Uncertainty Propagation in Combustion Kinetics

- Implication for Database Design and Computation

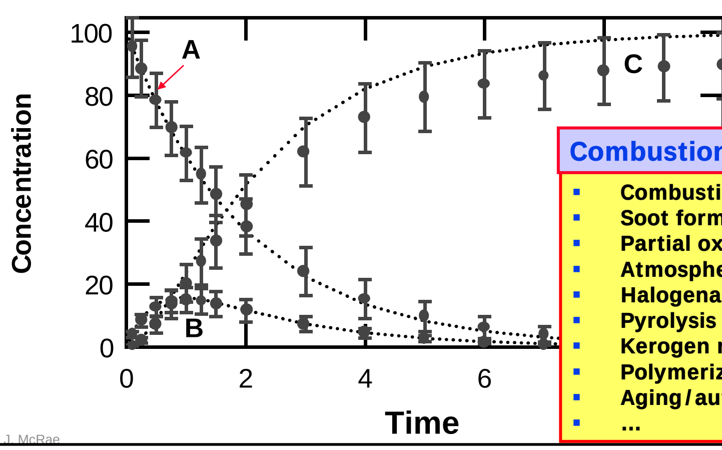
Gregory J. McRae

MIT Chemical Engineering

20th April 2006

The Problem – *Effect of Uncertainties*

$A \xrightarrow{k_1} pB + (1-p)D$	$p \sim U[0.9\text{-}1.0],\ [A(0)] \sim N[100,10]$
$B \xrightarrow{k_2} C$	$k_1 \sim N[0.5,0.1], k_2 \sim N[2.0,0.5]$



Combustion Problems

- Combustion
- Soot formation
- Partial oxidation
- Atmospheric chemistry
- Halogenation
- Pyrolysis
- Kerogen maturation
- Polymerization
- Aging / autoxidation
- ...

Outline for Presentation – *What do we need?*

- A systematic way to represent uncertainties in inputs and outputs from kinetic models
- Efficient computational algorithms that can propagate uncertainties through complex models
- Methods for identifying those parameters that control uncertainties in predictions

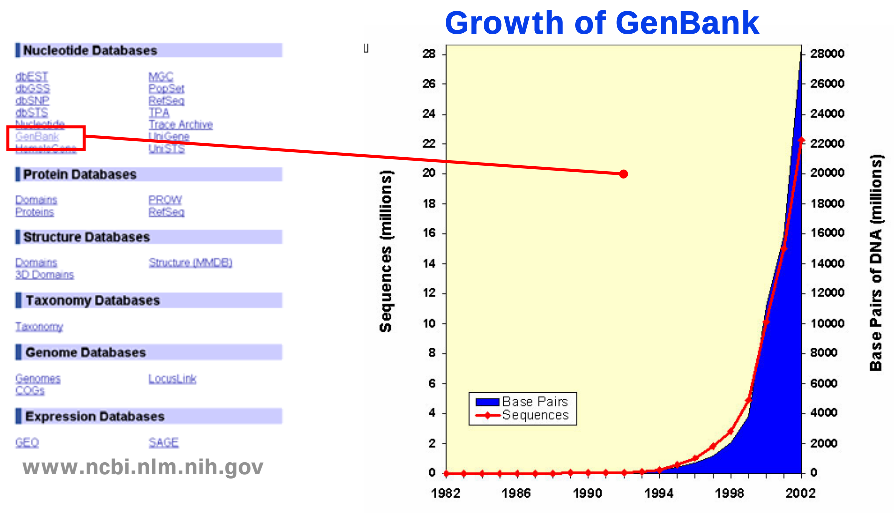
Key Message: We must become **much more systematic** about how we treat uncertainties

Current State of Uncertainty Representation

- **Thermophysical properties**
  - JANAF
  - NIST
  - CHEMKIN
  - DECHEMA
  - Group contribution, QM
  - ....
- **Kinetics**
  - NASA
  - Bielstein
  - DECHEMA
  - ....
- **Species and Mechanism**
  - LLNL
  - Literature, by-hand, GRI,...
  - Reaction path generation
  - ....
- **Etc.**

1. **Not integrated**
2. **Do not give useful uncertainty estimates**
3. **Varying levels of documentation**
4. **Few organized experimental data bases for evaluation**

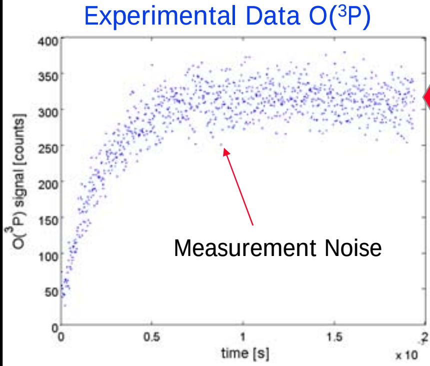
Growth – *Standards are a key to success*



There are approximately 63,183,065,091 bases in 12,465,546 sequence as of February 2006.

