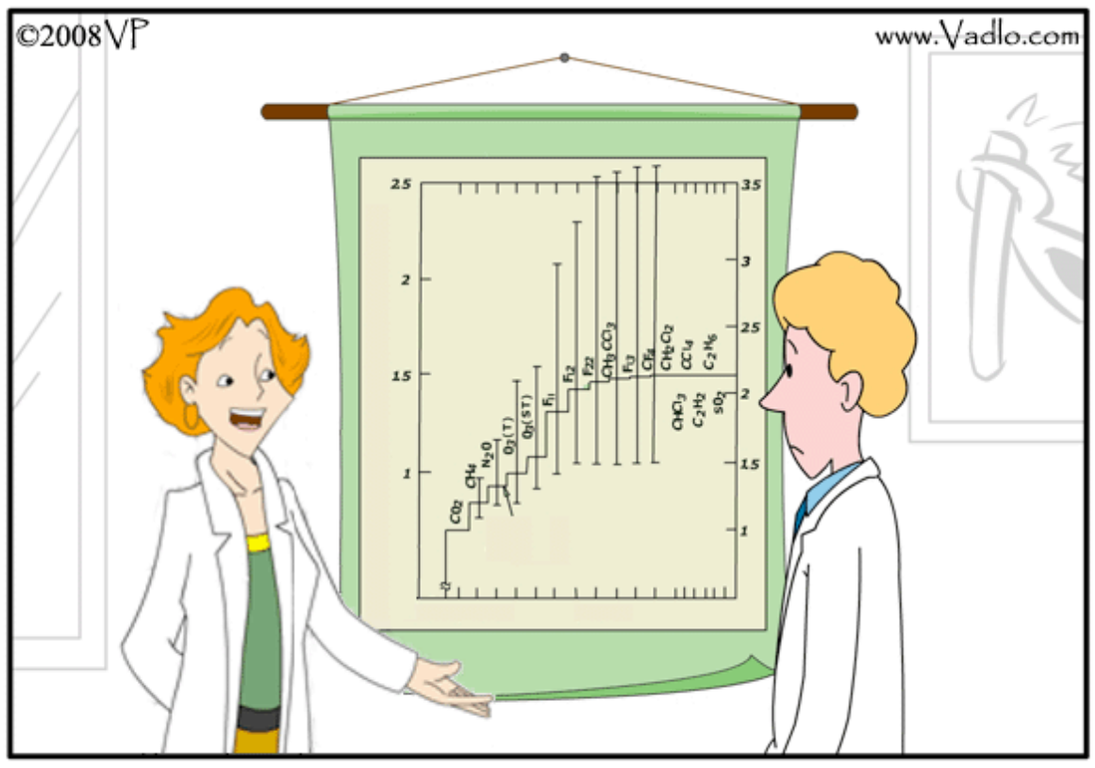
Uncertainty Quantification in Computational Science

Jan S Hesthaven
EPFL, Lausanne, CH
Jan.Hesthaven@epfl.ch



Did you really have to show the error bars?

Objective of lectures

The main objective of these lectures are

- To offer an overview of Uncertainty Quantification (UQ) and discuss its importance in modern predictive computational science.
- Focus is on the forward problem - uncertainty propagation.
- To introduce Polynomial Chaos (PC) as an attractive and efficient way of dealing with such challenges in the context of complex dynamical systems.
- To provide enough background to allow the audience to evaluate the importance within their own research area
- To suggest interesting research directions

The global picture

- Lecture I - Introduction to UQ

  Motivation, terminology, background, Wiener chaos expansions.

- Lecture II - Stochastic Galerkin methods

  Formulation, extensions, polynomial chaos, and examples.

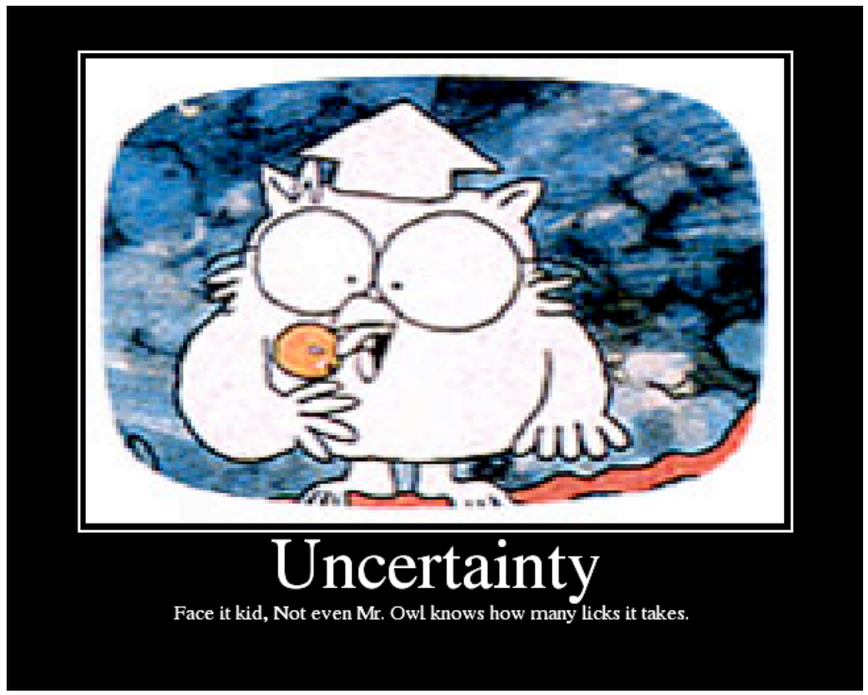
- Lecture III - Stochastic Collocation methods

  Motivation, formulation, high-d integration, and examples.

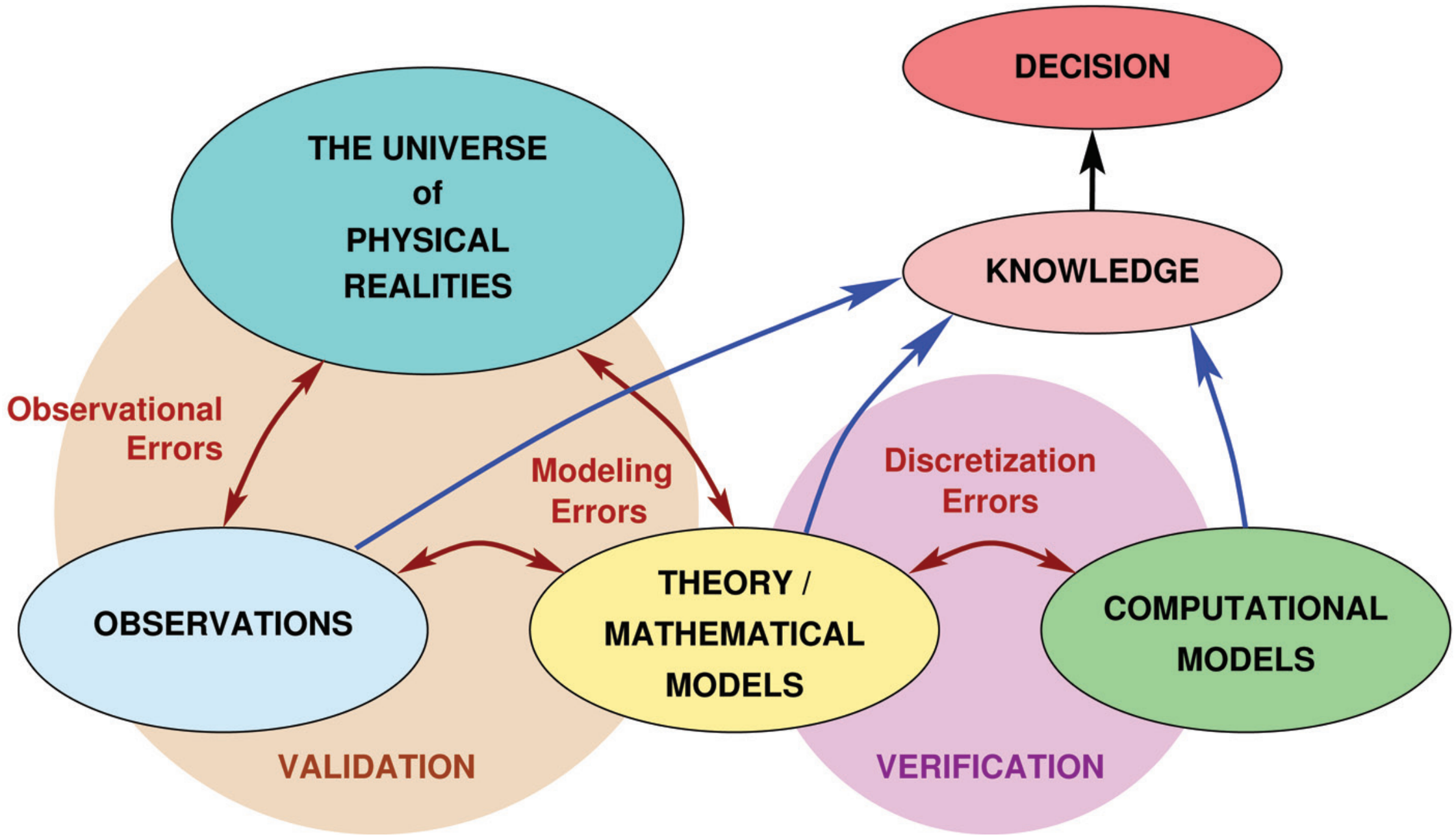
- Lecture IV - Extensions, challenges, reduced order modeling, and open questions

