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

OASCR Applied Mathematics PI Meeting
Lawrence Livermore National Laboratory
May 22-24, 2007

NNSA National Nuclear Security Administration

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.

Sandia National Laboratories

Outline of the Presentation

- **Background and perspectives of predictive capability**
- **Approaches to uncertainty quantification**
- **Distinction between aleatory and epistemic uncertainties**
- **Key areas of concern in extrapolation of models**
- **Concluding remarks**

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

and Scott Ferson and Jon Helton, consultants.

Sandia National Laboratories

What is Predictive Capability in Science and Engineering?

- Is it the speed of the computer?
- Is it the number of finite elements we have in a simulation?
- Is it the number of atoms/molecules we have in a simulation?
- From a science perspective, predictive capability could be viewed as the ability to generate new knowledge
- From an engineering perspective, I contend that predictive capability should be viewed by how well we answer the questions posed by Kaplan and Garrick (1981):
  - What can go wrong?
  - How likely is it to go wrong?
  - What are the consequences of going wrong?

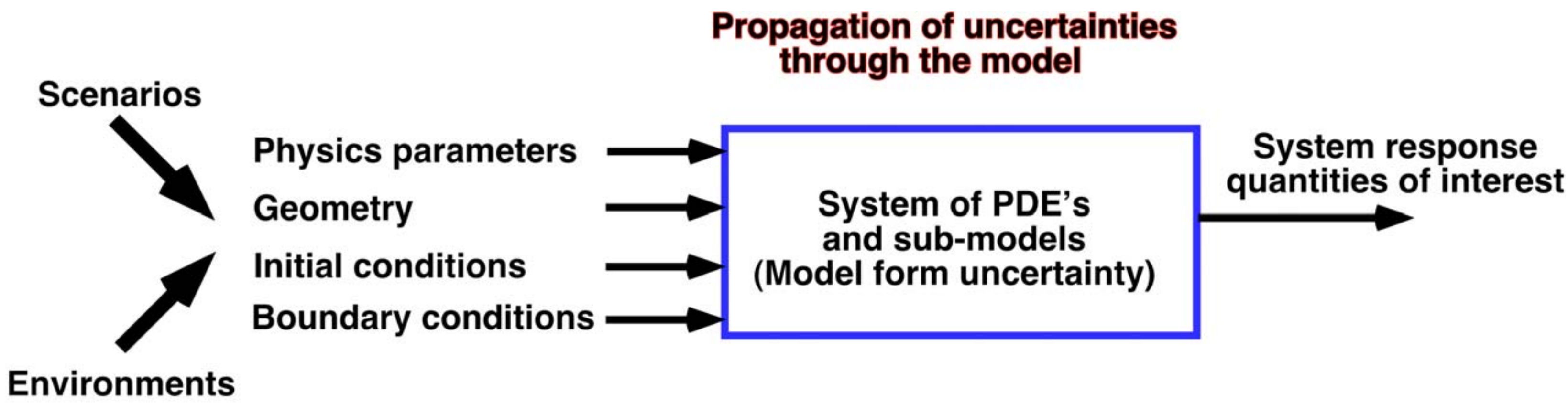
Approaches to Uncertainty Quantification

- **Risk assessment approach taken in:**
  - **Nuclear reactor safety**
  - **Underground storage of nuclear waste (Waste Isolation Pilot Plant and Yucca Mountain Project)**
- **Key steps in quantitative risk assessment (QRA):**
  - **Identify initiating events, fault trees, and event trees**
  - **Characterize all sources of uncertainty according to aleatory and epistemic**
  - **Propagate uncertainties through the computational model**
  - **Characterize system responses according to aleatory and epistemic uncertainty**
  - **Conduct sensitivity analysis to determine major sources of uncertainty in system responses**

Aleatory and Epistemic Uncertainty

- **Aleatory uncertainty is an inherent variation associated with the physical system or the environment**
  - **Also referred to as variability, irreducible uncertainty, and stochastic uncertainty, random uncertainty**
- **Examples:**
  - **Variation in weather conditions**
  - **Variation in manufacturing and assembly of systems**
- **Epistemic uncertainty is an uncertainty that is due to a lack of knowledge of quantities or processes of the system or the environment**
  - **Also referred to as subjective uncertainty, reducible uncertainty, and model form uncertainty**
- **Examples:**
  - **Lack of experimental data to characterize new materials and processes**
  - **Poor understanding of physics phenomena**
  - **Lack of experimental data/testing for complete systems**

