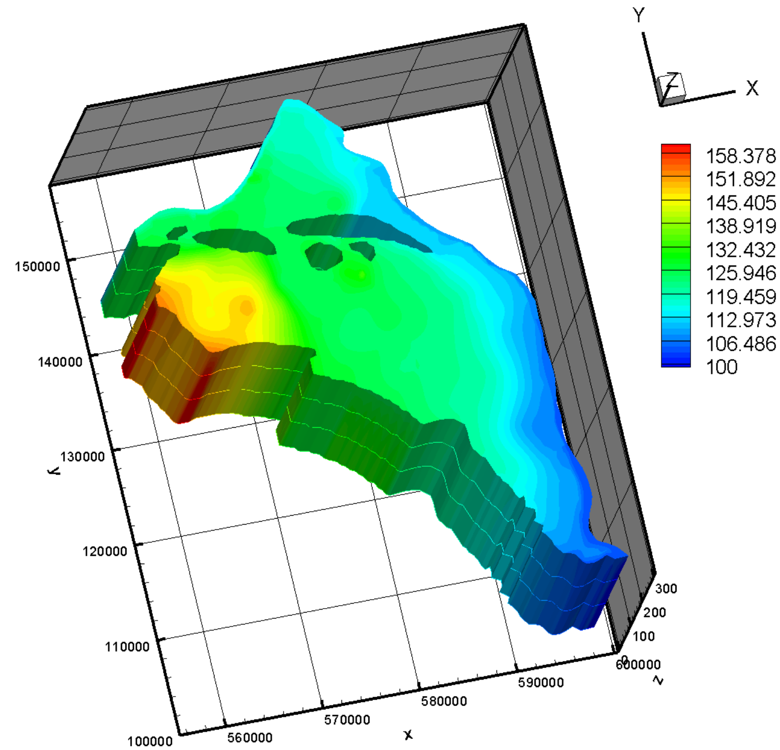
Uncertainty Quantification and Its Application in Energy and Environmental related complex systems

Guang Lin, Pacific Northwest National Laboratory



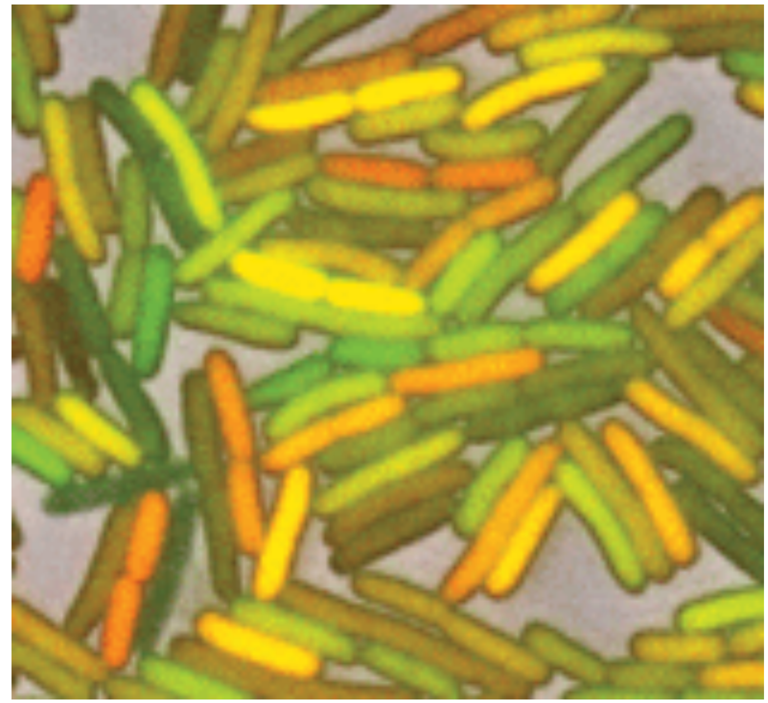
Pacific Northwest NATIONAL LABORATORY

Outline

- Motivation & Types of UQ
- VVUQ Applications
- Model Validation & UQ Algorithms
- Uncertainty Quantification Algorithms – generalized Polynomial Chaos
- Summary

Stochastic Gene Expression in a Single Cell

Eblowitz et al., Science, 2002



Bacterial cells expressing two different fluorescent proteins (red and green) from identical promoters.

Because of stochasticity (noise) in the process of gene expression, even two nearly identical genes often produce unequal amounts of protein.

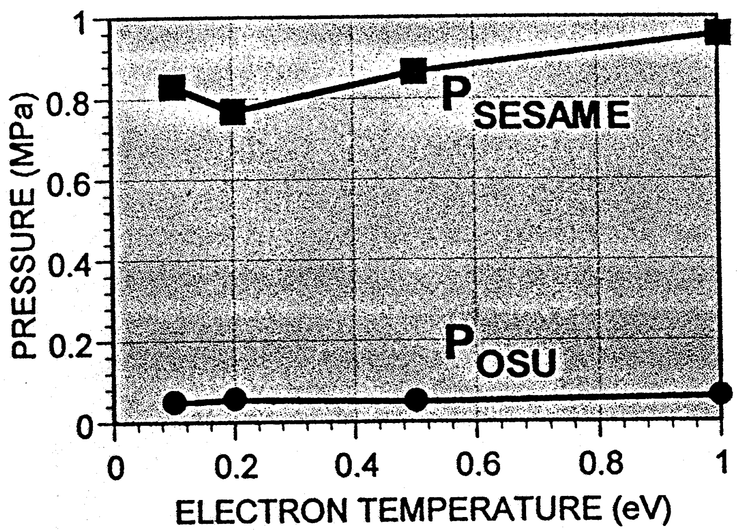
The resulting color variation shows how noise fundamentally limits the accuracy of gene regulation.

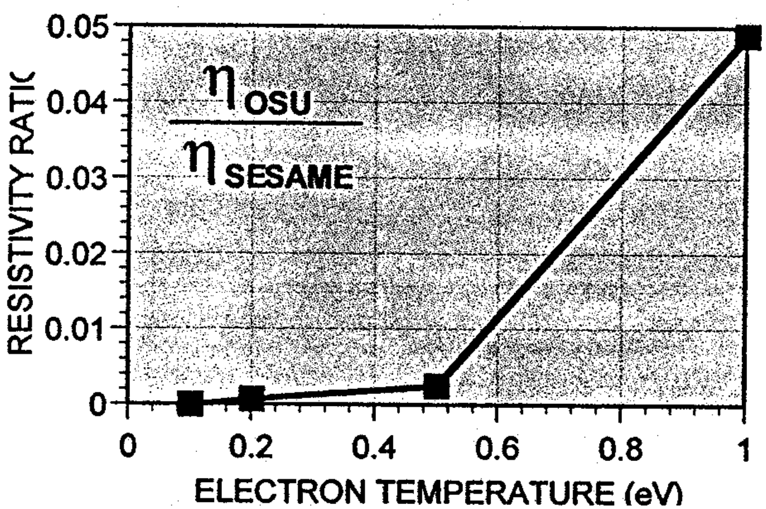
UNCERTAINTY QUANTIFICATION AND MANAGEMENT

Excerpts from published emails in connection with shuttle Columbia's last mission:

landing with 2 flat tires. Carlisle said that Howard Law had done an entry sim at Ames (the sim was evidently done o
Friday) and that sim showed that the landing with 2 flat tires was survivable. Bob Doremus and David Paternostro
expressed some skepticism as to the accuracy of the Ames sim in light of other data (Convair 990 testing), but
appreciated the information. All four agreed at the end of the discussion that we were doing a "what-if" discussion a
that we all expected a safe entry on Saturday.

Uncertainty in Equation of State: Teflon





Turchi et al.
AIAA 98-3807

MACH2 RESULTS

Need For:
Stochastic Simulations

Objective of Uncertainty Quantification

- Introduce error bars into numerical simulations.
- Understand the propagation of uncertainty in a dynamical system.
- Assessment of the stochastic response.
  - Desired statistics.
  - Reliability analysis.
  - Sensitivity analysis.
  - ......

Motivation of Uncertainty Quantification

Modeling and Simulation (M&S) plays an important role in scientific and engineering predictions

- **However, there are many cases where model inputs are subject to uncertainty with known probability distribution Uncertainties without precise probability distributions, e.g.,**
- **In M&S, we use PDE-based models that are simulated via approximate numerical methods**
  - **"All models are wrong, but some are useful." – G. Box**
  - **Numerical errors due to the mesh, iterations, etc.**

➢ ***How can we estimate total prediction uncertainty in M&S?***

Pacific Northwest
NATIONAL LABORATORY

Uncertainty versus Error

Uncertainty: A potential deficiency in any phase or activity of a modeling process that is due to lack of knowledge.

Aleatory (irreducible) Uncertainty
Epistemic (reducible) Uncertainty

Error: A recognizable deficiency in any phase or activity of modeling and simulation that is not due to lack of knowledge.