The local picture

- A few examples to motivate the need for UQ
- Classification of types of uncertainty
- Probability 101
- Overview of classic and some newer computational techniques to deal with uncertainty.
- The Wiener Chaos expansion
- Summary



Computational Science



Oden, Moser, Ghattas 2010, SIAM News

Motivation for UQ - V&V

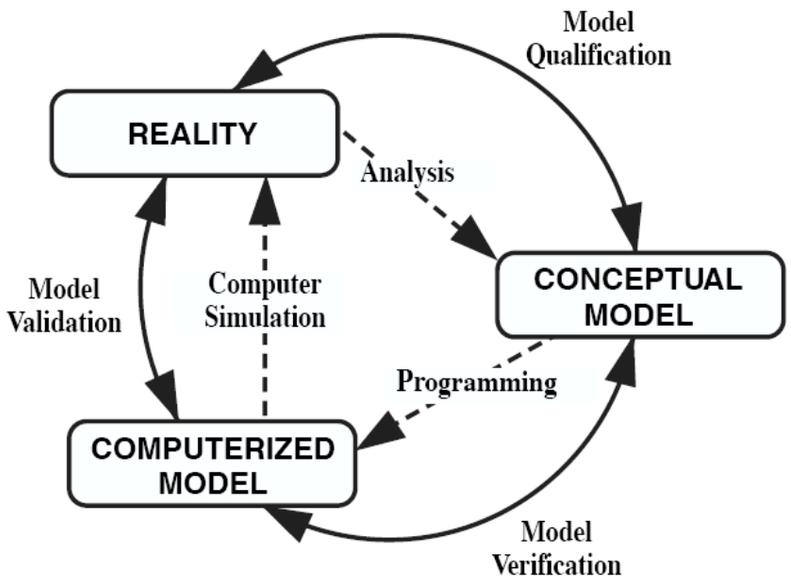
Consider the classic problem of Verification and Validation (V&V) in computational science

- Verification - the need to make sure the problem is solved correctly :

  Convergence, constructed solutions, analysis, stability etc

- Validation - the need to make sure the right problem is solved :

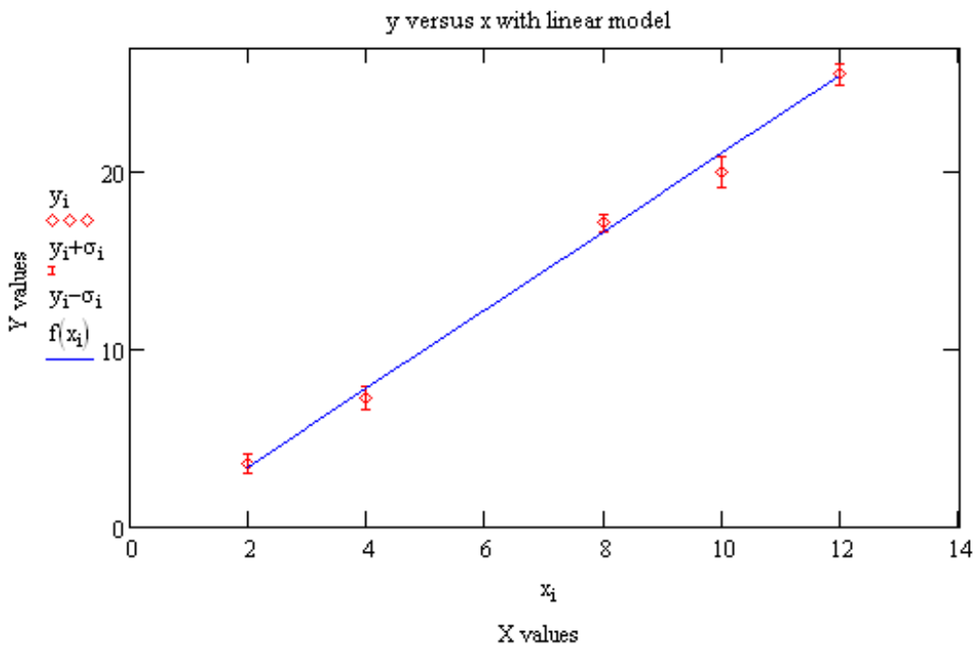
  Comparison against other codes, experimental data etc



