Assessing the Reliability of Complex Models: Mathematical Foundations of Verification, Validation, and Uncertainty Quantification

MARVIN L. ADAMS, Texas A&M University; DAVID M. HIGDON, Los Alamos National Laboratory; DEREK BINGHAM, Simon Fraser University; JAMES O. BERGER, Duke University; WEI CHEN, Northwestern University; ROGER GHANEM, University of Southern California; OMAR GHATTAS, University of Texas at Austin; JUAN MEZA, University of California, Merced; ERIC MICHIELSSEN, University of Michigan; VIJAYAN N. NAIR, University of Michigan; CHARLES W. NAKHLEH, Sandia National Laboratories; DOUGLAS NYCHKA, National Center for Atmospheric Research; STEVEN M. POLLOCK, University of Michigan (retired); HOWARD A. STONE, Princeton University; ALYSON G. WILSON, Institute for Defense Analyses; MICHAEL R. ZIKA, Lawrence Livermore National Laboratory

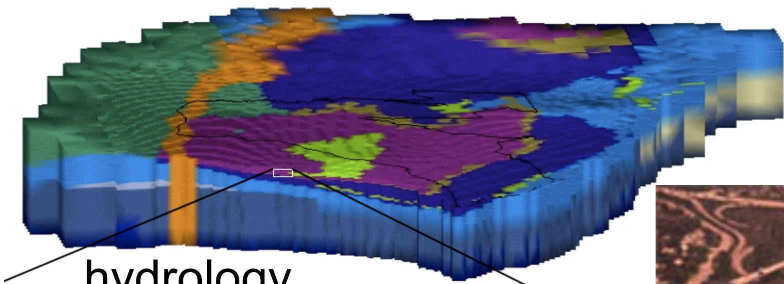
Charge to the Committee

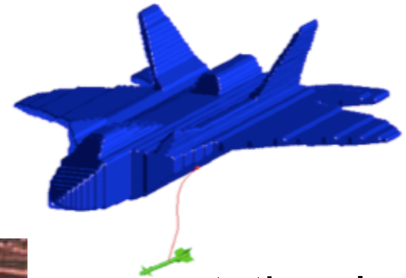
Sponsored by the National Nuclear Security Administration, Department of Energy; the Air Force Office of Scientific Research; and the National Science Foundation.

Study the mathematical foundations of VVUQ, and to recommend steps that will ultimately lead to improved processes. The specific tasking to the committee is as follows:

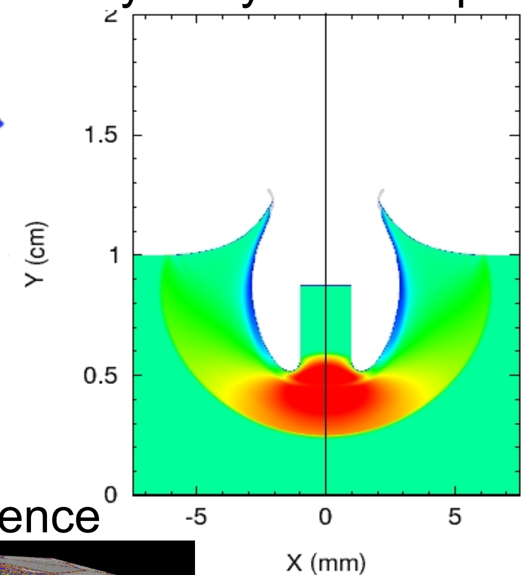
- A committee of the National Research Council will examine practices for VVUQ of large-scale computational simulations in several research communities.
- The committee will identify common concepts, terms, approaches, tools, and best practices of VVUQ.
- The committee will identify mathematical sciences research needed to establish a foundation for building a science of V&V and for improving the practice of VVUQ.
- The committee will recommend educational changes needed in the mathematical sciences community and mathematical sciences education needed by other scientific communities to most effectively use VVUQ.

