Title: Need for Uncertainty Quantification in Predictions

Author(s):   Srinivasan, Gowri

Intended for: Talk to LANL (ADTSC) summer students

Issued: 2015-07-06

Need for Uncertainty Quantification in Predictions

Gowri Srinivasan

T-05

Los Alamos National Lab, Los Alamos, NM, USA

July 7, 2015

– Typeset by FoilTEX –

Abstract

Natural and engineered systems are often modeled through mathematical equations using computer codes. Finite resources, inability to accurately control experimental conditions,encounter, or explore the entire parameter space of interest are a few reasons that scientists and engineers rely on numerical simulations for predictive capability. Associated with every prediction is an uncertainty, since our ability to simulate a physical process is limited by our knowledge of the physics, and the parameters. In the last two decades, the need for Uncertainty Quantification (UQ) has been recognized in several fields and we often ask questions about the risk or uncertainty associated with events and make decisions accordingly. In this talk, I will illustrate the effects of model and parametric uncertainties and the need to include them in our predictive capabilities.

1

Outline

- Motivation

- Introduction

- The Ball Drop example

- Uncertainties in Chemical Reactions

- Conclusions

2

Motivation

- Most natural and man-made phenomena are modeled mathematically.

- Deterministic models assume that knowledge of Initial Conditions (ICs) yield full knowledge of the system's behavior at all later times.

- However, this assumption ignores the fact that mathematical models are only approximations of the underlying physics.

- Additionally, there are random effects that act on systems which can significantly affect the behavior of the system, coupled with errors in measurements of ICs that can affect our ability to predict the system's behavior.

- Hence it is necessary to consider stochastic modeling.

Introduction to Quantifying Uncertainties

From National Research Councils 2012 report on VVUQ - Assessing the Reliability of Complex Models: Mathematical Foundations of Verification, Validation, and Uncertainty Quantification

- Determine and account for important sources of uncertainty

  - State which sources are included in the assessment, and which are not.
  - Errors (model form, parametric, numerical etc.) are sources of prediction uncertainty.
  - Process involves physical reasoning (and virtues), as well as mathematical/statistical reasoning.

4

Quantifying Uncertainties in Predictions

- List and evaluate caveats and assumptions that go with the prediction uncertainty assessment.

- Explicitly describe the assumed connection between computational model(s) and the true, physical system

  - Many ways to do this; different labs/systems POCs may choose different approaches
  - Depends on the required prediction, model maturity, availability and relevance of experimental data, etc.
  - Process involves physical reasoning (and virtues), as well as mathematical/statistical reasoning.

- Exploit historical experiments, as well as reduced complexity experiments.

Some Background on Examples

- The Ball Drop example: We will first look at a second order differential equation used to model the free fall of a spherical ball. This is a simple well-defined problem with uncertainty in parameters and the choice of 3 models to predict travel time. Practical applications are seldom as straightforward as this example, but we hope to learn some useful insights that we can apply in more complicated problems.

- The Chemical Reaction example: Chemical reactions are prevalent in a wide range of applications including environmental (contamination, natural gas extraction),electronics, manufacturing, pharmaceutical industry etc. Typically the kinetics are modeled as nonlinear differential equations, and uncertainties about the kinetics must be well quantified for optimal results.

6

Motion of an Object in Air:
The Ball Drop example

Collaborators :

David Higdon, Virginia Tech
Scott Vander Wiel, Los Alamos National Lab

Governing Equations

Here we consider the basic ball drop example, where a (nearly) spherical ball is dropped from rest from a given height h_0 and we estimate the time it will take to reach the ground.

The simplest model is one where the only force acting on the ball is the gravitational force.

$$\frac{d^2h}{dt^2} = -g \tag{1}$$

So if we dropped a ball in vaccuum from a given height, would it always take the same time to hit the ground?

What happens to our predictions if this experiment were performed at different lattitudes and altitudes?

Governing Equations

The acceleration due to gravity g varies from $9.78 - 9.83 m/s^2$ depending upon lattitude. g is also smaller at higher altitudes. So now our predictions have to be adjusted for these variations. Let us assume normally distributed g with mean $9.8065 m/s^2$ and standard deviation $0.05 m/s^2$

9
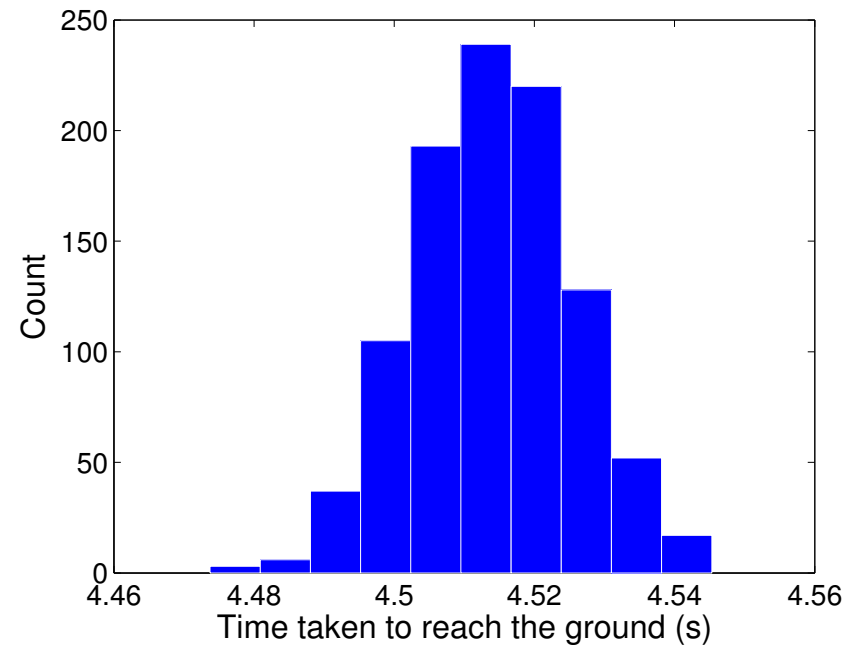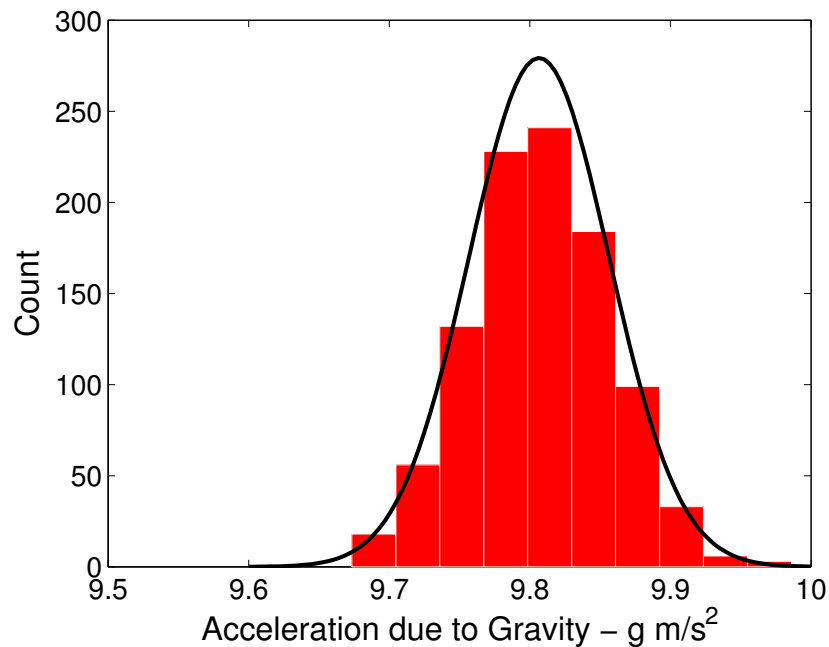
More practically...