While the former is well known, the latter quickly gets complex

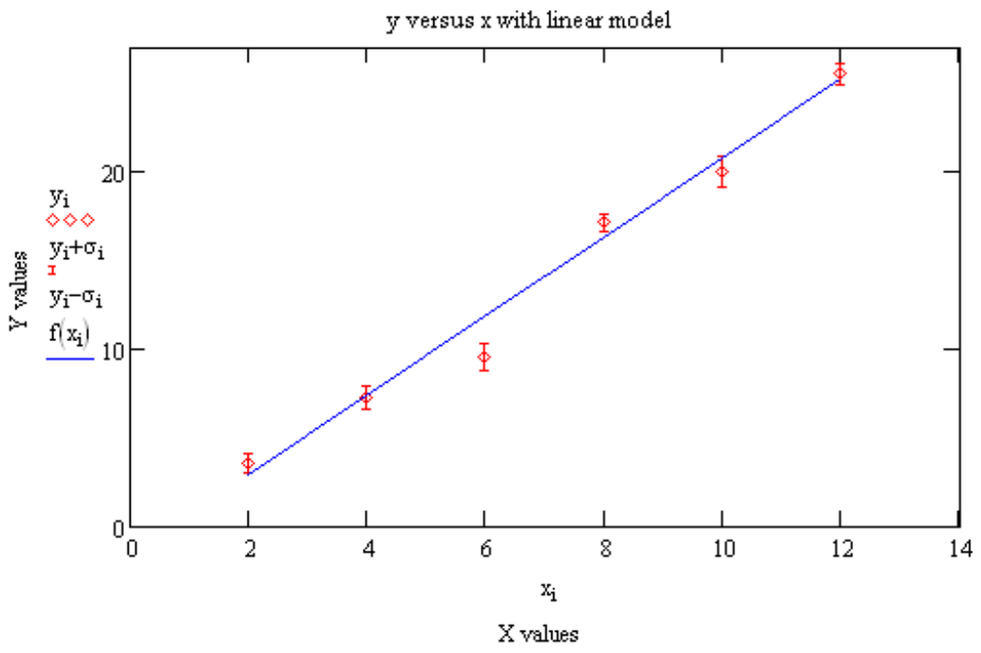
Motivation for UQ - V&V

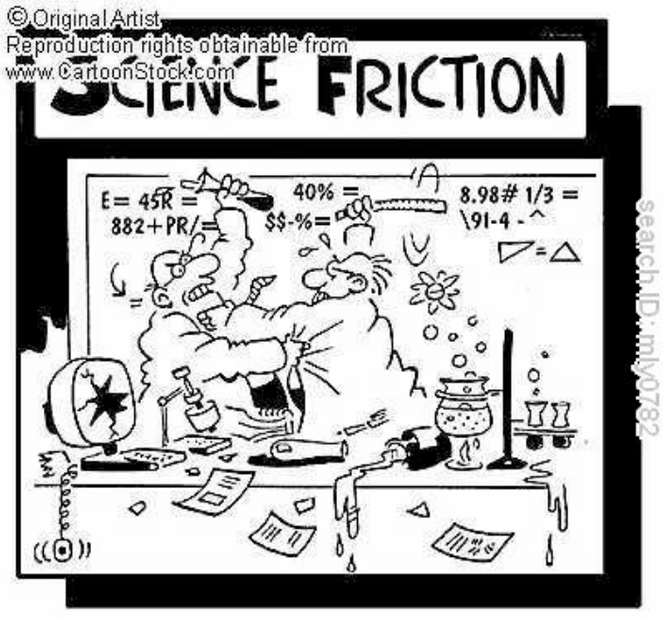
Imagine this outcome



All are pleased and happy

Imagine instead this outcome



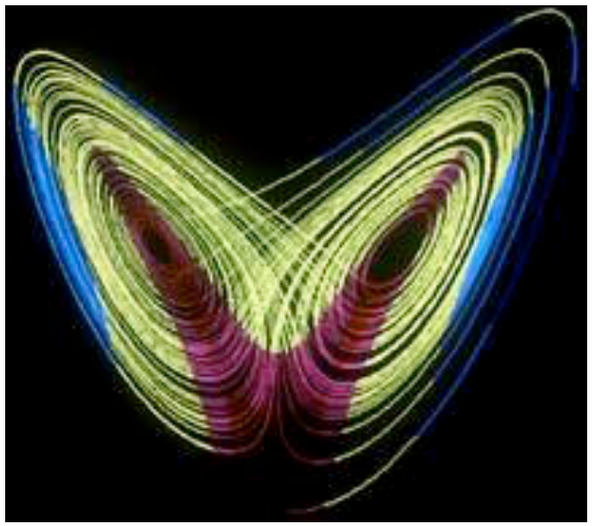


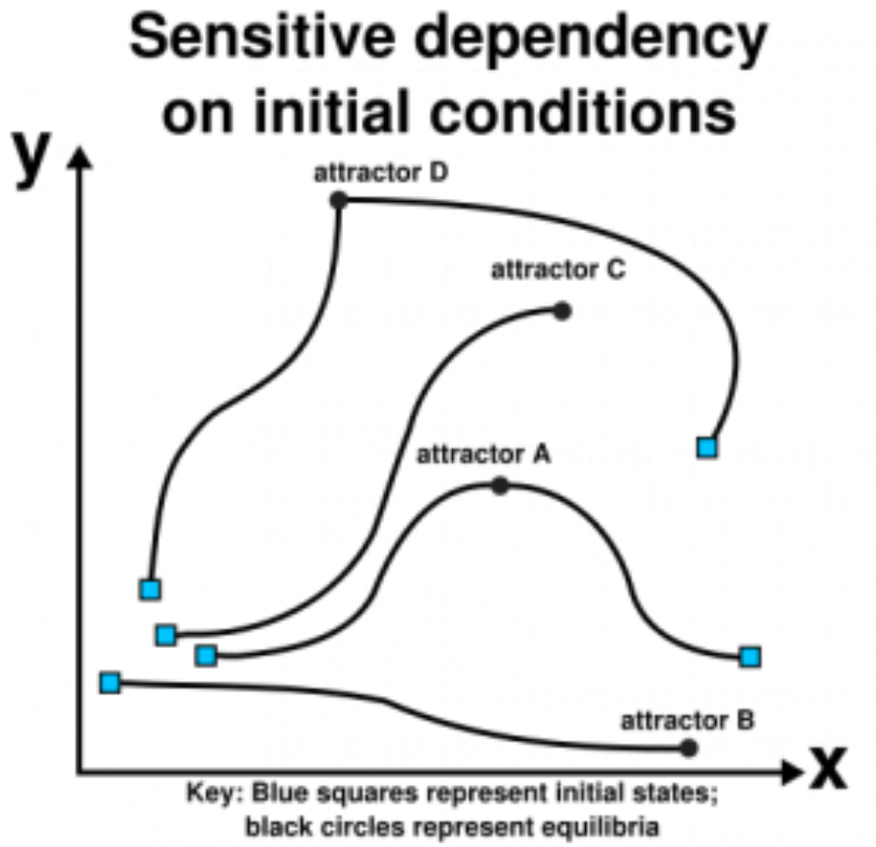
Let the blame game begin

Motivation for UQ - V&V

.. but what is/could really be going on ?

Imagine the problem is sensitive in some way





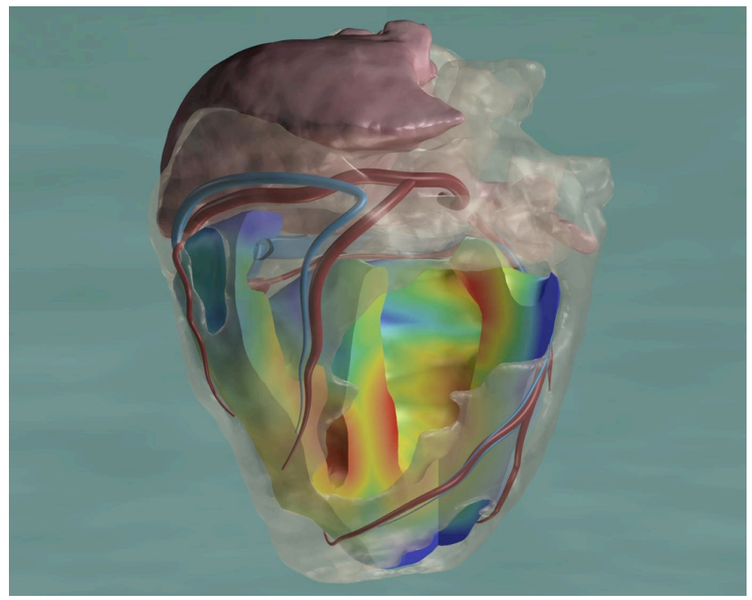
We could simply be solving different problem due to

- initial conditions
- boundary conditions
- parameters

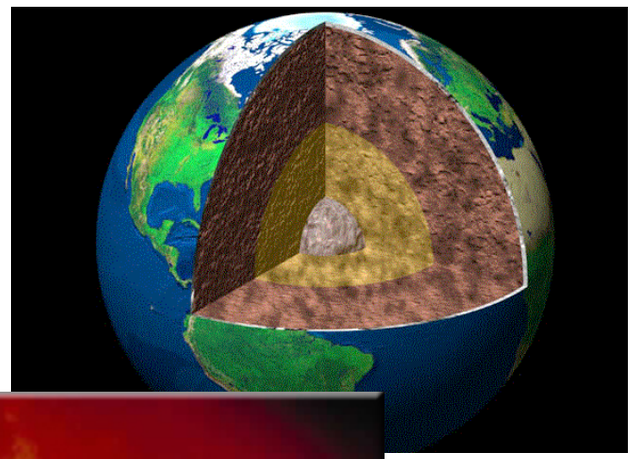
Motivation for UQ - unmeassuables

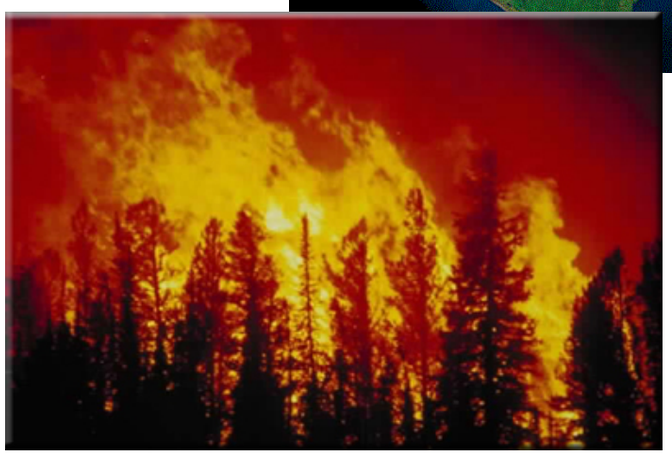
Many types of problem have inaccessible or unmeasurable parameters and characterization

Biological systems

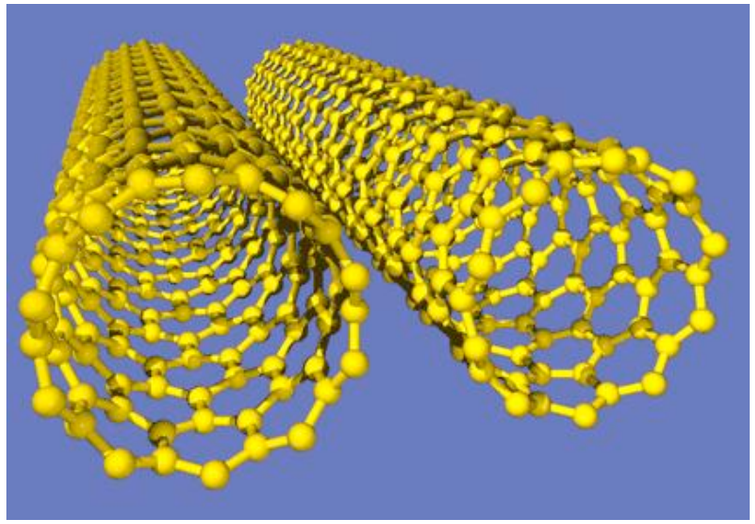


Modeling of complex environments





Micro/nano scale materials



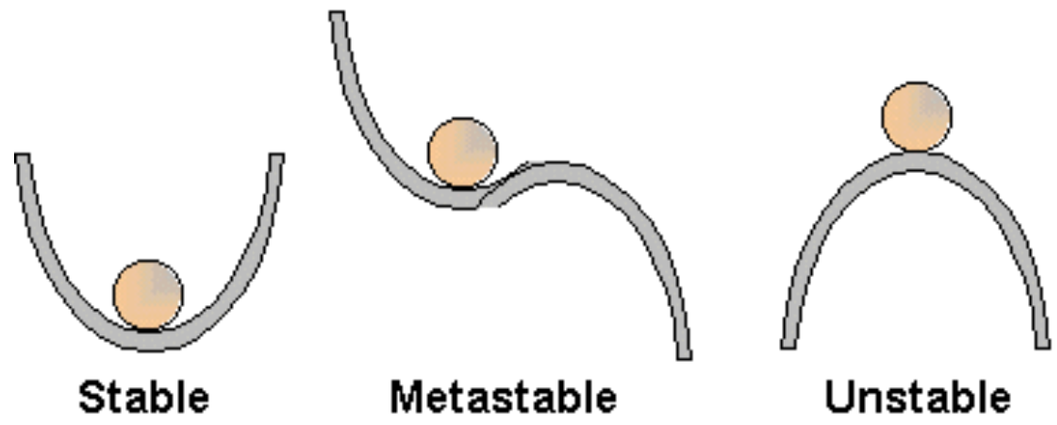
One can question the value of deterministic modeling of such systems - but then what ?

Motivation for UQ - optimization/design

In computer assisted optimization and design, one seeks to minimize a cost-function

$$\min_{x \in \Omega} J(x, \bar{\mu})$$

Why is this problematic ?



