



# PECOS

Predictive Engineering and Computational Sciences

Outside response of common concepts, terms, approaches, tools, and best practices of VVUQ

Robert Moser

The University of Texas at Austin

March 2012



# Outline

- 1 Verification
- 2 Introduction and Problem Statement
- 3 Towards Extrapolative Predictions
- 4 Current Tools
- 5 Data Reduction Modeling
- 6 Concluding Remarks

# Code and Solution Verification

## Code Verification

- Focuses on identifying failures of the code to correctly implement a desired numerical algorithm
- Analytical solutions to mathematical equations are used to calculate error in a corresponding approximate solution

## Solution Verification

- Process of quantifying the numerical errors (e.g. round-off, iterative, and discretization errors) that can cause the numerical solution to be an inadequate approximation of the correct solution
- One simulates the phenomenon of interest and has no a priori knowledge of the solution; in such cases error can only be estimated

# Code Verification

## Good Software Hygiene

- Unit Tests
- Regression Tests
- High Level Asserts
- Symmetry Tests
- Jacobian Tests
- Parametric Testing
- Exact Solutions
- Code-to-Code Comparison

# Code Verification

## Method of Manufactured Solutions (MMS)

- Exact solution typically not known
- Can **manufacture** a solution to generate source term from which one attempts to solve for manufactured solution
- Confirm convergence to solution at expected rate

## Pitfalls

- Explosion of terms in source
- Solutions need structure similar to application
  - ▶ E.g. Boundary layers
  - ▶ Exercise terms important to application
- Software reliability

## MASA Library

- Manufactured and Analytical Solution Abstraction Library
- Provides solutions and source terms for many operators
- Sources computed by AD
- Released under LGPL

# Maple MMS: 3D Navier-Stokes Energy Term

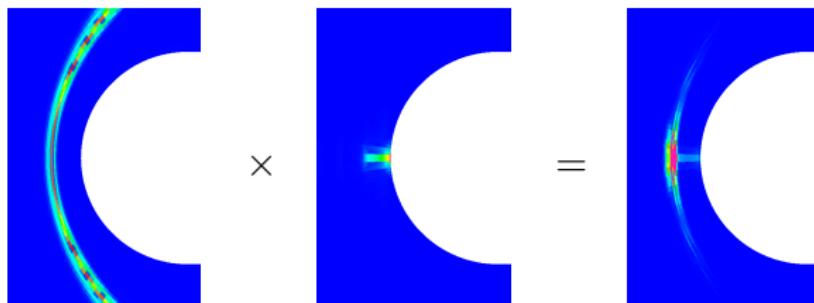
$$\begin{aligned}
Qe = & -\frac{a_{px}\pi p_x}{L} \frac{\gamma}{\gamma-1} \sin\left(\frac{a_{px}\pi x}{L}\right) \left[ u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] + \\
& + \frac{a_{py}\pi p_y}{L} \frac{\gamma}{\gamma-1} \cos\left(\frac{a_{py}\pi y}{L}\right) \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right] + \\
& - \frac{a_{pz}\pi p_z}{L} \frac{\gamma}{\gamma-1} \sin\left(\frac{a_{pz}\pi z}{L}\right) \left[ w_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] + \\
& + \frac{a_{px}\pi p_x}{2L} \cos\left(\frac{a_{px}\pi x}{L}\right) \left[ u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] \left[ \left( u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right)^2 + \right. \\
& \quad \left. + \left[ w_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right]^2 + \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right]^2 \right] + \\
& - \frac{a_{py}\pi p_y}{2L} \sin\left(\frac{a_{py}\pi y}{L}\right) \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right] \left[ \left( u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right)^2 + \right. \\
& \quad \left. + \left[ w_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right]^2 + \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right]^2 \right] + \\
& + \frac{a_{pz}\pi p_z}{2L} \cos\left(\frac{a_{pz}\pi z}{L}\right) \left[ w_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] \left[ \left( u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right)^2 + \right. \\
& \quad \left. + \left[ w_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right]^2 + \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right]^2 \right] + \\
& + \frac{a_{ux}\pi u_x}{2L} \cos\left(\frac{a_{ux}\pi x}{L}\right) \left( \left[ \left( u_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right)^2 + \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right]^2 \right. \right. + \\
& \quad \left. \left. + 3 \left[ u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right]^2 \right] \left[ \rho_0 + \rho_x \sin\left(\frac{a_{px}\pi x}{L}\right) + \rho_y \cos\left(\frac{a_{py}\pi y}{L}\right) + \rho_z \sin\left(\frac{a_{pz}\pi z}{L}\right) \right] + \right. \\
& \quad \left. + \left[ p_0 + p_x \cos\left(\frac{a_{px}\pi x}{L}\right) + p_y \sin\left(\frac{a_{py}\pi y}{L}\right) + p_z \cos\left(\frac{a_{pz}\pi z}{L}\right) \right] \frac{2\gamma}{(\gamma-1)} \right\} + \\
& - \frac{a_{uy}\pi u_y}{L} \sin\left(\frac{a_{uy}\pi y}{L}\right) \left[ v_0 + v_x \cos\left(\frac{a_{vx}\pi x}{L}\right) + v_y \sin\left(\frac{a_{vy}\pi y}{L}\right) + v_z \sin\left(\frac{a_{vz}\pi z}{L}\right) \right] \left[ \rho_0 + \rho_x \sin\left(\frac{a_{px}\pi x}{L}\right) + \rho_y \cos\left(\frac{a_{py}\pi y}{L}\right) + \rho_z \sin\left(\frac{a_{pz}\pi z}{L}\right) \right] \cdot \\
& \quad \cdot \left[ u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] + \\
& - \frac{a_{uz}\pi u_z}{L} \sin\left(\frac{a_{uz}\pi z}{L}\right) \left[ w_0 + w_x \sin\left(\frac{a_{wx}\pi x}{L}\right) + w_y \sin\left(\frac{a_{wy}\pi y}{L}\right) + w_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] \left[ \rho_0 + \rho_x \sin\left(\frac{a_{px}\pi x}{L}\right) + \rho_y \cos\left(\frac{a_{py}\pi y}{L}\right) + \rho_z \sin\left(\frac{a_{pz}\pi z}{L}\right) \right] \cdot \\
& \quad \cdot \left[ u_0 + u_x \sin\left(\frac{a_{ux}\pi x}{L}\right) + u_y \cos\left(\frac{a_{uy}\pi y}{L}\right) + u_z \cos\left(\frac{a_{uz}\pi z}{L}\right) \right] +
\end{aligned}$$