Sources of Uncertainty: Initial and Boundary Conditions, Thermo-physical/Structural Properties, Geometric Roughness, Interaction Forces, Background Noise, …

Sandia National Laboratories

Pacific Northwest NATIONAL LABORATORY

Types of Uncertainty

Classification of uncertainties

- **Aleatory uncertainty – inherent variation in a quantity**
  - **Given sufficient samples, it can be characterized with a precise probability distribution**
- **Epistemic uncertainty – due to a lack of knowledge**
  - **There is insufficient information to specify either a fixed value or a probability distribution**
  - **More difficult to treat mathematically**
  - **Can be eliminated by adding sufficient information**

➢ **Uncertainties can be aleatory, epistemic, or a mixture of the two**

Characterization and Propagation of Aleatory Uncertainty

Aleatory (stochastic) uncertainty

- **Characterized by precise PDF or CDF**
- **Aleatory uncertain model inputs must be propagated through the model to determine effects on a System Response Quantities (SRQ)**
- **Pure aleatory uncertainty can be propagated by:**
  - **Sampling (MCS, LHS, etc.)**
  - **Perturbation methods**
  - **Polynomial chaos**
  - **Stochastic collocation**

Characterization and Propagation of Epistemic Uncertainty

Epistemic uncertainty – due to lack of knowledge

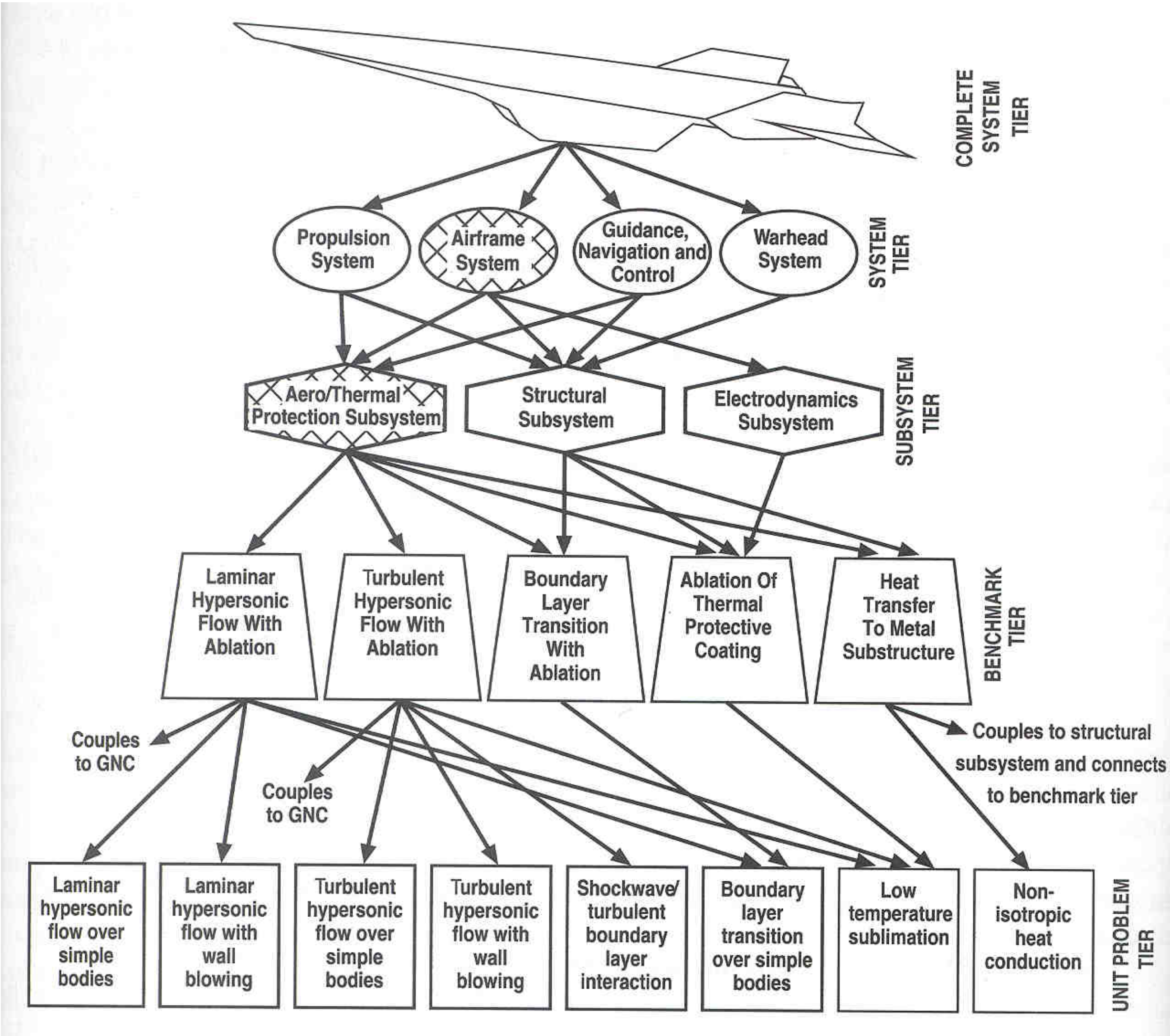
- **Sometimes characterized as a probability distribution representing the degree of belief**
- **Our position is that pure epistemic uncertainties should be represented as intervals with no associated probability distribution, e.g., $x \in [0.45, 0.55]$**
  - **If deterministic, the true value can be any value in the range**
  - **No likelihood/belief that any value is more true than another**
  - **Different than assuming all values are equally likely (uniform)**
- **Interval uncertainties can be propagated by:**
  - **Sampling (MCS, LHS, etc.)**
  - **Interval arithmetic (non-naïve implementations)**
  - **Optimization: given interval-valued inputs, find the min and max output SRQ**

Characterization of Mixed Aleatory and Epistemic Uncertainty

In practice, uncertain inputs are often a mixture of aleatory and epistemic uncertainty (e.g., random variables w/ few samples)

- **Can be characterized by imprecise probability theory**
  - **Probability Bounds Analysis (PBA)**
  - **Evidence theory (Dempster-Shafer)**

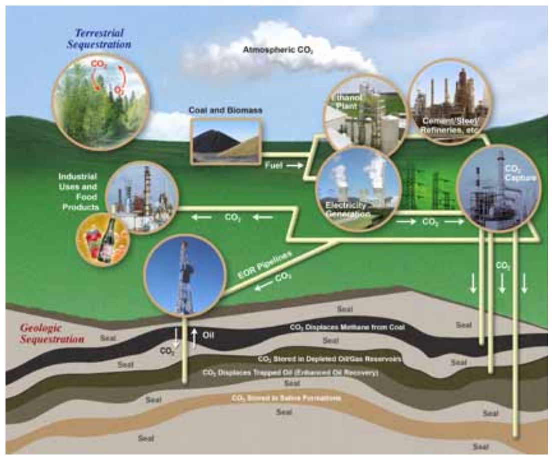
Validation Hierarchy for a Hypersonic Cruise Missile



Uncertainty Quantification for Predictive Modeling of Complex Systems

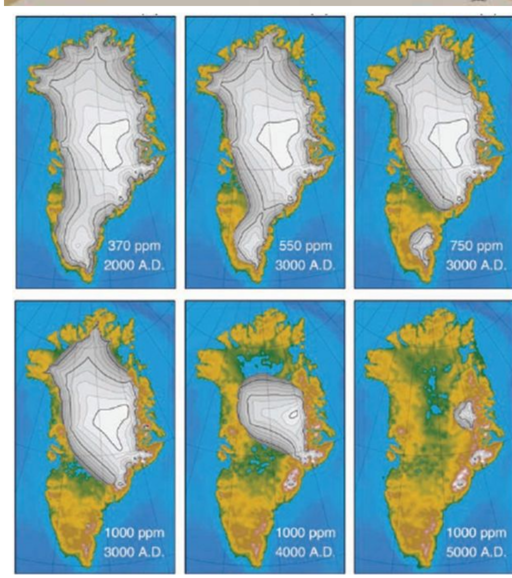
- Vision - Transform our ability to uncertainty quantification, model verification, validation and calibration of complex systems
- Outcomes - Provide fundamental understanding to enable

Safe and efficient remediation and CO2 sequestration strategies



Better prediction of climate changes

Better understand the ice sheet dynamics and its interaction with climate



Better prediction and control of power system stability and reliability

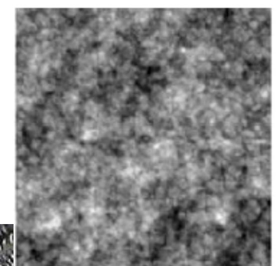
Why uncertainty and multiscaling?

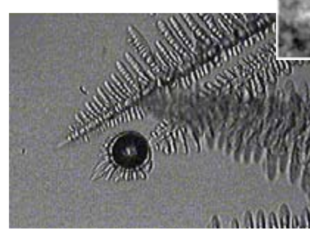
- All physical systems have inherent associated randomness
- Uncertainties introduced across various length scales have a non-trivial interaction

Micro

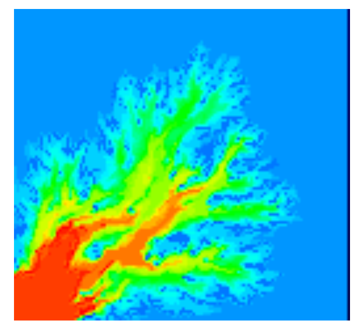
- Physical properties, structure follow a statistical description

Meso





- Use micro averaged models for resolving physical scales



Macro

- Imprecise boundary conditions
- Initial perturbations

Schematic Diagram of UQ & Multi-scale Modeling of Complex Heterogeneous Reaction System