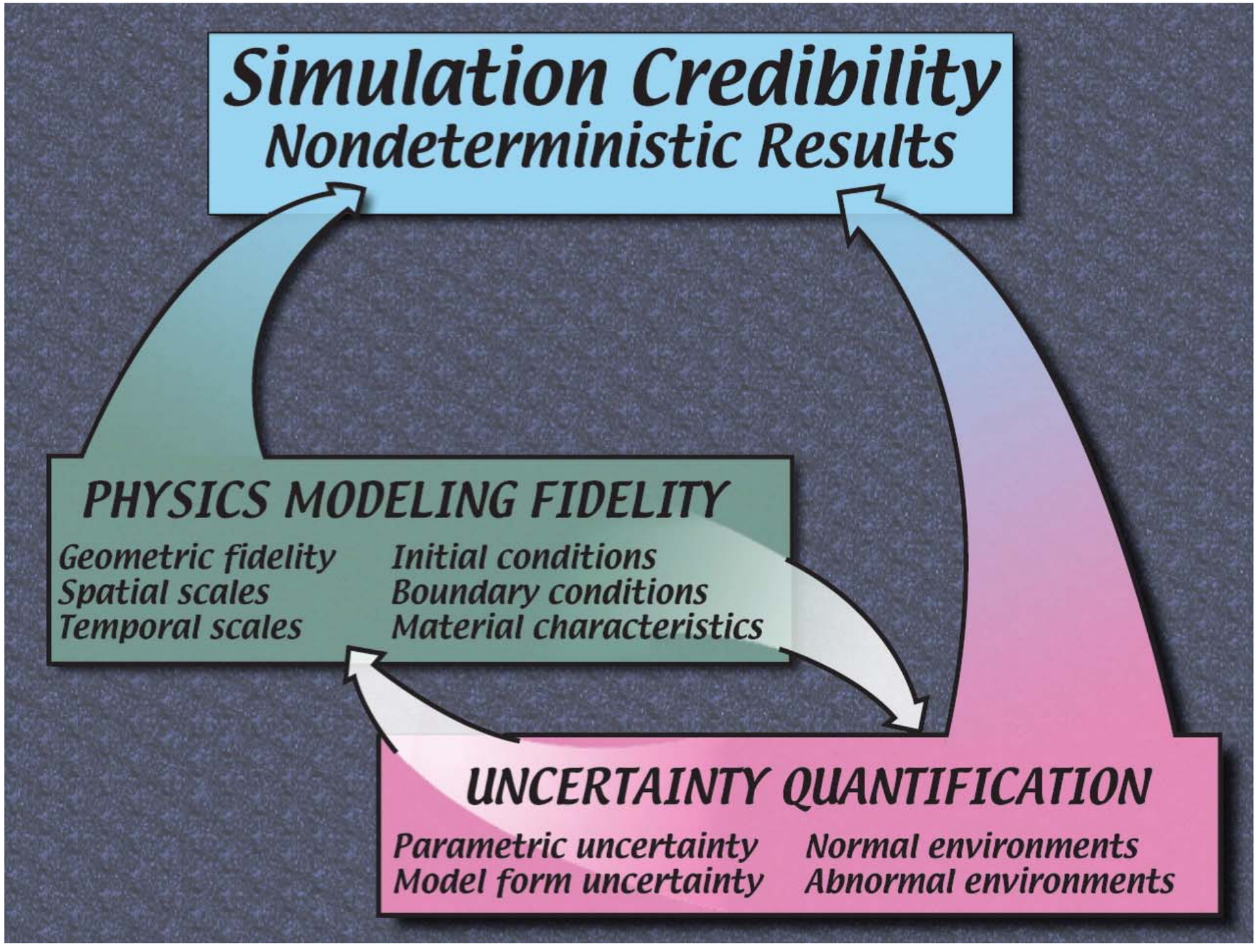
A better approach may be

$$\min_{x \in \Omega} \mathsf{E}[J(x, \mu)] + \kappa \mathrm{var}(J(x, \mu))$$

.. but this requires us to be able to evaluate the impact of the parametric uncertainty - quickly.

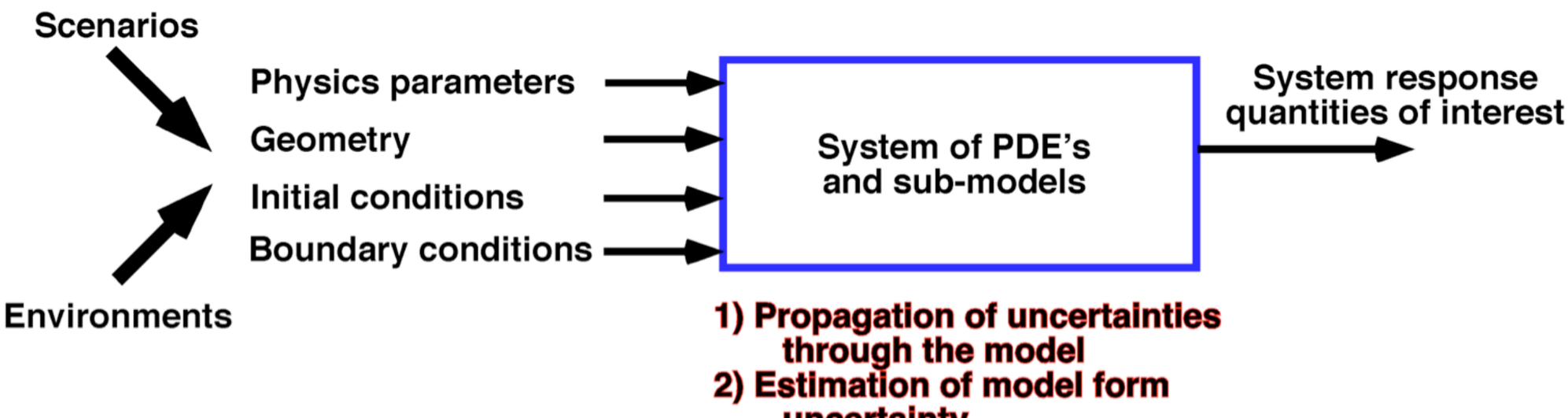
Motivation for UQ

Suggests that we need to re-evaluate our computational approach to achieve a true predictive capability



Motivation for UQ

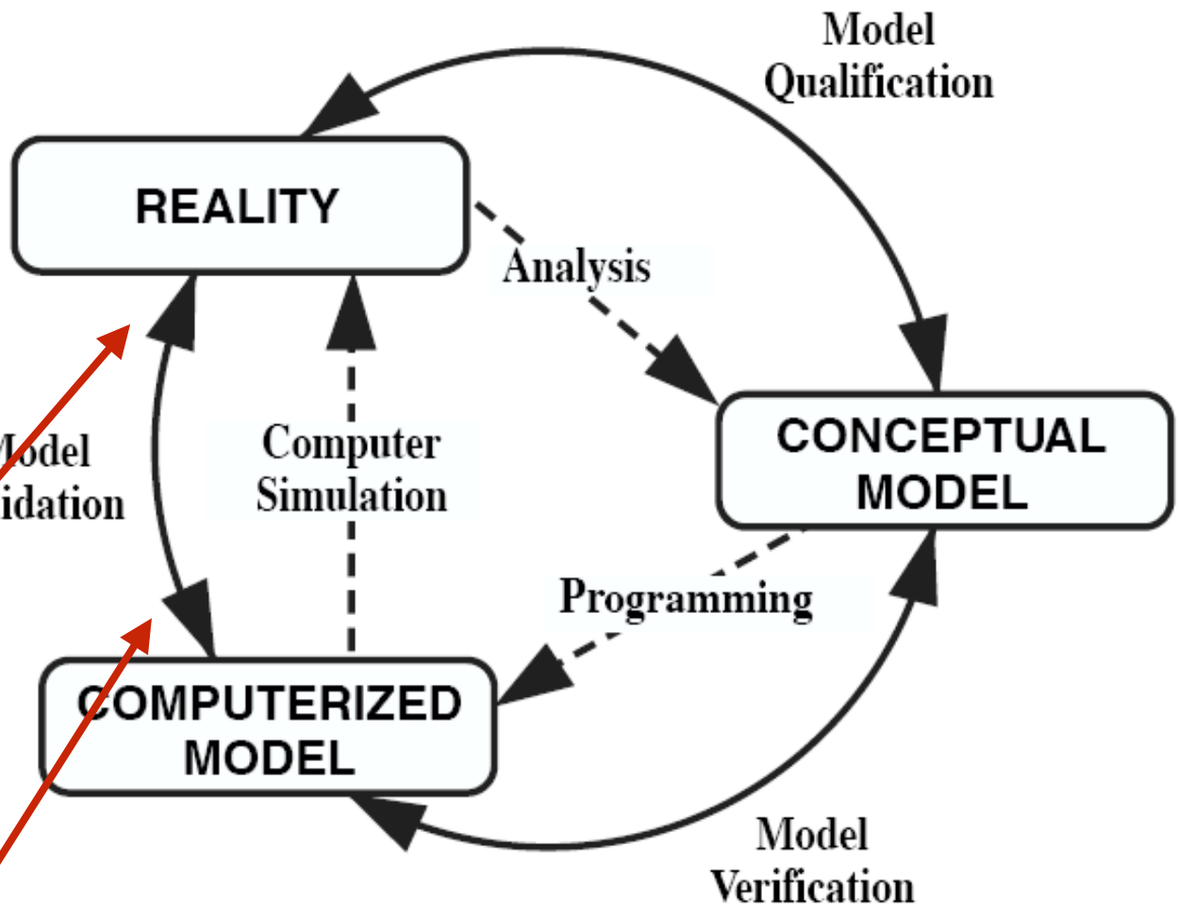
We need to consider a more complex problem



This raises important questions such as

- How do we do this reliably ?
- What is the cost ?
- Do we need to develop everything from bottom up ?

The other side of the story



The forward problem - today

The inverse problem - tomorrow

A little terminology

Before embarking on this, let us revisit the sources of uncertainty and how we can hope to control them

Uncertainty can be caused by a number of things

- Initial and boundary conditions
- Geometries
- Parametric variations
- Modeling errors
- Sources
- etc

A little terminology

- Aleatory uncertainty is inherent variation of the physical system and the environment
  - variability, irreducible uncertainty, random uncertainty etc

  Examples: Rapid variations in parameters, inherent randomness in a microstructure etc

- Epistemic uncertainty is caused by insufficient knowledge of parameters or processes
  - subjective uncertainty, reducible uncertainty, model uncertainty

  Examples: Insufficient experimental results, poor understanding of system, microstructure etc

A little terminology

With the need to consider systems subject to both types of uncertainty, there is no alternative but to model the uncertainty in some way

- Aleatory uncertainty is often modeled by some assumed probability measure.