We know however that in reality, a feather and bowling ball don't hit the ground at the same time when dropped from the same height. This is due to the force of air resistance acting on objects. Let us consider two separate models that are typically used to describe this $drag$ force. Here C_D is the drag co-efficient, ρ is the density and R is the radius.
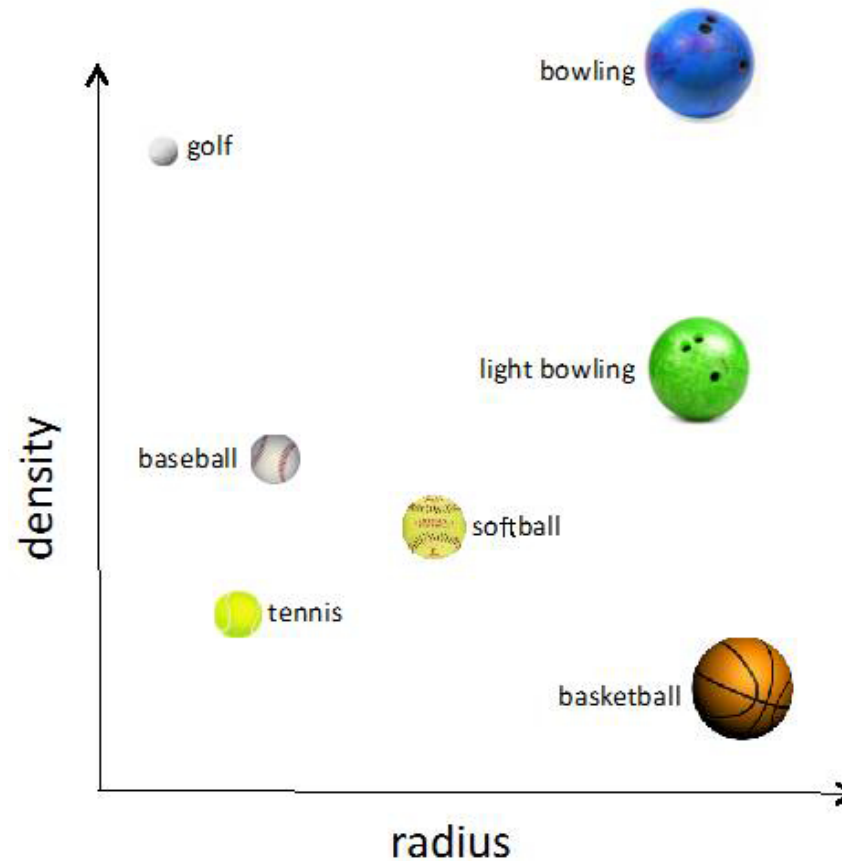
Model 1
$$\frac{d^2h}{dt^2} = -g + \frac{3}{8}\frac{C_D\rho_{air}}{R_{ball}\rho_{ball}}\left(\frac{dh}{dt}\right)^2 \tag{2}$$