But wait, there's more!

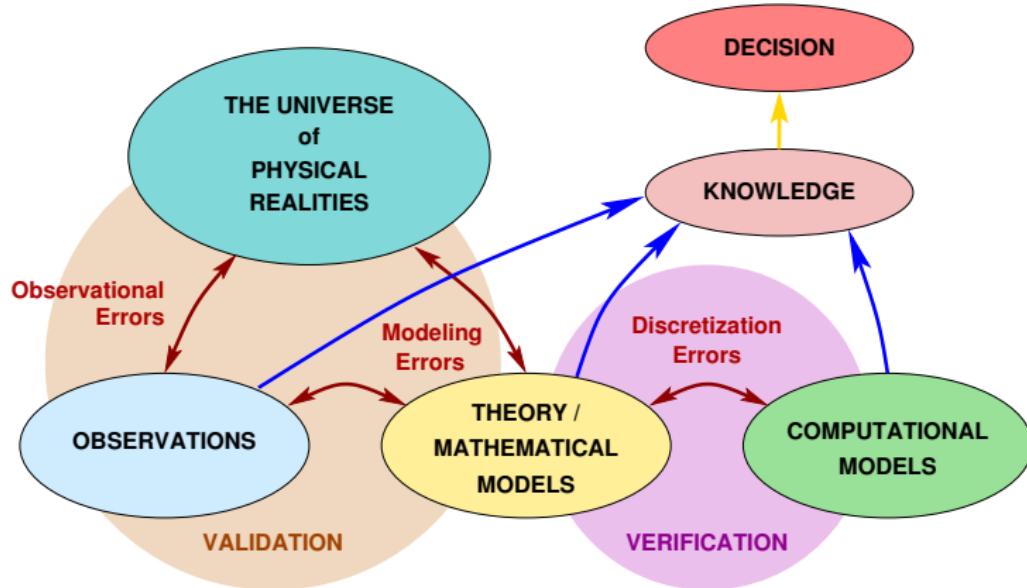
# Solution Verification

## Goal-Oriented Error Estimation & Adaptivity

- Typically want to control the error in some **functional** of the solution  $u$ , say  $Q(u)$  and not  $u$
- Error representation:  $Q(u) - Q(u^h) = R(u^h; p) + \Delta$ 
  - ▶  $p$  is the solution to the adjoint problem
  - ▶  $\Delta$  higher-order remainder term
- Use adjoint and residual to drive mesh adaptivity
- Under best conditions rigorous bounds on error, but useful indicator regardless