- Epistemic uncertainty is more problematic as it is grounded in insufficient knowledge of the system.

  It is often modeled by intervals of possible values

  ... but it is, in principle, reducible at added cost.

Probability 101

Before we continue, let us make sure we recall the necessary background and terminology.

- The outcome of experiment is an event - $\omega \in \Omega$

  Ex: Flipping a coin gives head or tails. $\Omega = \{\text{head}, \text{tail}\}$

- We assign a number to the outcome to recover a Random variable - $X = X(\omega)$

  Ex: $X(\omega) \in [0, 1]$

- The event space, the empty set, and a number of set combinations is called the σ-field - called $\mathcal{F}$

- We assign probabilities to measure likelyhood of the outcome of the random variables - $P(\omega : X(\omega)) \in [0, 1]$

  Ex: $P(\omega : X(\omega) = 0) = P(\omega : X(\omega) = 1) = 0.5$

Probability 101

Definition 2.4 (Probability space). *A probability space is a triplet* $(\Omega, \mathcal{F}, P)$ *where* Ω *is a countable event space,* $\mathcal{F} \subset 2^{\Omega}$ *is the* σ*-field of* Ω*, and* P *is a probability measure such that*

1. $0 \leq P(A) \leq 1, \forall A \in \mathcal{F}$.
2. $P(\Omega) = 1$.
3. *For* $A_1, A_2, \ldots \in \mathcal{F}$ *and* $A_i \cap A_j = \emptyset, \forall i \neq j$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

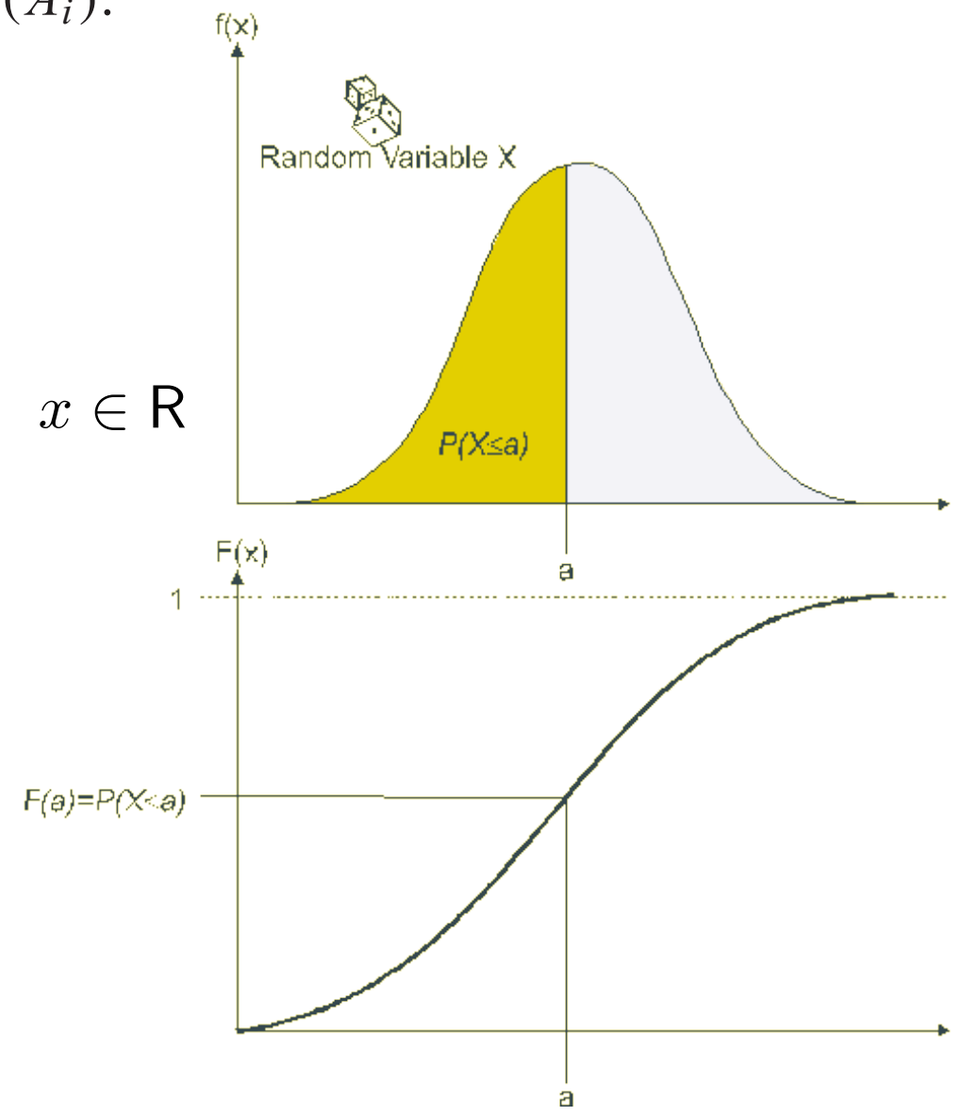
- The probability distribution

$$F_X(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}), \quad x \in \mathrm{R}$$

- The probability density

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy$$

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(y)\, dy = 1$$



Probability 101

We can now more appropriately define

- The mean or expectation

$$\mu_X = \mathsf{E}[X] = \int_{-\infty}^{\infty} x f_X \, dx \qquad \mathsf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X \, dx$$

- The variance

$$\sigma_X^2 = \operatorname{var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X \, dx$$

- The m'th moment

$$\mathsf{E}[X^m] = \int_{-\infty}^{\infty} x^m f_X \, dx$$

- The standard deviation

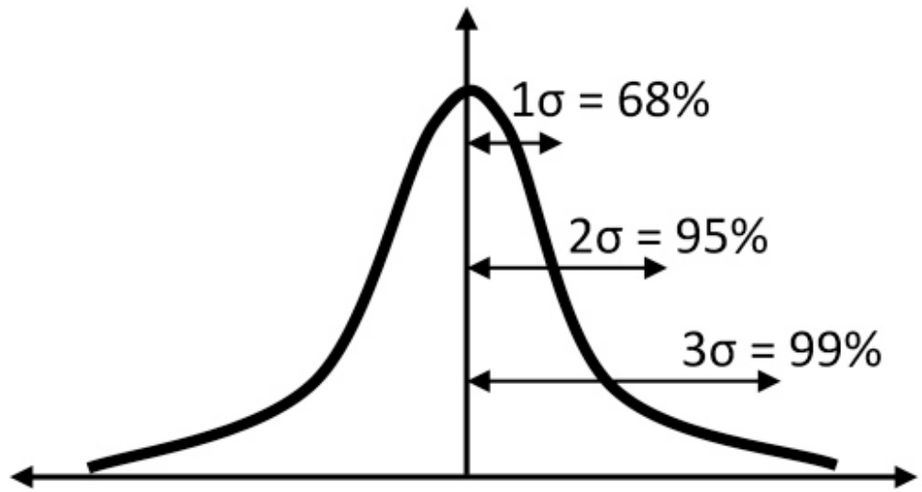
$$\sigma_X^2 = \mathsf{E}[X^2] - \mu_X^2$$

Probability 101

Let us recall a couple of widely used densities

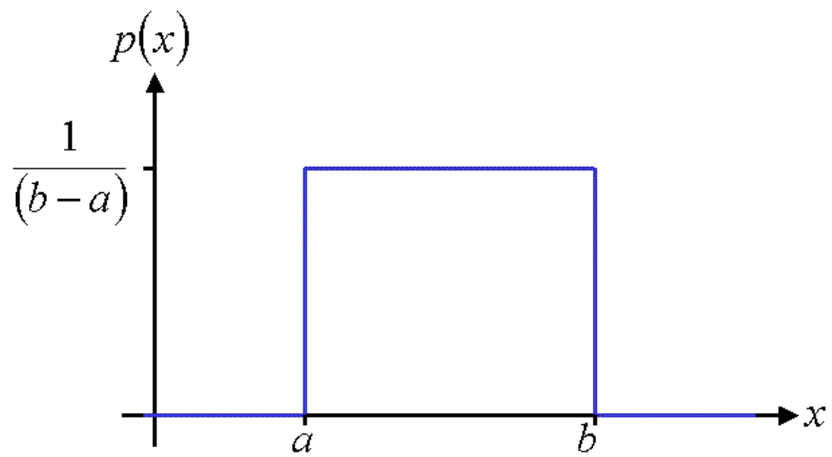
- The normal/Gaussian distribution - $N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],$$



- The uniform distribution - $U(a, b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & \text{otherwise}, \end{cases}$$



Probability 101

The last concepts we will recall are

- The correlation between two random variables is

$$\text{corr}(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \qquad -1 \leq \text{corr}(X_1, X_2) \leq 1.$$

$$\text{cov}(X_1, X_2) = \mathsf{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]$$

They are uncorrelated if $\text{corr}(X_1, X_2) = 0$

- The variables are independent if $P(A_1 \cap A_2) = P(A_1)P(A_2)$

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

$$\text{Independence} \Rightarrow \text{Uncorrelated}$$

(Some) classic methods for UQ

Let us now - finally - return to the quest at hand:

How do we account for the impact of the uncertainty on the output of a dynamical system

