SAND2007-2316C

Predictive Capability in Computational Science and Engineering

William L. Oberkampf

Distinguished Member Technical Staff
Validation and Uncertainty Quantification Department
Sandia National Laboratories, Albuquerque, New Mexico
505-844-3799, wloberk@sandia.gov

2007 Salishan Conference
Confidence in Predictive Simulations
April 23 - 26, 2007

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.

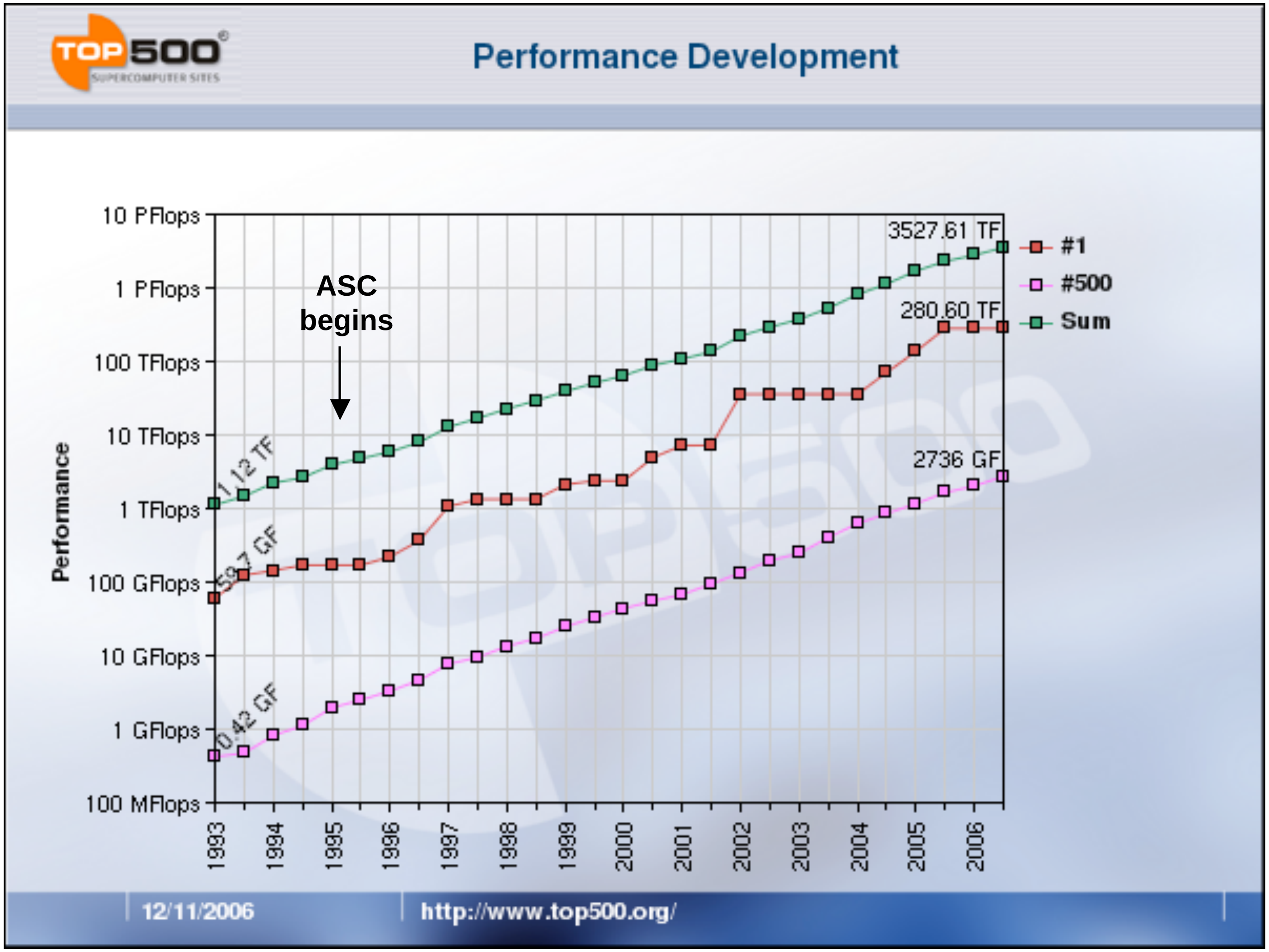
Outline of the Presentation

- **Background and perspectives of predictive capability**
- **Proposed perspective**
- **Validation metrics and predictive uncertainty**
- **Closing Remarks**

Work in collaboration with Marty Pilch and Tim Trucano, SNL,

and Scott Ferson and Jon Helton, consultants.

Progress in Computer Speed



Sandia National Laboratories

How do We Measure Progress in Predictive Capability?

- By the number of finite elements/volumes we have in a simulation?
- By the number of atoms/molecules we have in a simulation?
- By the size of the vortices we can resolve in a turbulent flow simulation?
- I contend that predictive capability for a system should be measured by how well we answer the questions posed by Kaplan and Garrick (1981):
  – What can go wrong?
  – How likely is it to happen?
  – What are the consequences?
- And “What is the maturity of numerical simulation activities?”

Our View of the Elements Contributing to Predictive Capability

- Identification of the scenarios, or initiating events, under which the system must operate, perform, fail safe, etc
- Fidelity of modeling of the physics, geometry, initial condition, boundary conditions, etc
- Completeness of the software quality and code verification activities
- Quantification of numerical accuracy of the discretized solutions
- Assessment of the accuracy of the simulation results by comparison with experimental measurements
- Estimation of the uncertainty in system responses due to all plausible sources of uncertainty
- Understanding the sensitivities of the system responses to all sources of uncertainty

Predictive Capability Maturity Model (Pilch, Oberkampf, Trucano)