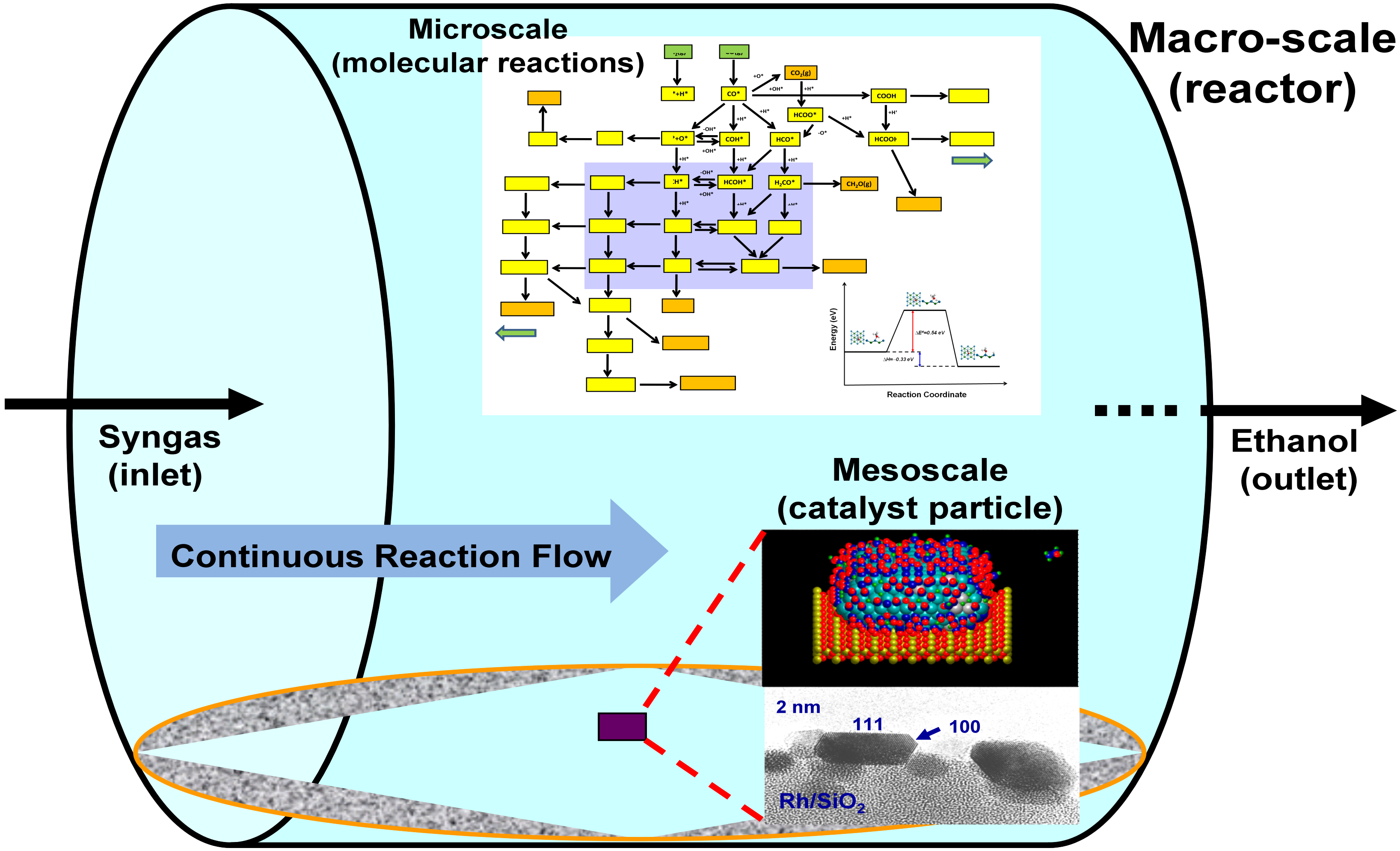
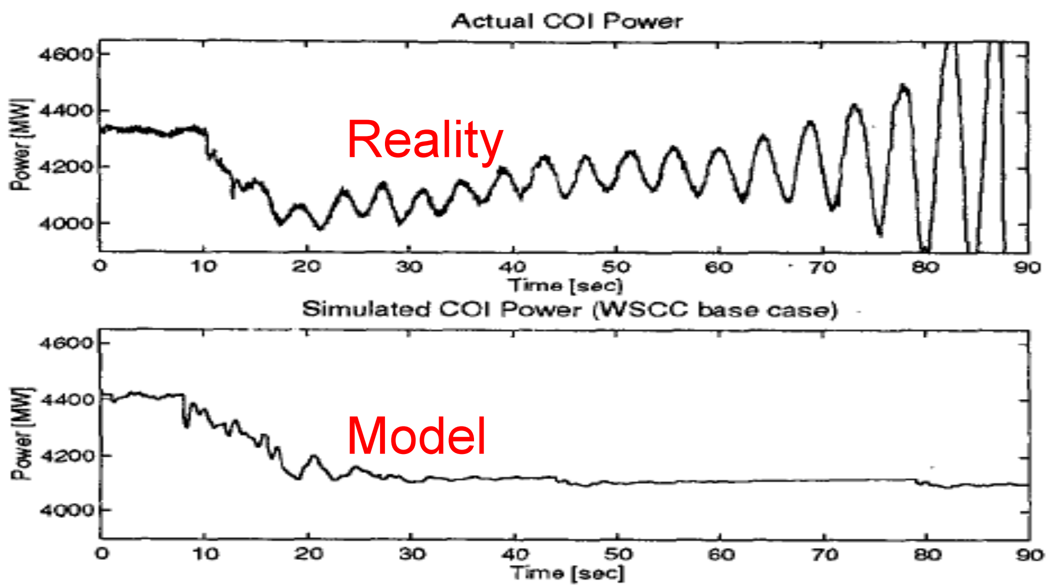


Fig. 1 Multi-scale computational modeling: from intrinsic molecular reactivity to macroscopic catalytic kinetics in the reactor under operating reaction and flow conditions

west
NATIONAL LABORATORY

Background: Model Validation Need and Challenges

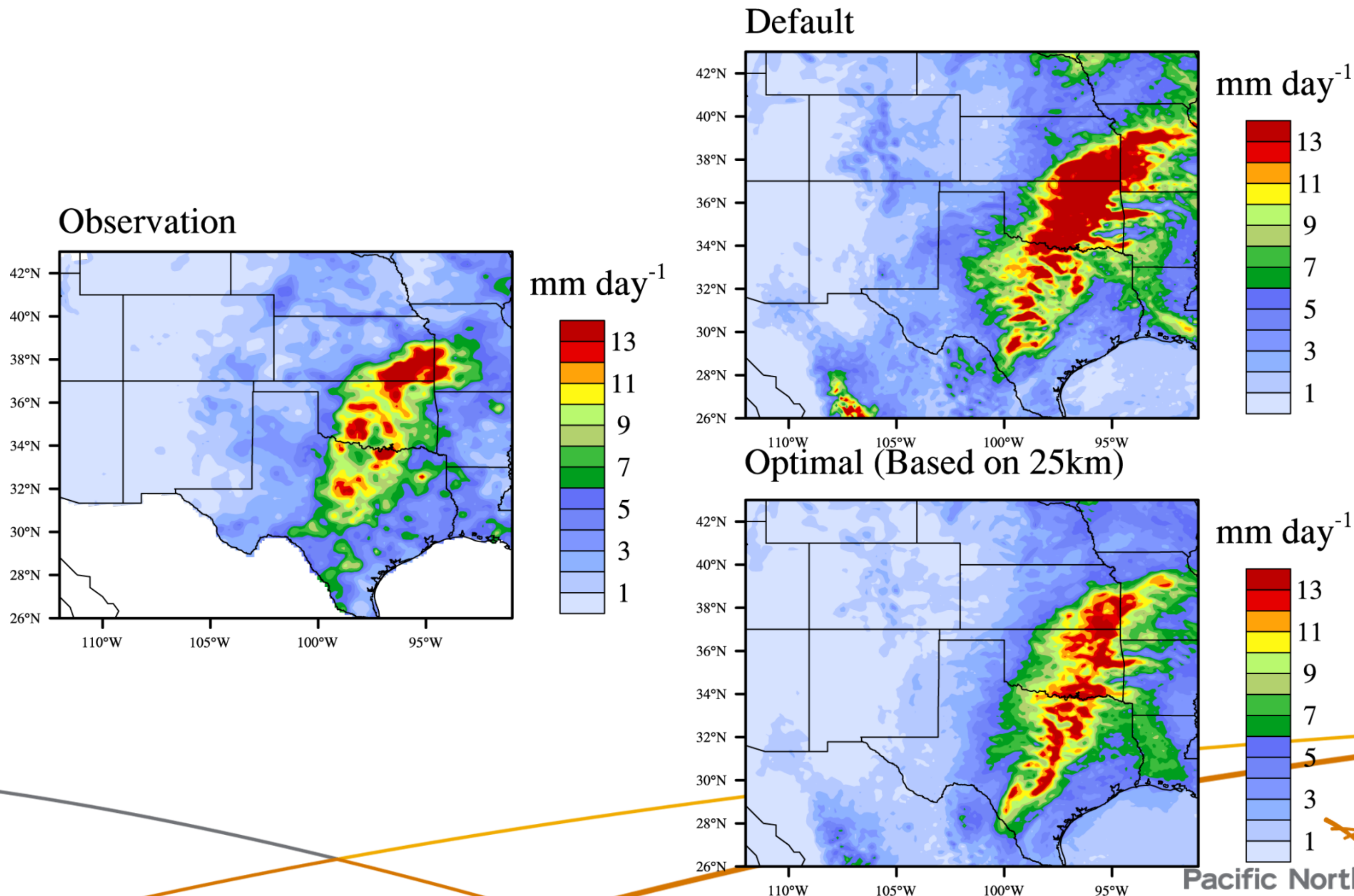


Recorded system dynamics vs. simulation results: California and Oregon Intertie (COI) real power flow during the August 10, 1996 event (Kosterev et al.1999)

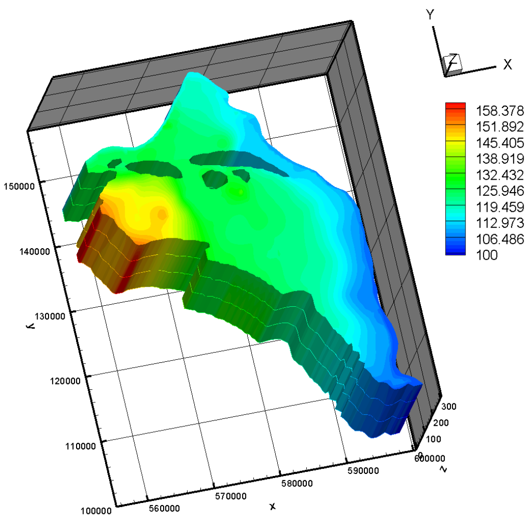
Pacific Northwest
NATIONAL LABORATORY

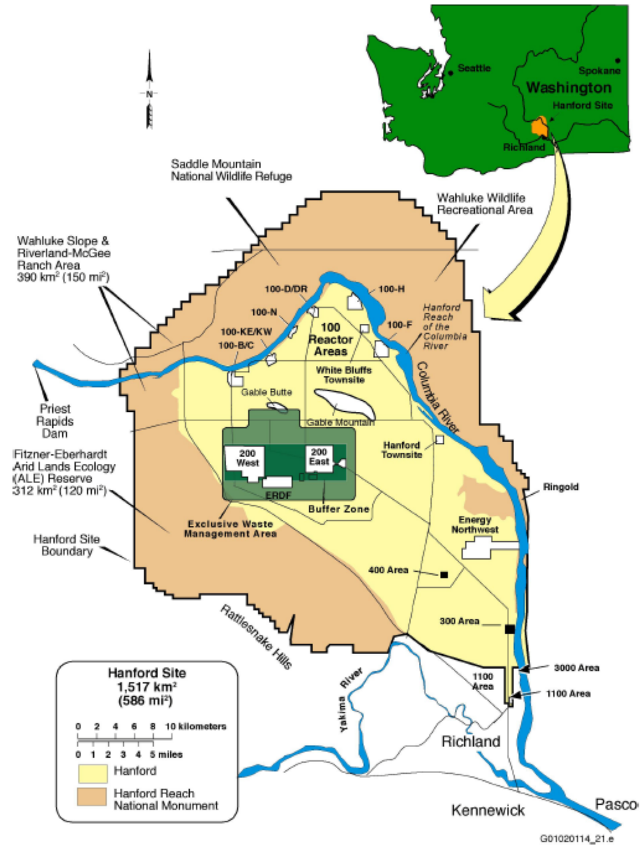
Climate Model Calibration using Observation Data Sets