To briefly recall a few classic methods, let us consider the following problem

$$\frac{du}{dt}(t,\omega) = -\alpha(\omega)u, \qquad u(0,\omega) = \beta(\omega),$$

$$u(t,\omega) : [0,T] \times \Omega \to \mathsf{R}$$

- (α, β) **independent then** $u(t, X_1, X_2) : [0,T] \times \mathsf{R}^2 \to \mathsf{R}$

$$\frac{du(t,\omega)}{dt} = -X_1 u, \quad u(0,\omega) = X_2$$

- (α, β) **dependent then** $u(t, X, g(X)) : [0,T] \times \mathsf{R} \to \mathsf{R}$

$$\frac{du(t,\omega)}{dt} = -X u, \quad u(0,\omega) = g(X)$$

Monte Carlo methods

The truly classic and simple approach

1. Create M iid's from the assumed distribution - $(\alpha^{(i)}, \beta^{(i)})$
2. Solve problem for each set of iid's - $u^{(i)}(t) = u(t, X^{(i)})$
3. Compute required statistics - $\bar{u}(t) = \frac{1}{M} \sum_{i=1}^{M} u(t, X^{(i)}) \approx \mathsf{E}[u]$

As samples are based on iid's, the central limit theorem implies

$$|\bar{u}(t) - \mathsf{E}[u]| \approx \frac{1}{\sqrt{M}}$$

This result is both the curse and the strength

- The convergence is slow, i.e., expensive for good accuracy
- The convergence does not depend on dimension

Note: Lots of games in town to improve

A few remarks on MLMC

A powerful approach has emerged within the last few years to accelerate the classic MC - known as Multi-Level MC

Consider

$$E[u] = \frac{1}{M}\sum_{i=1}^{M} u_i$$

A standard way to accelerate convergence is to consider

$$E[u] = \frac{1}{M}\sum_{i=1}^{M} (u_i - \lambda(g_i - E[g]))$$

Where g is supposed to well correlated with f

This reduces variance and increases convergence

A few remarks on MLMC

Instead, consider

$$E[u] = E[u_1] + E[u - u_1]$$

Then, if

- $E[u_1]$ is cheap to compute
- $u \simeq u_1$

We can compute

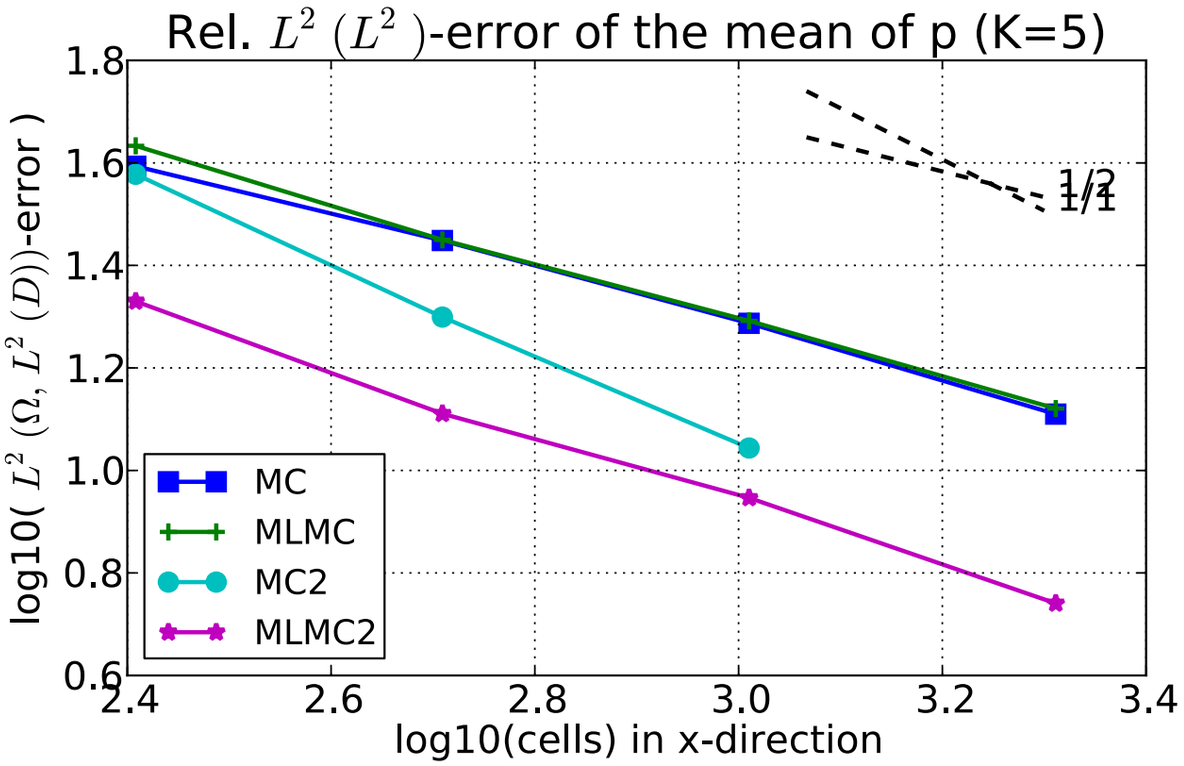
$$E[u] = \frac{1}{N_1} \sum_{i=1}^{N_1} (u_1)_i + \frac{1}{N_0} \sum_{i=1}^{N_0} (u_i - (u_1)_i)$$

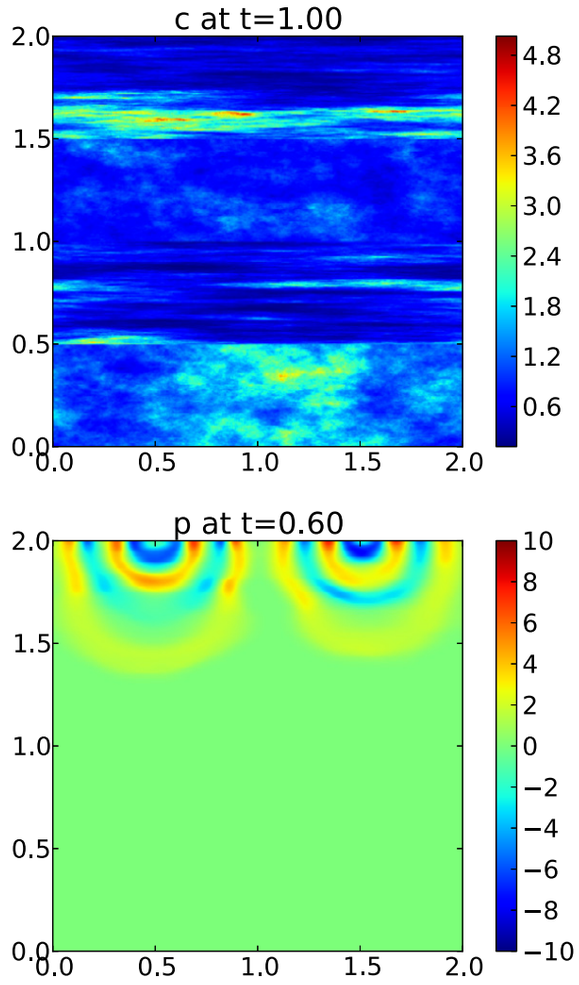
Fast Quick convergence

A few remarks on MLMC

Hence, there is a potential for substantial savings - in particular for multiple levels

Main challenge is to determine number of samples at





Accelerates MC - but convergence remains the same

Mishra et al, JCP 2016

Moment methods

Consider the following equation

$$\mu(t) = \mathsf{E}[u] \qquad \frac{d\mu}{dt} = -\mathsf{E}[\alpha u], \;\; \mu(0) = \mathsf{E}[\beta]$$

The unknown can be computed

$$\frac{d\mathsf{E}[\alpha u]}{dt} = -\mathsf{E}[\alpha^2 u], \;\; \mu(0) = \mathsf{E}[\alpha\beta]$$

Tempting

$$\frac{d\mu}{dt} = -\mathsf{E}[\alpha]\mu, \;\; \mu(0) = \mathsf{E}[\beta]$$

.. and can be continued

$$\mu_k = \mathsf{E}[\alpha^k u] \qquad \frac{d\mu_k}{dt} = -\mu_{k+1}, \;\; \mu_k(0) = \mathsf{E}[\alpha^k \beta]$$

but the 'closure problem' remains - often solved by assuming

$$\mu_{k+1} = g(\mu_0, \ldots, \mu_k)$$

- Accuracy is unclear in model
- Complex for a large problem

Perturbation methods

Let us assume that $\epsilon = O(\alpha(\omega)) \sim \sigma_\alpha \ll 1$

Then we can express the solution as

$$u(t,\omega) = u_0(t) + \alpha(\omega)u_1(t) + \alpha^2(\omega)u_2(t) + \cdots,$$

Matching orders in the expansion yields

$$O(1): \quad \frac{du_0}{dt} = 0, \qquad O(\epsilon): \quad \frac{du_1}{dt} = -u_0, \qquad O(\epsilon^2): \quad \frac{du_2}{dt} = -u_1,$$

- Only applies for small variance(s)
- Derivation is problem specific

Challenges in classic methods

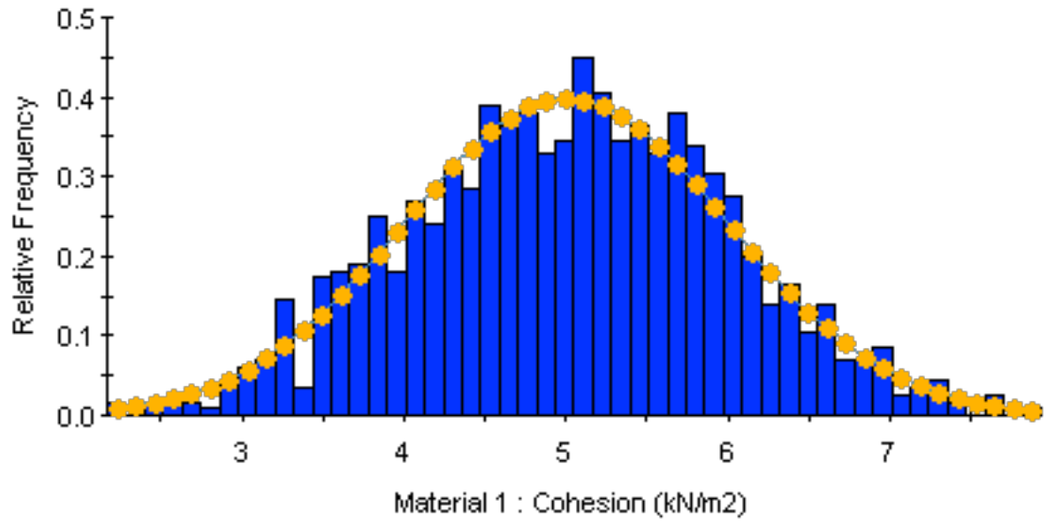
There are serious problems with these methods

- Moment/Perturbation methods are too complex and intrusive to serve as a general tool.