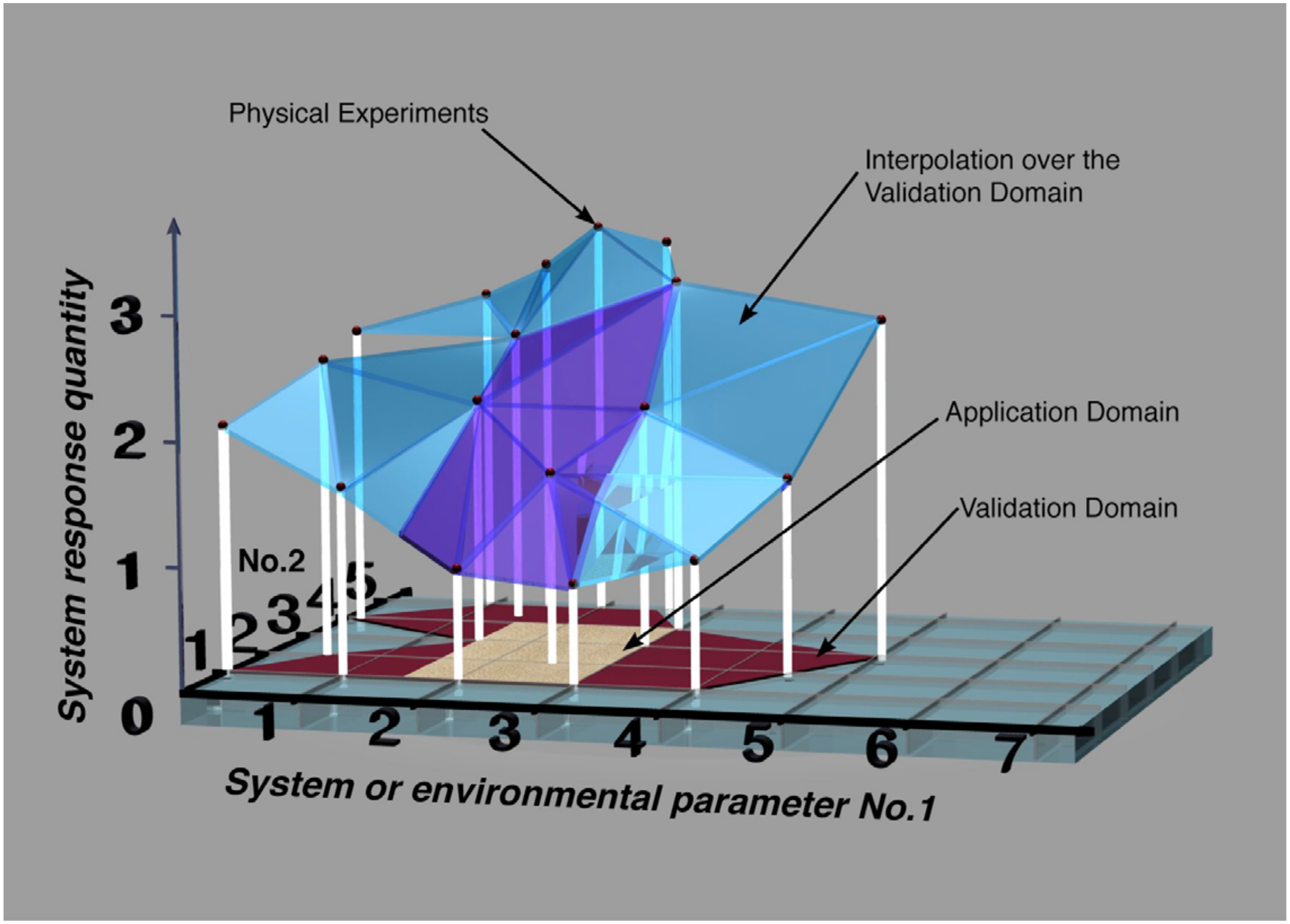
ATTRIBUTE \ MATURITY	Maturity Level 0 Low Consequence, Minimal M&S Impact, e.g. Scoping Studies	Maturity Level 1 Moderate Consequence, Some M&S Impact, e.g. Design Support	Maturity Level 2 High-Consequence, High M&S Impact, e.g. Qualification Support	Maturity Level 3 High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification
Representation and Geometric Fidelity Are important features neglected because of simplifications or stylizations?	• Judgment only • Little or no representational or geometric fidelity for the system and BCs	• Significant simplification or stylization of the system and BCs • Geometry or representation of major components is defined	• Limited simplification or stylization of major components and BCs • Geometry or representation is well defined for major components and some minor components	• Essentially no simplification or stylization of components in the system and BCs • Geometry or representation of all components is at the detail of "as built", e.g., gaps, material interfaces, fasteners, welds, adhesive bonding, surface finish
Physics and Material Model Fidelity How fundamental are the physics and material models and what is the level of model calibration?	• Model forms are either unknown or completely empirical • Few, if any, physics-informed models • No coupling of models	• Some models are physics-based and are calibrated using data from related systems • Minimal or ad hoc coupling of models	• Physics-based models for all important physics • Significant calibration needed using Separate Effects Tests (SET) and Integral Effects Tests (IET) • One-way coupling of models	• All models are physics-based • Minimal need for calibration using SETs and IETs • Sound physical basis for extrapolation and coupling of models • Full, two-way, coupling of models
Code Verification Are software errors and algorithm deficiencies corrupting the simulation results?	• Judgment only • Minimal testing of any software elements • Little or no SQE procedures specified or followed	• Most codes managed by SQE procedures • Unit and regression testing conducted with significant code coverage	• All codes managed by SQE procedures • Verification test suites regularly used for key algorithms and coverage of key Features & Capabilities (F&C) used	• SQE procedures reviewed by independent, external panel • Test suites conducted for all important algorithms, all important F&Cs used, all important coupled physics, and all important coupled codes
Solution Verification Are human procedural errors or numerical solution errors corrupting the simulation results?	• Judgment only • Numerical errors have an unknown or large effect on simulation results	• Effect of numerical errors and parameters is small for some relevant SRQs • Input/output verified only by the analysts	• Numerical effects are quantitatively estimated to be small on most relevant SRQs • Some input/output data verified by experts internal to the organization	• Numerical effects are quantitatively estimated to be small on all important SRQs for all codes and code couplings • All input/output data verified by independent, external experts
Model Validation How accurate are the simulation results at various tiers in a validation hierarchy?	• Judgment only • Few, if any, comparisons with measurements from similar systems	• Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest • Large or unknown experimental uncertainties	• Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs • Experimental uncertainties are well characterized for most SETs, but poorly known for IETs	• Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application • Experimental uncertainties are well characterized for all IETs and SETs
Uncertainty Quantification and Sensitivity Analysis What is the impact of variabilities and uncertainties on system performance and margins?	• Judgment only • Only deterministic analyses conducted for system margins • Informal "what if" analyses conducted for system margins	• Aleatory and epistemic (A&E) uncertainties represented and propagated without distinction • Sensitivities to some uncertainties and conditions are explored	• A&E uncertainties segregated, propagated and properly interpreted • Quantitative sensitivity analyses conducted for some uncertainties • Some environments and scenarios of the system are analyzed • Minimal estimation of margins due to extrapolation of models	• A&E uncertainties due to all plausible environments and scenarios of the system are analyzed • Comprehensive sensitivity analyses conducted for parameters and models • Extensive estimation of system margins due to extrapolation of models and physics-coupling effects