Total Precipitation in June 2007



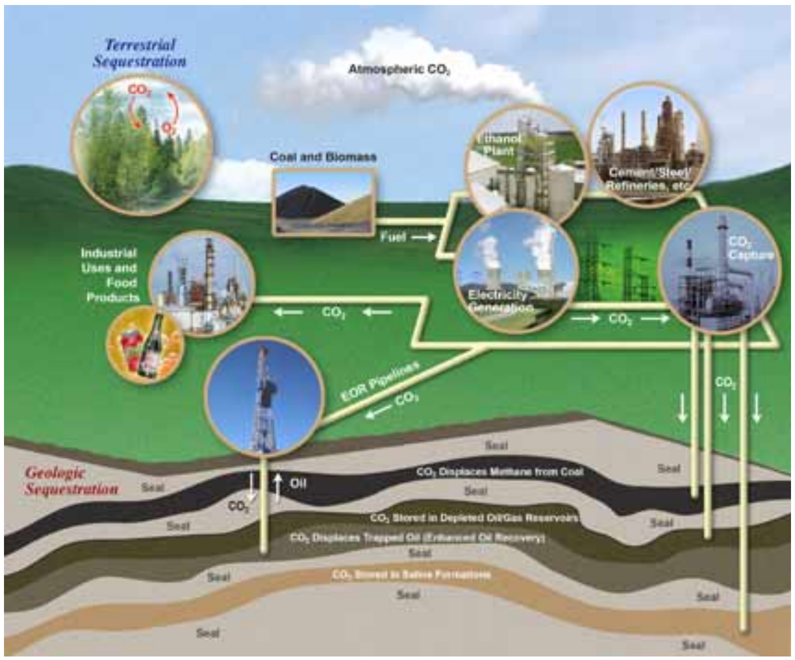
Uncertainty Quantification for Nuclear Contaminant Flow and Transport

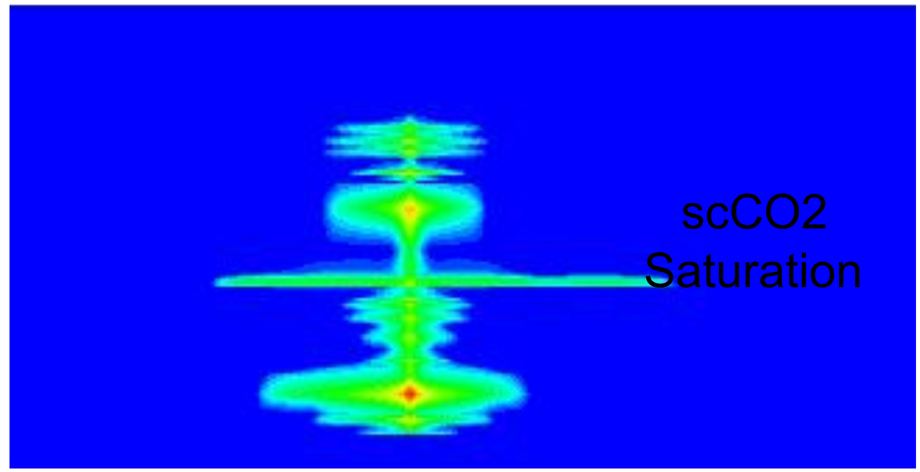




Pacific Northwest
NATIONAL LABORATORY

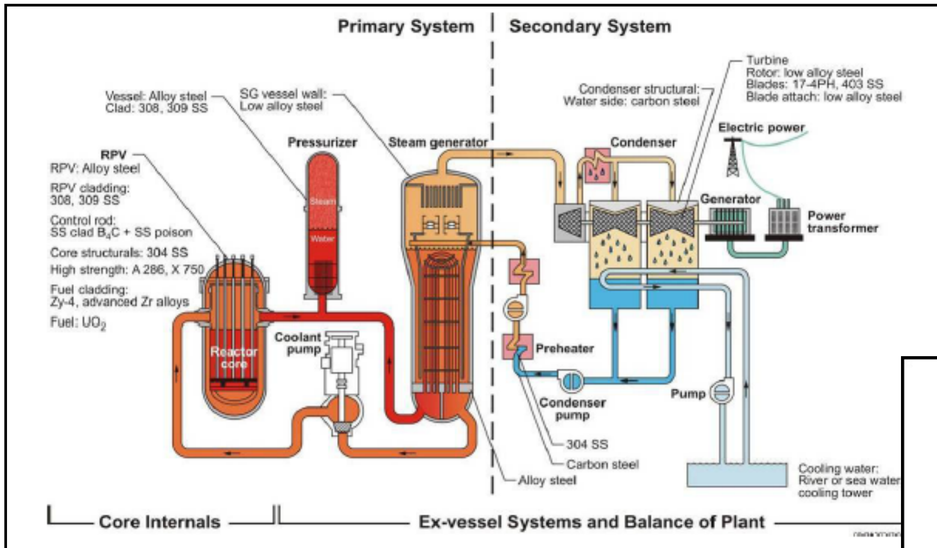
UQ for CO2 Sequestration

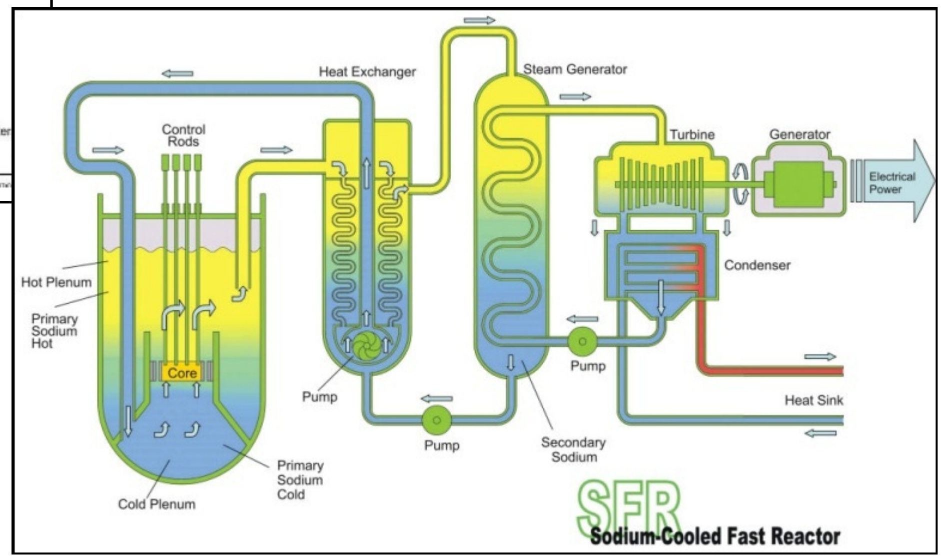




Pacific Northwest
NATIONAL LABORATORY

Uncertainty Quantification





Pacific Northwest
NATIONAL LABORATORY

What are Verification, Validation, Uncertainty Quantification?

- Verification: Are the requirements *implemented correctly*?
  - *Are we solving the equations correctly?*
  - *Are we solving the equations to sufficient accuracy?*
- Validation: Are the requirements *correct*?
  - *Are we solving the right equations?*
- Uncertainty Quantification: The end-to-end study of the *reliability of scientific inferences*.
  - *Uncertainty and error affect every scientific analysis or prediction.*

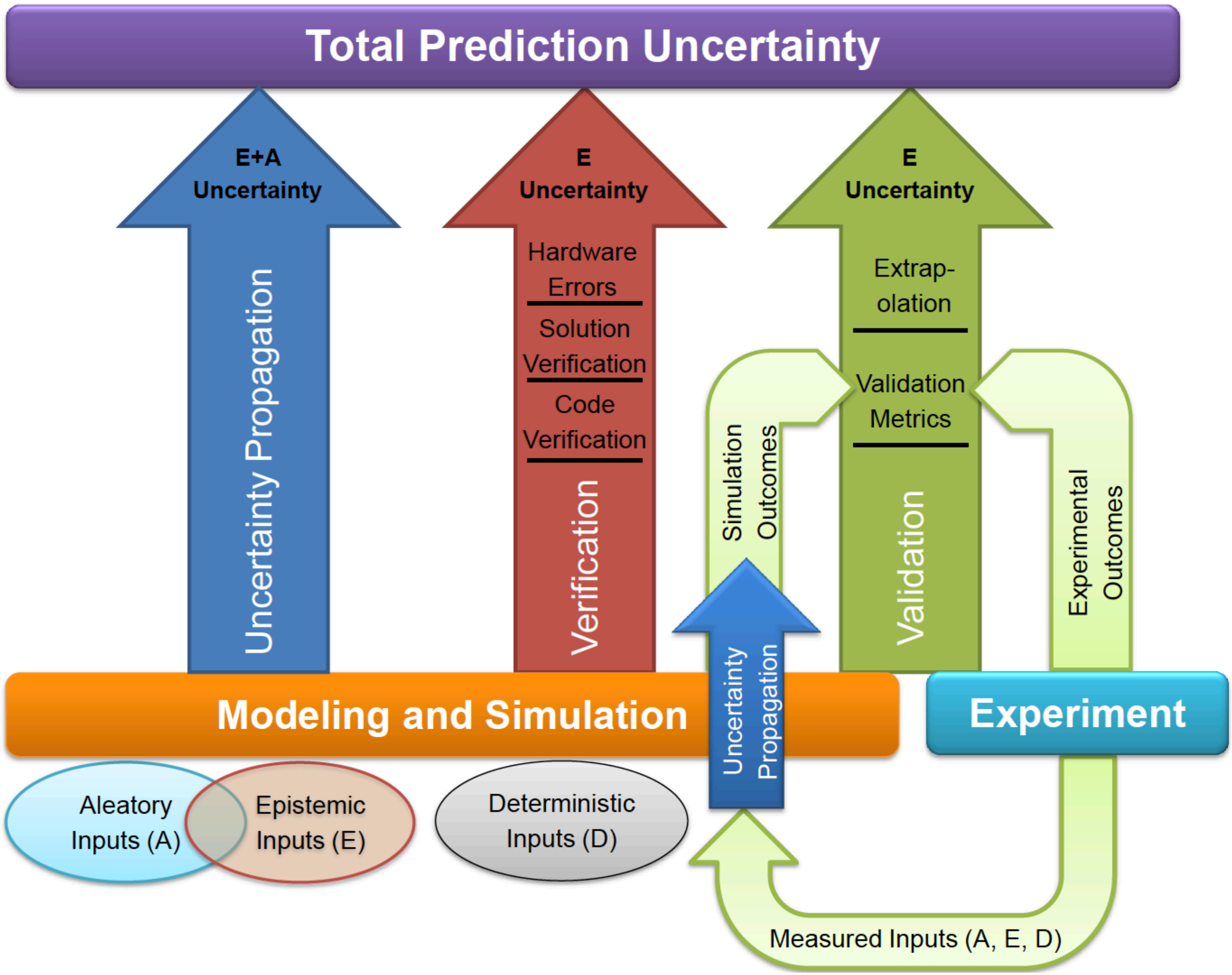
Collectively known as "VVUQ"

Pacific Northwest
NATIONAL LABORATORY

What Will VV UQ Do?

