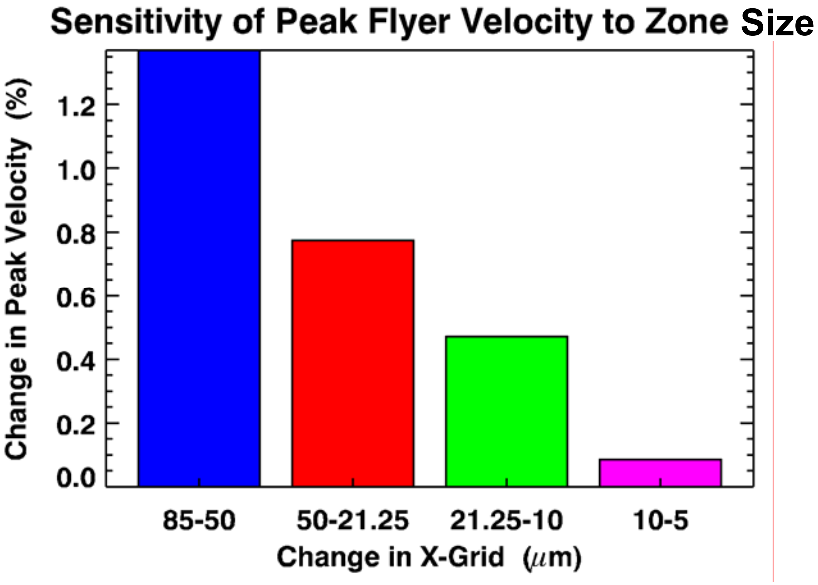


- A 1-D calibration
  - Experimental boundary data for the calculation is not measured accurately enough to improve the previous calculation
- This calibration is now viewed as a way of reducing the experimental boundary data uncertainty.

SAND2006-2723C

Lemke example: Uncertainty quantified

- "Numerical convergence: peak flyer velocity varies by < 0.1% with grid change of 10 to 5 μm."

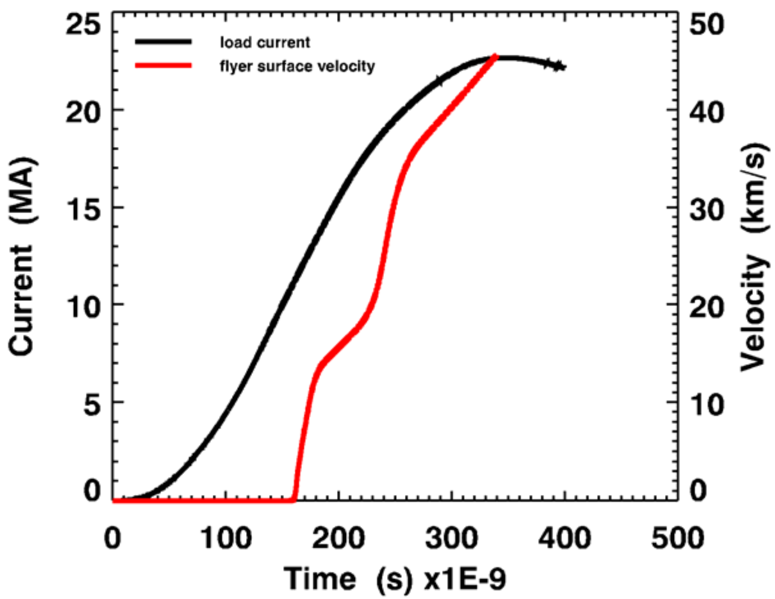


- Flyer temperature & density vs. X show significant change with dx
- Magnetic solve convergence tolerance varied by 4 orders of magnitude
- Sensitivity of flyer velocity and state to EOS: six different EOSs used for aluminum flyer
- Sensitivity of flyer velocity and state to EOS
- Sensitivity of flyer velocity and state to electrical conductivity model
- Radiation + thermal conductivity has no significant effect