Sandia National Laboratories

Approaches to Predictive Capability

- **Traditional (risk assessment) approach:**
  - Characterize all sources of uncertainty, aleatory and epistemic
  - Calibrate deterministic model parameters
  - Use the model to extrapolate in space, time, boundary conditions, forcing functions, loading conditions, etc. to the application of interest
- **Bayesian approach:**
  - Assume prior distributions for uncertain parameters in the model
  - Update the prior distributions for uncertain parameters using available experimental data and Bayes formula
  - Use the updated parameters in the model to make predictions for the application of interest
  - Disadvantages:
    - Assumes the key issue is calibrating parameter distributions
    - Assumes the model form is accurate
    - Is computational very expensive

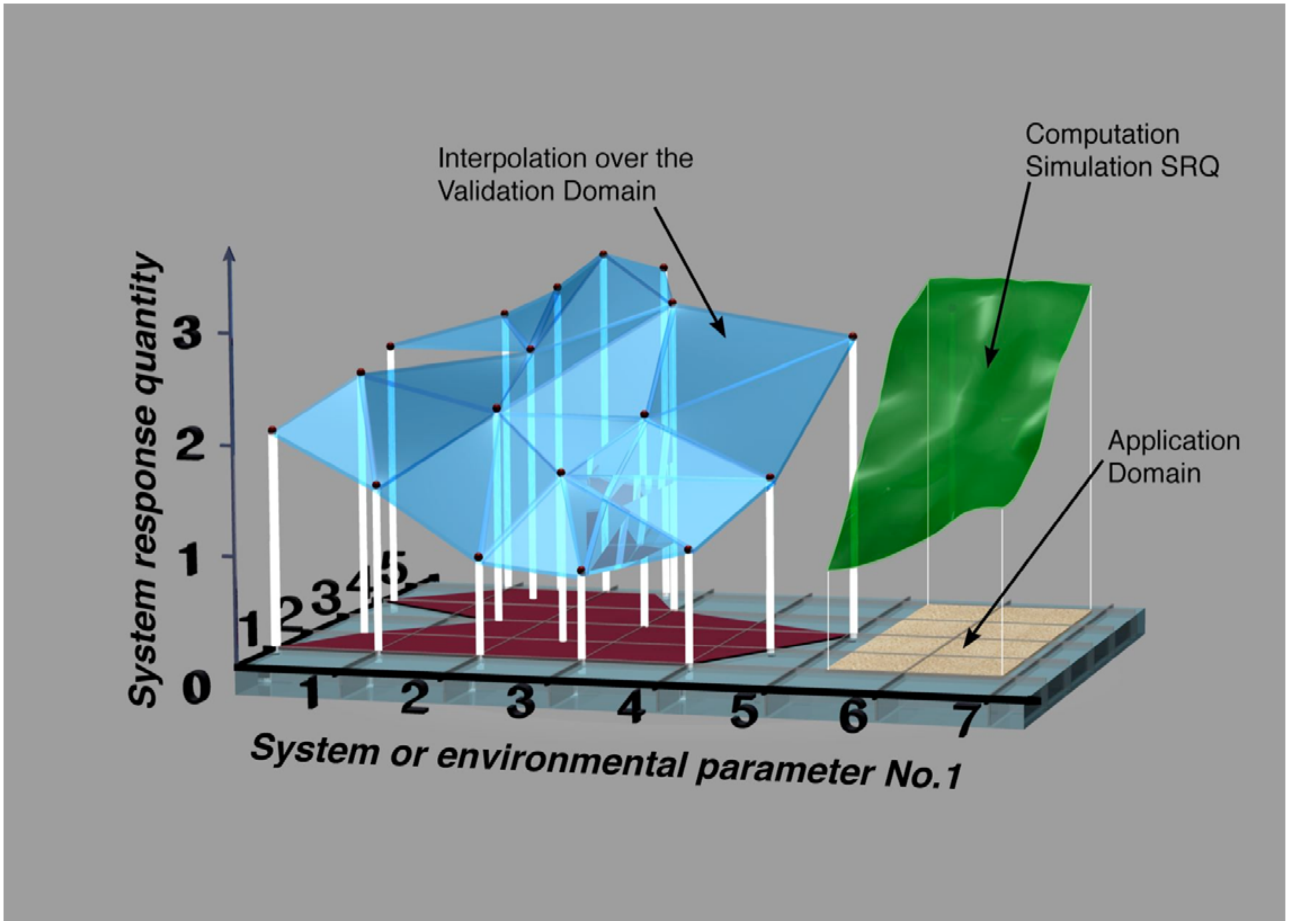
Interpolation: Application Domain Within the Validation Domain