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**

Approaches to Representation of Aleatory and Epistemic Uncertainties

- **Second-order probabilistic analysis:**
  - Use a two step process separating epistemic and aleatory uncertainties
  - Treat the range all epistemic uncertainties as possible realizations with no probability associated with realizations from sampling
  - Treat aleatory uncertainties as random variables
- **Robust Bayesian inference:**
  - Investigate the effect of different assumptions of prior distributions
  - Investigate the effect of partitioning the available data
- **Evidence theory:**
  - Can represent aleatory and epistemic uncertainties within one framework
  - Early criticism misdirected at Dempster's rule of aggregation of evidence
  - Early applications have been very successful

Mathematical Structure of Evidence Theory

- **Let the universal set (or sample space) be defined as**

$$\mathscr{X} = \{x : x \ is \ a \ possible \ value \ of \ the \ uncertain \ quantity\}$$

- **Based on the information available concerning uncertain quantities, a basic probability assignment (BPA) can be defined as**

$$m(\mathscr{E}) \geq 0 \ for \ \mathscr{E} \subset \mathscr{X}$$

$$\sum_{\mathscr{E} \subset \mathscr{X}} m(\mathscr{E}) = 1$$

- **Then the plausibility function can be defined as**

$$Pl(\mathscr{E}) = \sum_{\mathscr{U} \cap \mathscr{E} \neq \varnothing} m(\mathscr{U})$$

- **And the belief function can be defined as**

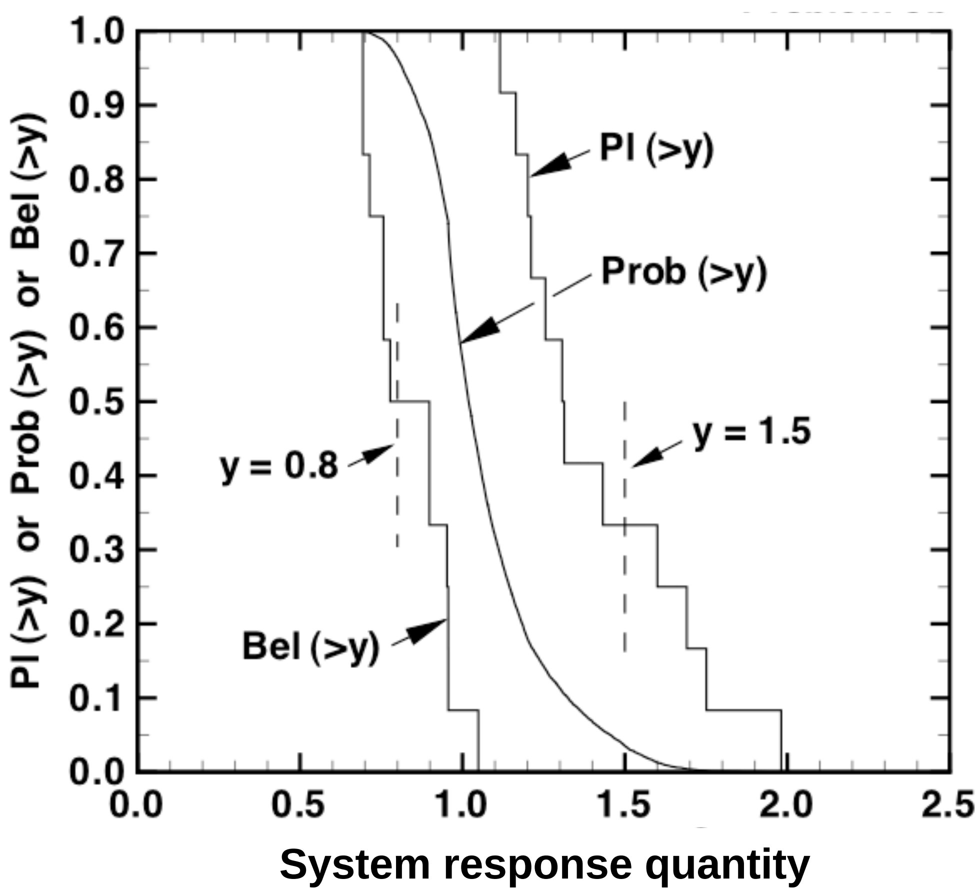
$$Bel(\mathscr{E}) = \sum_{\mathscr{U} \subset \mathscr{E}} m(\mathscr{U})$$

- **Plausibility and belief are super-additive and sub-additive, respectively**

$$Pl(\mathscr{E}) + Pl(\mathscr{E}^c) \geq 1 \qquad\qquad Bel(\mathscr{E}) + Bel(\mathscr{E}^c) \leq 1$$