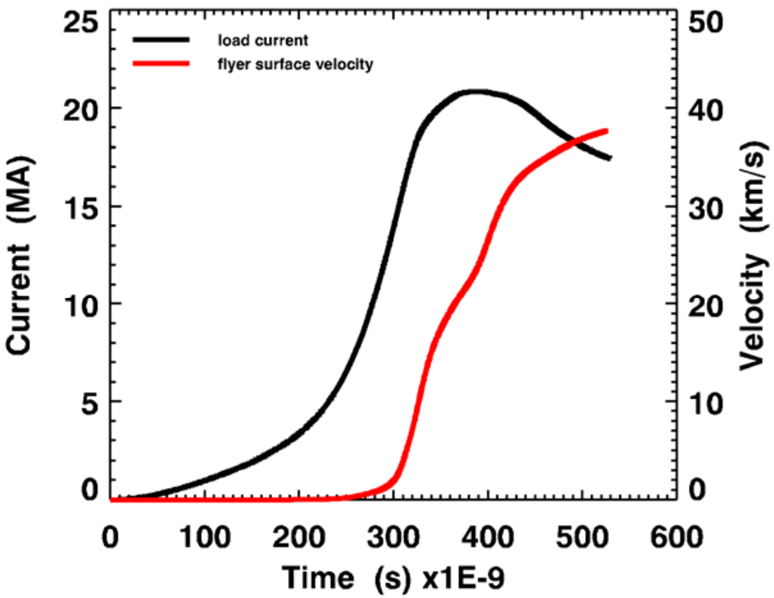
Lemke example: Prediction – ZR Hardware Change Recommendation

- ***"Shaped pulse on ZR must be shortened, or must use thicker flyer and longer flight distance: ensures flyer survival; constant velocity; reduces performance."***

Standard pulse; 850 μm Al flyer; 5 mm flight distance



Shaped pulse; 900 μm Al flyer; 6 mm flight distance



The National Agenda for high consequence "predictive" computational science

rests on the three major credibility elements I have highlighted in this talk:

- Is the solution of the PDEs numerically accurate?

  Verification
- Is the physics represented by the PDEs accurate?

  Validation
- Is the use of the M&S results correct?

  Decisions

+ "Community of Practice" issues

"Community of Practice" ?

- Where do we stand on defining and measuring the credibility of our work for important applications?
  - "Ground-water models cannot be validated."
- How do we define, use and empower benchmarks?
  - "My code passes more benchmarks than your code."
- Are "standards" needed? Appropriate? Frightening?
  - "I feel free to ignore those DMSO Guidelines."
- Are journals helping or hurting?
  - "Good enough for a journal does not imply good validation."
- Will we ever solve enough of the technical problems to make the above questions reasonable?
- How can education help?

These issues lie at the core of what we term "The National Agenda for V&V".

SAND2006-2723C

Of interest … but reserved for breakfast and lunch talks in the future:

- V&V in climate modeling (sometimes called verification; sometimes called evaluation)
- V&V in social modeling (including economics)
  - "I would argue that accuracy is not only impossible to obtain (let alone assess) in practice, but also undesirable in principle…"

SAND2006-2723C

A few references

1. ASME (2006), "Guide for Verification and Validation in Computational Solid Mechanics," to be published.
2. Klein, et al (2006), "ASC Predictive Science Academic Alliance Program Verification and Validation Whitepaper," UCRL-TR-220342-Rev, to be released.
3. National Science Foundation (2006), "Simulation-Based Engineering Science: Revolutionizing Engineering Science through Simulation," Report of the National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science.
4. Oberkampf and Trucano (2002), "Verification and validation in computational fluid dynamics," Progress in Aerospace Sciences, Vol. 38, No. 3, 209-272. (Review)
5. Oberkampf, Trucano, and Hirsch (2004), "Verification, validation, and predictive capability in computational engineering and physics," Applied Mechanics Reviews, Vol. 57, No. 5, 345-384. (Review)
6. Pilch, et al (2000), "Guidelines for Sandia ASCI Verification and Validation Plans – Content and Format: Version 2.0," SAND2000-3101.
7. Trucano et al (2001), "Description of the Sandia Validation Metrics Project," SAND2001-1339.
8. Trucano et al (2002), "General Concepts for Experimental Validation of ASCI Code Applications," SAND2002-0341.
9. Trucano et al (2003), "On the Role of Code Comparisons in Verification and Validation," SAND2003-2752.

SAND2006-2723C

A recent presentation with more details:

"Uncertainty in Verification and Validation: Recent Perspective," Trucano, 2005 SIAM Conference on Computational Science and Engineering, February 12-15, 2005, Orlando, Florida.

Available at CCIM Publications Page:

http://gaston.sandia.gov/ccim_pubs_prod/main.cfm

SAND2006-2723C