Proposed Perspective to Predictive Capability

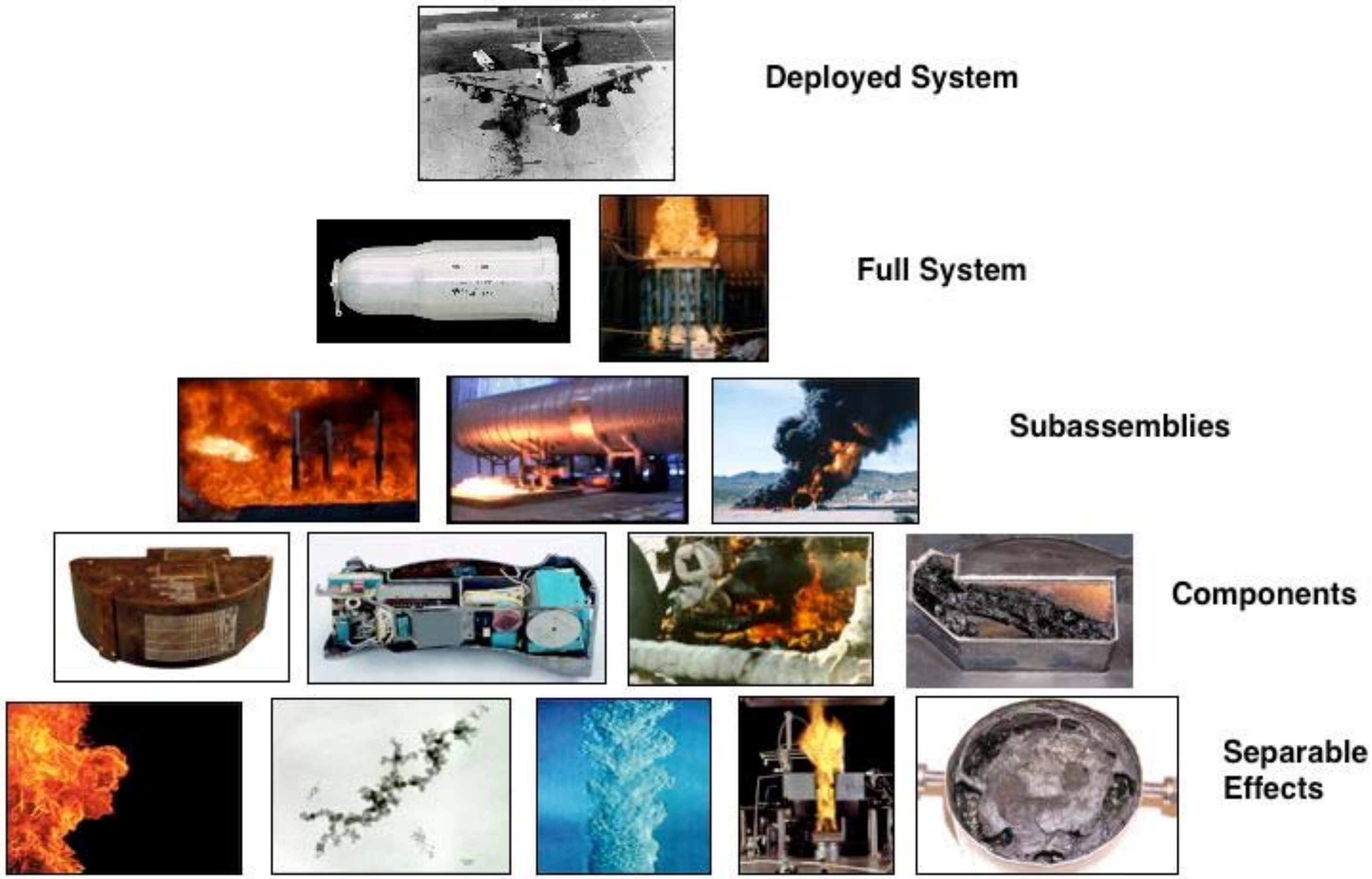
- Characterize all of the uncertainties:
  - Aleatory: inherent variation associated with the parameter
  - Epistemic: uncertainty due to lack of knowledge of the quantity
- Calibrate uncertain model parameter distributions before model validation activities
- Assess the model accuracy by quantitative comparisons with experimental validation data, i.e., compute a validation metric
- Use the model to extrapolate to the application of interest:
  - Extrapolate in space, time, boundary conditions, forcing functions, loading conditions, etc. to the application of interest
  - Model-form inaccuracies directly estimated from validation experiments
- Advantages over traditional and Bayesian approaches:
  - Proven to be very effective in identifying weaknesses in models
  - Better able to estimate model-form uncertainty when using the model:
    - Far from the conditions of the validation experiments
    - When the complete system can not be tested

Large Extrapolation Beyond the Validation Domain



Sandia National Laboratories

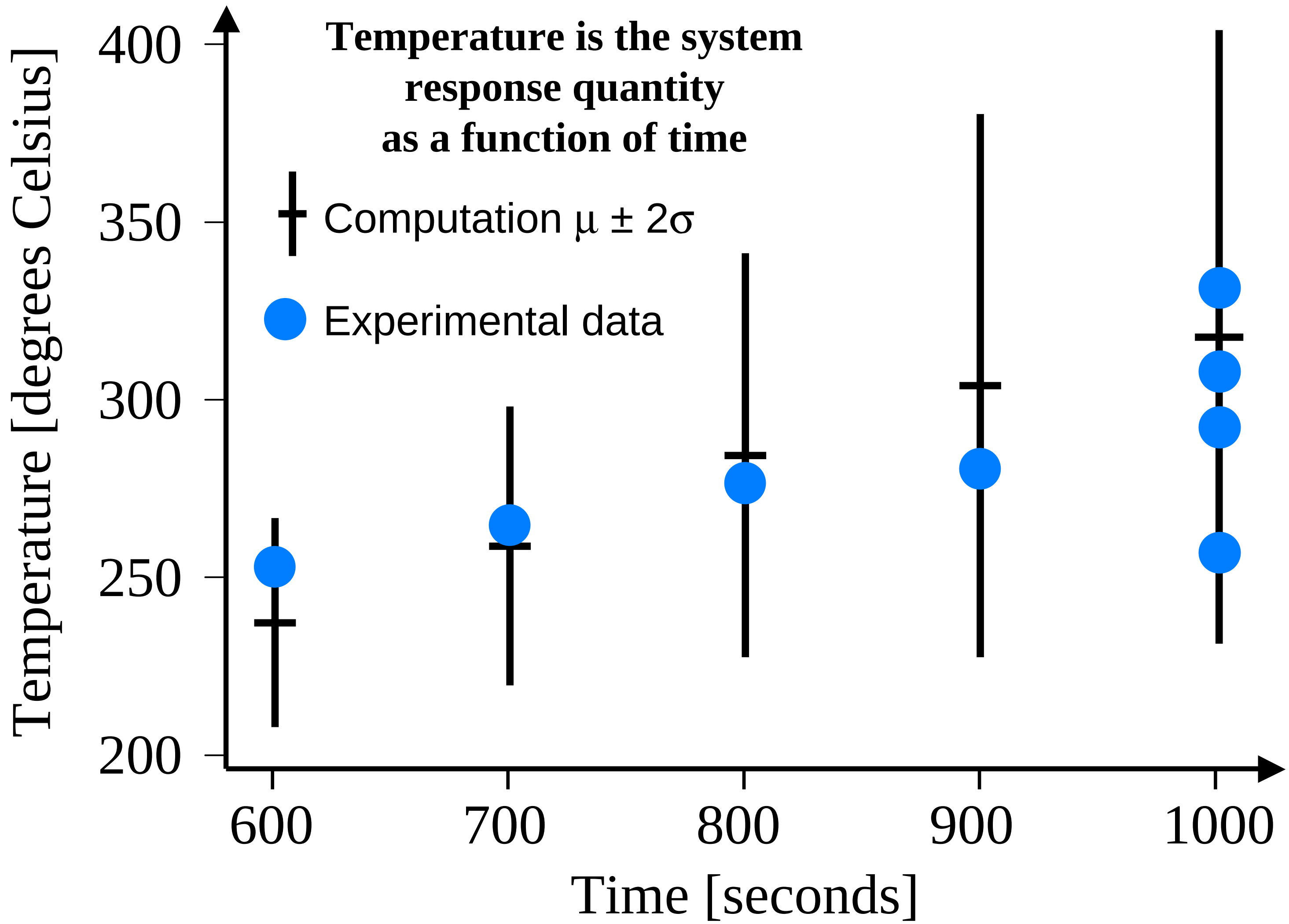
Example of Extrapolation Within a Validation Hierarchy (Weapon in a Fire)