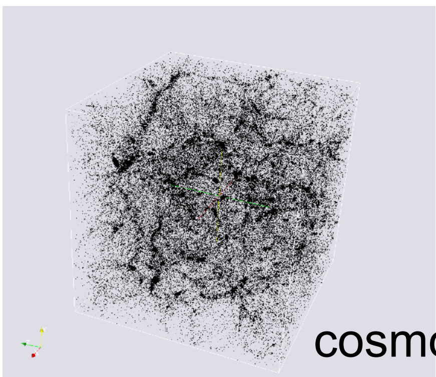
Computational Models & Physical Systems

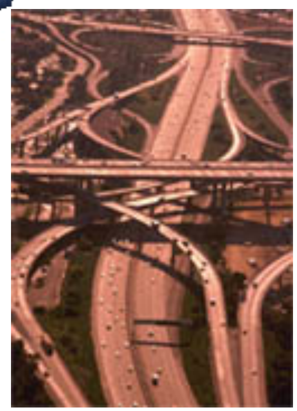

hydrology


computational fluid dynamics

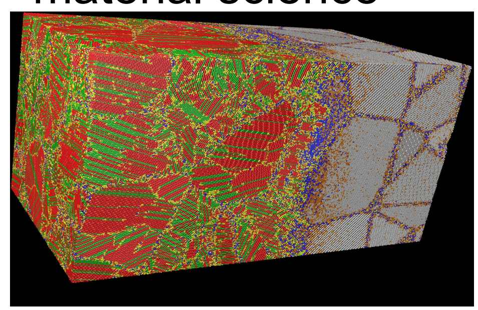
hydrodynamic impact

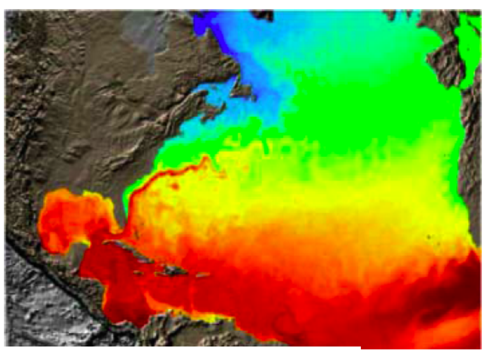



cosmology

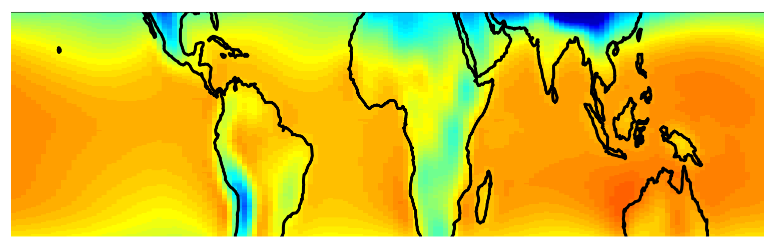

agent-based models

material science

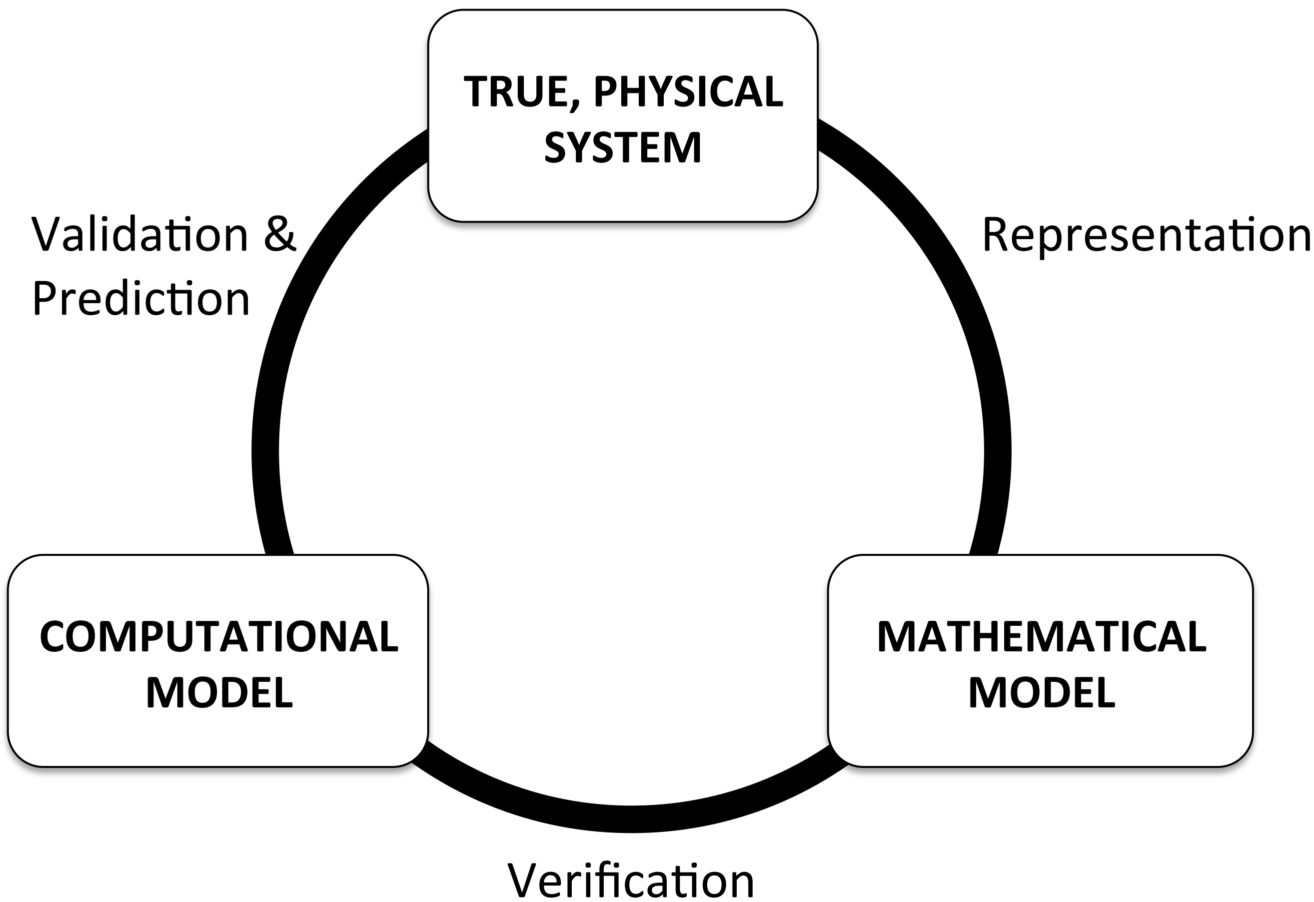



ocean dynamics


physics under extreme conditions

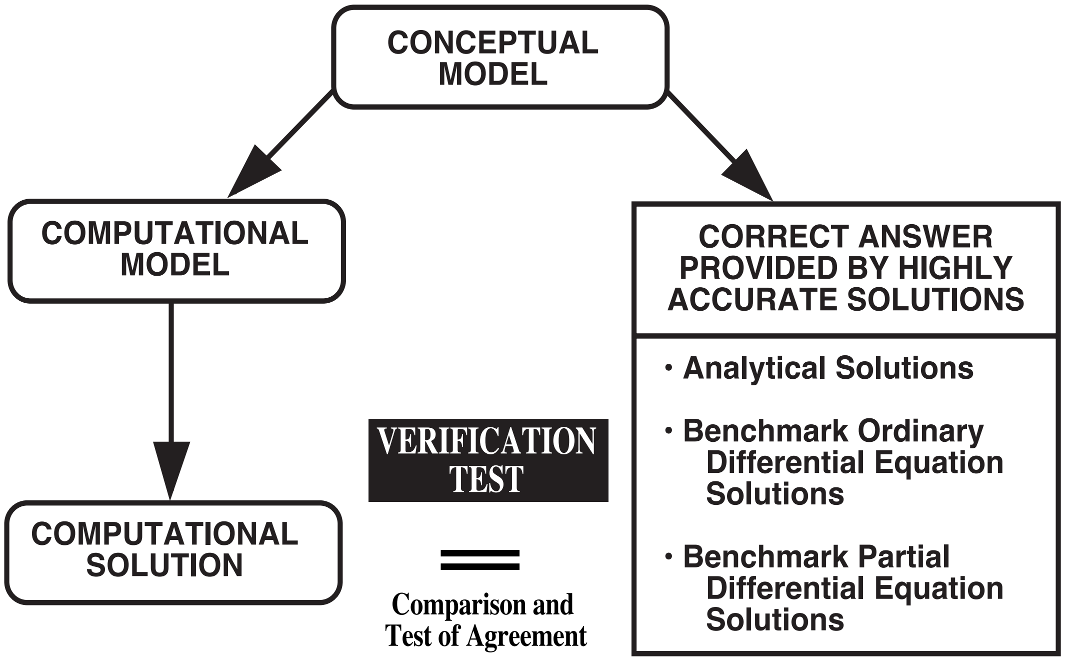

general circulation models

Verification, Validation and Uncertainty Quantification



Verification

[AIAA 1998]

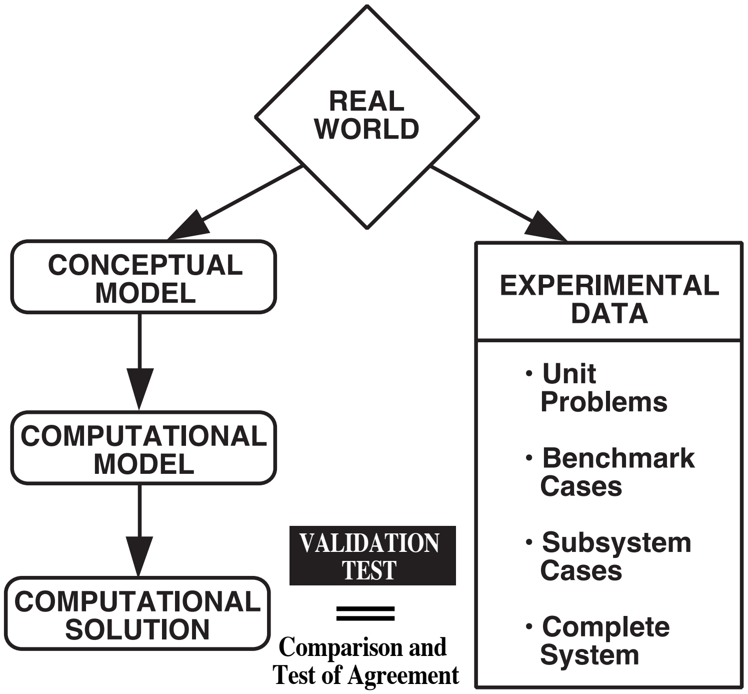


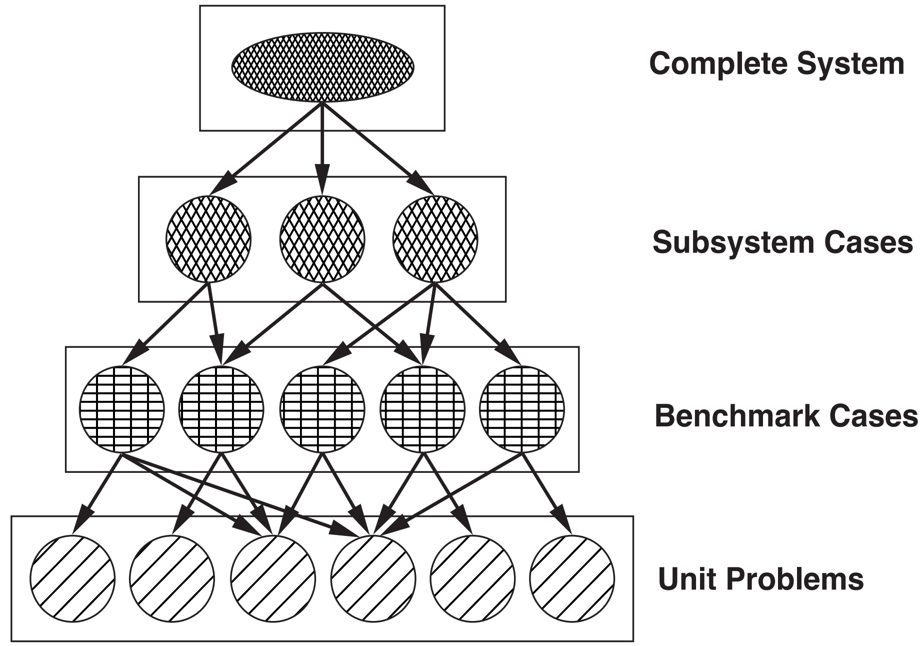
<u>Verification:</u> the process of determining how accurately a computer program ("code") correctly solves the equations of the mathematical model. This includes code verification and solution verification.

- Code verification
  - Determining whether the code correctly implements the intended algorithms
- Solution verification
  - Determining the accuracy with which the algorithms solve the mathematical model's equations for specified QOIs.

- Interest focuses on key outputs or "Quantities of Interest" (QOIs).
- Highly accurate solutions are typically only available for simple or specialized problems.
- Assessing accuracy of QOIs can be challenging.
- Understand what can be done to improve accuracy

Validation





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.

Breaking complex system into a hierarchy of successively less complicated subsystems, and eventually separated effects.
Physical experiments at different levels of this hierarchy for calibration, estimating prediction accuracy, and assessing quality of model-based predictions.

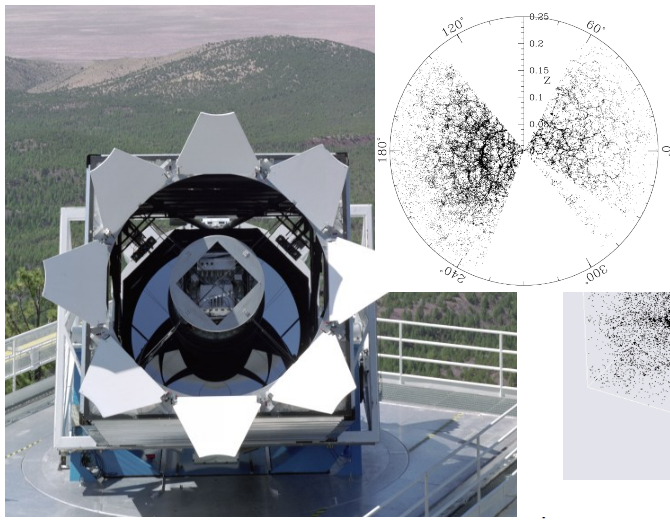
[from AIAA Guide on V&V]

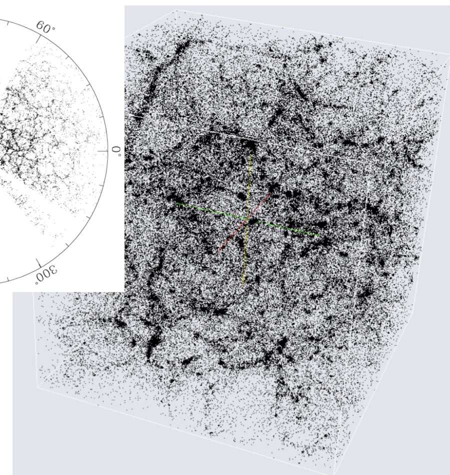
A Hierarchy of Multiple Types of Experiments in Important for Modeling Physical Systems

Observation/experiment

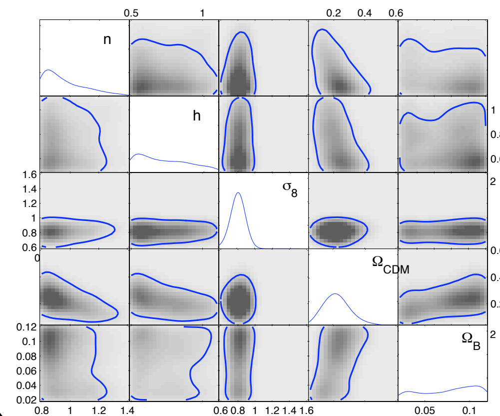
simulations

large scale structure of universe



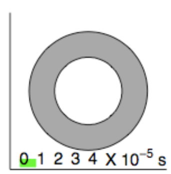


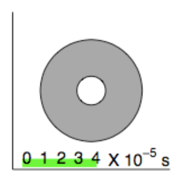
Calibration: finding parameter settings consistent with observations



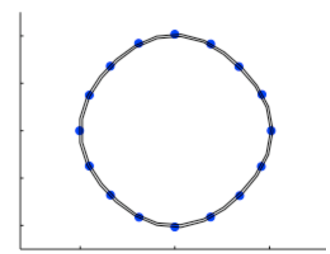
Hydrodynamic behavior

notional data & simulations





prediction uncertainties



Uncertainty Quantification

The process of quantifying uncertainties associated with model calculations of true physical QOIs, with the goals of accounting for all important sources of uncertainty and quantifying the contributions of specific sources to the overall uncertainty.

Ideally, the UQ assessment includes:

- an inventory of possible sources of error and uncertainty in the inferences and predictions
- an inventory of the sources of error and uncertainty accounted for in the assessment
- an inventory of assumptions on which the assessment is based.

Uncertainty Quantification

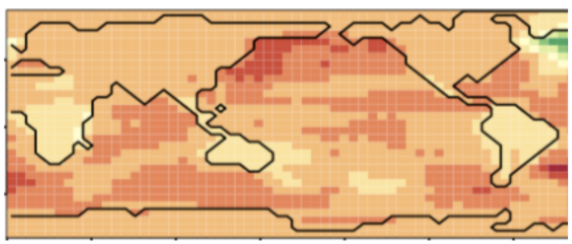
More generally, UQ is the discipline that develops theory, methods and tools to assess the reliability of scientific inferences, often aided by computational models.