- **Verification:** Develop test problems, new methods, and software tools
- **Validation:** VVUQ will collect validation datasets and identify database gaps as required by the born-assessed and licensing missions
- **Calibration, SA, UQ:** Develop and deploy new capabilities and software tools for the NEAMS IPSCs

Pacific Northwest
NATIONAL LABORATORY


Total Prediction Uncertainty
E+A
Uncertainty
Uncertainty Propagation
E
Uncertainty
Hardware
Errors
Solution
Verification
Code
Verification
Verification
E
Uncertainty
Extrap-
olation
Validation
Metrics
Validation
Simulation
Outcomes
Experimental
Outcomes
Uncertainty
Propagation
Modeling and Simulation
Experiment
Aleatory
Inputs (A)
Epistemic
Inputs (E)
Deterministic
Inputs (D)
Measured Inputs (A, E, D)

Two Approach to Achieve the Best Prediction:

- Model Calibration
  - Adjust the values of the model parameters to reduce the uncertainties associated with parameter specification, output measurement error, etc.
- Data assimilation
  - Use observed data for system state and output variables to update the system state variables

Data Assimilation & Model Calibration

Data Assimilation:

a. Kalman Filter, Kalman–Bucy filter

b. Nonlinear Kalman Filter: (Ensemble Kalman Filter, Extended Kalman filter, Unscented Kalman filter)

c. Particle Filter (sequential Monte Carlo methods)

Model Calibration:

a. Least Squares Parameter Estimation

b. Bayesian Parameter Estimation

Pacific Northwest
NATIONAL LABORATORY

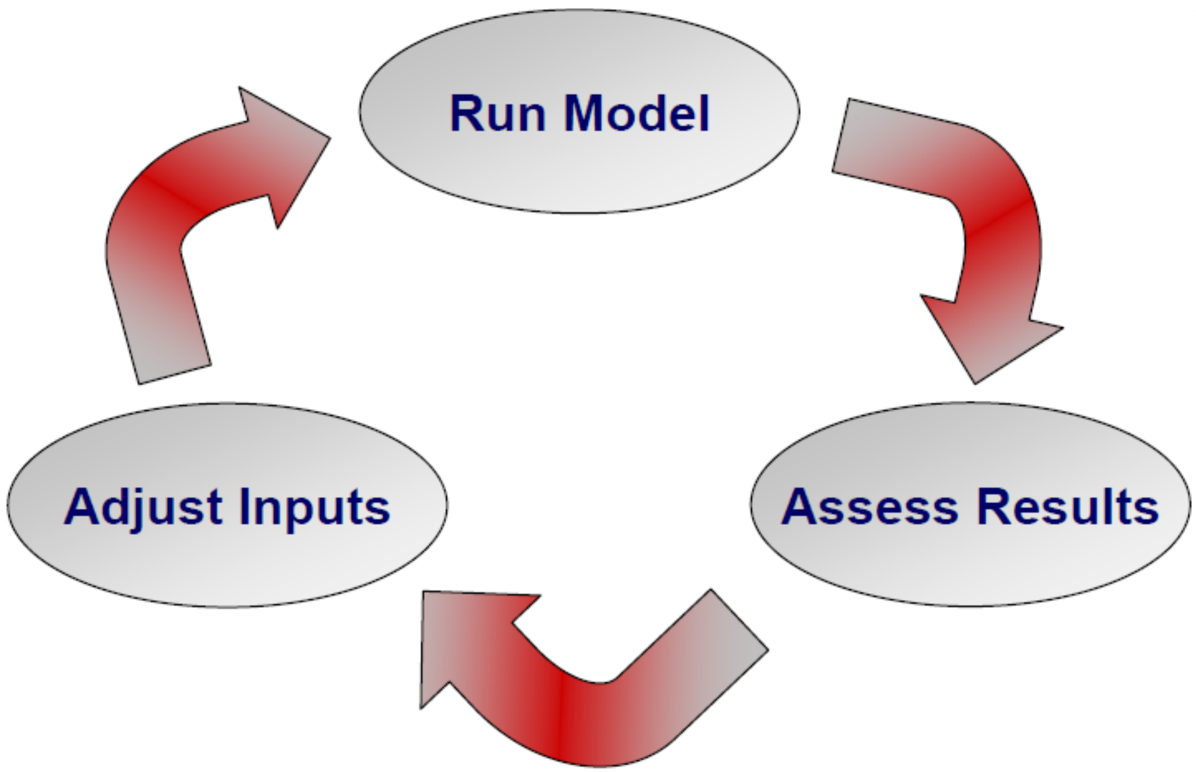
What is Calibration?

- A systematic adjustment of model parameters
- Establishes predictive validity of a model
- Model outputs govern model inputs

When is Calibration Needed?

- Whenever predictive validity of model is in question
- When data are inadequate to estimate model inputs

Calibration Cycle



How do we perform model calibration?

- Obtain calibration data sets
- Select an optimization scheme
- Optimize selected model parameters

Pacific Northwest
NATIONAL LABORATORY

Methods for Uncertainty Quantification

❖Monte Carlo Method

❖Quasi-Monte Carlo Method

❖Multi-level Monte Carlo Method

❖Pdf Method

❖Moments Approach

❖Latin Hypercube Sampling

❖Fuzzy Logic

❖Evidence Theory

❖Generalized Polynomial Chaos

a. Intrusive Approach – Galerkin Projection Method

b. Non-intrusive Approach – Probabilistic Collocation Method

Special Techniques to achieve fast convergence:

❖Important Sampling Method

❖Variance Reduction Method

Representation of a Random Process

$$T(x,t;\omega)=\sum_{j=0}^{\infty} T_j(x,t)\Phi_j(\xi(\omega))$$

- $\mathrm{T}(\mathbf{x},\mathrm{t};\theta)$ - Random process
  - $(\mathbf{x},\mathrm{t})$ - Spatial/temporal dimension
  - θ - Random dimension
- $\mathrm{T_i}(\mathbf{x},\mathrm{t})$ - Deterministic coefficients
- $\Psi_\mathrm{i}(\xi(\theta))$ - ***Generalized*** Polynomial Chaos

Classical polynomial chaos – Wiener 1938, Ghanem & Spanos 1991

Generalized Polynomial Chaos (gPC)

Xiu & Karniadakis, SIAM, J. Sci. Comp., vol. 24, 2002

$$T(x,t;\omega) = \sum_{j=0}^{\infty} T_j(x,t)\Phi_j(\xi(\omega))$$

Polynomials of random variable $\xi(\omega)$

Orthogonality : $\langle \Phi_i \Phi_j \rangle = \langle \Phi_i^2 \rangle \delta_{ij}$

$$\langle \mathrm{f}(\xi)\mathrm{g}(\xi) \rangle = \int \mathrm{f}(\xi)\mathrm{g}(\xi)\mathrm{W}(\xi)\mathrm{d}\xi \qquad \langle f(\xi)g(\xi) \rangle = \sum_i f(\xi_i)g(\xi_i)w(\xi_i)$$

Weight function determines underlying random variable
(*not necessarily Gaussian*)

Complete basis from *Askey scheme*

Each set of basis converges in L^2 sense

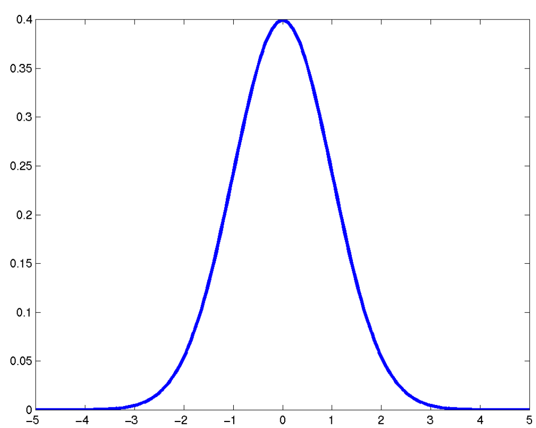
Computational Speed-Up

Lucor & Karniadakis, Generalized Polynomial Chaos and Random Oscillators
Int. J. Num. Meth. Eng., vol. 60, 2004

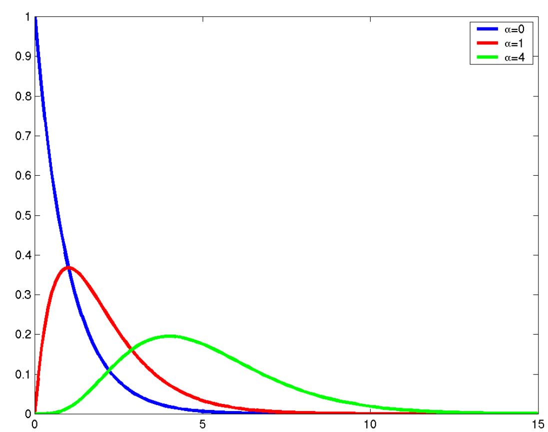
PDF	Error (mean)	Monte-Carlo: M	GPC: (P+1)	**Speed-Up**
Gaussian	2% 0.8% 0.2%	350 2,150 33,200	56 120 220	6.25 18 151
Uniform	0.2% 0.018% 0.001%	13,000 1,580,000 610,000,000	10 20 35	13,000 79,000 17,430,000