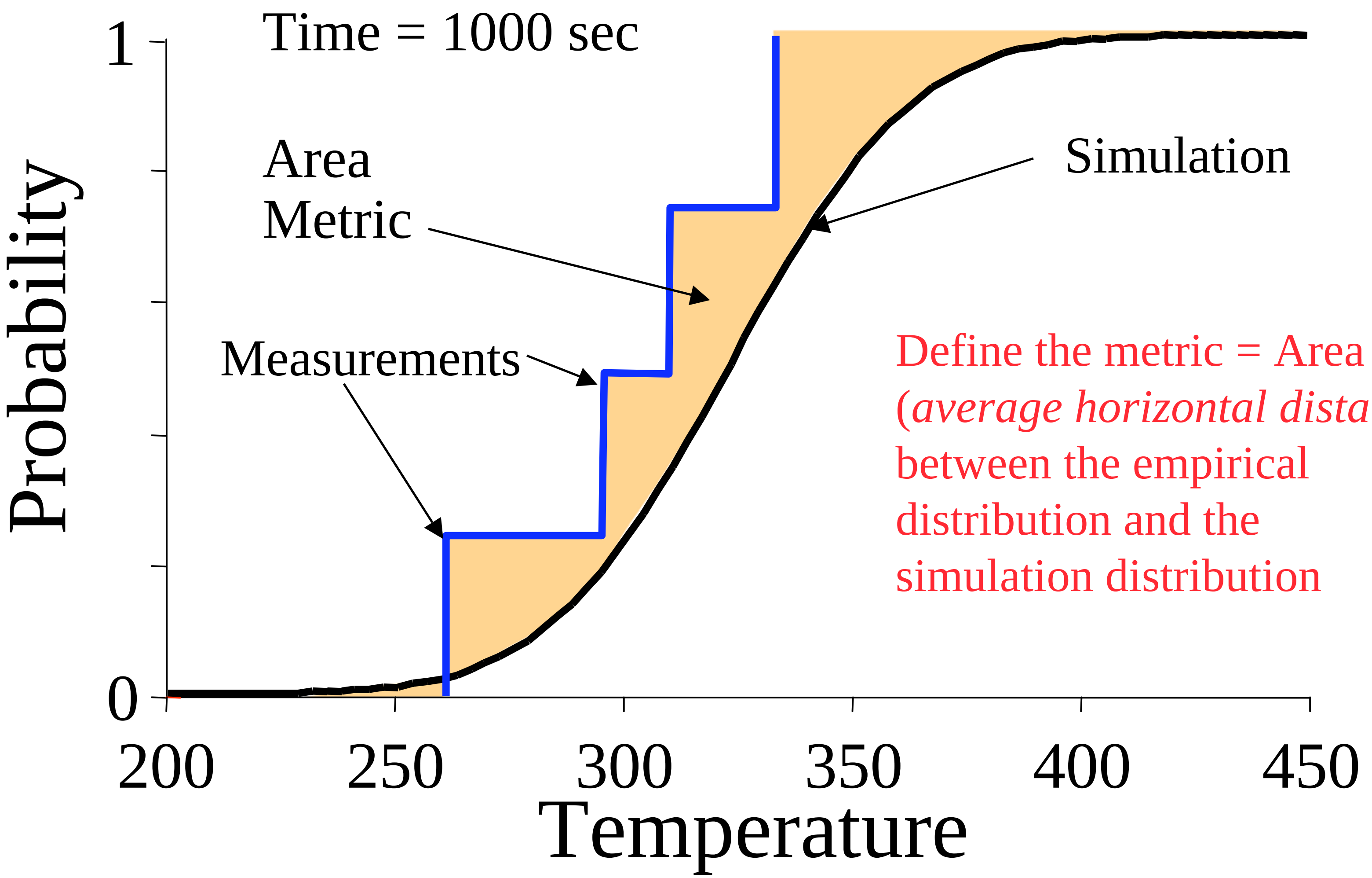
Desirable Validation Metric Characteristics

- Validation metric is a measure of the mismatch between the model prediction and the experimental data
- Should be a statistical “distance” between the distribution of the prediction and distribution of the experimental data
- Should be expressed in physical units, not normalized relative to some statistical measure
- Should not mix calibration of the model and accuracy assessment of the model
- Should be sensitive to how many function evaluations (numerical solutions to the computational model) are available
- Would be very useful if the validation metric could be computed when only one experimental realization is available