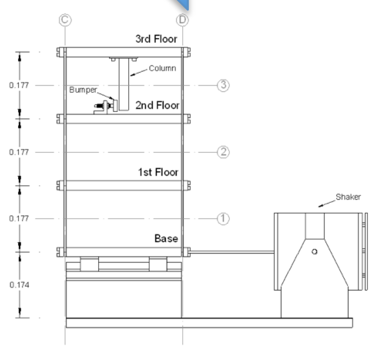
This includes:

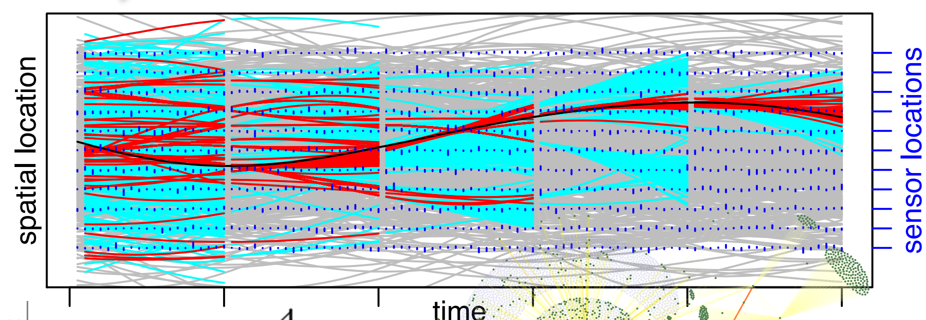
- Sensitivity analysis, emulation, construction of reduced models;
- Forward propagation of input uncertainties through large-scale computational models;
- Inverse problems and parameter estimation;
- Quantifying model discrepancy/structural error;
- Aggregating these various sources of uncertainty to estimate prediction uncertainty.
- Informing about steps to take to reduce prediction uncertainty

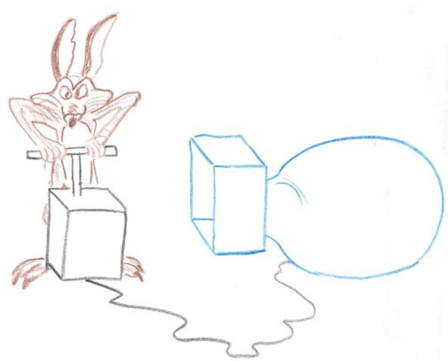
VVUQ Approaches depend on the application

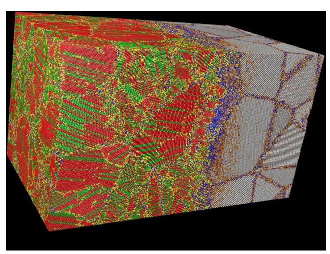
physical	model	empirical
complex, nonlinear	physical system	simple, few interactions
limited	observations	plentiful and relevant
extrapolative	prediction	interpolative
demanding, slow	computation	light, fast

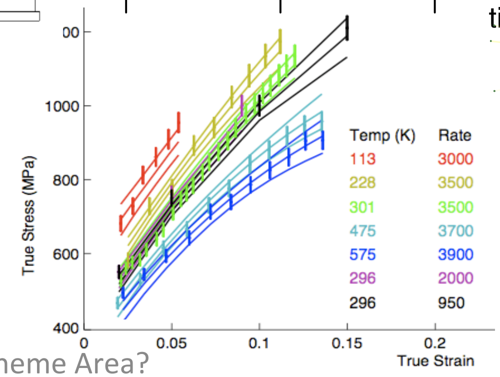


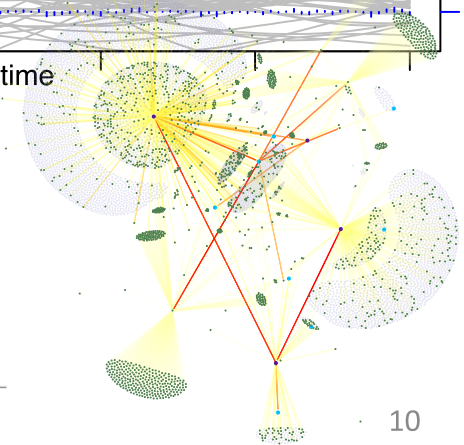








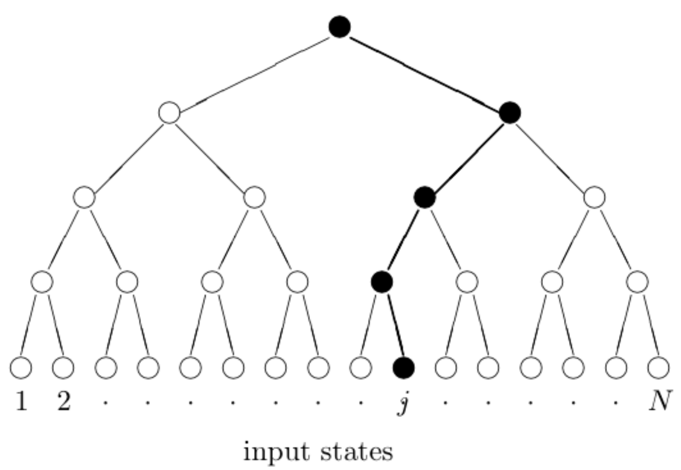




Verification tasks

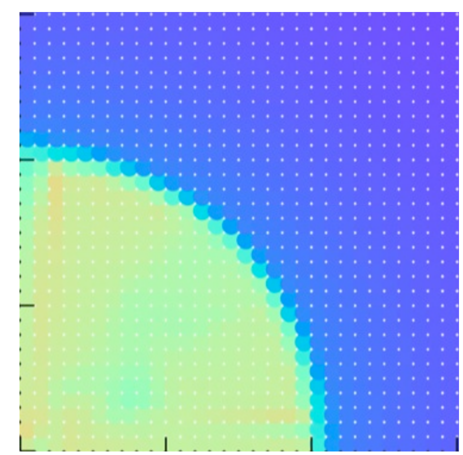
Code Verification

- Software Quality Assurance practices
- Regression tests
- Comparison to high quality, analytic solutions (e.g. manufactured solutions)
- Coverage?
- Algorithm checks – Is convergence rate as expected?

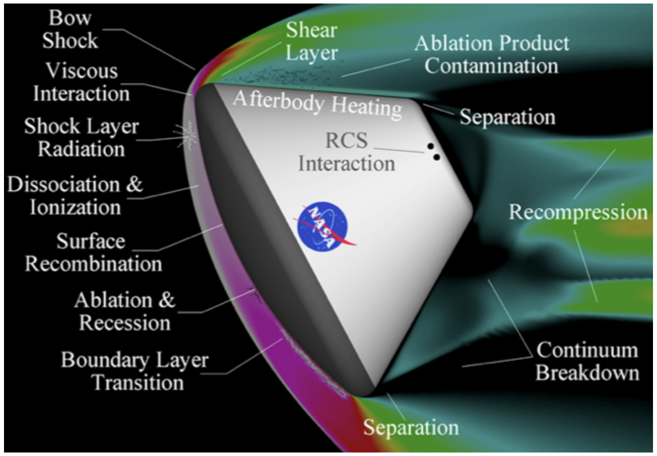


Solution Verification

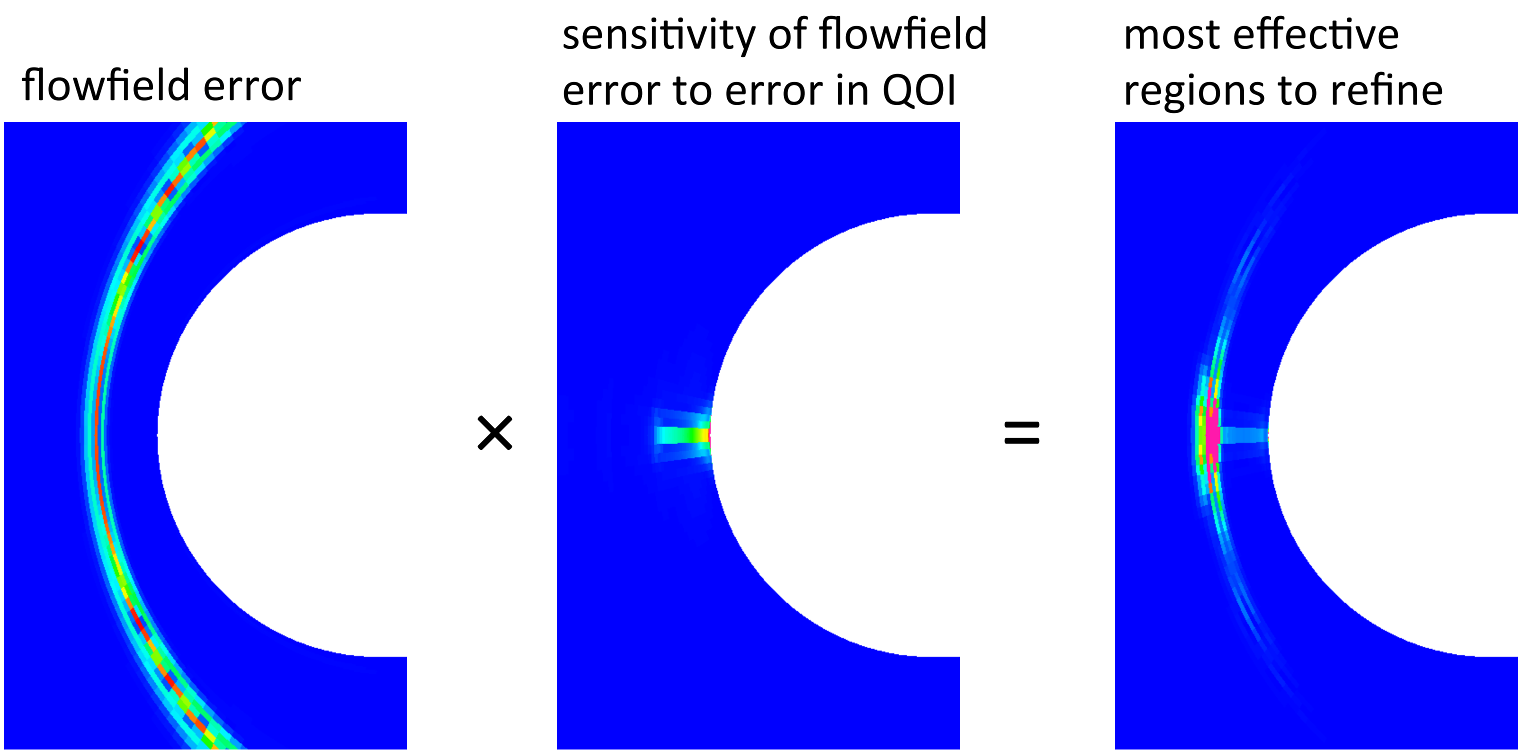
- Assessing/estimating solution error
- Convergence studies
- *A posteriori* – as the calculations are being carried out
- Goal oriented – with focus on QOI, guiding calculations