- Monte Carlo methods are flexible/ general - the convergence rate is problematic: 1 digit requires 100 simulations

How can be strive to improve this ?

MC approximates the density using piecewise constant samples - resulting in the slow pointwise convergence



Challenges in classic methods

But recall the context here - we will

- Make assumption on the nature of the aleatory uncertainty through the input variables.

- Make some assumptions on the character of the epistemic uncertainty, possibly using just uniformly distributed variables.

These are random variables with a smooth density

From an approximation standpoint, using piecewise constant functions to represent a smooth function is 'a poor choice'

Detour on global expansions

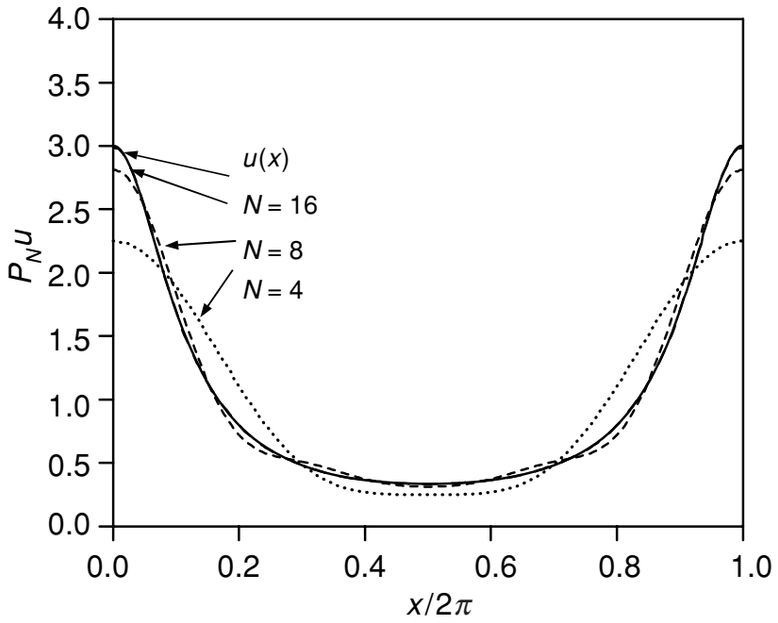
Consider the smooth period function

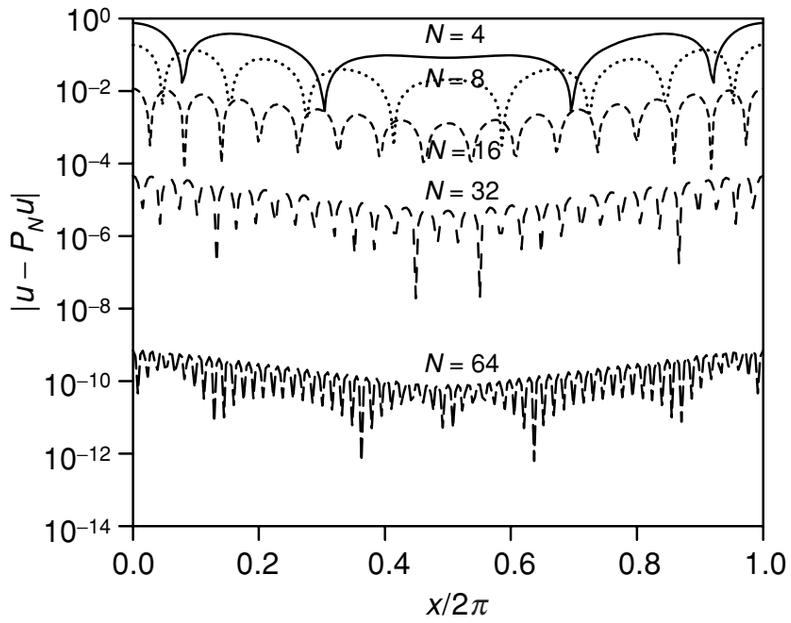
$$u(x) = \frac{3}{5 - 4\cos(x)}.$$

Using a Fourier series

$$\mathcal{P}_N u(x) = \sum_{|n| \le N/2} \hat{u}_n e^{inx},$$

$$\hat{u}_n = \frac{1}{2\pi} \int_0^{2\pi} u(x) e^{-inx}\, dx$$





Rigorous theory confirms that

$$\|u - \mathcal{P}_{2N} u\|_{W_p^q[0,2\pi]} \le \frac{C}{N^{r-q}} \|u\|_{W_p^r[0,2\pi]}. \qquad \|u - \mathcal{P}_{2N} u\|_{L^\infty} \le C \frac{1}{N^{q-\frac{1}{2}}} \left\|u^{(q)}\right\|_{L^2[0,2\pi]}.$$

Challenge: How to take advantage of this for UQ ?

The Wiener Chaos expansion

Define the space of square integrable functions

$$L^2_{F_X} = \{f : I \to \mathsf{R} \,|\, \mathsf{E}[f^2] < \infty\}$$

Provided $f \in L^2_{F_X}$ we have

$$f(X) = \sum_{n=0}^{\infty} \hat{f}_n \Phi_n(X) \qquad \hat{f}_n = \frac{1}{\gamma_n} \mathsf{E}[f(X)\Phi_n(X)]$$

$$\gamma_n = \mathsf{E}[\Phi_n^2(X)]$$

Where the basis - the Chaos Polynomial - satisfies

$$\mathsf{E}[\Phi_m(X)\Phi_n(X)] = \int \Phi_m(X(x))\Phi_n(X(x))\, dF_X(x) = \gamma_n \delta_{mn}$$

Basis depends on the distribution of the random variable !

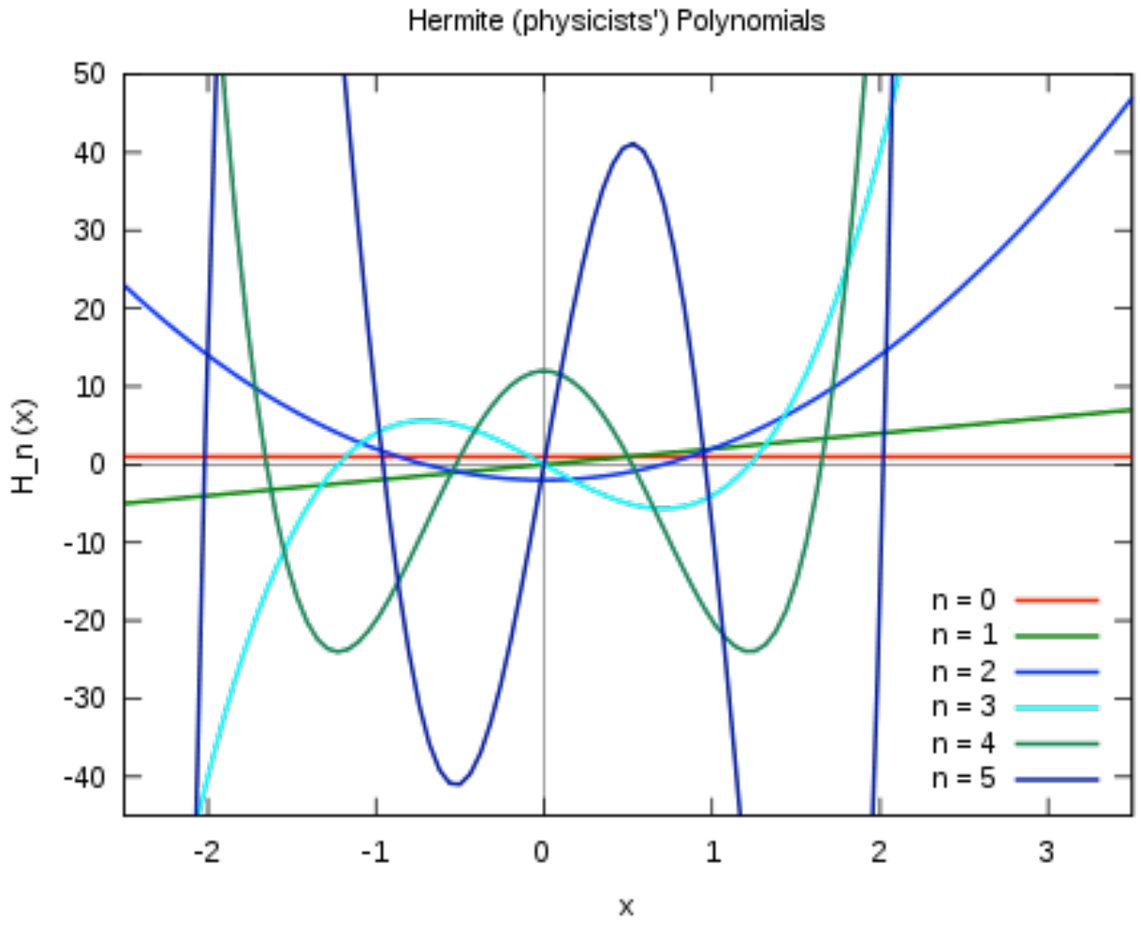
The Wiener Chaos expansion

Consider Gaussian variables with

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The corresponding polynomials are the Hermite Polynomials

$$H_0(X) = 1,\ H_1(X) = X,\ H_2(X) = X^2 - 1,\ H_3(X) = X^3 - 3X$$



The Wiener Chaos expansion

If we now consider the truncated expansion

$$\mathcal{P}_N f(X) = \sum_{n=0}^{N} \hat{f}_n \Phi_n(X)$$

then strong convergence follows directly from classic theory

$$\|f - \mathcal{P}_N f\|_{L^2_{F_X}} \to 0, \quad N \to \infty$$

If we now consider a more general problem

$$Z_N = \sum_{n=0}^{N} \hat{a}_n \Phi_n(X) \qquad \begin{matrix} X \in L^2_{F_X} \\ Z \in L^2_{F_Z} \end{matrix}$$