Typical Method of Comparison of Computation and Experimental Data

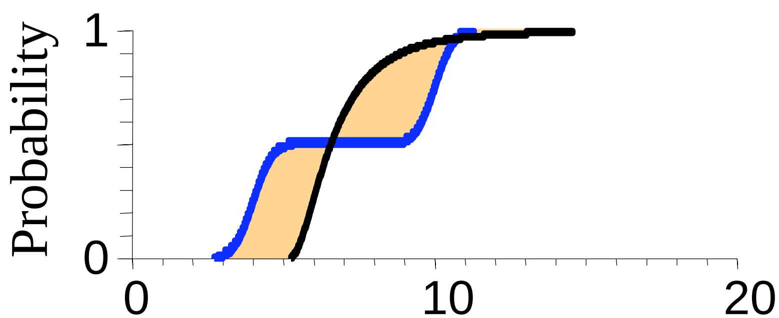


Compare the Simulation and Data Using the Cumulative Distribution Function

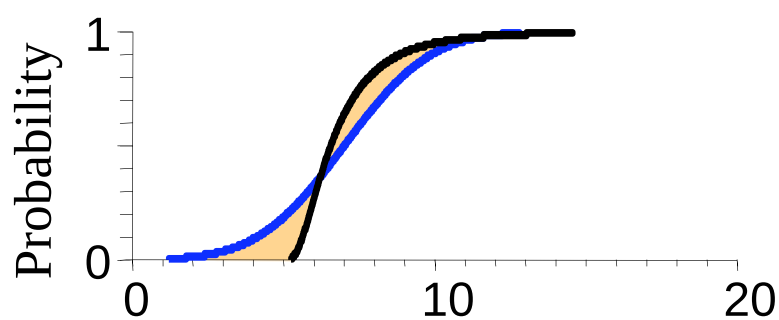


Define the metric = Area (*average horizontal distance)* between the empirical distribution and the simulation distribution

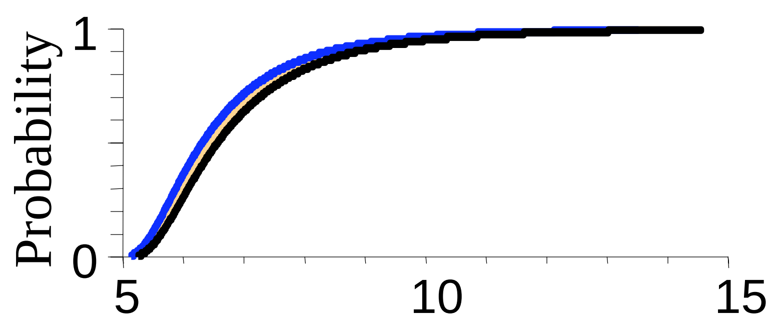
Validation Metric Reflects the Difference Between the Full Distributions