Model 2
$$\frac{d^2h}{dt^2} = -g + \frac{300}{8}\frac{C_D\rho_{air}}{R_{ball}(\rho_{ball})^{1.5}}\left(\frac{dh}{dt}\right) \tag{3}$$
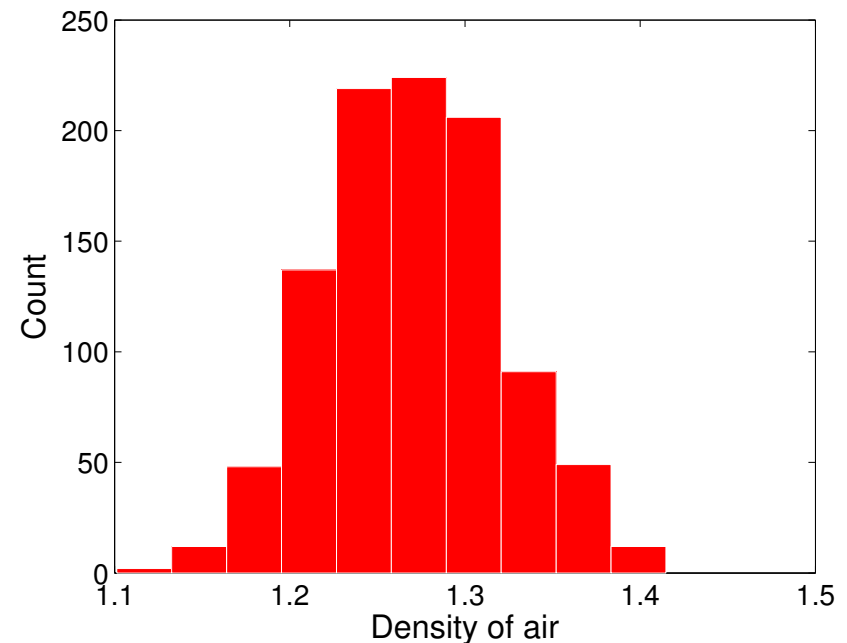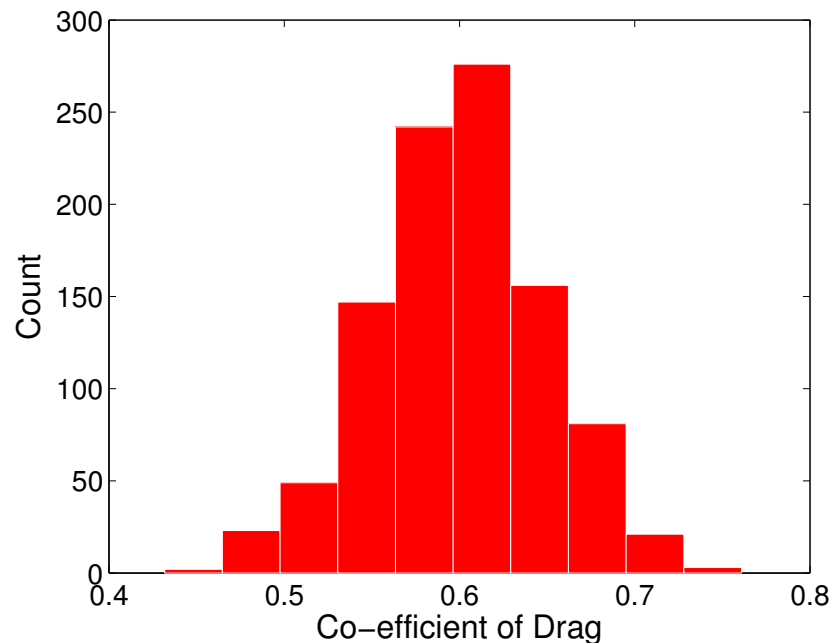
If you could perform experiments with only the bowling ball, could you predict the behavior of the others?
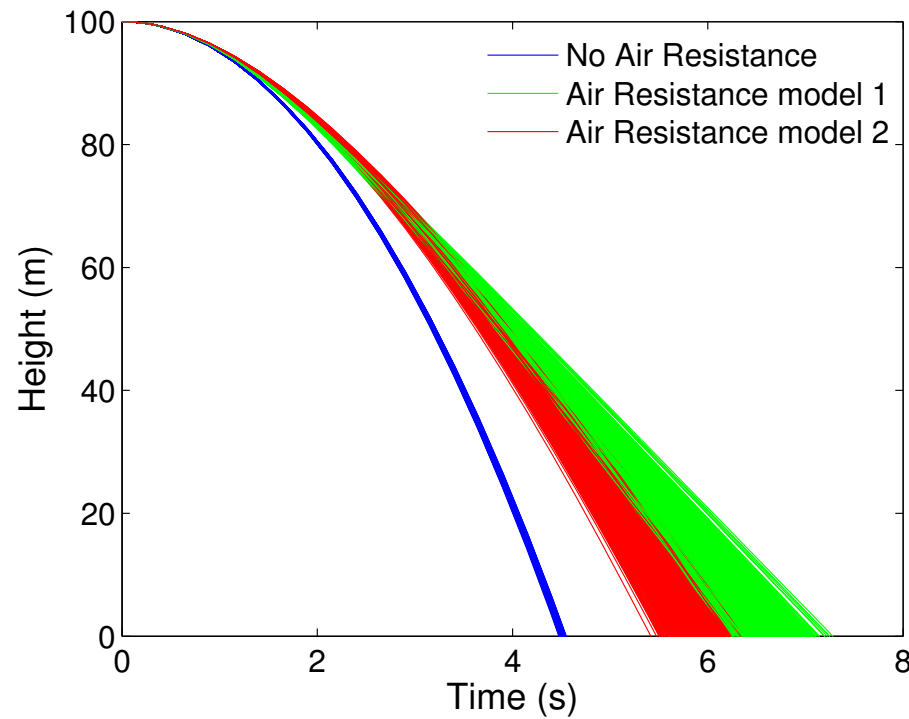

11

More parameters, more uncertainty

In additional to normally distributed g as before, we have variations in C_D and ρ_{air}, all represented as normal distributions.

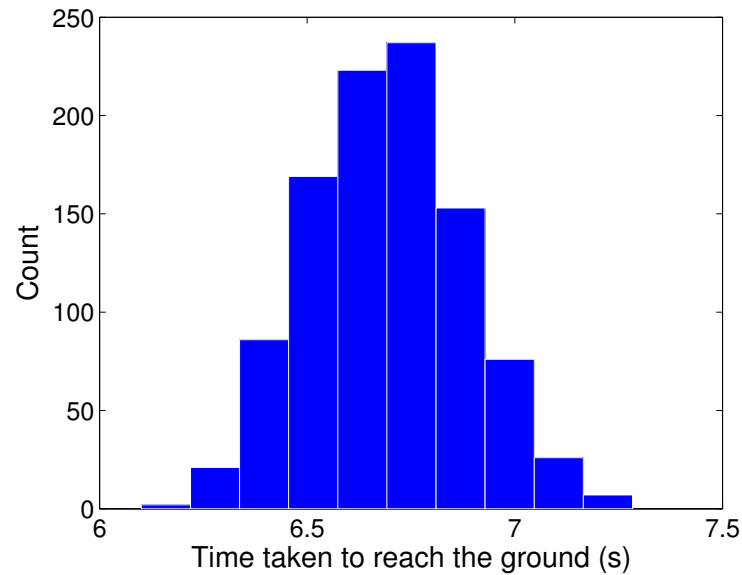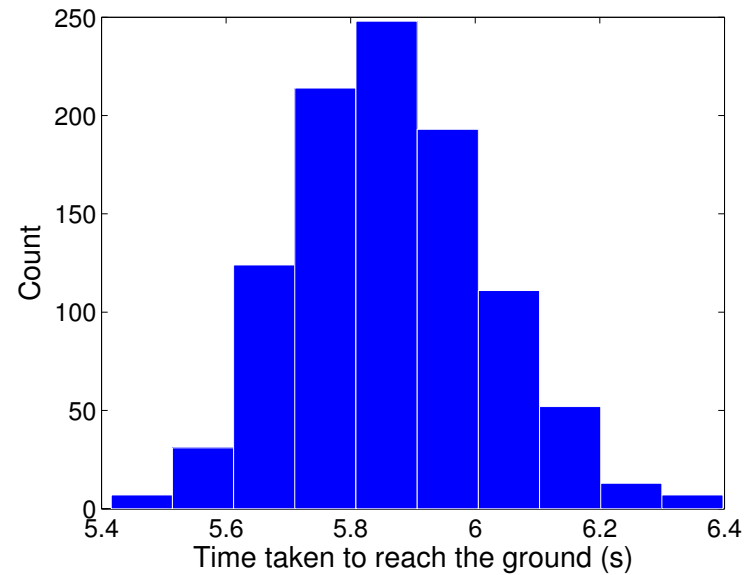# Comparison of the three models