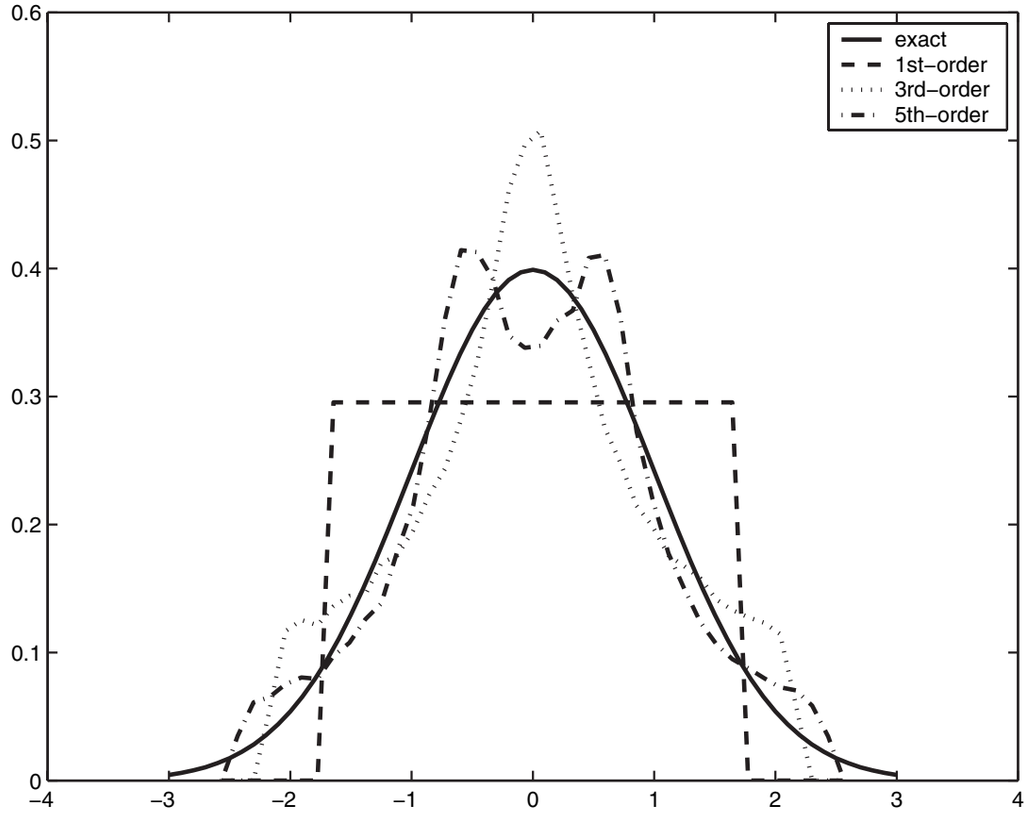
Weak convergence can be achieved by defining

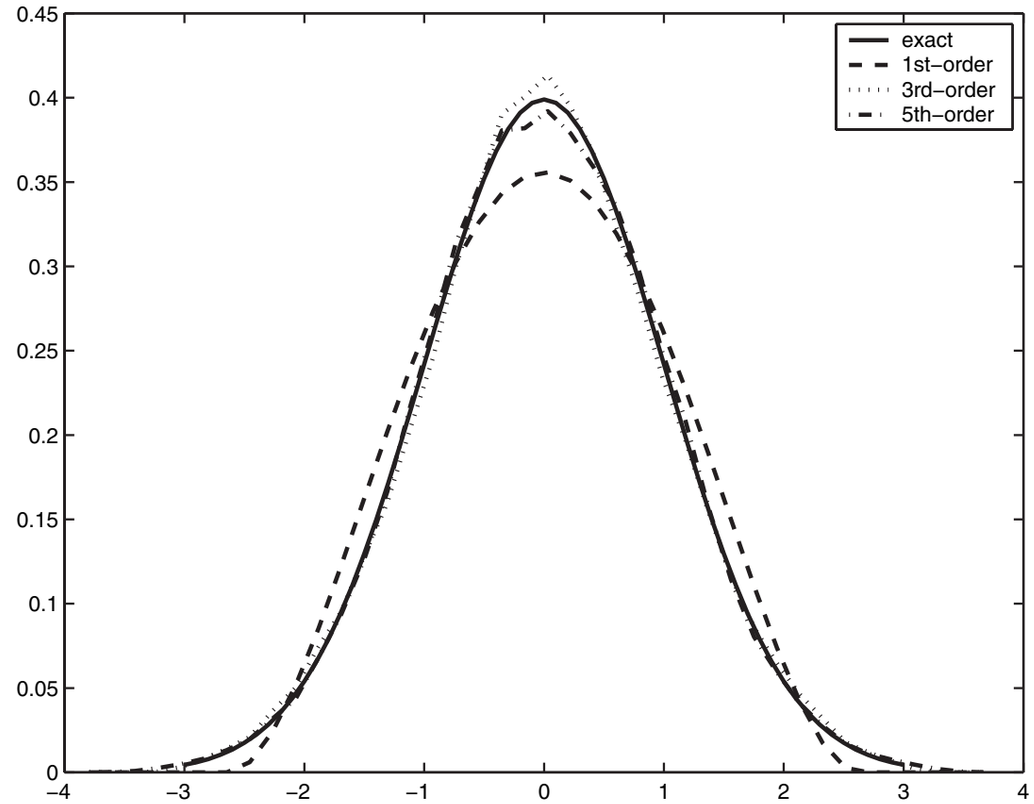
$$\hat{a}_n = \frac{1}{\gamma_n} \mathsf{E}_X[F_Z^{-1}(F_X(X))\Phi_n(X)]$$

We can use one random variable to approximate another

The Wiener Chaos expansion

Xiu, 2010





It is clear that choosing the right basis -- associated with the nature of the random variable -- is key to performance

This the advantage and the curse -- as we shall see

The Wiener Chaos expansion

The extension to multiple random variables follows

Define

$$F_{X_i}(x_i) = P(X_i \leq x_i) \qquad x_i \in I_{X_i}$$

$$\mathbf{X} = (X_1, \ldots, X_d) \qquad F_X = F_{X_1} \times \ldots \times F_{X_d}$$

and the multi-dimensional polynomial chaos

$$\Phi_i(\mathbf{X}) = \Phi_{i_1}(X_1) \times \ldots \times \Phi_{i_d}(X_d) \qquad |\mathbf{i}| \leq N$$

$$\mathsf{E}[\Phi_i(\mathbf{X})\Phi_j(\mathbf{X})] = \int \Phi_i(\mathbf{x})\Phi_j(\mathbf{x})\, dF_X(x) = \gamma_i \delta_{ij} \qquad \gamma_i = \mathsf{E}[\Phi_i^2]$$

The *homogeneous Chaos expansion* is

$$f_N(\mathbf{X}) = \sum_{|i|=0}^{N} \hat{f}_i \Phi_i(\mathbf{X}) \in \mathsf{P}_N^d \qquad \dim \mathsf{P}_N^d = \left(\begin{array}{c} N+d \\ N \end{array} \right) = \frac{(N+d)!}{N!d!}$$

The ‘curse of dimension’ shows its face !

The Wiener Chaos expansion

Assuming the Chaos expansion is known, we need statistics

$$f_N(\mathbf{X}) = \sum_{|i|=0}^{N} \hat{f}_i \Phi_i(\mathbf{X}) \in \mathsf{P}_N^d$$

The expectation follows from

$$\mu = \mathsf{E}[f] \approx \mathsf{E}[f_N] = \int \sum_{|i|=0}^{N} \hat{f}_i \Phi_i \, dF_X = \hat{f}_0$$

In a similar fashion, the variance is

$$\mathrm{var}(f) = \mathsf{E}[(f-\mu)^2)] \approx \sum_{|i|>0}^{N} \gamma_i \hat{f}_i^2$$

Other moments can be obtained in a similar fashion.

Functions of the expansion can also be estimated through Monte Carlo sampling

Summary

We have achieved quite a bit

- Motivated the need for UQ in Computational Science
- Discussed in some detail the shortcomings of classic methods such as Monte Carlo methods.
- Realizing that smoothness in the behavior of the random variables should be explored
- .. and introduced the Chaos expansion to achieve this

We have still to use this insight to solve differential equations - and demonstrate the promised benefits