Characterization of System Response Quantity

Complementary Cumulative Plausibility and Belief over system response



- **It can be shown that**

$$CCBF(\mathscr{Y}_v) \le CCDF(\mathscr{Y}_v) \le CCPF(\mathscr{Y}_v)$$

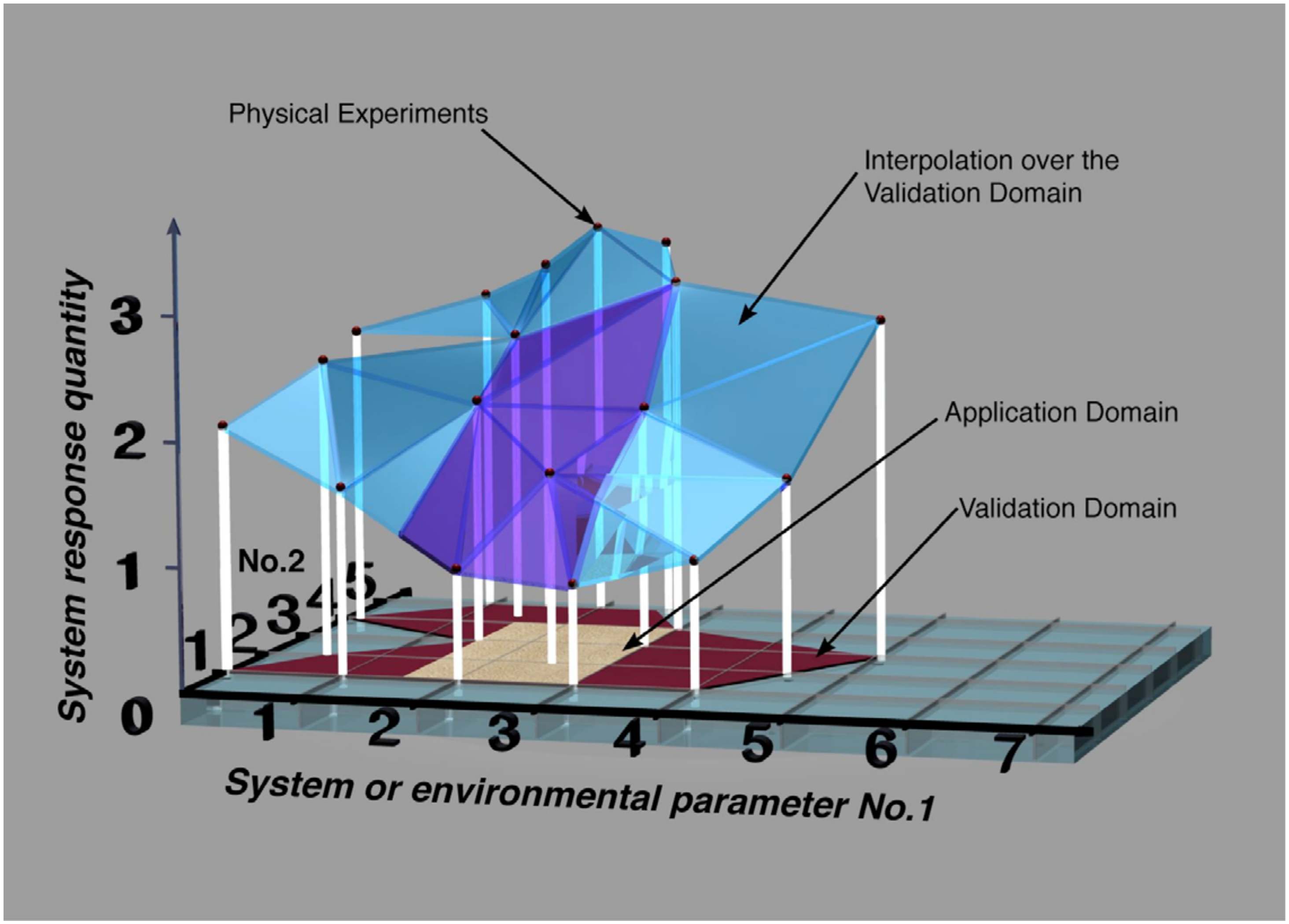
- **Given the epistemic uncertainties, the probability of a given system response value can only be given as an interval-valued probability**
- **Second-order probability yields an ensemble of CCDFs**