Comparison of Ball Drop Times for 1000 simulations

13

Comparison of the air resistance models

Model 1

Model 2

14

Summary of Results

1. The assumptions in the governing equations can significantly affect the drop time for the ball.

2. Varying the parameters in the governing equation gives us a distribution of drop times.

3. The governing physics is sometimes represented by more than one model form, and complete uncertainty analysis must include all known model forms and parametric uncretainties in order to obtain a full statistical distribution of resulting quantities, in this case, the drop time.

Quantification of Uncertainties in Geochemical Reactions

Collaborators :

Daniel Tartakovsky, UC San Diego

Bruce Robinson, Los Alamos National Lab

Alejandro Aceves, Southern Methodist University

16

Groundwater Transport Problem

- Contaminants are transported in groundwater by many physical processes.

- Advection : Transport along with flow of groundwater.

- Diffusion : Movement from areas of higher concentration to lower concentration.

- Dispersion : Movement along and perpendicular to flow due to mixing of flowpaths with varying velocities. Diffusion and Dispersion are represented together.

- Advection Dispersion Equation

$$\frac{\partial C}{\partial t} = -v\frac{\partial C}{\partial x} + \frac{\partial}{\partial x}\left(D(x)\frac{\partial C}{\partial x}\right)$$

17

Groundwater Transport Problem

- Sorption : Process by which solutes sometime attach to rocks and other minerals.

- Reaction : Interaction with other minerals in the ground to yield other chemicals.
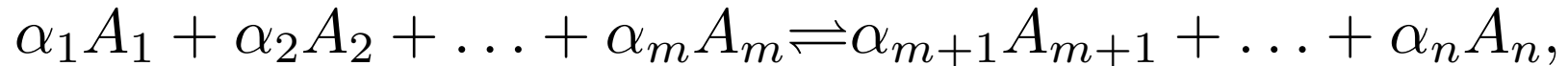
$$\frac{dC}{dt} = F(C, C_i)$$

- Advection Dispersion Reaction Equation

$$\frac{\partial C}{\partial t} = -v \cdot \nabla C + \nabla \cdot D \nabla C + F(C, C_i)$$

- We will first consider uncertainties in reactions and look at the transport problem separately.

Chemical Reactions

Consider a (reversible) chemical reaction between n species $A_1, A_2, ... A_n$ that can be represented as

$$\alpha_1 A_1 + \alpha_2 A_2 + \ldots + \alpha_m A_m \rightleftharpoons \alpha_{m+1} A_{m+1} + \ldots + \alpha_n A_n,$$

where $\{\alpha_i\}_{i=1}^n$ are stoichiometric coefficients. The concentration $C_i(t)$ of a reacting species A_i at time t is described by a nonlinear rate equation,

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = F_i(C_1, C_2, ... C_n), \qquad i = 1, \ldots, n$$

subject to appropriate initial conditions.

A Simple Chemical Reaction

Consider an irreversible chemical reaction involving three reacting species, two reactants A_1 and A_2 forming a species A_3, such that

$$A_1 + A_2 \rightarrow A_3.$$
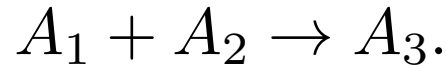
The reaction rate equations for the concentrations C_i of each species can be written as

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = -kC_1C_2, \quad i = 1, 2, \qquad \frac{\mathrm{d}C_3}{\mathrm{d}t} = kC_1C_2,$$

where the reaction rate constant k, initial conditions are specified

$$C_1(0) = C, \qquad C_2(0) = C, \qquad C_3(0) = 0,$$

C and k have precisely determined values.

20

Deterministic Model

The system of reaction equations described in the previous slide can be solved exactly by introducing new variables

$$u_1 = C_1 - C_2, \quad u_2 = C_1 + C_3, \quad u_3 = C_1 + C_2 + 2C_3$$

which yields a set of decoupled equations

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = 0, \qquad i = 1, \ldots, 3.$$

Due to the Initial Conditions, $C_1(t) = C_2(t)$, which leads to

$$C_1(t) = C_2(t) = \frac{C}{Ckt + 1}, \qquad C_3(t) = \frac{Ckt}{Ckt + 1}.$$

Uncertainties in Chemical Reactions

- Parametric uncertainty : Refers to the imperfect knowledge about the coefficients entering the functions F_i $(i = 1, \ldots, n)$ and/or initial concentrations.

- Parametric uncertainties relating to reaction rate constant are studied here.

- Model uncertainty : Refers to the imperfect knowledge about the functional forms of F_i $(i = 1, \ldots, n)$.

- Model Uncertainty relating to imperfect molecular collisions and imperfect mixing are considered.

- The solution in the determinsitic case is compared with those that consider each or both types of uncertainty.

Quantification of Parametric Uncertainty

This uncertainty can be quantified by treating the reaction rate constant k as a random process $k(t)$ with a mean $\bar{k}$, variance σ_k^2.

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = -\bar{k}C_1C_2 - \delta k C_1 C_2, \quad i = 1, 2, \qquad \frac{\mathrm{d}C_3}{\mathrm{d}t} = \bar{k}C_1C_2 + \delta k C_1 C_2,$$

The numerical solution is given (Kloeden and Platen,)by

$$C_1(t + \Delta t) = C_1(t) - \bar{k}C_1(t)C_2(t)\left[\Delta t + \Delta W\right],$$

where Δt is a time step and ΔW is a Gaussian variable with zero mean and the variance proportional to the variance of k.