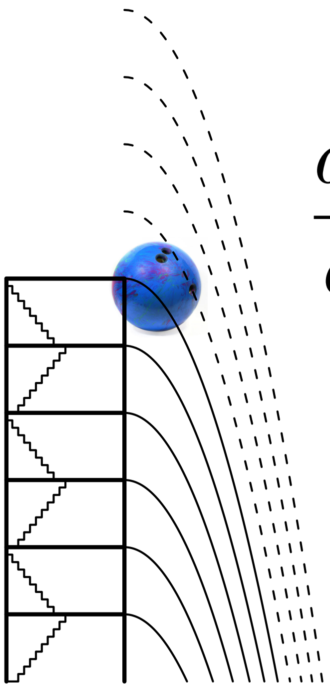
Accuracy of QOI can be used to drive adaptive mesh refinement for solution verification



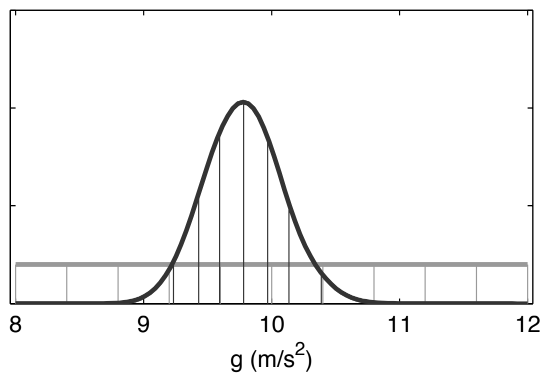
- Goal is accuracy regarding the QOI: peak ablation rate
- Refine mesh to minimize error in QOI while also minimizing computational effort
- Use the adjoint to estimate the impact of flowrate error on the QOI

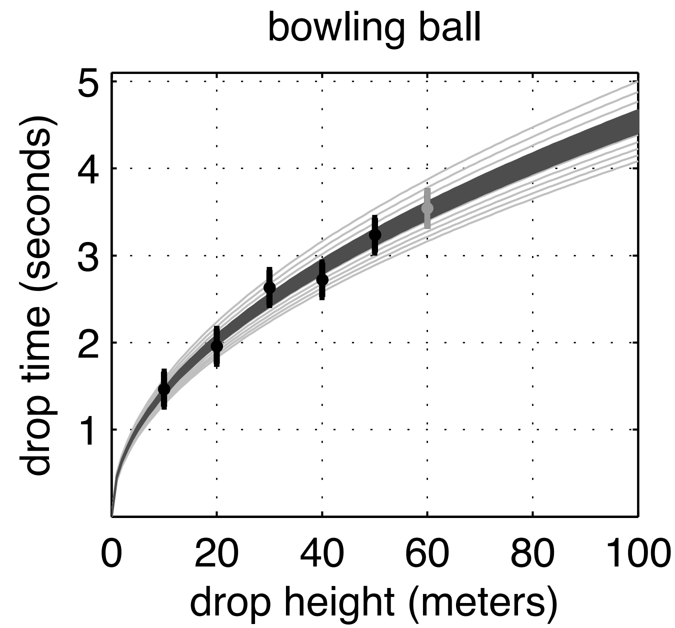


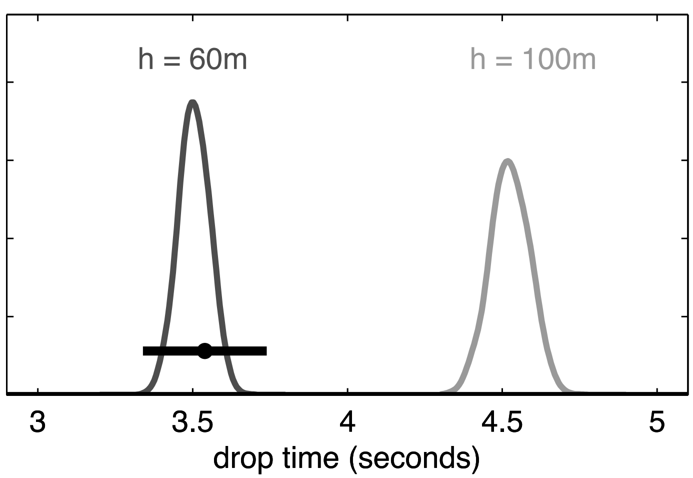
Basic Ball Drop Example

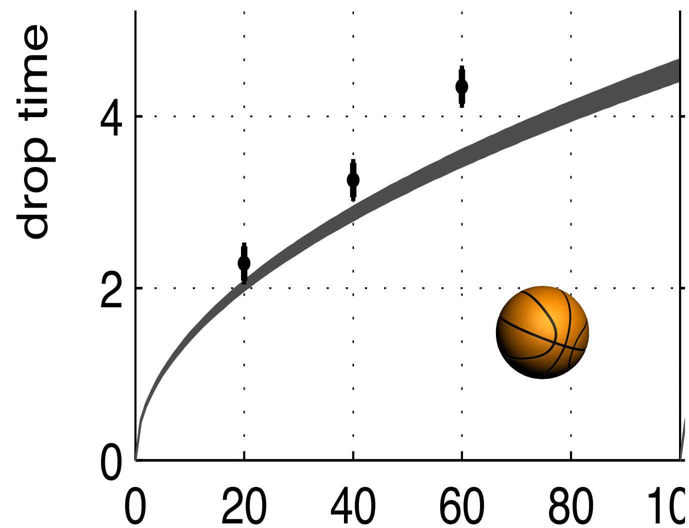


$$\frac{d^2 h}{dt^2} = g$$

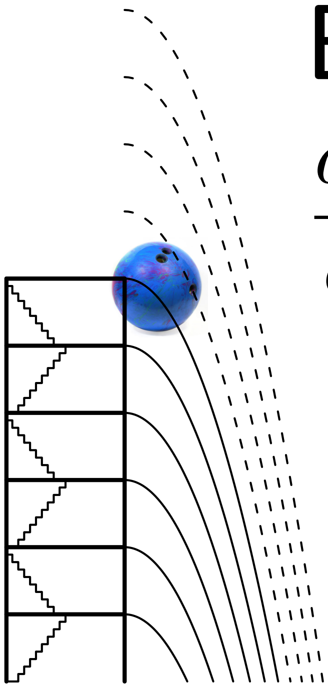






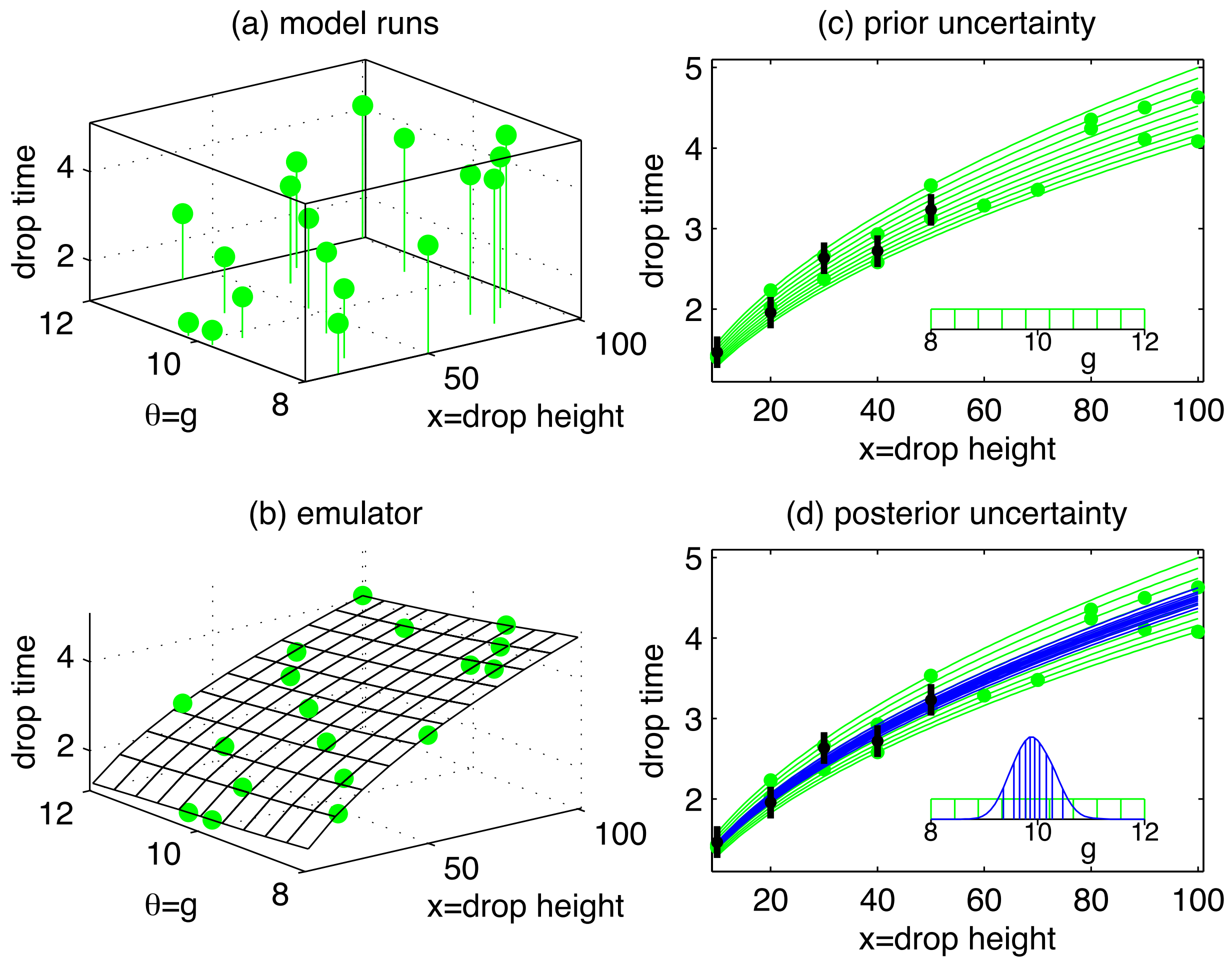


Ball Drop, with Emulator



$$\frac{d^2h}{dt^2} = g$$

Emulating the computer model response



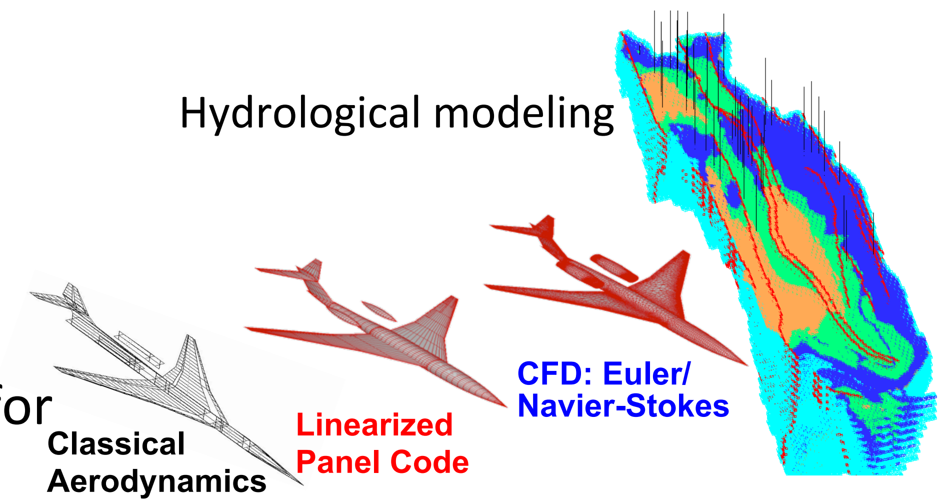
Emulators and Reduced Models

Goal is faster model evaluations for sensitivity analysis, forward propagation, and parameter estimation.