# Imperfect Paths to Knowledge and Predictive Simulation



***Predictive Simulation:*** the treatment of model and data uncertainties and their propagation through a computational model to produce predictions of **quantities of interest with quantified uncertainty**.

# Models are Imperfect

Mathematical models for complex multi-scale multi-physics systems are not usually posed as general truth statements about physical reality

## Examples of Useful Imperfect Models

- Newtonian mechanics
- Continuum mechanics
- Chemical reaction mechanisms
- RANS turbulence models
- Homogenization of complex materials

How do the imperfections of the mathematical models impact the reliability of the simulations in which they are used?

Validation processes are designed to find out

# Quantities of Interest

Simulations have a purpose: to inform a decision-making process

- Quantities are predicted to inform the decision
- These are the Quantities of Interest (QoIs)
- Models are not (evaluated as) scientific theories

Acceptance of a model is conditional on:

- its purpose
- the QoIs to be predicted
- the required accuracy

# What are Predictions?

## Prediction

Purpose of predictive simulation is to predict QoIs for which measurements are **not** available (otherwise predictions not needed)

Measurements may be unavailable because:

- instruments unavailable
- scenarios of interest inaccessible
- system not yet built
- ethical or legal restrictions
- it's the future

How can we have confidence in the predictions?

# Posing a “Predictive Validation” Process

## Predictive Validation Question

Does the combination of physical models, uncertainty models and supporting data yield acceptable credible predictions of the QoIs?

## Validation Activities

- **Inform:** Calibrate to match observations
  - ▶ What parameter values, model errors, etc. are plausible given the data?
- **Challenge:** Check that model output consistent with observations
  - ▶ Are discrepancies explained by plausible errors/uncertainties (in light of uncertainty models)?
- **Assess:** Determine impact of uncertainty/error on QoI's
  - ▶ Are observed discrepancies between model & data significant to QoI's?
  - ▶ Are the QoI's sensitive to models & uncertainties to which the observations are not?

# Validation Expectations are Model Dependent

## Interpolation “models”: simple fit to data

- Test for missing dependencies
- Test accuracy of fitting function
- Check that use is in the range of training data

## Physics-based models: formulated from theory

- This is what allows extrapolation
- Check that used in domain of applicability
- May include less reliable components:
  - ▶ Embedded (semi-) empirical models
  - ▶ Simplifying assumptions
  - ▶ Inadequacy models
  - ▶ A validation assessment is needed for these

# Predictive Validation and Uncertainty

## Treating Uncertainty is Integral to Predictive Validation

- Uncertainty in data and parameters limit the sensitivity of the validation process
- Uncertainty from model inadequacy enables assessment of impact of inconsistencies with data on QoI's

## Need Mathematical Treatment of Uncertainty

- We represent with Bayesian probability
- Probabilistic “models” of knowledge/uncertainty
  - ▶ Data & its uncertainty
  - ▶ Prior knowledge
  - ▶ Knowledge of model inadequacy
- Bayesian inference for calibration & model selection

# Outline

- 1 Verification
- 2 Introduction and Problem Statement
- 3 Towards Extrapolative Predictions
- 4 Current Tools
- 5 Data Reduction Modeling
- 6 Concluding Remarks

# An Abstract Setting for Predictions

Consider a model of the form:

$$\mathcal{R}(u, \tau(\theta); \xi) = 0$$

- $\mathcal{R}$  Physics-based system model (e.g. momentum conservation)
- $u$  System state
- $\tau(u; \theta, \xi)$  embedded model (e.g. constitutive) with parameters  $\theta$
- $\xi$  scenario parameters

- Observables given by  $d(u, \tau(\theta); \xi)$
- Qols given by  $q(u, \tau(\theta); \xi)$
- $\mathcal{R}$ ,  $d$  and  $q$  considered reliable (no uncertainties)