Relevance of Uncertainty in Kinetics

Experimental Data $O(^3P)$



Postulated Reaction Mechanism

$$O(^1D) + N_2 \xrightarrow{k_1} O(^3P) + N_2$$

$$O(^1D) + O_3 \xrightarrow{k_2} 2O(^3P) + O_2$$

$$O(^1D) + O_3 \xrightarrow{k_3} 2O_2$$

$$O(^1D) + He \xrightarrow{k_4} O(^3P) + He$$

- Parameter/state estimation – *k's, I.C.'s,...*
- Model based experimental design – *Stopping rules*
- Control/optimization – *Instrumental control*
- Robust statistics – *Outlier detection*

Representation of Uncertainties

- **Limits** – *Upper and lower limits*
- **Confidence Intervals** – *e.g. 95% limits*
- **Fuzzy Sets** – *Uncertain ranges*
- **Histograms** – *Data derived distributions*
- **pdf's** – *e.g. Normal, Gamma, Weibull,…*

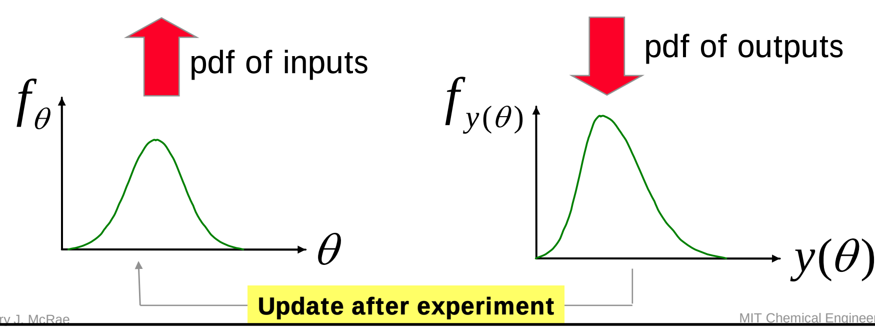
Key Point: We need to be able to store and calculate with probability density functions for both the parameters and the model outputs

The Basic Problem – *Propagation of Uncertainty*

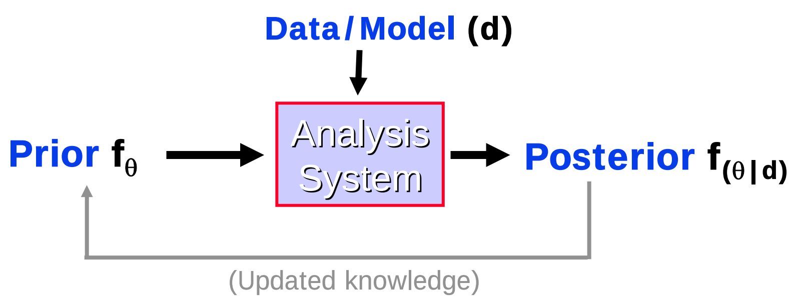
Parameters → **Model** → **Outputs**

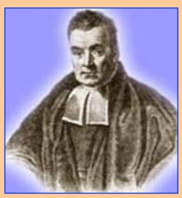
$$\theta \longrightarrow M : \theta \rightarrow y(\theta) \longrightarrow y(\theta)$$