- Polynomial chaos representations
  - Efficiently represents output uncertainty given specified input uncertainty.
  - Can also emulate model response.
  - Makes use of functional analysis theory.
- Reduced, or low fidelity models
  - Capture necessary features for application.
  - Discard unnecessary complexity.

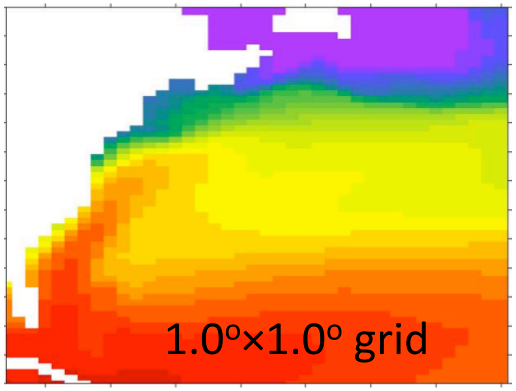
$$\eta(t,x,Z) \approx \sum_{|k|=0}^{N} \hat{\eta}_k(t,x)\Phi_k(Z), \; \# \text{ of bases} = \binom{n_z+N}{n_z}$$

$$\hat{\eta}_k = E[\eta_(Z)\Phi_k(Z)] = \int \eta(Z)\Phi_k(Z)\rho(Z)dZ, \; 0 \le |k| \le N$$



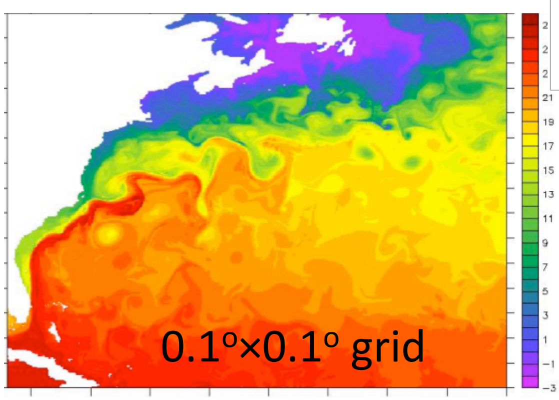
Bridging Resolutions & Model Combination

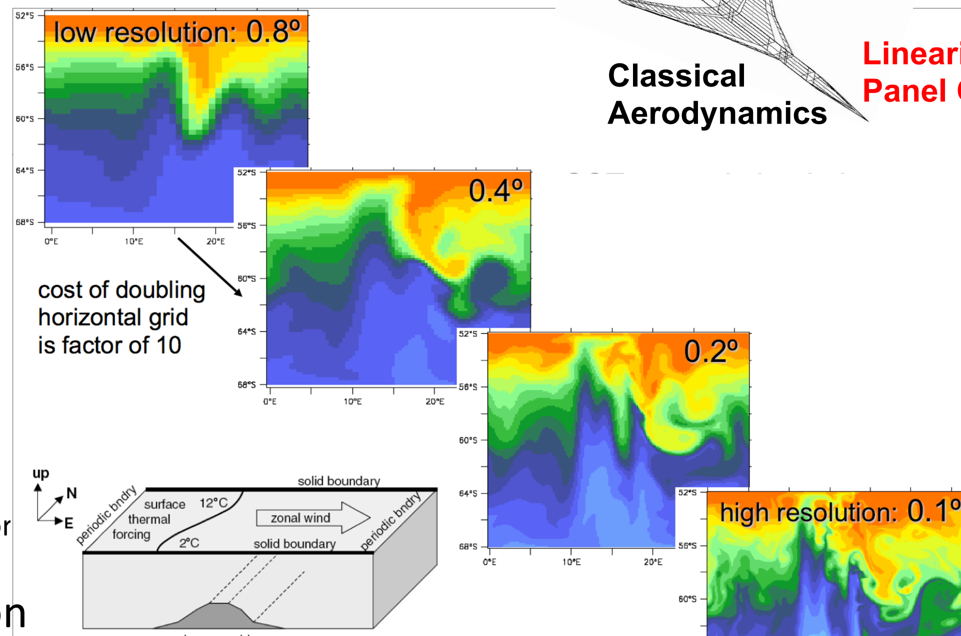
IPCC simulation



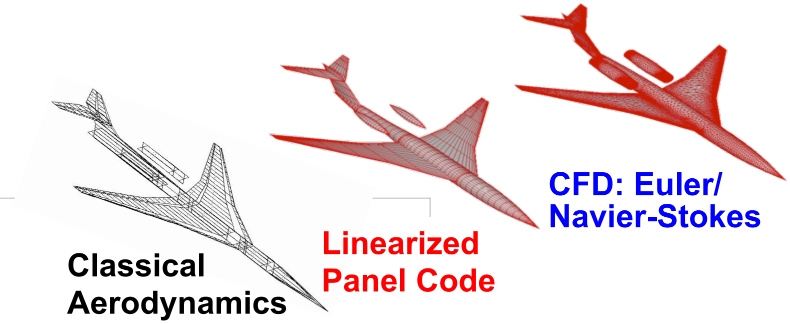
~10× additional computational effort for each doubling of resolution

Strongly eddying simulation

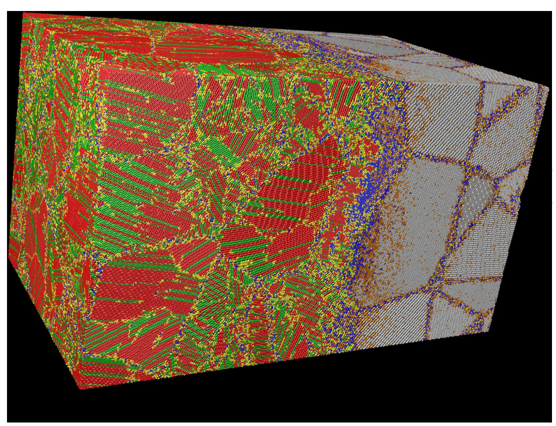


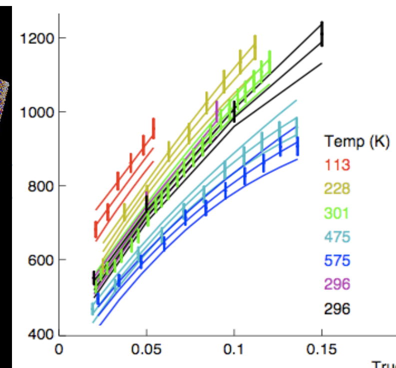


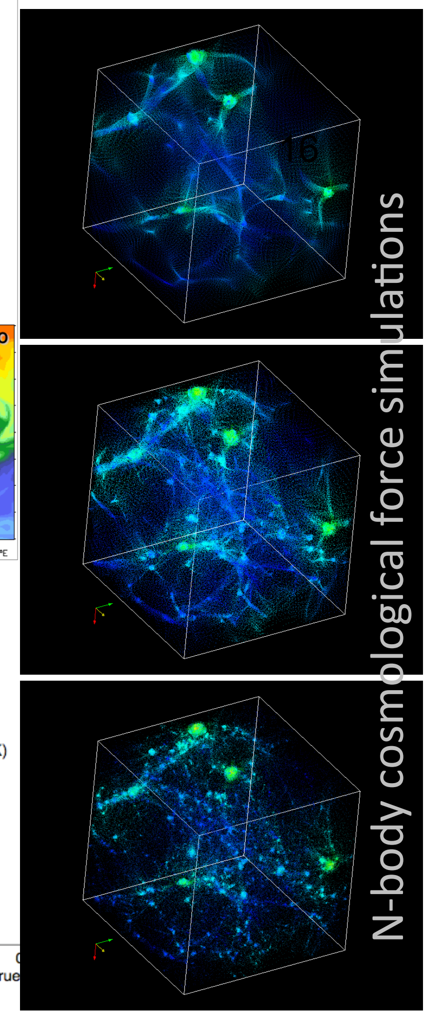
Highly idealized Southern-Ocean-like periodic channel flow



Material properties



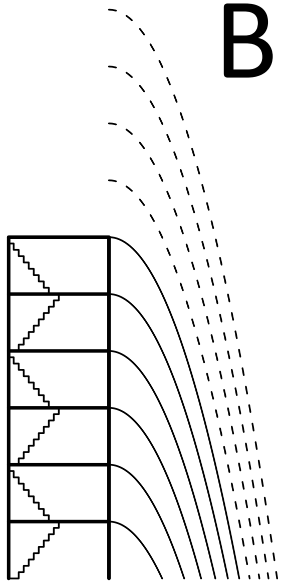


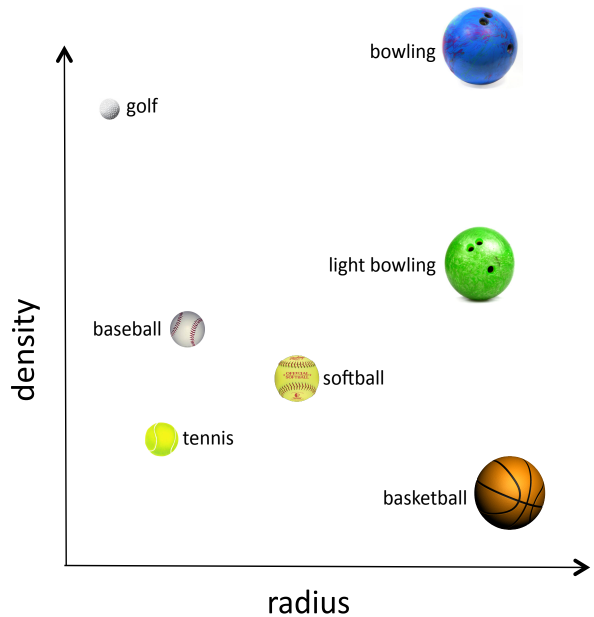


Potential Research Directions for Emulation, Reduced Models & Forward Propagation

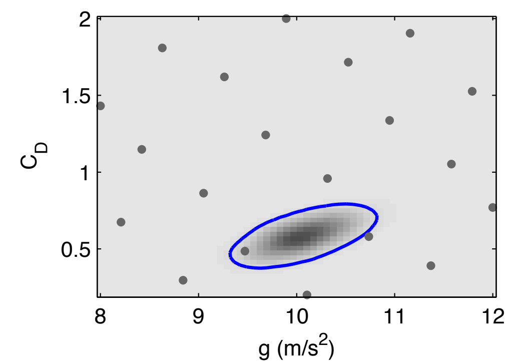
- Approaches for dealing with high-dimensional spaces of uncertain inputs.
  - Use process knowledge to help counter the curse of dimensionality.
- Phenomena aware emulation
  - Make use of adjoint and/or derivative information.
  - Make use of other "intrusive" UQ methods.
  - Use knowledge of system to model discontinuities and represent uncertainty.
- Emulation, sensitivity analysis, and uncertainty propagation across hierarchies of models.
  - Multiple fidelities, model hierarchies, multiple resolutions.
- Efficient exploitation of modern and future massively parallel architectures.
  - Rethink the development of computational models with UQ in mind
  - Co-develop UQ methods and computational models with computing architecture in mind.