$\tau$  may be inadequate and  $\theta, \xi$  may be uncertain

# Data Uncertainty and Model Inadequacy

$$0 = \mathcal{R}(u, \tau(\theta) + \epsilon_{\text{mod}}; \xi)$$

$$D = d(u, \tau(\theta) + \epsilon_{\text{mod}}; \xi) + \epsilon_{\text{exp}}$$

$$Q = q(u, \tau(\theta) + \epsilon_{\text{mod}}; \xi)$$

$$0 = \mathcal{R}(u, \tau(\theta); \xi)$$

$$D = d(u, \tau(\theta); \xi) + \epsilon_{\text{exp}} + \tilde{\epsilon}_{\text{mod}}$$

$$Q = ?$$

## Predictive Uncertainty

## Kennedy & O'Hagen

Uncertainty in predictions,  $q$ , arise from uncertain parameters,  $\theta, \xi$   
**AND** model inadequacy  $\epsilon_{\text{mod}}$

### Some Caveats:

- all sources of uncertainty have been identified
- the data are accurate with well characterized uncertainties
- computational models are reliable (verified)
- numerical solutions are well resolved

# A Processes for Predictive Validation

## 1. Inform models $\tau^m$ and $\epsilon_{\text{mod}}$

- Use data for observables  $D_c$  from scenarios  $\xi_c$
- Bayesian inference to calibrate  $\theta_i$  for models  $\tau_i^m$  and meta-parameters for  $\epsilon_{\text{mod}i}$  for model classes  $i$
- Bayesian model selection among model classes  $i$

## 2. Challenge selected models

- Use data for observables  $D_v$  from scenarios  $\xi_v$  (include  $D_c$  from  $\xi_c$ )
- Are physics + uncertainty models consistent with observations?  
(Bayesian hypothesis testing?)
- Uncertainty models **must** account for all discrepancies between physics models and observations

# A Processes for Predictive Validation

## 3. Asses validity of predictions

- Does scenario  $\xi_p$  exercise  $\tau^m$  outside the conditions in which it has been challenged? (requires characterization of relevant “conditions”)
- Are prediction quantities  $q$  sensitive to uncertainties to which observed quantities are not?
- Are prediction uncertainties in  $q$  too large for decision maker

Entitled to make predictions **only** if answers to questions in (3) are “no”

# Outline

- 1 Verification
- 2 Introduction and Problem Statement
- 3 Towards Extrapolative Predictions
- 4 Current Tools
- 5 Data Reduction Modeling
- 6 Concluding Remarks

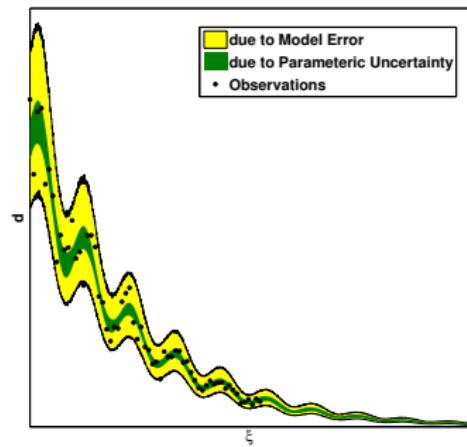
# Tools for Calibration

- Bayesian model calibration
- Bayesian model selection

## Determine:

- Values of the parameters that are consistent with the calibration data
- Includes learning about the model error.

$$\underbrace{\pi(\theta|D_{cal})}_{\text{posterior}} = \frac{\overbrace{\pi(D_{cal}|\theta) \pi(\theta)}^{\text{likelihood prior}}}{\underbrace{\pi(D_{cal})}_{\text{evidence}}}$$



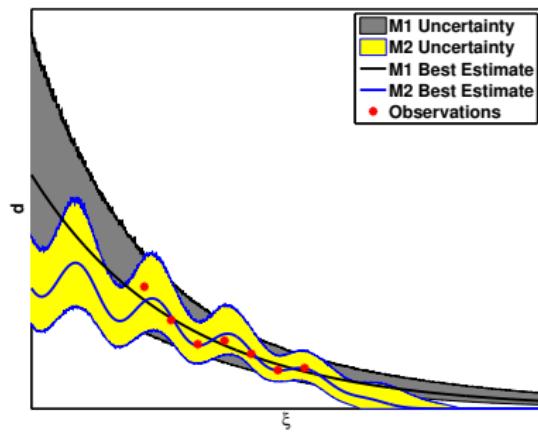
# Tools for Calibration

- Bayesian model calibration
- **Bayesian model selection**

## Identify:

- The most plausible error structure for model error
- The best among a set of approximate or phenomenological models.