f_θ $\qquad$ $f_{y(\theta)}$



Bayes Theorem *– Learning from Data/Models*





T. Bayes (1702-1761)

$$f_{(\theta|d)} = \frac{f_{(d|\theta)} f_\theta}{f_d}$$

Value of Information

Shannon Information

$$H[f_\theta] = \int f_\theta \log f_\theta d\theta$$

Information Gain from Experiment

$$I(\theta) = H[f_\theta^0] - H[f_\theta]$$

Prior Posterior

Example *– Normal Distribution*

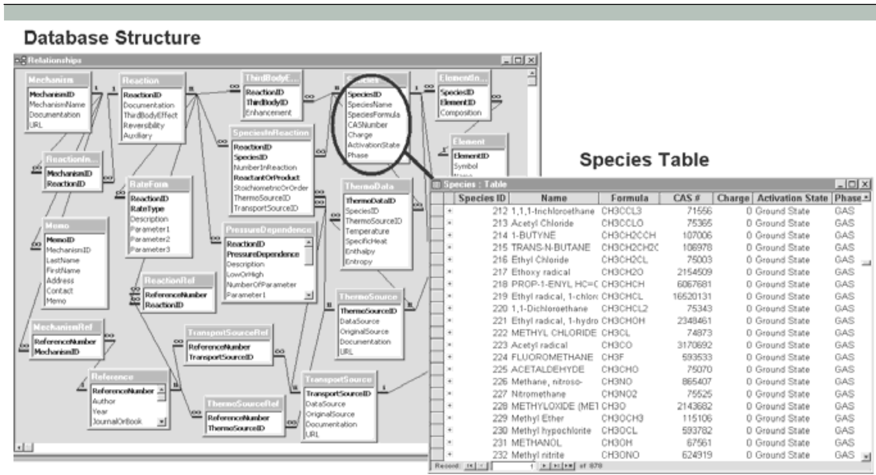
$$f_\theta \sim N[\mu_\theta, \sigma_\theta]; \quad H(\theta) \propto \sigma_\theta$$

Advantages of a Bayesian Approach

1. Can use prior knowledge and physical constraints in the analysis
2. Provides a formal framework for combining measurements of different quality
3. Gives the pdf's of the solution
4. New algorithms (MCMC) can solve non-linear problems
5. Broad applications including decision analysis

… Both Bayesian and Frequentist views are useful in practice

Sample Relational Database *-- Elements of PrIME*



Key Point: We need community consensus about how to represent data (Analogies with TCP/IP, Genomics,...)

XML Representation – Example Normal Distribution

pdf for parameter

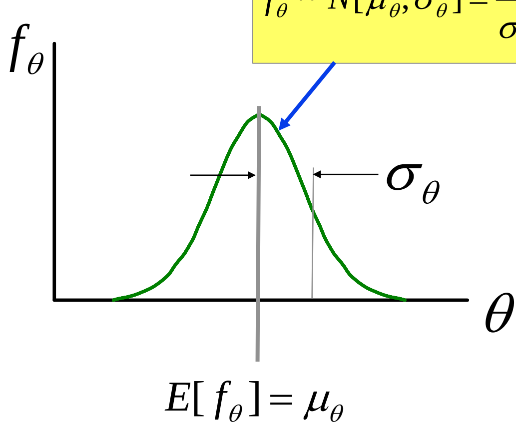
$$f_\theta \sim N[\mu_\theta,\sigma_\theta] = \frac{1}{\sigma_\theta\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2\right\}$$