Matches in mean



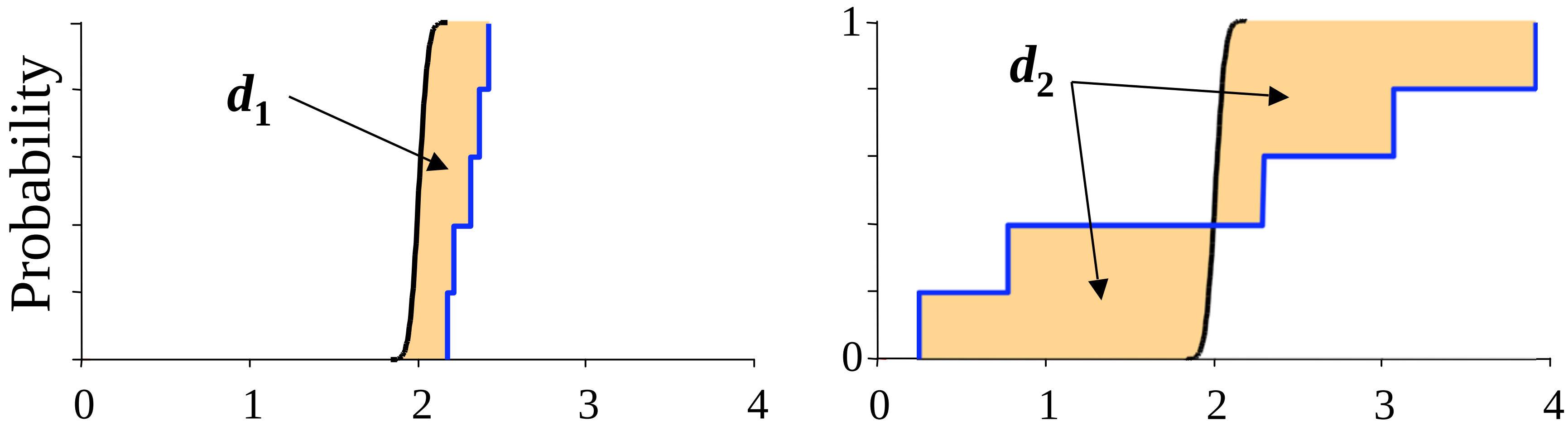
Both mean and variance



Matches well overall

Sandia National Laboratories

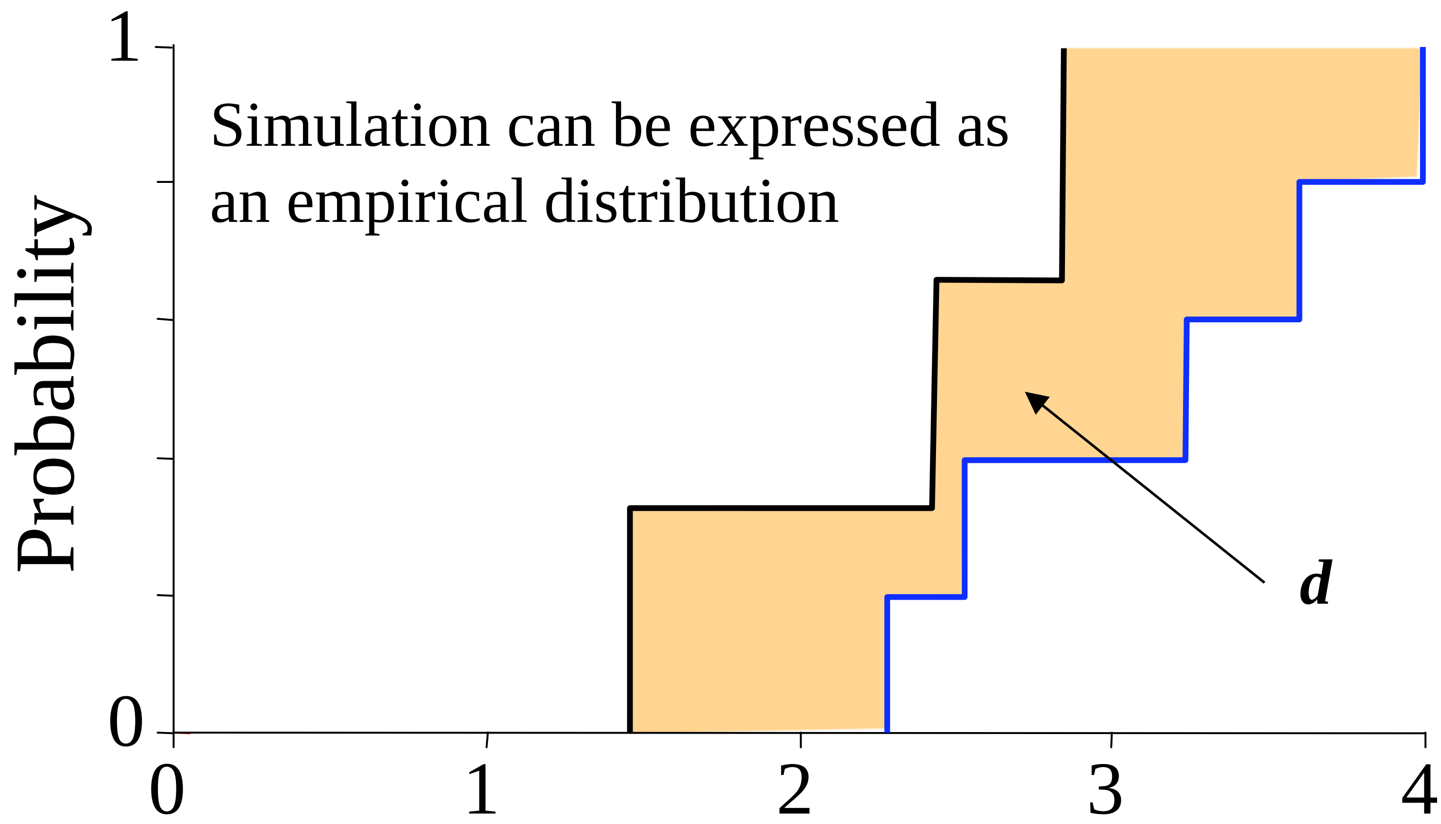
Why Require Physical Units for the Validation Metric?



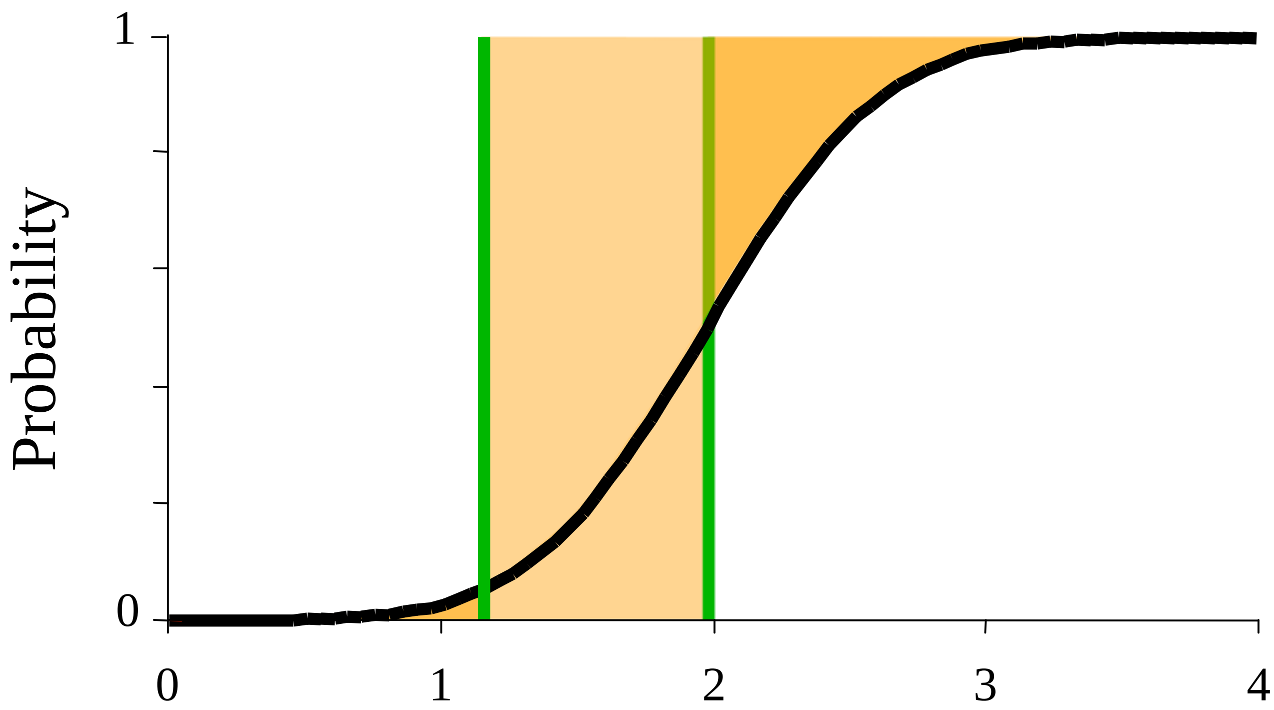
The simulation on the left is much closer to the experimental data than the simulation on the right