$$\underbrace{\pi(M_i|D_{cal}, \mathcal{M})}_{\text{posterior plausibility}} = \frac{\overbrace{\pi(D_{cal}|M_i, \mathcal{M})}^{\text{evidence}} \overbrace{\pi(M_i|\mathcal{M})}^{\text{prior plausibility}}}{\pi(D_{cal}|\mathcal{M})}$$



# Tools for Validation

- **Bayesian model selection**
- Consistency metric

- Reject models that are less consistent with the calibration data
- Keep the models that best trades-off the principle of parsimony with the goodness-of-fit

$$M_1 \succ M_2 \iff \pi(M_1 | D_{cal}, \mathcal{M}) > \pi(M_2 | D_{cal}, \mathcal{M})$$

$$\underbrace{\frac{\pi(D_{cal}|M_1)}{\pi(D_{cal}|M_2)}}_{\text{Bayes factor}} \underbrace{\frac{\pi(M_1, \mathcal{M})}{\pi(M_2, \mathcal{M})}}_{\text{prior odds}} > 1$$

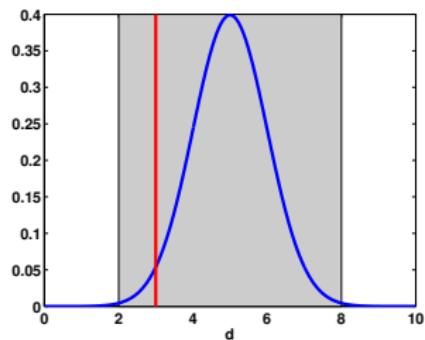
$$\underbrace{\ln[\pi(D_{cal}|M_1)]}_{\text{log evidence}} = \underbrace{E[\ln[\pi(D_{cal}|\theta, M_1)]]}_{\text{data fit}} - \underbrace{\text{KL}\left(\pi(\theta|D_{cal}, M_1) \parallel \pi(\theta|M_1)\right)}_{\text{model complexity}}$$

$\left( \text{model complexity ; goodness of fit ; prior} \right)$

# Tools for Validation

- Bayesian model selection
- **Consistency metric**

Need a quantitative characterization of consistency of model predictions with validation data.



- This is different from accuracy which assess whether observations and model predictions are “close enough”
- To entitle prediction, uncertainty models must plausibly account for all discrepancies between physics models and observations.

**Consistent**

# Outline

- 1 Verification
- 2 Introduction and Problem Statement
- 3 Towards Extrapolative Predictions
- 4 Current Tools
- 5 Data Reduction Modeling
- 6 Concluding Remarks

# Data Reduction Modeling

Assessing uncertainties in data is **NOT** primarily statistical

## Why is it needed?

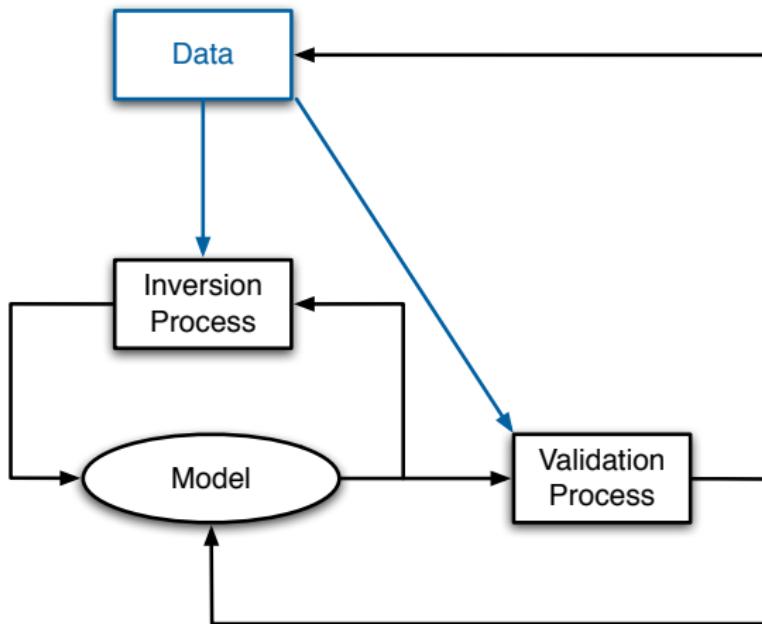
- Very likely that quantities we wish to measure are not directly measurable in an experiment
- Have to infer the values from other measurements using a mathematical model
- Estimate/recover uncertainties in legacy experimental data