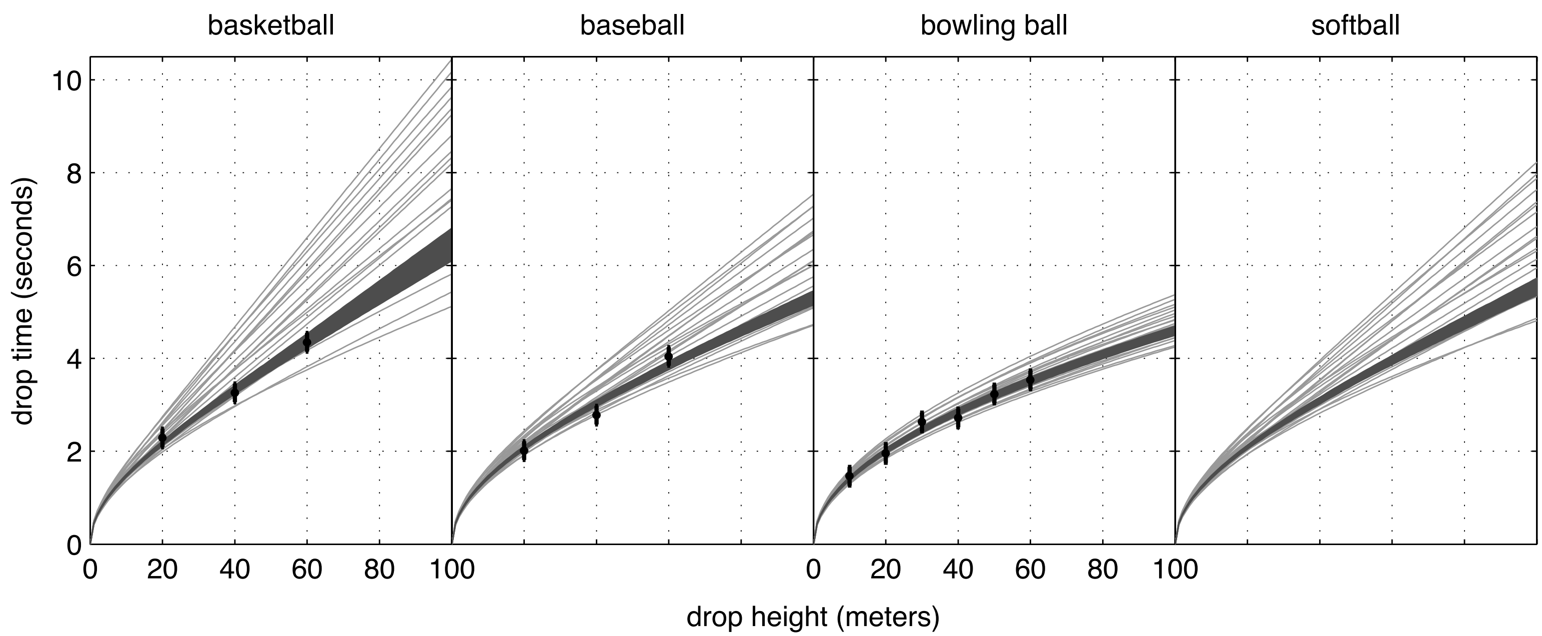
Ball Drop with Air Resistance



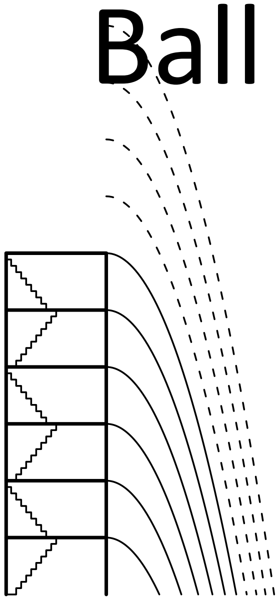


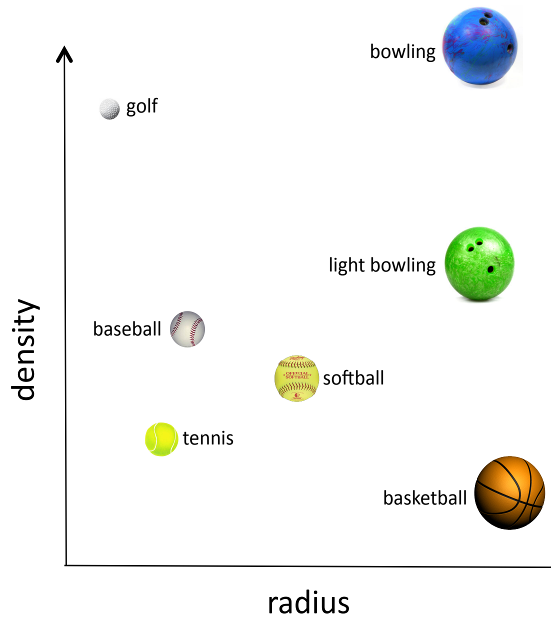
$$\frac{d^2h}{dt^2} = g - \frac{C_D}{2}\frac{3\rho_{\mathrm{air}}}{4R_{\mathrm{ball}}\rho_{\mathrm{ball}}}\left(\frac{dh}{dt}\right)^2$$