Orthogonal Polynomials and Probability Distributions

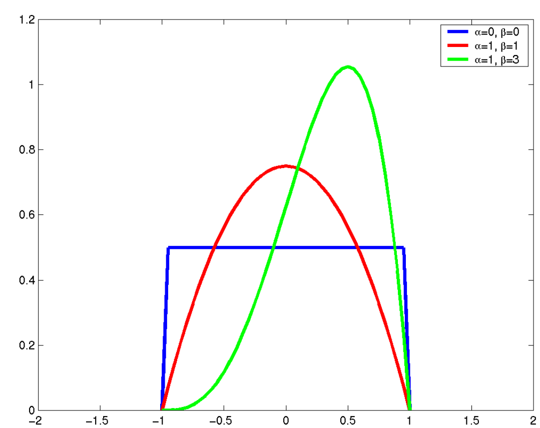
- **Continuous Cases**:
  - *Hermite* Polynomials ⟷ *Gaussian* Distribution
  - *Laguerre* Polynomials ⟷ *Gamma* Distribution
    (special case: *exponential* distribution)
  - *Jacobi* Polynomials ⟷ *Beta* Distribution
  - *Legendre* Polynomials ⟷ *Uniform* Distribution



Gaussian distribution



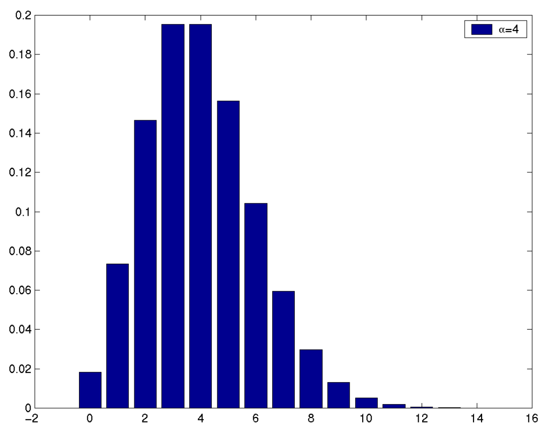
Gamma distribution



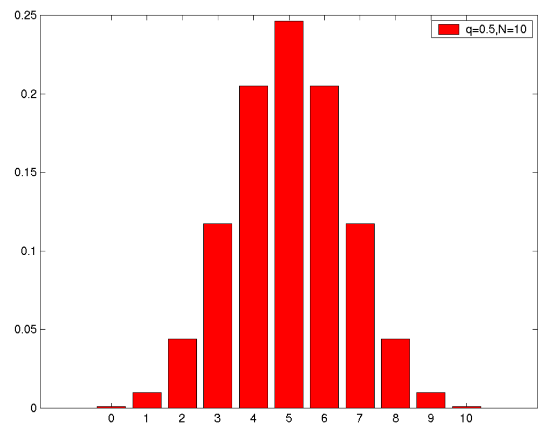
Beta distribution

Orthogonal Polynomials and Probability Distributions

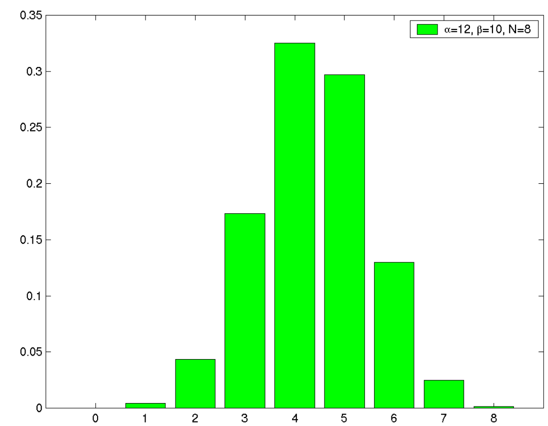
- **Discrete Cases :**
  - *Charlier* Polynomials ⟷ *Poisson* Distribution
  - *Krawtchouk* Polynomials ⟷ *Binomial* Distribution
  - *Hahn* Polynomials ⟷ *Hypergeometric* Distribution
  - *Meixner* Polynomials ⟷ *Pascal* Distribution



Poisson distribution



Binomial distribution



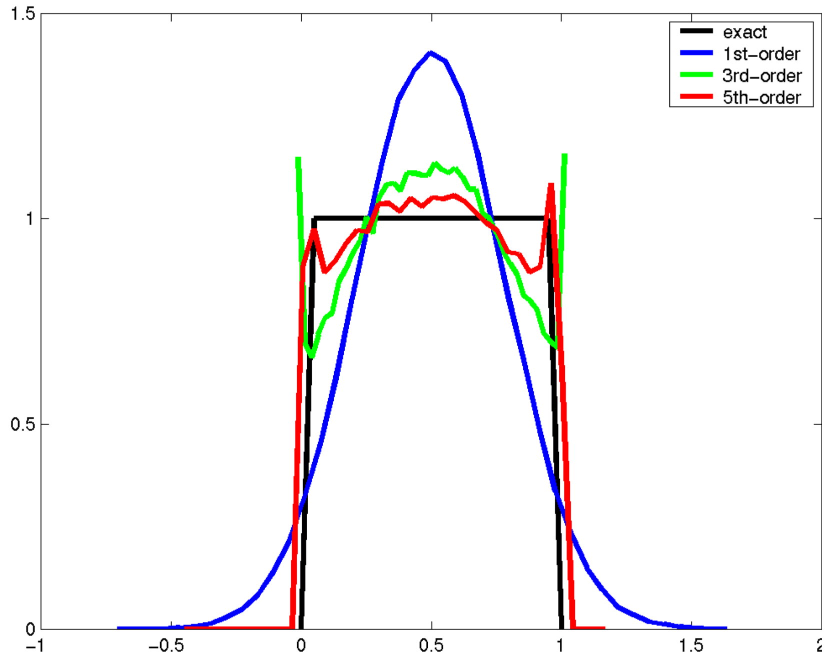
Hypergeometric distribution

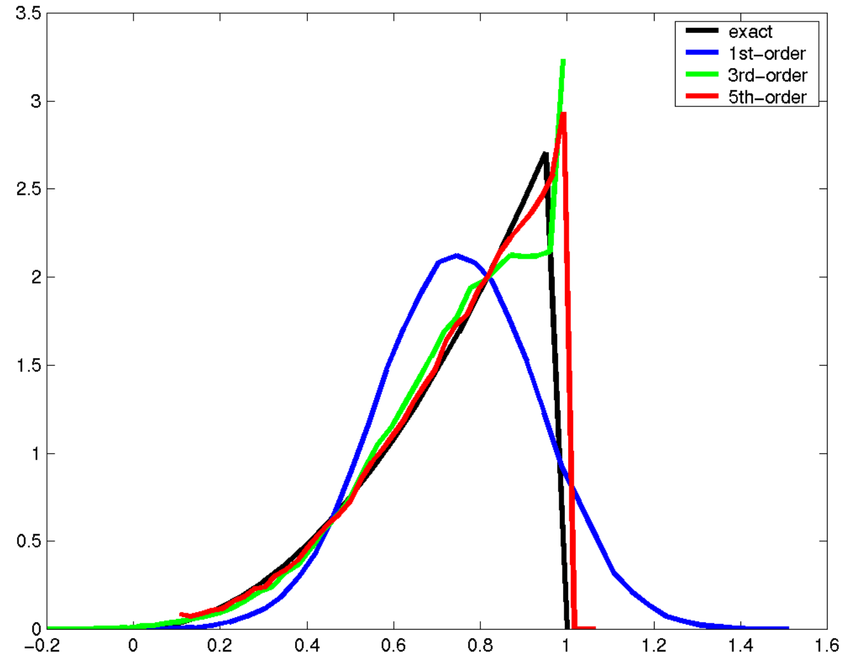
Hermite-Chaos Expansion of Beta Distribution

$$\mathrm{PDF}:\ \ f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},\quad \alpha,\beta>0,\quad 0\le x\le 1$$

Uniform distribution : $\alpha = 1,\ \ \beta = 1$

$\alpha = 3,\ \ \beta = 1$





Exact PDF and PDF of 1^{st}, 3^{rd}, 5^{th}-order Hermite-Chaos Expansions

- Equation :

$$\frac{dy}{dt} = -ky, \qquad y\big|_{t=0} = \hat{y}.$$

k is the decaying coefficient with given probability distribution.

- Chaos expansion :

$$y(x,t;\theta) = \sum_{i=0}^{P} y_i(x,t)\Psi_i(\xi(\theta)), \quad k(\theta) = \sum_{i=0}^{P} k_i\Psi_i(\xi(\theta))$$

- Galerkin projection :

$$\frac{dy_i}{dt} = -\frac{1}{\left\langle \Psi_k^2 \right\rangle} \sum_{i=0}^{P} \sum_{j=0}^{P} \left\langle \Psi_i \Psi_j \Psi_k \right\rangle k_i y_j, \quad k = 0, 1, 2, \ldots, P$$

- The Chaos will be chosen according to the distribution of *k*.

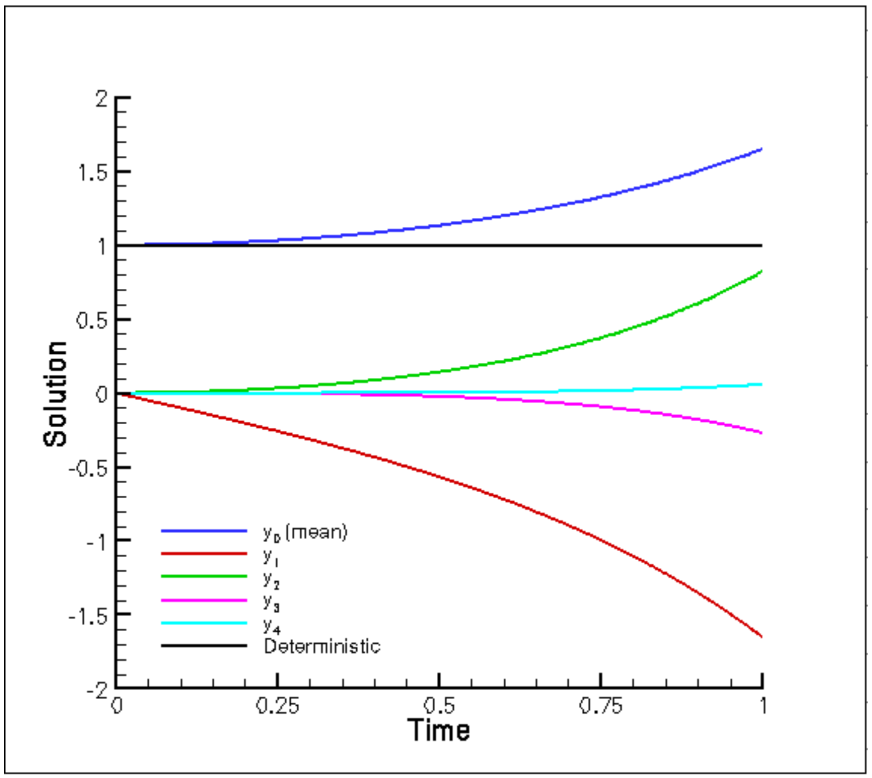
- L^{inf} error :

$$\left| \frac{\overline{y}_{chaos}(t) - \overline{y}_{exact}(t)}{\overline{y}_{exact}(t)} \right|$$