## Impact on Validation and UQ

- Our philosophy: All mathematical models must be validated
- Must incorporate uncertainty of both the measured data and the data reduction model into the final uncertainty quantification of the data

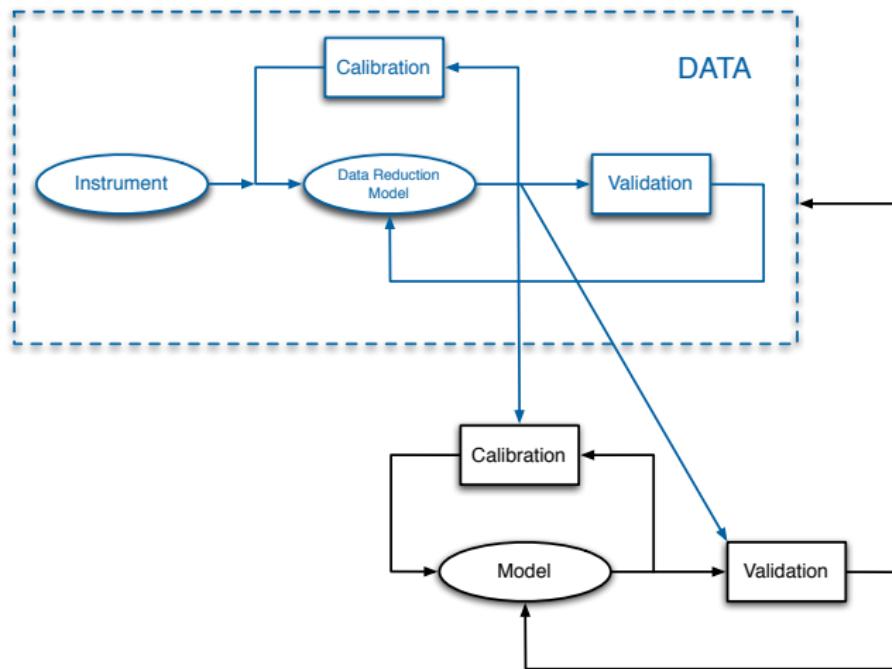
# Data Reduction Modeling

Traditional calibration schematic:



# Data Reduction Modeling

Incorporation of data reduction model:



# Challenges

## Complexity of Analysis

- DRM's may be complex multi-physics models in their own right
- Need very reliable validation and uncertainty analysis
- Logical dependencies of measurements

## Cooperation of Experimental and Computational Scientists

- As data consumers, computational scientists must be able to properly characterize uncertainty in the data, including any data reduction models
- Requires many details of experimental procedures - some data producers may be reluctant to share such details or may simply be unavailable (legacy data)

# Outline

- 1 Verification
- 2 Introduction and Problem Statement
- 3 Towards Extrapolative Predictions
- 4 Current Tools
- 5 Data Reduction Modeling
- 6 Concluding Remarks

# Final Remarks

## Verification

- Techniques of code verification are well established but not as widely used as they should be
- Techniques of solution verification can be further developed
  - ▶ Strengthen *a posteriori* error estimates of QoI's for complex problems
- Bigger issue is adoption, e.g.:
  - ▶ Few application codes support adjoints
  - ▶ Even error estimates based on grid refinement are often not used
- Distinct from Validation and UQ

# Final Remarks

## Validation and UQ

- Involve much larger conceptual & research issues
  - ▶ Unobserved prediction QoI's (predictive validation)
  - ▶ Importance of reliable physics models enabling extrapolation
  - ▶ Critical role of inadequacy models esp. for “embedded models”
  - ▶ Need observational data with well characterized uncertainties, but in many problems, this is not available...
  - ▶ Uncertainty modeling: mathematically encoding the often qualitative generally incomplete physical information that we have
    - Priors, model inadequacy and data uncertainties

## Important, but I did not Discuss

- Algorithms & software
- Decision making
- Rare events
- Education & socialization