XML Descriptor

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE state SYSTEM
“Uncertainty_Parameter.dtd">
<state>
<name> PDF </name>
<description> Normal </description>
<data name="Mean"> 0.200 </data>
<data name="Stdv"> 0.001 </data>
</state>
```


Example of an Uncertainty Distribution

$$f_\theta \sim N[\mu_\theta,\sigma_\theta] = \frac{1}{\sigma_\theta\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\theta-\mu_\theta}{\sigma_\theta}\right)^2\right\}$$



Question: But is it the ‘right’ one to choose?

Example – *Chemical Kinetics A → B*

Model

$$\frac{dy(t)}{dt} = -k\, y(t) \quad ; y(0) = y_0, \qquad y(t) = [A(t)]$$

Solution

$$y(t) = y_0 e^{-kt}$$

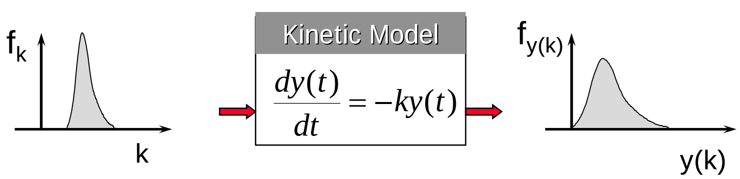
Sensitivity to parameter variations

$$S = \left.\frac{\partial y(t)}{\partial k}\right|_{\bar{k}} = -t\, y_0 e^{-\bar{k}\, t}$$

But, what if k is uncertain?

Solution in Presence of Uncertainty

Uncertain Rate Constant **Uncertain Output**



Normal distribution for rate constant (k) **But** → Lognormal distribution for solution at each time

$$f_k = \frac{1}{k_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-k_0}{k_1}\right)^2}$$

$$f_{y(k)} = \frac{1}{k_1 t y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(y/y_0)+tk_0}{tk_1}\right)^2}$$

$$E[y(t)] = \int_{-\infty}^{+\infty} y(t)\, f_k\, dk = y_0 e^{\left(-tk_0 + \frac{t^2 k_1^2}{2}\right)}$$

Example – *Simple Equilibrium Problem*

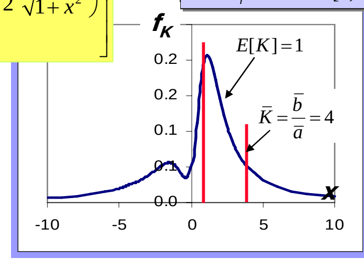
Problem

$$A \Leftrightarrow B; \quad K = \frac{k_f}{k_r}; \quad k_f \sim N[\bar{a},1], \quad k_r \sim N[\bar{b},1]$$

Solution (f_K pdf for K !!)

$$f_K = \frac{e^{-\frac{\bar{a}^2+\bar{b}^2}{2}}}{\pi(1+x^2)}\left[1+\frac{\frac{\bar{b}+\bar{a}x}{\sqrt{1+x^2}}}{\frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}\left(\frac{\bar{b}+\bar{a}x}{\sqrt{1+x^2}}\right)^2\}}\frac{1}{2}Erf\left(\frac{1}{\sqrt{2}}\frac{\bar{b}+\bar{a}x}{\sqrt{1+x^2}}\right)\right]$$

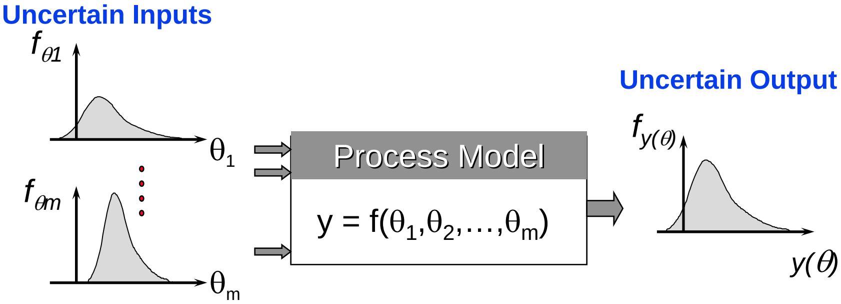
$$K = \frac{k_f}{k_r} = \frac{2+N[0,1)}{0.5+N[0,1]}$$



Solution is typically <u>not normal</u>, often bimodal and E[b/a] ≠ E[b]/E[a]

The General Problem

Uncertain Inputs



Measures of Uncertainty (Expected value, variance, pdf, etc.)

e.g. $$E[y(\boldsymbol{\theta})] = \int_y y(\boldsymbol{\theta})\, f_{y(\boldsymbol{\theta})}\, dy(\boldsymbol{\theta}) \equiv \underbrace{\int \cdots \int_{\boldsymbol{\theta}} y(\boldsymbol{\theta})\, f_{\boldsymbol{\theta}}\, d\theta_1 \cdots d\theta_m}$$

Perturbation, number theoretic, Monte-Carlo, pattern searches, semigroup methods etc. for multi-dimensional integrals are computationally <u>very expensive</u>

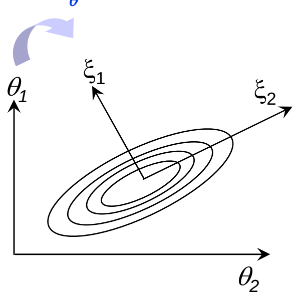
Computational Aspects of Combustion

- **Large scale** – *O(10^4+) points, 100's state variables*