Sandia National Laboratories

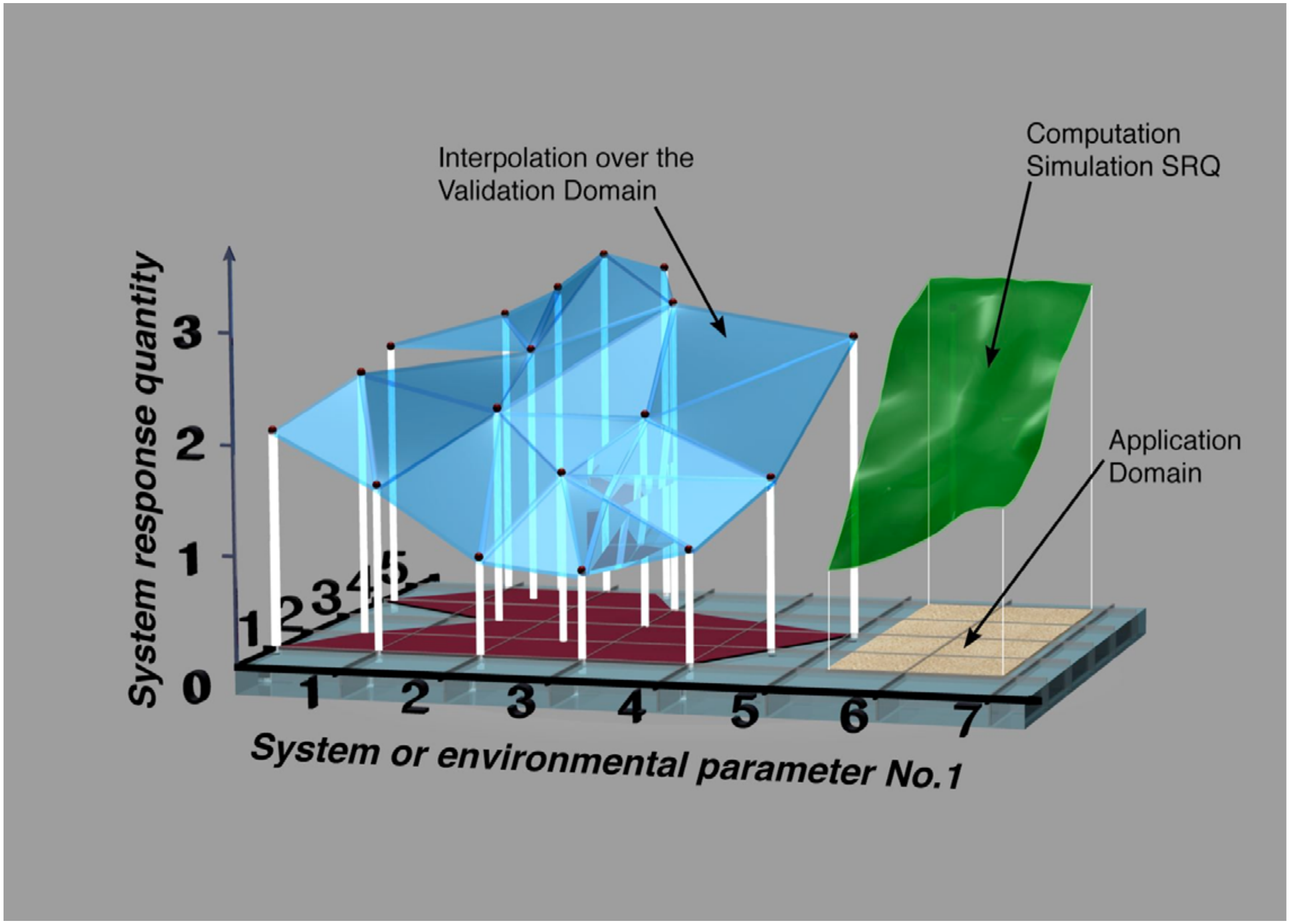
Bayesian Approach to Uncertainty Quantification