Solutions of FPE - Reaction problem

The stochastic rate equations can be written as follows:

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = -\bar{k}C_1C_2 - \delta k C_1 C_2, \quad i = 1, 2, \qquad \frac{\mathrm{d}C_3}{\mathrm{d}t} = \bar{k}C_1C_2 + \delta k C_1 C_2, \tag{4}$$

Close inspection of the Langevin-type equations reveals that they have only one independent variable, say concentration C_1. Now, the probability density function for C_1, $p(C_1; t)$ satisfies the Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial v p}{\partial C_1} + \frac{\partial^2}{\partial C_1{}^2}\left(Dp\right), \tag{5}$$

24

where the drift and diffusion coefficients are given according to Ito calculus by

$$v = \bar{k}C_1^2 \quad \text{and} \quad D = \sigma_k^2 C_1^4, \tag{6}$$

respectively. An initial condition for (5) reflects the degree of certainty in the initial concentration C and, in general, is expressed by a probability density function, $p(C_1; 0) = p_0(C_1)$. If it is known precisely, the initial condition takes the form $p(C_1; 0) = \delta(C_1 - C)$, where $\delta(x)$ is the Dirac delta function.

We solved the Fokker-Plank equation (5)–(6) both analytically and numerically with a forward Euler algorithm. The analytical solution for the Fokker-Planck equation can be obtained by setting the probability current S defined below to 0
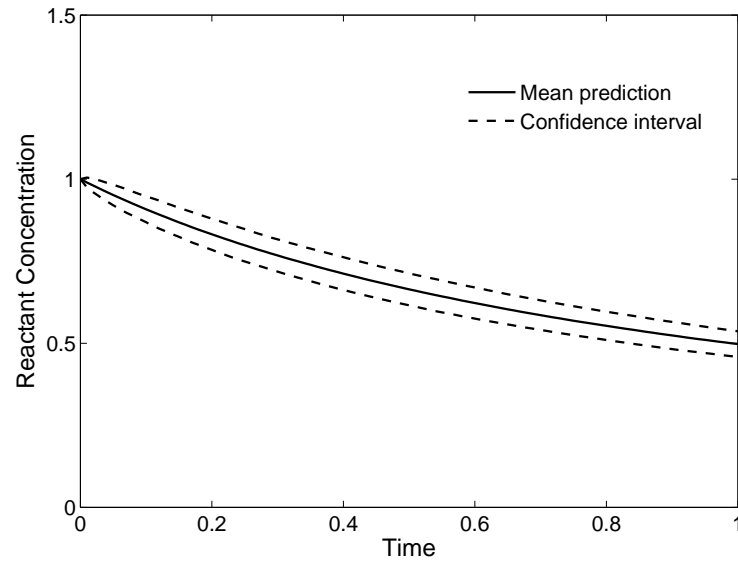
$$\left[v - \frac{\partial}{\partial C_1} D \right] p = 0 \tag{7}$$

This is valid assumption for a stationary process where the drift and diffusion coefficients do not depend explicitly upon t. The probability density can be solved for as follows:
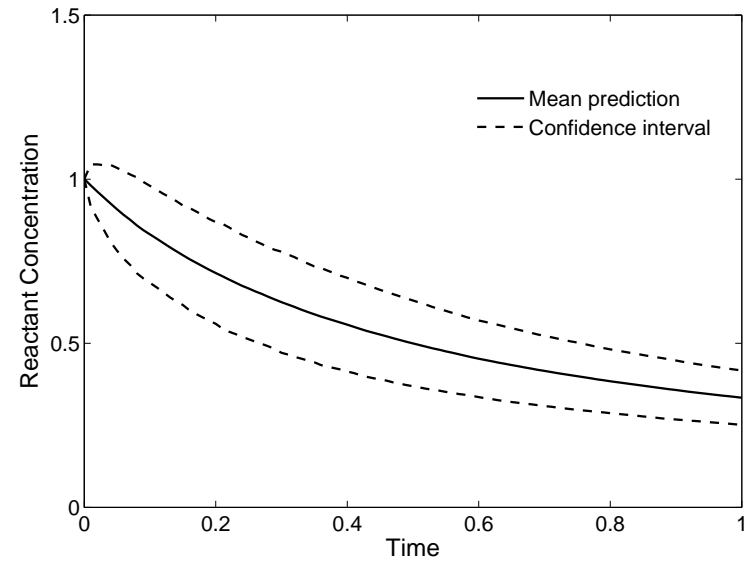
$$p(C_1) = p_0 exp \left(\frac{\bar{k}}{\sigma_k^2 C_1} \right) \tag{8}$$

26

99.7% Confidence Interval

Normal distribution

Log-Normal Distribution

Quantification of Model Uncertainty

Modified Gillespie algorithm: Reaction PDF $P(\tau, \mu)$ for reaction μ to occur in the infinitesimal time interval $[t + \tau, t + \tau + \Delta\tau]$ given a certain state at time t.

Residence time τ, during which no reactions occur, depends upon the total molecular population of all reacting species and reflects the randomness of collisions.

A constant deterministic value τ corresponds to standard reaction rate equations

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = F_i(C_1, C_2, ...C_n), \qquad i = 1, \ldots, n$$

# Quantification of Model Uncertainty (cntd.)

Modified SSA:

1. Compute the total number of reacting pairs of molecules available for each reaction a_i, and compute their sum $a_0 = \sum a_i$

2. Generate random numbers r_1 and r_2 on the uniform unit interval and m uniformly random on the interval $[1, 10]$

3. Compute $\tau = -m a_0^{-1} \ln r_1$

4. Determine which reaction μ occurs by taking μ to be that integer for which $\sum_{j=1}^{\mu-1} a_j < r_2 a_0 \leq \sum_{j=1}^{\mu} a_j$