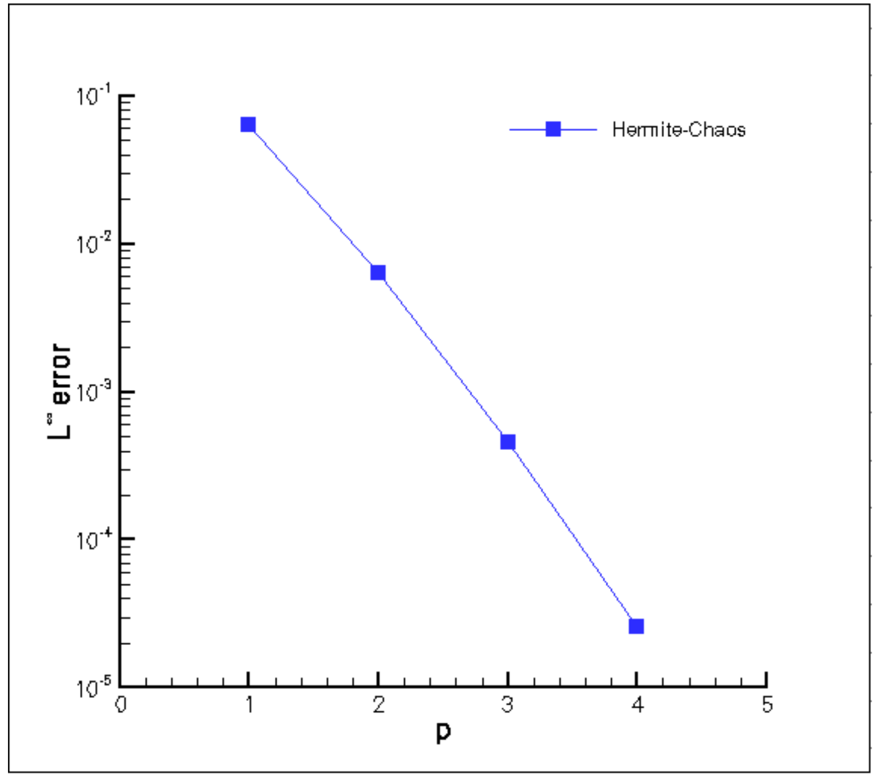
Continuous Distribution : Gaussian (Hermite-Chaos)

- *dy/dt = - k y, y(t=0)=1*
- *k* is a **Gaussian** random variable :

$$\text{PDF}: \quad f_k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



Solution of expansion modes



Convergence w.r.t. expansion terms

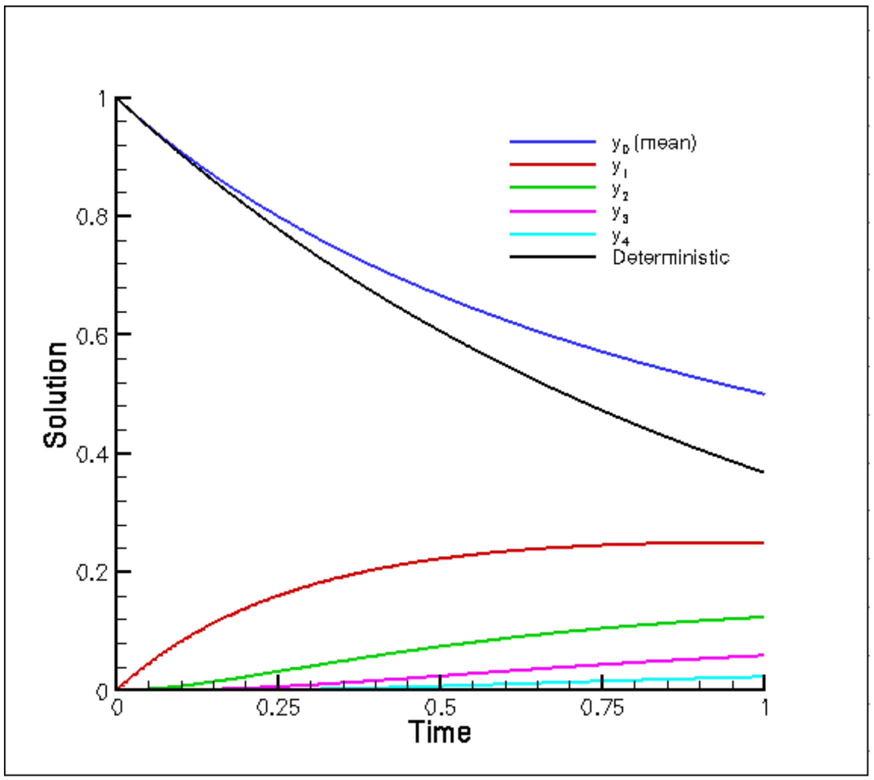
- 4th-order **Hermite-Chaos** expansion
- Exponential convergence rate

Continuous Distribution : Gamma (Laguerre-Chaos)

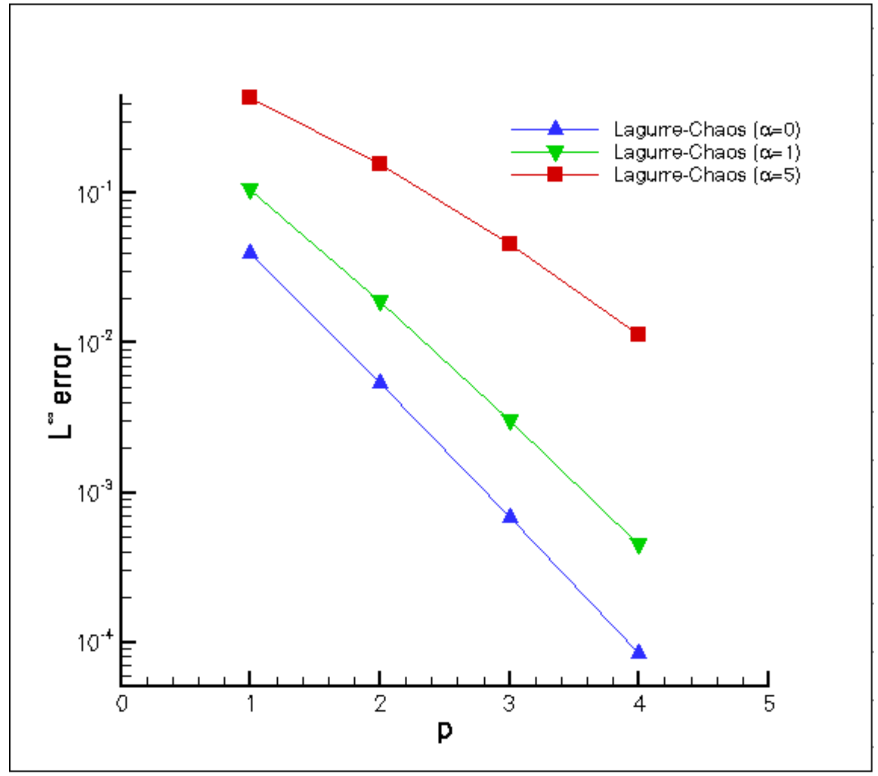
- *dy/dt = - k y, y(t=0)=1*
- *k* is a **Gamma** random variable :

$$\text{PDF}: \quad f_k(x) = \frac{e^{-x}x^{\alpha}}{\Gamma(\alpha+1)}$$

- **Exponential** distribution : $\alpha = 0$



Solution of expansion modes : $\alpha = 0$



Convergence w.r.t. expansion terms

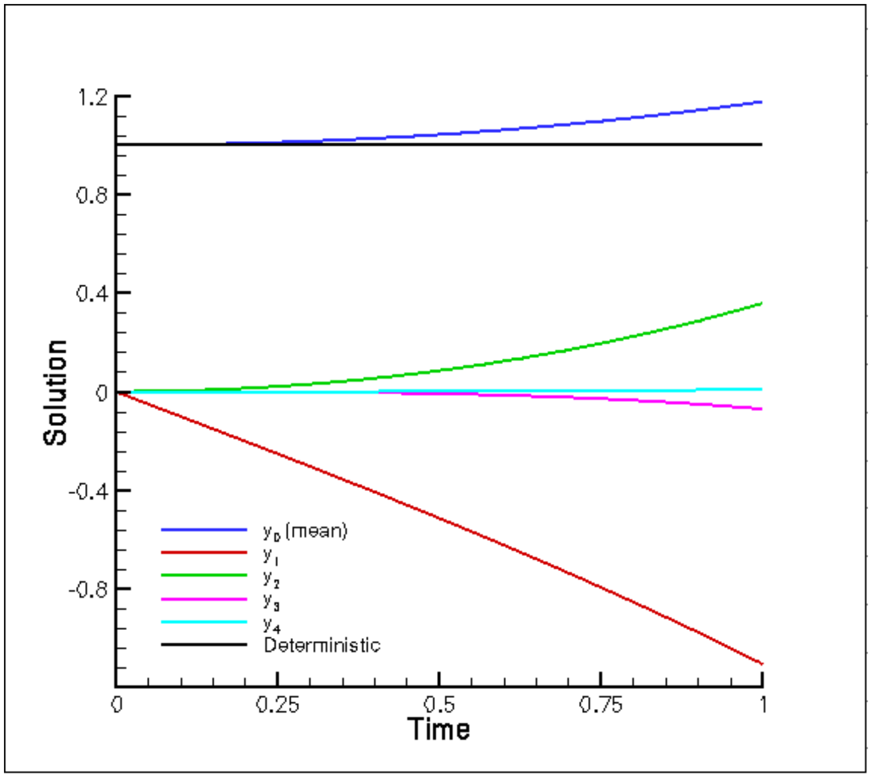
- 4^{th}-order **Laguerre-Chaos** expansion
- Exponential convergence rate