- **Key steps in 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**

Typical Application of Bayesian Inference: Interpolation



Key Area of Concern: Large Extrapolation in a Model Parameter

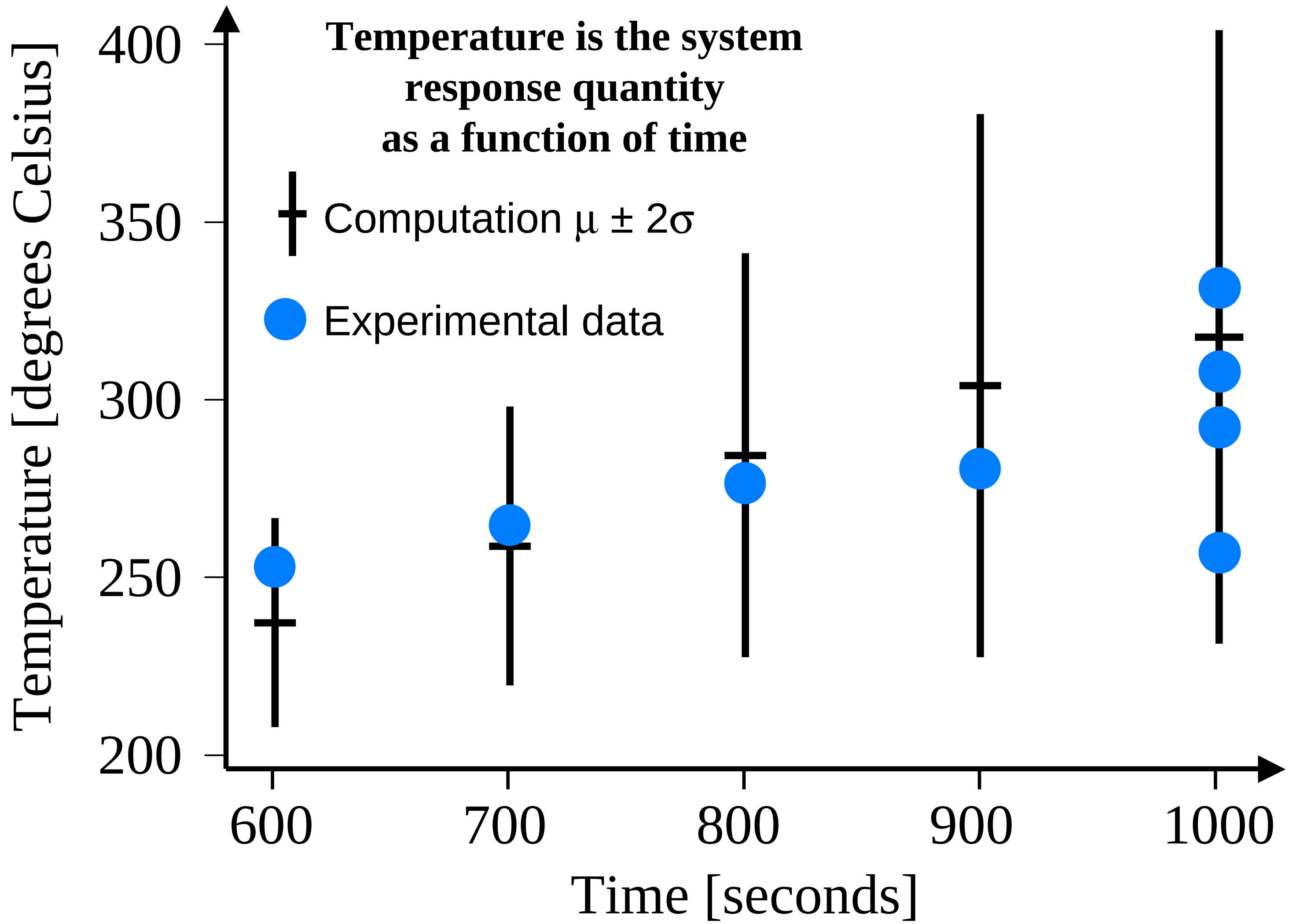


Sandia National Laboratories

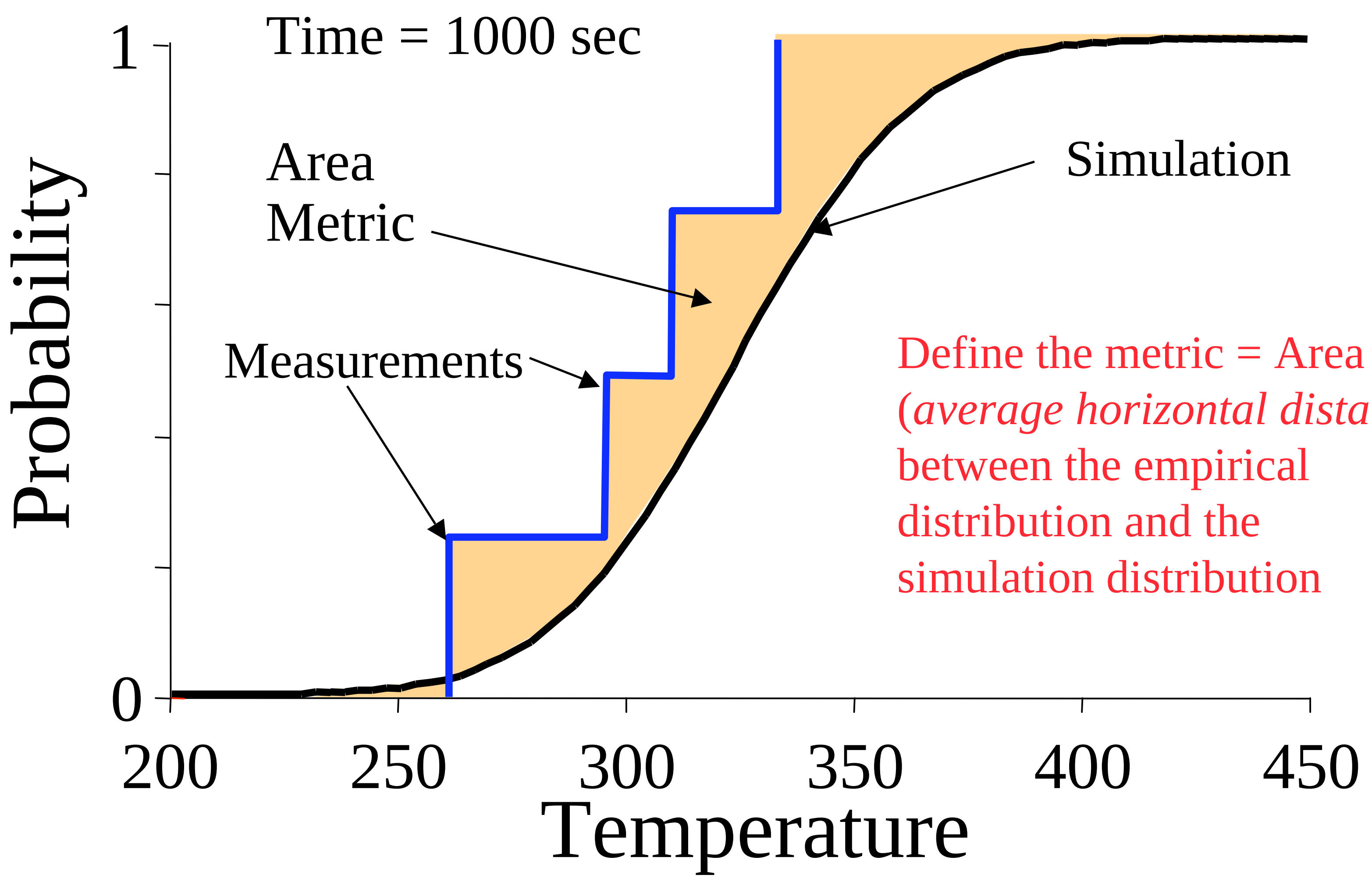
Key Area of Concern: Extrapolation of a Validation Metric Result