5. Update time by τ and molecular levels for reaction μ (decrease reactants by 1 and increase products by 1)

6. Repeat steps 1-5 until either of the reactant population goes to zero or steady state is reached

To account for both sources of uncertainty, we modify the first step of the algorithm by replacing the constant value k with its random counterpart whose mean and variance are $\bar{k}$ and σ_k^2, respectively.
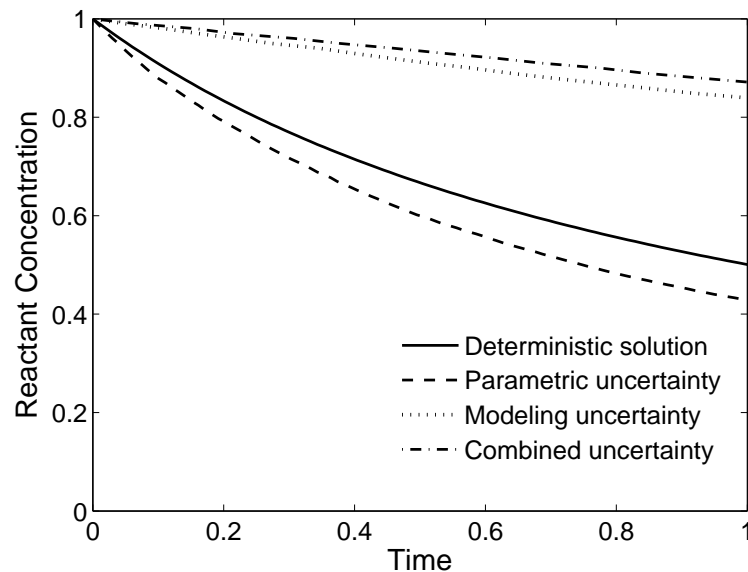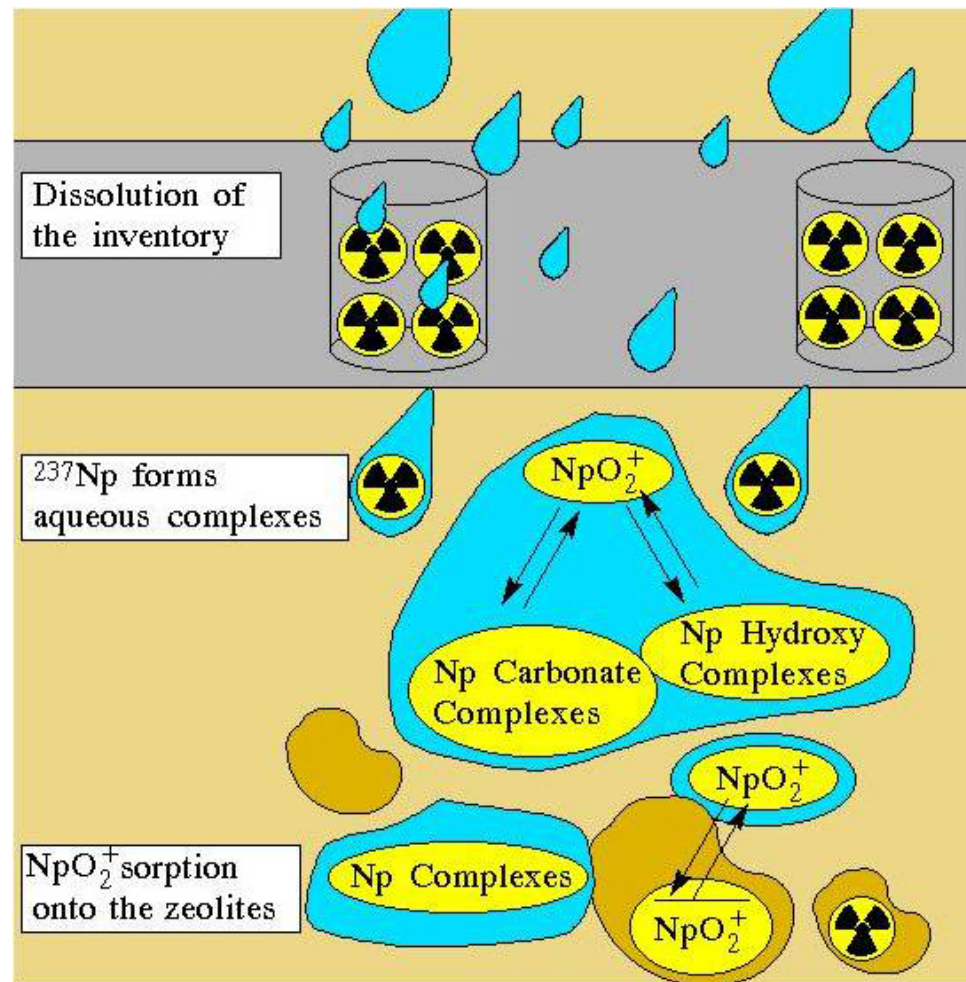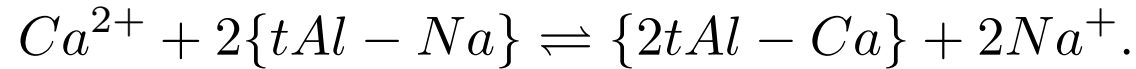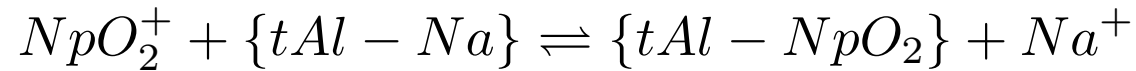

Reactant Concentration
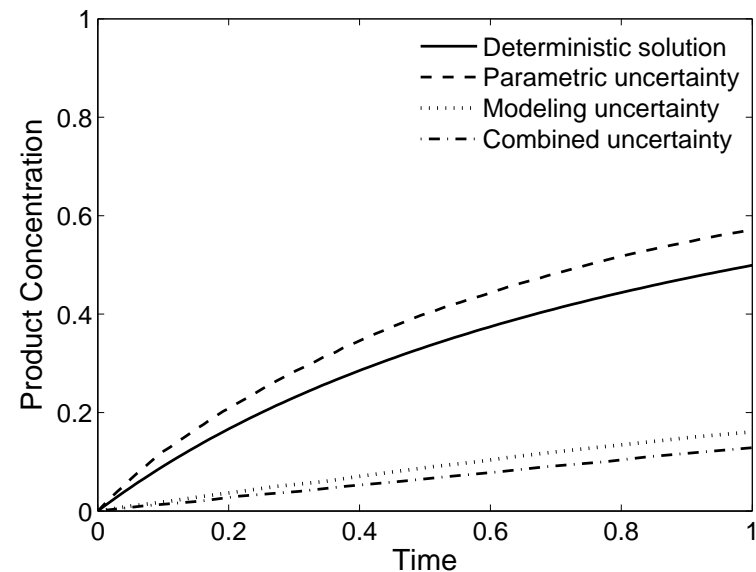
Product Concentration