Continuous Distribution : Beta (Jacobi-Chaos)

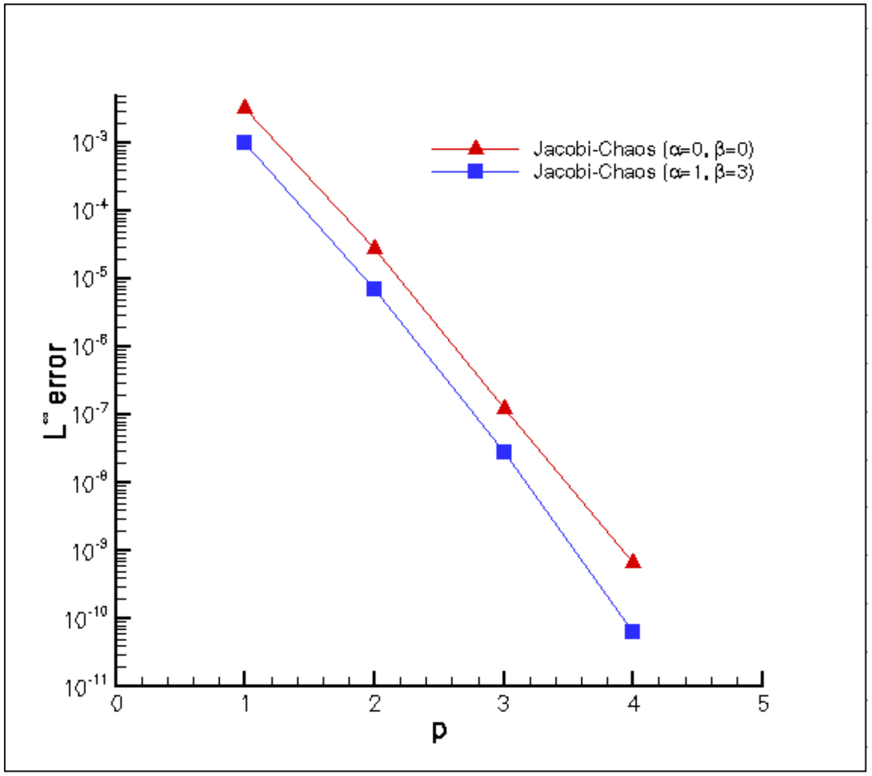
- *dy/dt = - k y, y(t=0)=1*
- *k* is a **Beta** random variable :

$$\text{PDF}: \quad f_k(x) = \frac{(1-x)^{\alpha}(1+x)^{\beta}}{2^{\alpha+\beta+1}B(\alpha+1,\beta+1)}$$

- **Uniform** distribution : $\alpha = 0, \quad \beta = 0$



Solution of expansion modes : $\alpha = 0, \quad \beta = 0$



Convergence w.r.t. expansion terms

- 4^{th}-order **Jacobi-Chaos** expansion
- Exponential convergence rate

Stochastic Spectral Methods

Galerkin projection (GPC)

$$u(\mathbf{x}, t; \omega) = \sum_{i=0}^{P} u_i \Psi_i,$$

$$\boxed{\left\langle L\left(\mathbf{x}, t, \theta; \sum_{i=0}^{P} u_i \Psi_i\right) \Psi_k \right\rangle = \langle f \Psi_k \rangle \quad k = 0, 1, \cdots, P:}$$

Solve coupled system for coefficients.

Moment estimation through exploiting orthogonality of basis.

References: e.g.

- Ghanem & Spanos, 91
- Xiu & Karniadakis, 02
- Schwab & Todor, 03
- Matthies & Keese, 05
- Deb, Babuska & Oden, 00
- Le Maitre et al, 04
- Wan & Karniadakis, 05

Collocation projection (PCM)

$$\langle \cdot, \delta(y_k) \rangle, k = 1, \cdots, M$$

$\{y_k\}_{k=1}^{M}$ a set of collocation points on Γ, coincides with quadrature rule

$$\boxed{L(\mathbf{x}, t, y_k; u) = f(\mathbf{x}, t, y_k)}$$

Solve M decoupled equations.

$$\hat{u}(x, t, y) = \sum_{k=1}^{M} u(x, t, y_k) l_k(y)$$

Lagrange interpolating polynomials

Moment estimation:

$$E[\hat{u}^2](\mathbf{x}, t) = \sum_{k=1}^{M} u^2(\mathbf{x}, t, y_k) w_k$$

Tatang & McRae, 94; Isukapalli et al, 00; Xiu & Hesthaven, 05; Babuska et al, 05

Advantage of gPC

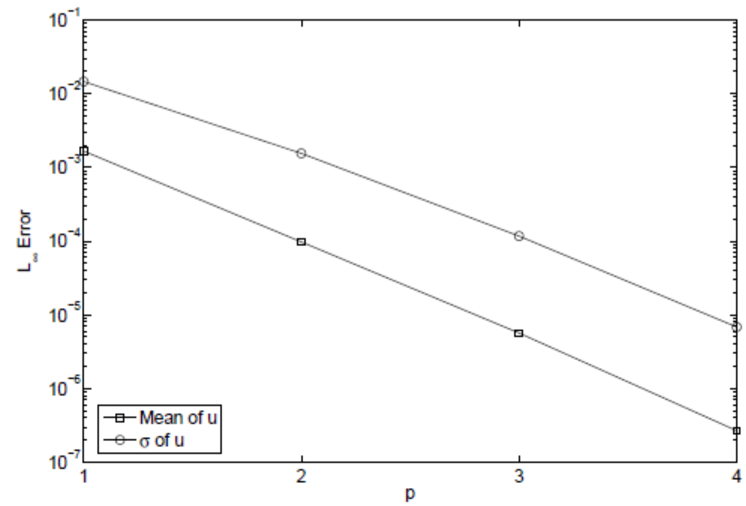
- Fast convergence due to spectral expansion.
- Efficiency due to orthogonality.

$$\begin{aligned} \frac{\partial u}{\partial t} + (u \cdot \nabla) u &= -\nabla p + \nu(1 + \delta \xi) \nabla^2 u \\ \nabla \cdot u &= 0 \end{aligned}$$

Kovasznay Flow:

$$\begin{aligned} u &= 1 - \mathrm{e}^{\lambda x} \cos 2\pi y \\ v &= \frac{\lambda}{2\pi} \mathrm{e}^{\lambda x} \sin 2\pi x \\ \lambda &= \frac{Re(\xi)}{2} - \left(\frac{Re^2(\xi)}{4} + 4\pi\right)^{1/2} \end{aligned}$$

ξ : random variable of Beta(1,1).

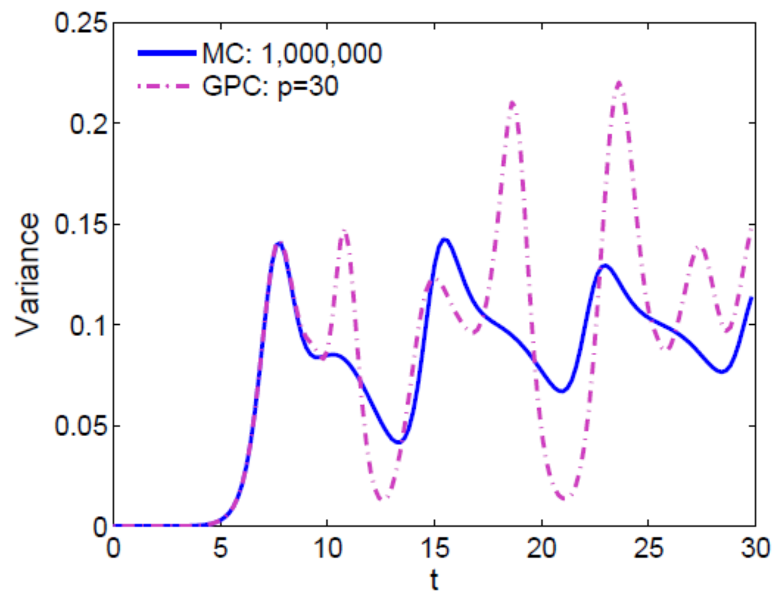


Limitations of gPC

- Inefficient for problems with low regularity in the parametric space.
- May diverge for long-time integrations.

Kraichnan-Orszag three-mode model:

$$\begin{cases} \frac{\mathrm{d}Y_1}{\mathrm{d}t} = Y_2 Y_3 \\ \frac{\mathrm{d}Y_2}{\mathrm{d}t} = Y_1 Y_3 \\ \frac{\mathrm{d}Y_3}{\mathrm{d}t} = -2 Y_2 Y_3 \\ \text{random intial conditions.} \end{cases}$$



Pacific Northwest
NATIONAL LABORATORY

Comments on Polynomial Chaos

- *Advantages of gPC:*
  - Fast convergence due to spectral expansion.
  - Efficiency due to orthogonality.
- *Disadvantages of gPC:*
  - Efficiency decreases as the number of random dimensions increases.
  - Inefficient for problems with low regularity in the parameter space.
  - May diverge for long time integration.

Stochastic Sensitivity Analysis

Motivation:

- Rank all inputs and parameters in order of their significance to output variation
- Reduce dimension of parametric space in experiments or simulations

Sensitivity Algorithms:

- Approximated Gradient Method

Morris, QMC, MC, Multi-Element Sparse Collocation

$$EE_i^j(x_1^0,\dots,x_d^0) = \frac{\left|y_j(x_1^0, x_2^0, \dots, x_{i-1}^0, x_i^0 + \Delta, x_{i+1}^0, \dots, x_d^0) - y_j(x_1^0, \dots, x_d^0)\right|}{\Delta}$$

Input X_i : i=1:d

Output y_j : j=1:n

Lin & Karniadakis, AIAA-2008-1073, 2008; IJNME 2009

Pacific Northwest NATIONAL LABORATORY

Summary

- Forward Uncertainty Quantification can provide an error bar to the model simulation results.

- Model calibration and data assimilation quantify uncertainty and bridge the gap between Simulation-Experiment.

- Data assimilation is a technique that is used to correct the errors in state variables.

- Sensitivity analysis can reduce dimension of parametric space in experiments or simulations

Uncertainty Quantification and Its Application in Energy and Environmental related complex systems

"…Because I had worked in the closest possible ways with physicists and engineers, I knew that our data can never be precise…"

Norbert Wiener

Pacific Northwest
NATIONAL LABORATORY

Questions