- What is a validation metric?
- A quantitative measure of the mismatch between the CDFs from the computational model and the experimental data
- A “distance” between the CDFs measured in terms of dimensional units of the system response quantity
- The primary purpose of the validation metric is measure the predictive accuracy of the physics model, not calibration of the model
- If experimental data is limited, the validation metric results can either:
  - Increase
  - Remain the same and decrease the confidence in the validation metric result

Typical Method of Comparison of Computation and Experimental Data

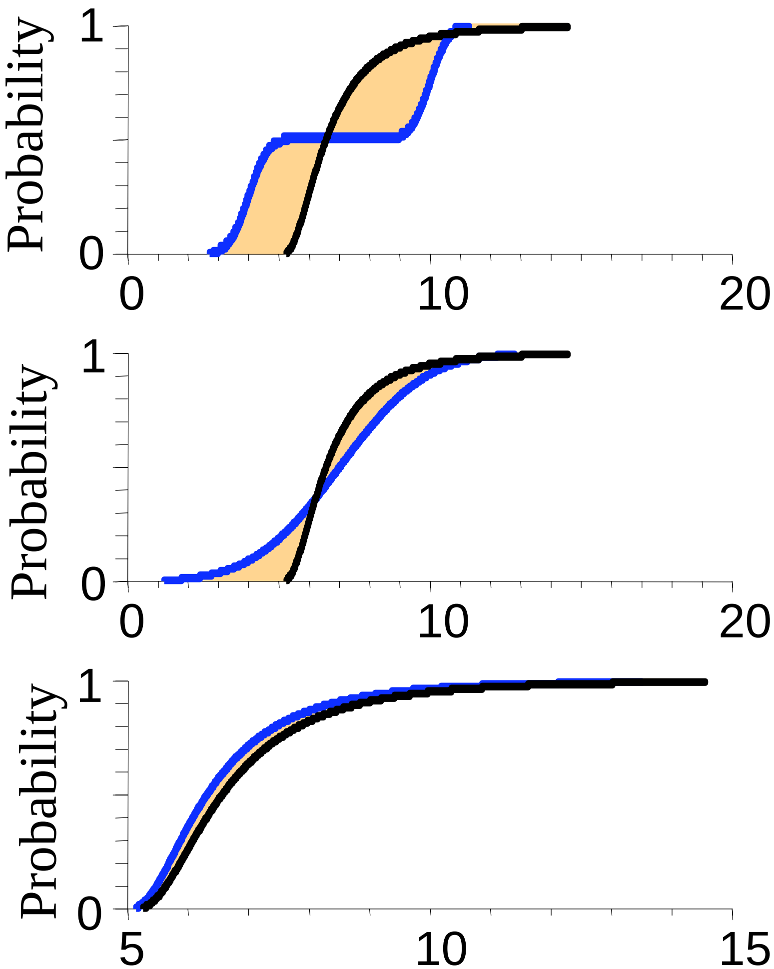


Compare the Simulation and Data Using the Cumulative Distribution Function



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

Validation Metric Reflects the Difference Between the Full Distributions



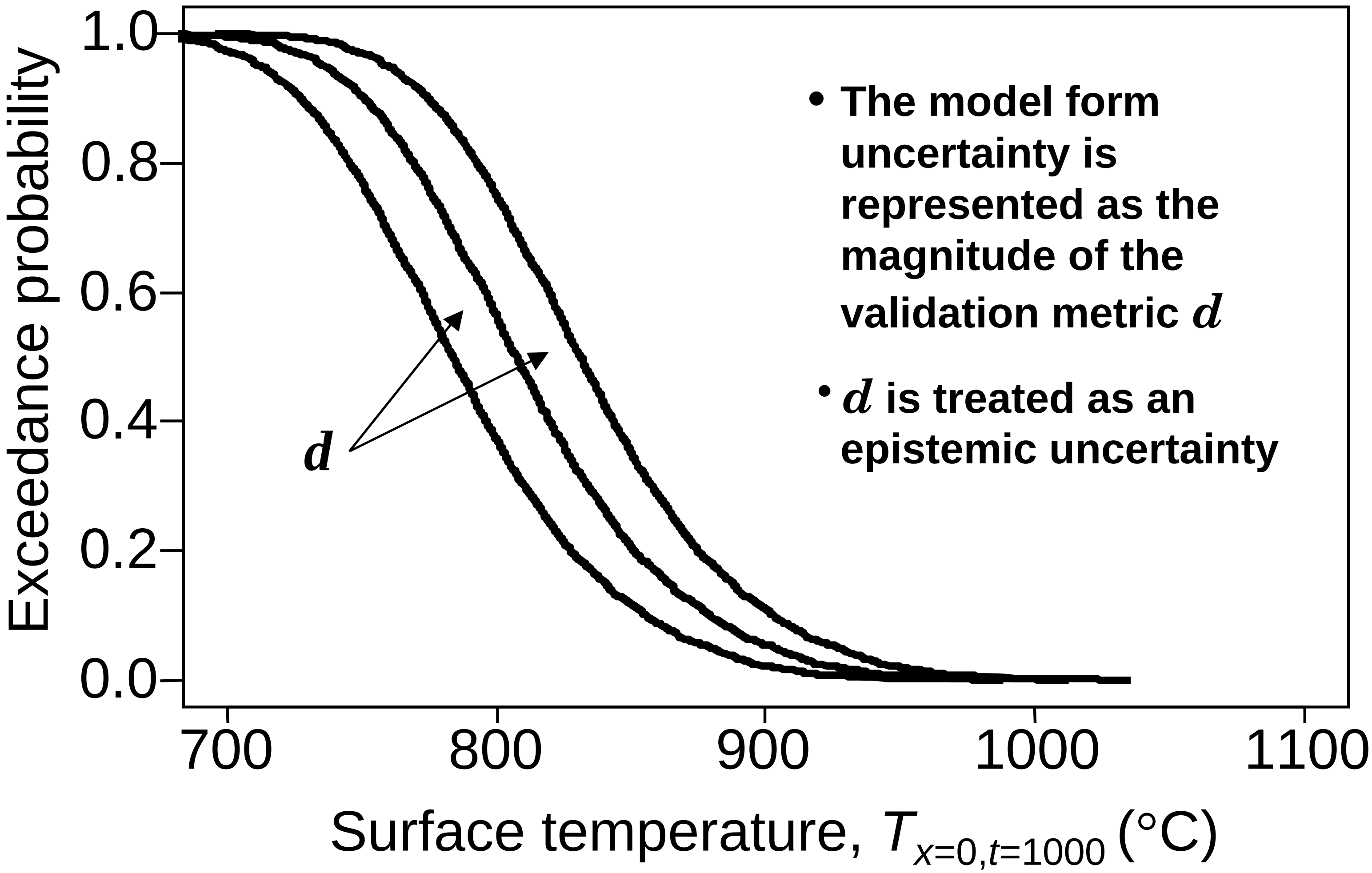
Matches in mean

Both mean and variance

Matches well overall

Sandia National Laboratories

Prediction with Extrapolation of Aleatory and Epistemic Uncertainties

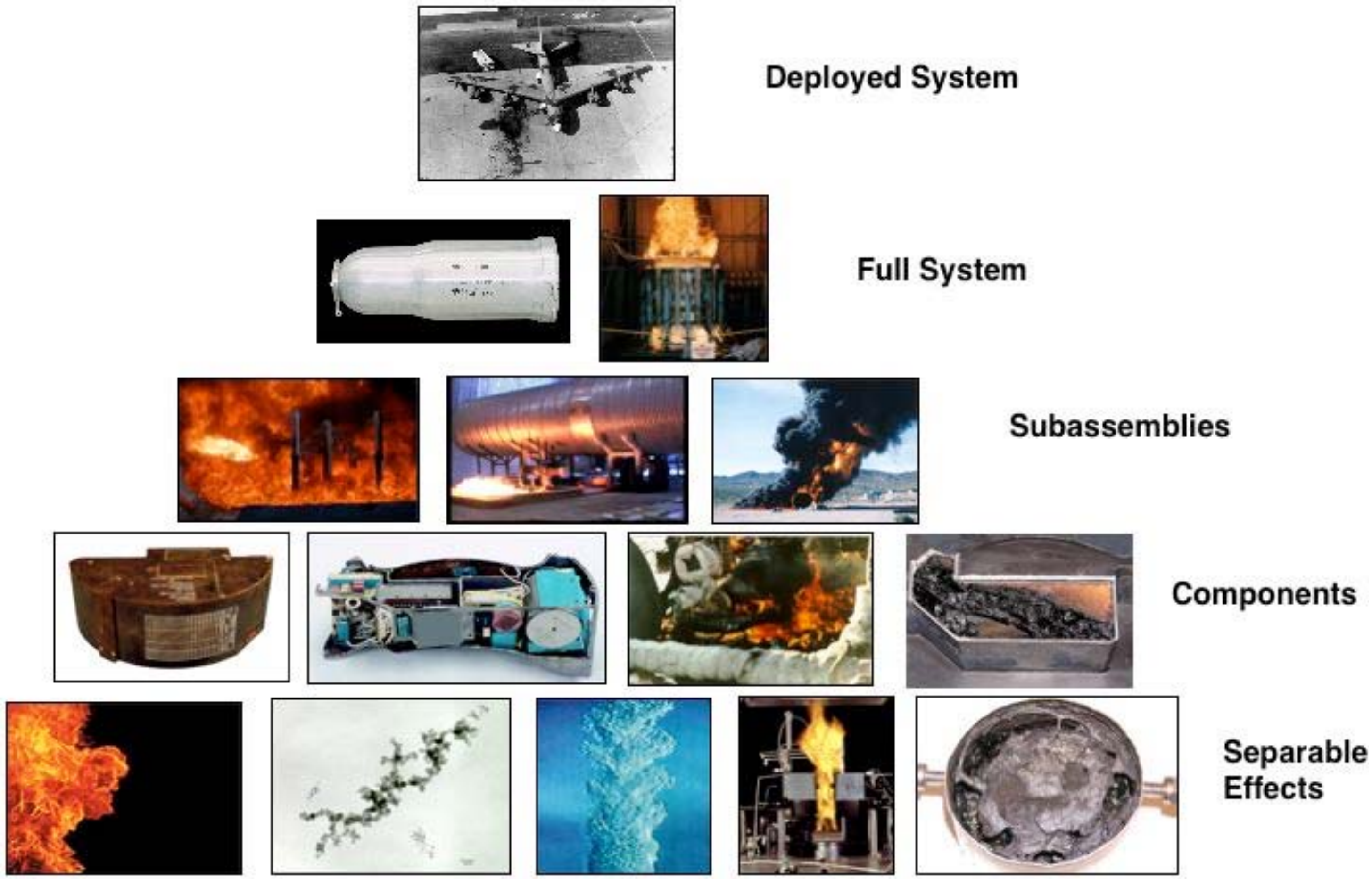


- The model form uncertainty is represented as the magnitude of the validation metric d
- d is treated as an epistemic uncertainty

Key Area of Concern: No Experimental Data on Coupled Physics

- No experimental data, and no validation metric result, is available for:
  - Physics that exist at the same level in the validation hierarchy as where other physics models can be evaluated
  - Coupled physics that only exists at higher levels in the validation hierarchy
- Sandia experience for both of these situations has shown that model accuracy is commonly poor
- This is a model form inaccuracy due to coupled physics
- Possible approaches to estimate this epistemic uncertainty:
  - Alternate physics modeling approaches
  - Hierarchical physics models

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



Concluding Remarks

- Predictive capability in engineering decision making relies on a clear representation of aleatory and epistemic uncertainties
- Improvements needed in evidence theory:
  - Understanding of dependence between epistemic uncertainties
  - Understanding of sensitivity analysis for epistemic uncertainties
- Improvements needed in Bayesian inference:
  - Develop better methods to separate parameter estimation and model bias error identification
  - Develop methods to better estimate uncertainty in predictions
- Improvements needed in uncertainty quantification due to:
  - Extrapolation of a validation metric result
  - No experimental data for coupled physics