Schematic of containers with Neptunium stored underground

31

Example: Neptunium Ion Exchange

Reacting system:

$$NpO_2^+ + \{tAl - Na\} \rightleftharpoons \{tAl - NpO_2\} + Na^+$$

$$Ca^{2+} + 2\{tAl - Na\} \rightleftharpoons \{2tAl - Ca\} + 2Na^+.$$

Standard deterministic model:

$$\frac{dC_1}{dt} = -k_1 C_1 C_4 + k_2 C_2 C_3 - 2k_3 C_1^2 C_6 + 2k_4 C_2^2 C_5,$$

$$\frac{dC_2}{dt} = k_1 C_1 C_4 - k_2 C_2 C_3 + 2k_3 C_1^2 C_6 - 2k_4 C_2^2 C_5,$$

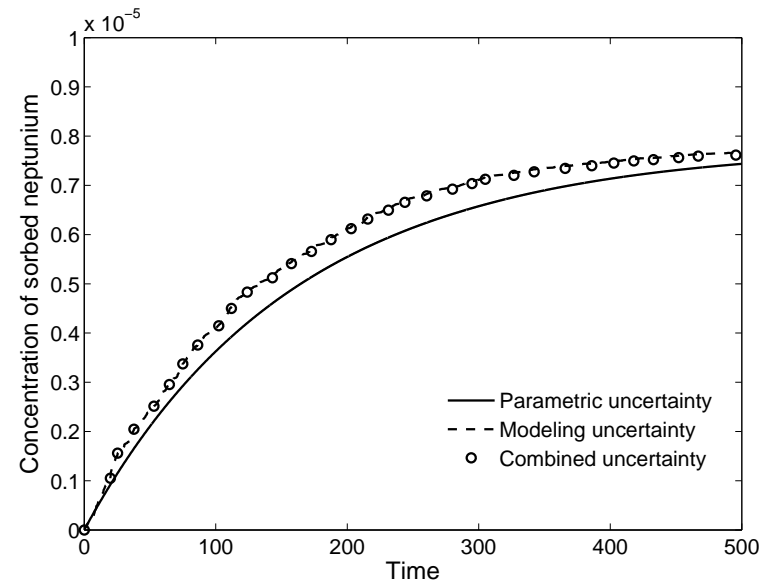$$\frac{dC_3}{dt} = k_1 C_1 C_4 - k_2 C_2 C_3, \qquad \frac{dC_4}{dt} = -k_1 C_1 C_4 + k_2 C_2 C_3,$$

$$\frac{dC_5}{dt} = k_3 C_1^2 C_6 - k_4 C_2^2 C_5, \qquad \frac{dC_6}{dt} = -k_3 C_1^2 C_6 + k_4 C_2^2 C_5$$

32

Neptunium Ion Exchange : Results

To account for uncertainty in the reaction rate constants k_1, k_2, k_3, and k_4, we treat them as normal white noise, whose mean values were determined from experimental data (Viswanathan et al, 1998).
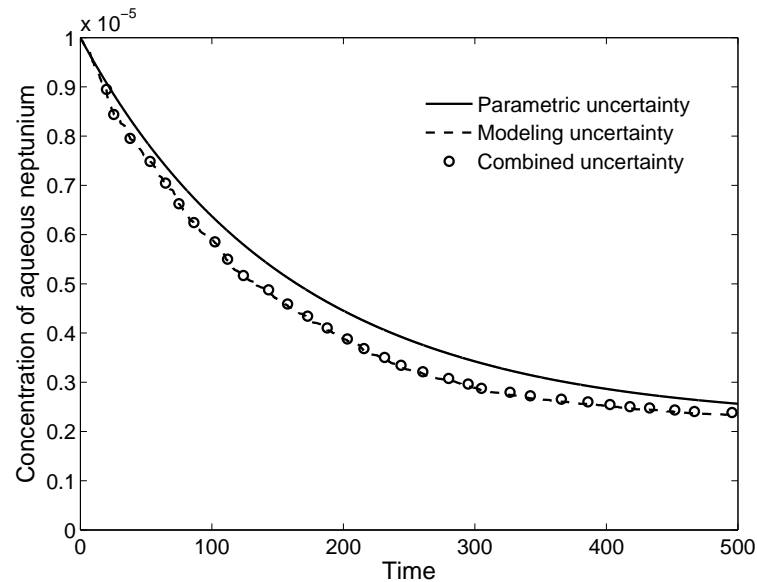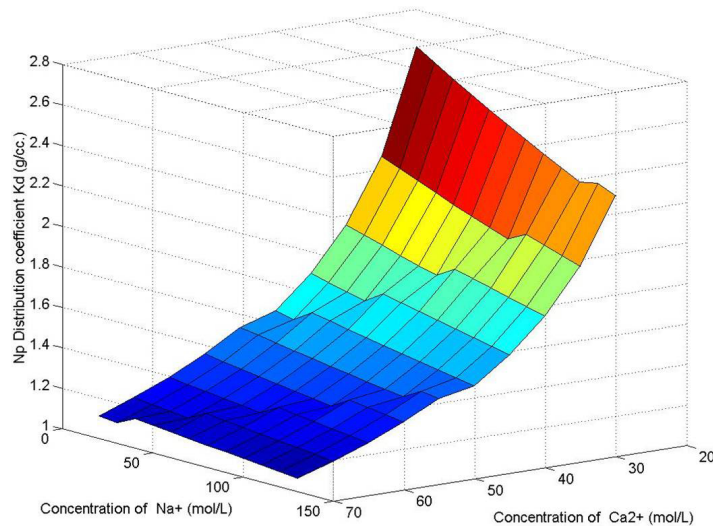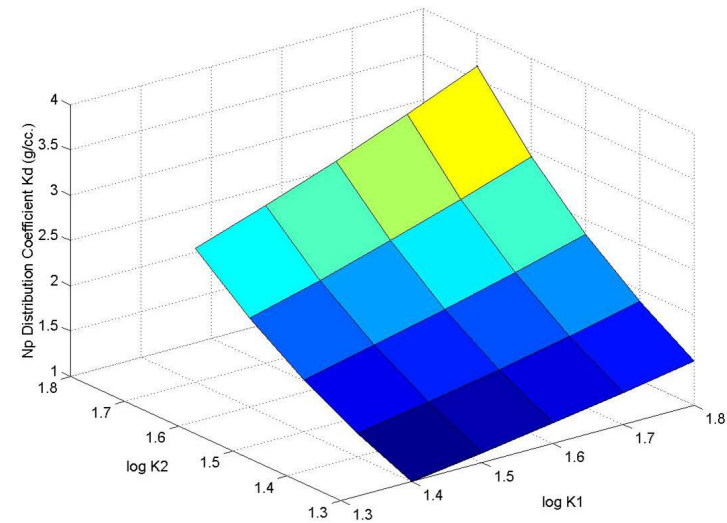