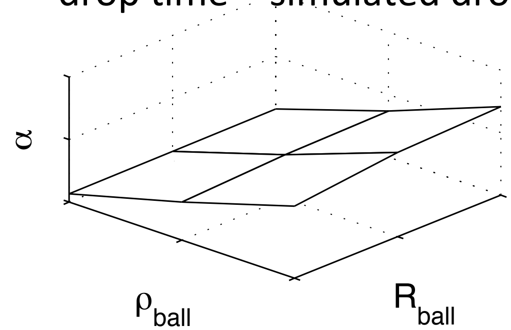
Ball Drop with Model Discrepancy

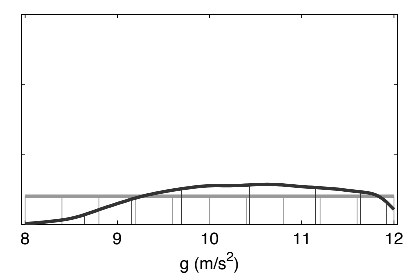


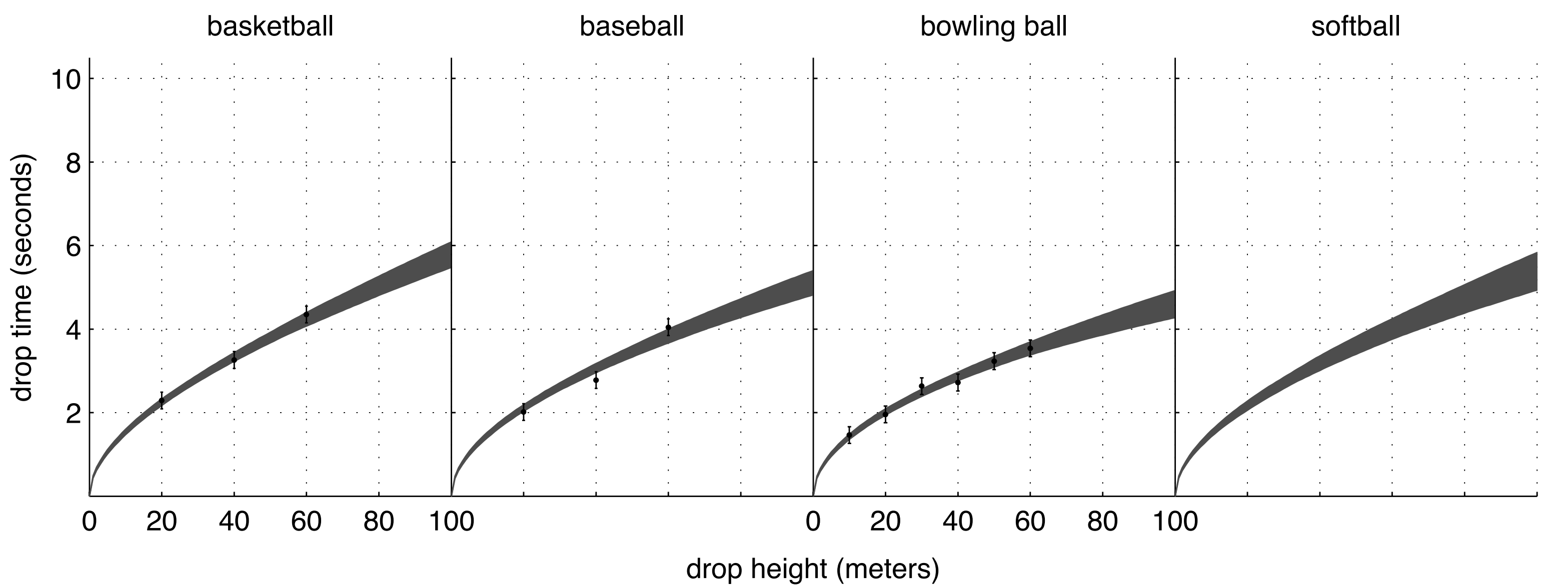


$$\frac{d^2h}{dt^2} = g$$

drop time = simulated drop time + α × drop height



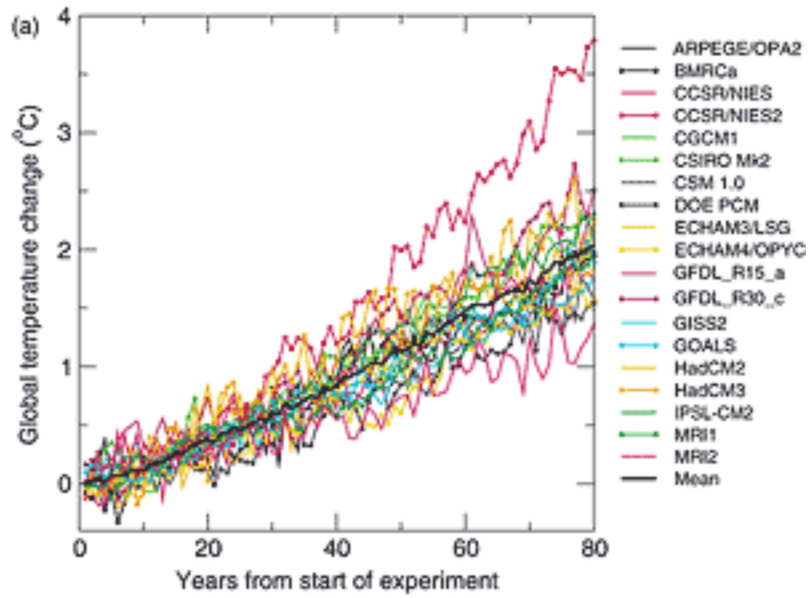




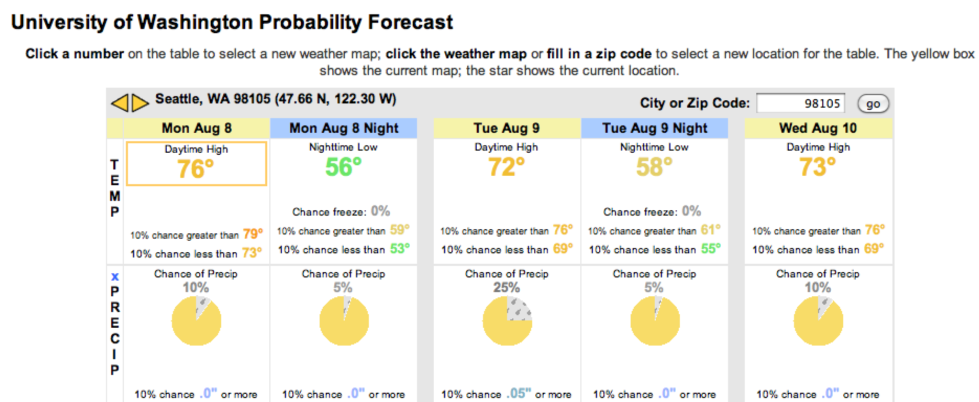
Multi-model ensembles

- To help estimate structural error (discrepancy or model-form error)
- Typically combined using Bayesian hierarchical modeling, model averaging, mixture of experts, …
- Notable successes in data assimilation applications
- Difficult in extrapolations – How do differences between model predictions inform about the difference between model and reality?
- Can ensembles of models be constructed with assessing model error in mind?

IPCC

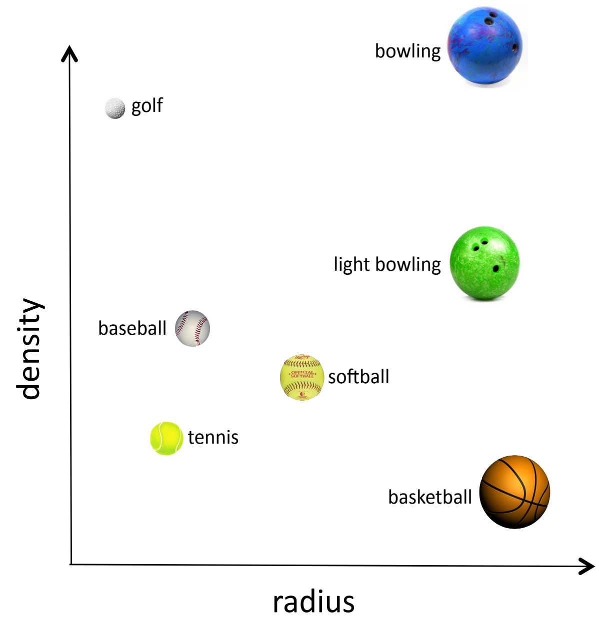


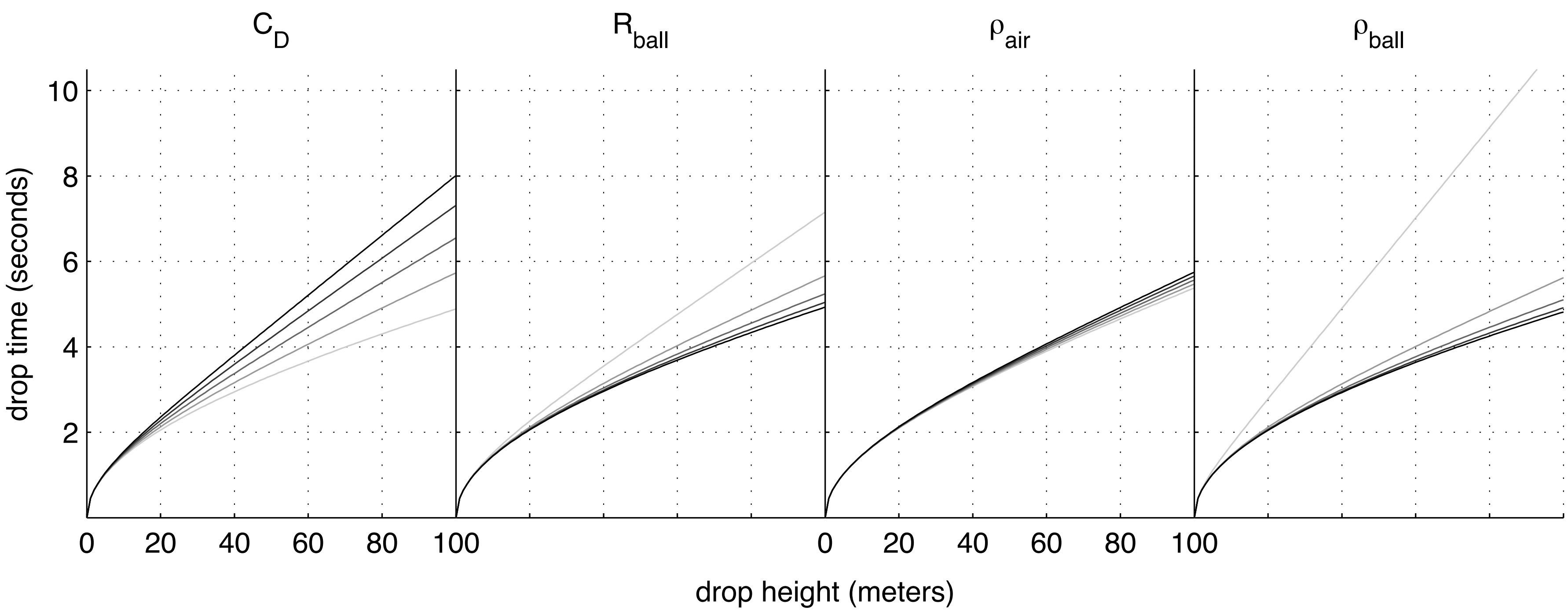
Probcast: Gneiting and Raftery (2005+)



Sensitivity Analysis for the Ball Drop Example

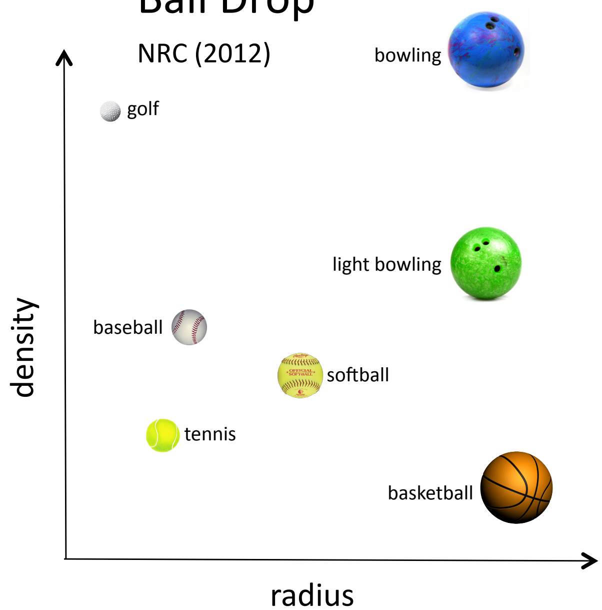
$$\frac{d^2h}{dt^2} = g - \frac{C_D}{2}\frac{3\rho_{\text{air}}}{4R_{\text{ball}}\rho_{\text{ball}}}\left(\frac{dh}{dt}\right)^2$$





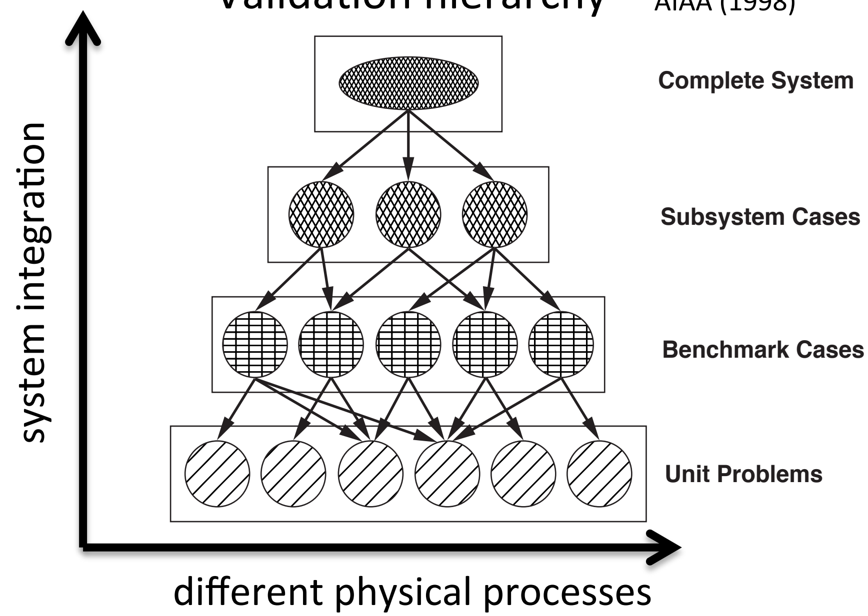
Domain space – describing the prediction scenario