$$d_1 << d_2$$

Effect of Few Function Evaluations on the Validation Metric

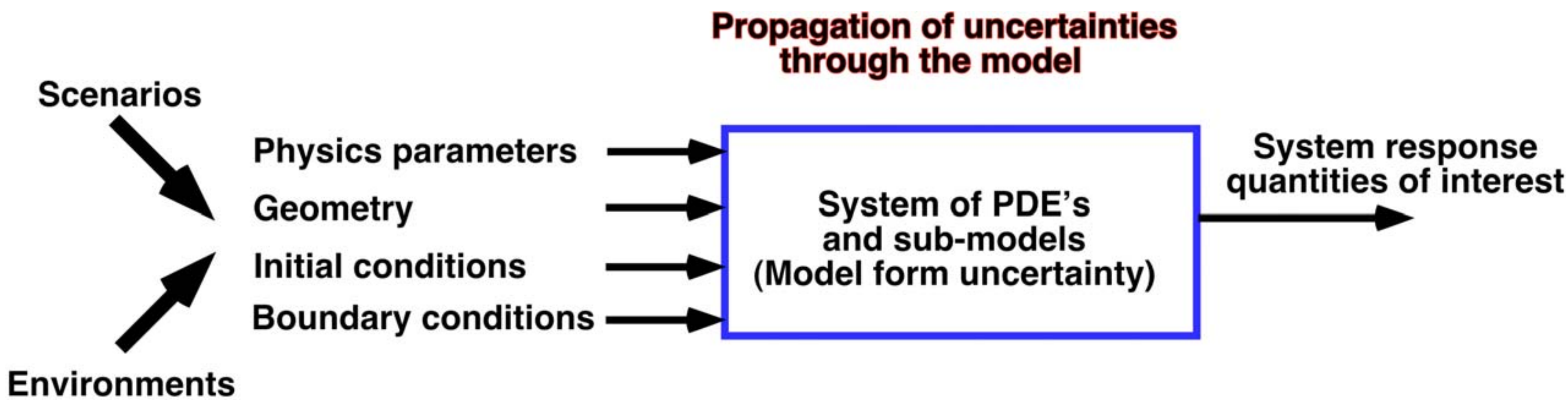


Single Observation (two of them)



- **A single datum can never match the entire predicted distribution, $d \neq 0$**
- **Single datum has a minimum value of d when it matches the median of the predicted distribution**

Propagation of Uncertainties

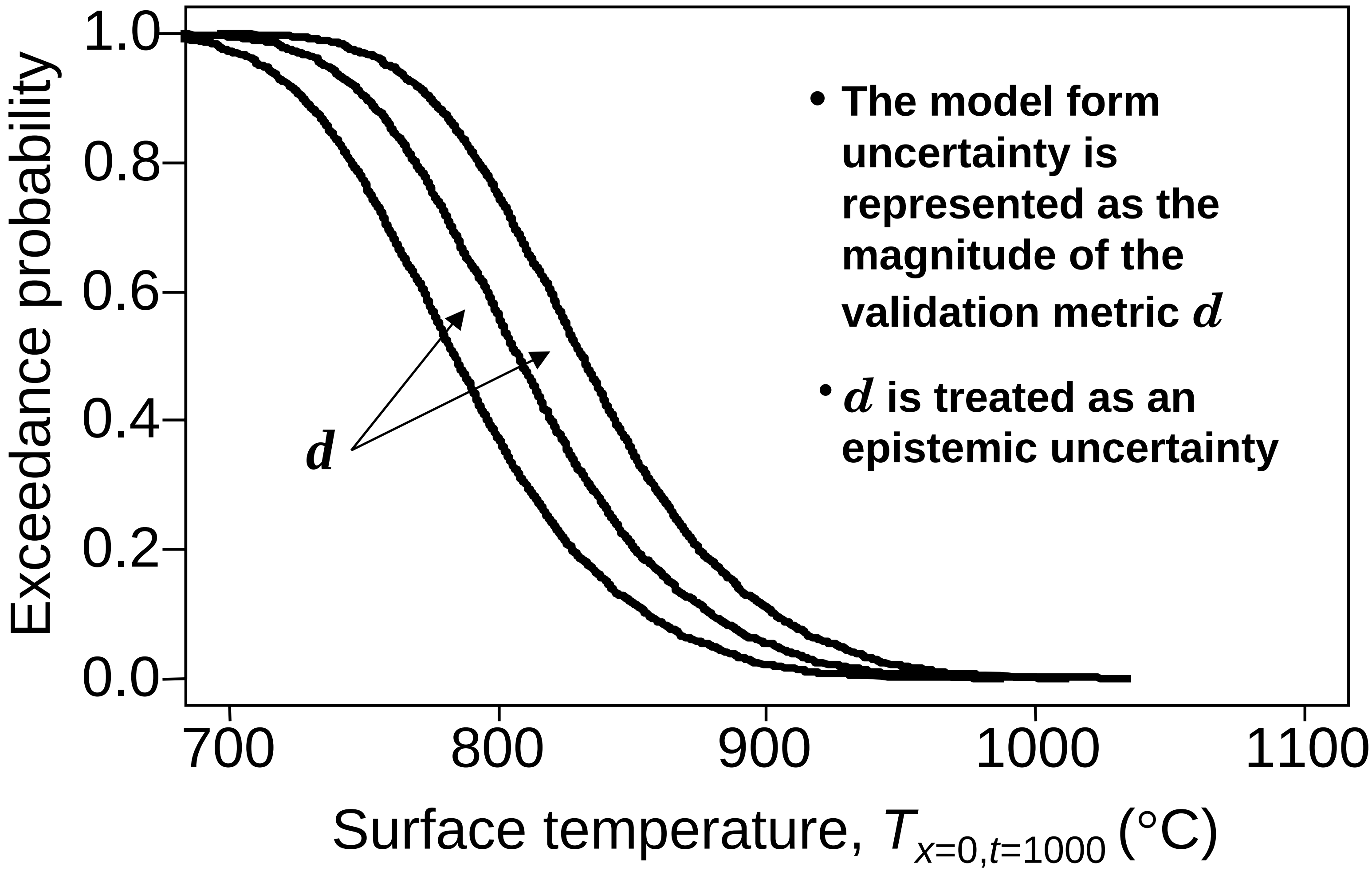


The propagation of uncertain input quantities through a mathematical model to obtain outputs can be written as

$$y = f(\vec{x}_a, \vec{x}_e)$$

- y **is a system response quantity of interest**
- f **is the mathematical model of the physical process of interest**
- $\vec{x}_a = x_1, x_2, \cdots x_m$ **is the vector of all aleatory uncertainties**
- $\vec{x}_e = x_{m+1}, x_{m+2}, \cdots x_n$ **is the vector of all epistemic uncertainties**

Prediction with Extrapolation of Aleatory and Epistemic Uncertainties



- **The model form uncertainty is represented as the magnitude of the validation metric** d
- d **is treated as an epistemic uncertainty**

Sandia National Laboratories

Concluding Remarks

- **Bayesian estimation improvements needed:**
  - **Develop better methods to separate parameter estimation and model bias error identification**
  - **Develop methods to better estimate uncertainty in predictions**
- **Improvements needed in the present approach:**
  - **Improve methods for extrapolation of the validation metric, d**
  - **Devise methods for estimating the uncertainty due to physics couplings that have not been assessed by experimental validation**
- **We must continue to find ways of testing our predictive capability by "blind" comparisons with experiments**