Aqueous Neptunium

Sorbed Neptunium

33

Neptunium Ion Exchange: Results

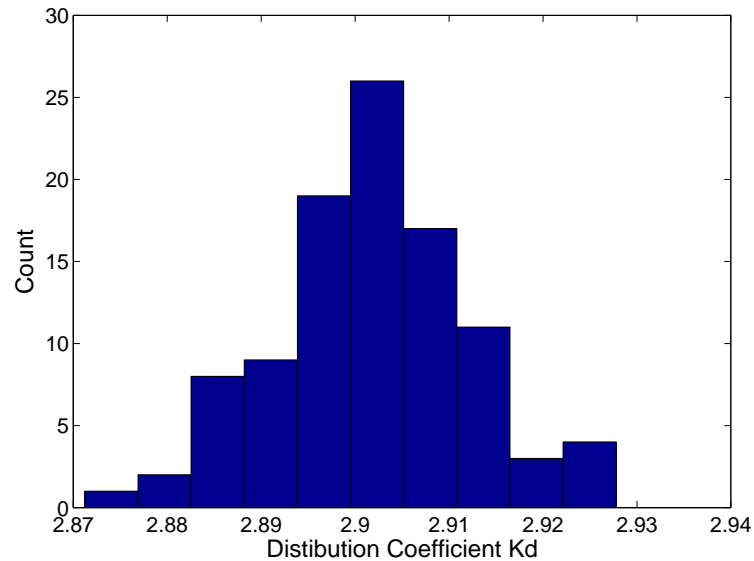Effects of varying parameters (Initial Conditions and Reaction Rates) on Distribution coefficient $K_d = C_3/C_4$:
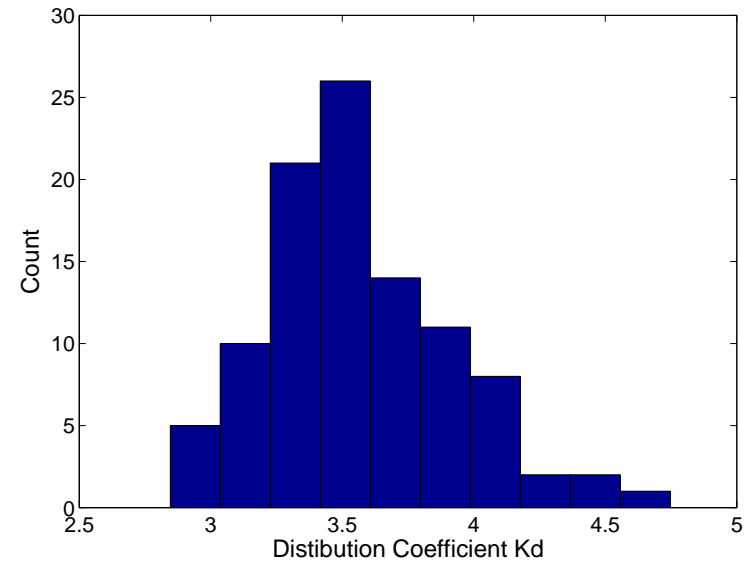

Varying Initial Conditions

Varying Reaction Rates

34

Neptunium Ion Exchange: Results

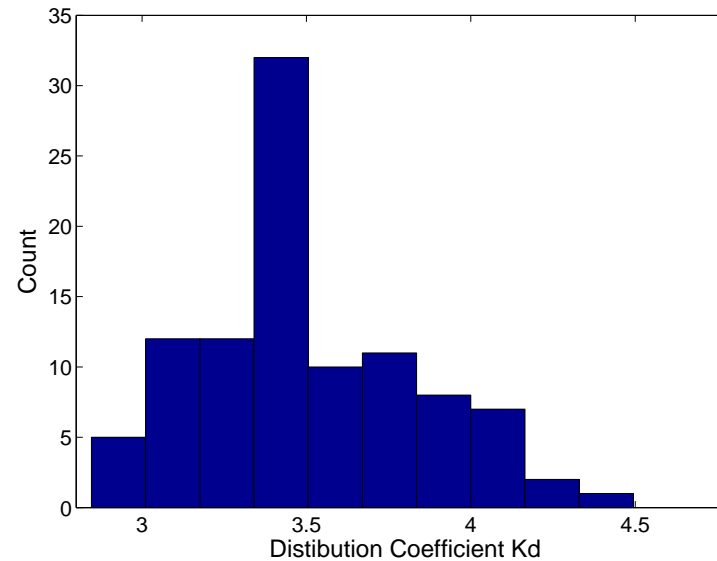Distribution coefficient $K_d = C_3/C_4$:

Parametric Uncertainty

Model Uncertainty

Neptunium Ion Exchange: Results

Distribution coefficient $K_d = C_3/C_4$:

Joint Uncertainty

36

Summary of Results

1. The rate at which a reaction occurs varies greatly due to the inclusion of various sources of uncertainty.

2. The range of distribution coefficients obtained for the neptunium competitive ion exchange process considered underscores the need for uncertainty quantification. It is expected that the same behavior will be present in more complex, multicomponent systems involving more chemical species.

3. The proposed approach yields a complete probabilistic description of the reaction rates and distribution coefficient, key parameters affecting the fate and migration of neptunium in the subsurface. This is important, since these distributions are highly skewed.

Summary of Results

4. The uncertainty quantification tools employed in this study are fully scalable and can be used to investigate any number of coupled reversible or irreversible geochemical reactions.

5. Transport of radionuclides such as neptunium could be greatly affected by the reactions that take place in the subsurface and hence quantification of modeling and parametric uncertainties is crucial in describing the overall transport.

Conclusions

- Uncertainties in parameters are important, but do not present a complete picture of the uncertainties in predictions.

- Sensitivity analysis can be performed to isolate those parameters contributing the biggest source of uncertainty.

- Model form uncertainty is often ignored, but is key to understanding errors in predictions when simulation results deviate from observations.

- Uncertainty under extrapolation is challenging, especially for high dimensional systems.

39

Questions?

Thank you!