Ball Drop

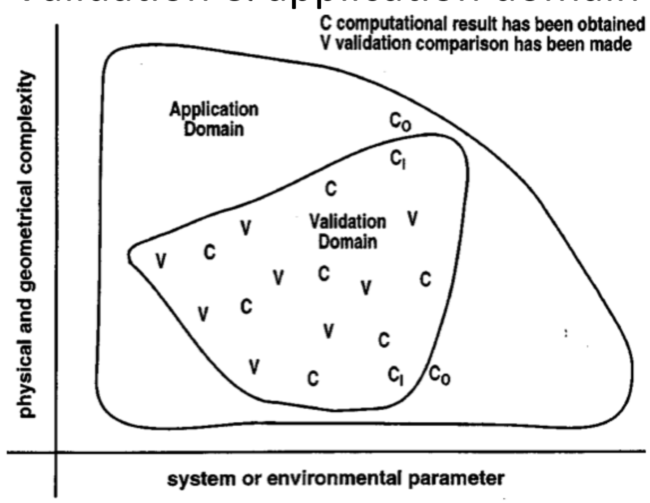


Validation hierarchy

AIAA (1998)

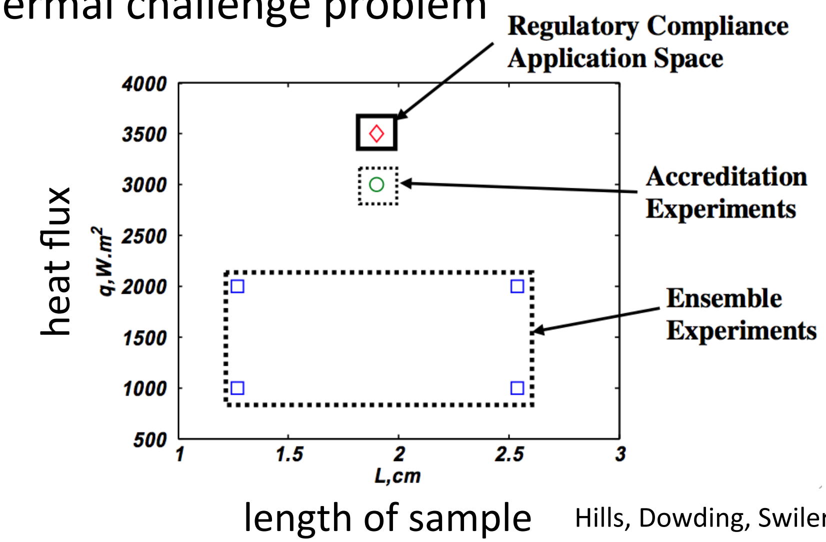


Validation & application domain



Oberkampf, Trucano & Hirsch (2004)

Thermal challenge problem

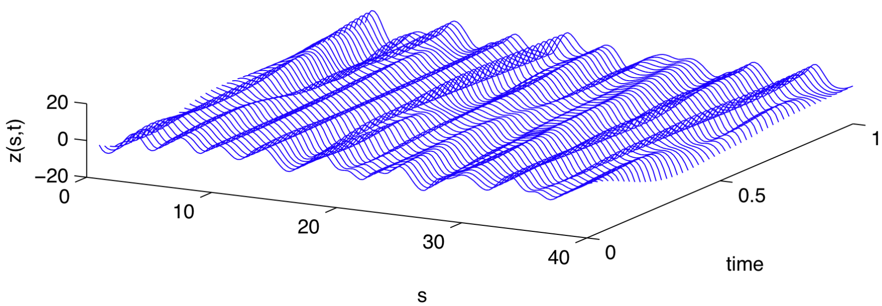


Hills, Dowding, Swiler (2008)

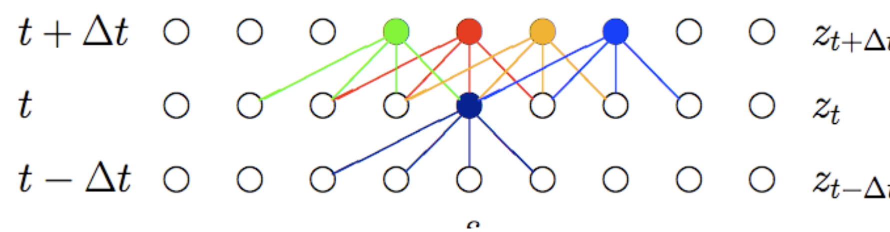
Exchange traditional computational models for VVUQ friendly implementations

PDE Solver

$$\frac{\partial}{\partial t} z(s,t) = [z(s+1,t) - z(s-2,t)]z(s-1,t) - z(s,t) + F$$



Markov Random Field



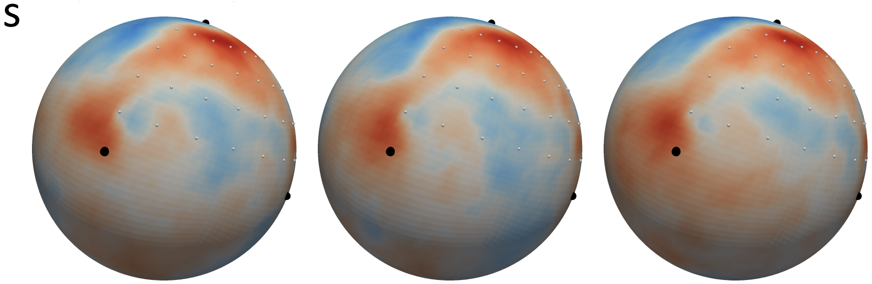
[Wikle and Hooten, 2010]

Constructing forward models and adjoints

- Very informative in high-dimensional problems
- Can help in building response surface
- Can facilitate MCMC or other UQ algorithms

Exploiting new, heterogeneous, high performance computing architectures

Adjoint enabled posterior sampling for a 30K-dimensional inverse problem



$$\boldsymbol{\Gamma}_{\text{post}}^{1/2}\boldsymbol{x} = \boldsymbol{\Gamma}_{\text{pr}}^{1/2}\{\boldsymbol{V}_r[(\boldsymbol{\Lambda}_r + \boldsymbol{I}_r)^{-1/2} - \boldsymbol{I}_r]\boldsymbol{V}_r^T + \boldsymbol{I}\}\boldsymbol{x}$$

[Martin, Wilcox, Burstedde, and Ghattas, 2012]

Potential Research Directions for Validation and Prediction

- Combining models
  - With different resolutions, spatial and temporal scales, and/or of different features of the physical process.
- Constructing computational models with VVUQ in mind.
  - Availability of adjoints, embedding terms to account for model discrepancy, …
- Designing VVUQ systems that consider the computational model, high-performance computing, the application area, in the context of VVUQ needs.
- Assessing the reliability of extrapolative predictions.
  - Idea of a domain space in which problems reside.
  - Designing ensembles of models for this purpose?

Rare, high consequence events

- PRA ideas:
  - scenarios
  - consequences
  - chances
- Often involves catastrophic failure – a difficult process to model
- Rare ➔ estimation of small probabilities
- Rare ➔ difficult to compare to reality
- Difficult in extrapolation applications – are all important processes in the model?
- Build models to cater to such events?

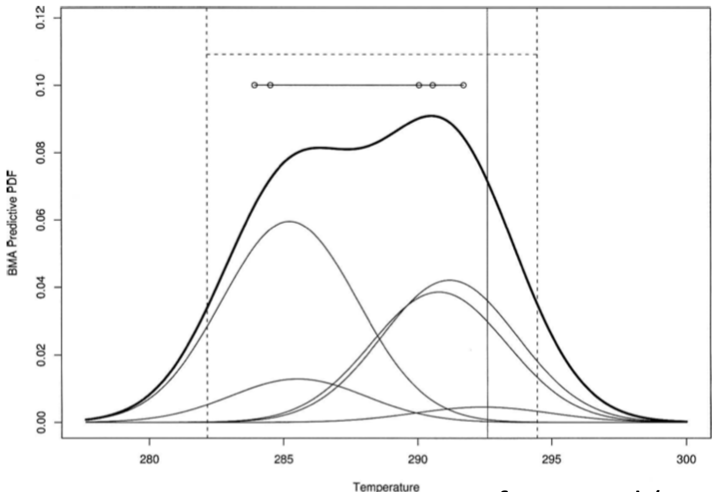
Educational changes that can facilitate advances in VVUQ

- Consider ideas and principles of VVUQ early in students' experience.
  - Confront and reflect on ways that knowledge is acquired, used, and updated
  - Statistical thinking, physical-systems modeling, numerical methods, and computing as core curriculum.
- VVUQ methods + computational modeling + HPC to empower the next generation of researchers to exploit opportunities at the interface of these fields.
- Develop examples of VVUQ done well for students, as well as practitioners.

Inference with multiple computational models

Bayesian Model Averaging

48 hour ahead temperature forecast



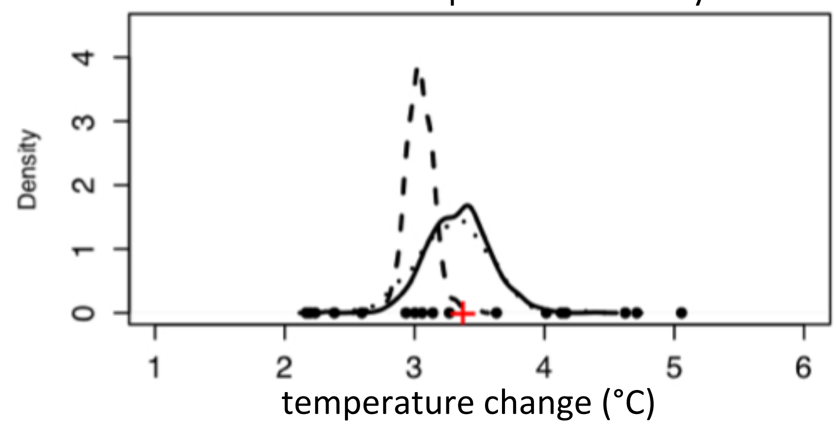
Raftery et al (2005)

$$\pi(T|\eta_1, \ldots, \eta_M) = \sum w_k \pi_k(T|\eta_k)$$

- Assumes physical observations come from the one, true physical process
- Resulting prediction uncertainty is a mixture of separate prediction uncertainties
- Estimation with historical data, typically via EM
- Very successful in data-rich settings

Hierarchical Model

Eastern North America – predicted change in mean summer temperature in 100 years



Tebaldi & Smith 2010

$$\pi(\Delta T|\eta_1, \ldots, \eta_M) \propto \prod \pi_k(\eta_k|T)$$

- Assumes each model prediction is centered about the truth
- Resulting prediction uncertainty depends on agreement between separate predictions and uncertainties
- Estimation typically via MCMC
- Predictions often better, but uncertainties can be misleading – depends on amount of shrinkage

Uncertainty Quantification

A definition from an extreme computing workshop Oct, 2009

UQ is the end-to-end study of the reliability of scientific inferences.

Ideally, UQ results in:

- a quantitative assessment of that reliability
- an inventory of possible sources of error and uncertainty in the inferences and predictions
- an inventory of the sources of error and uncertainty accounted for in the assessment
- an inventory of assumptions on which the assessment is based.

Note: nothing about computational models