- **Nonlinear** – *Chemistry, IPDE operators,...*
- **Wide range of time scales** – *$|\lambda_{max}/\lambda_{min}| \sim 10^{14}$*
- **Many uncertainties** – *Parameters, inputs,..*
- **Common issues** – *Resource allocation*
- **High payoffs** – *Fuel efficiency, energy security,...*

Goal is to minimize the elapsed time to solution

Some Initial, Obvious Ideas

Orthogonal transformation of f_θ



$$E[y(\boldsymbol{\theta})] = \int \cdots \int_{\boldsymbol{\theta}} y(\boldsymbol{\theta})\, f_{\boldsymbol{\theta}}\, d\theta_1 \cdots d\theta_m$$

$$= \int y(\xi)\, f_{\xi_1} d\xi_1 \cdots \int y(\xi)\, f_{\xi_m} d\xi_m$$

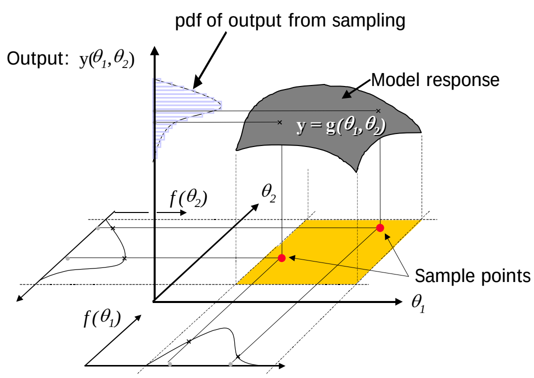
$$= \prod_{i=1}^{m} \left(\int y(\xi)\, f_{\xi_i} d\xi_i \right)$$

$$\approx \prod_{i=1}^{m} \left(\sum_{k=1}^{p} w_i\, y(\xi) \right)$$

Product of 1-D quadratures (embarrassingly parallel)

But this only gives *E[y(θ)]* we would like $f_{y(\theta)}$

Conventional Approach



Samples → Model → Histogram → pdf

How to short circuit the process?

Desirable Characteristics of New Algorithms

- **<u>Very fast</u>, accurate, cheap!!**
- **Able to get pdf's as well as *E[],…***
- **Be able to treat "black box" models**
- **Take advantage of distributed computing**
- **Ease of use without compromising complexity**

Expert user required

Modern Aircraft

Expert is under the hood

Toyota Prius

Incorporating Uncertainty at the Beginning

Fourier Series Representation of f(x)

$$f(x) = a_0 + \sum_{i=1}^{\infty} a_i \sin(\omega_i x) + b_i \cos(\omega_i x)$$

What happens if x is a random variable?

Polynomial Chaos Representation of *f(ω)* (Wiener, 1938)

$$f(\omega) = \sum_{i=1}^{\infty} a_i \, H_i[\xi_1(\omega), \cdots, \zeta_m(\omega)]$$

- Coefficients of expansion (a_i)
- Orthogonal Functionals (e.g. Hermite Polynomials) (H_i)
- Known probability distributions (e.g. unit Normal N[0,1]) ($\xi_1(\omega), \cdots, \zeta_m(\omega)$)

Example – *PCE Galerkin Fomulation*

Model

$$\frac{dy(t)}{dt} = -ky(t); \quad k = k_0 + k_1\xi; \quad \xi \sim N[0,1]$$

Expansion of y(t)

$$y \sim \hat{y}(w) = \sum a_i H_i[\xi] \approx a_0(t) + a_1(t)\xi + a_2(t)(\xi^2 - 1) + \cdots$$

Residual

$$R(t,\xi) = \frac{dy(t,\xi)}{dt} - k(\xi)\hat{y}(t,\xi)$$

Residual Minimization (Galerkin or Collocation)

$$\int_{-\infty}^{+\infty} R(t,\xi) w_j(\xi) d\xi = 0 \;\; ; j = 0,1,\cdots$$

Uncertain coefficients (Galerkin $w_j = H_j$)

$$\mathbf{A}\frac{d\mathbf{a}(t)}{dt} + \mathbf{B}\mathbf{a}(t) = 0$$

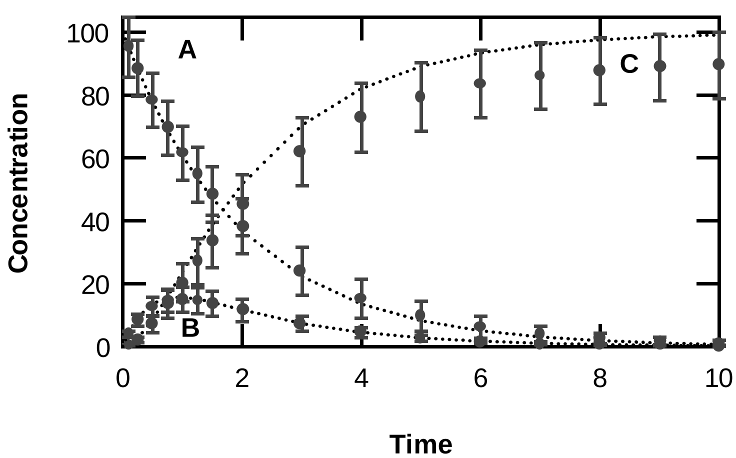
Advantages of PCE Algorithms

- Compact way to represent non-standard uncertainty distributions
- Gives pdf's of the solution
- 3-4 Orders of magnitude faster than Monte Carlo, Latin Hypercube,...
- Identifies those parameters that contribute to uncertainties in prediction
- Can treat model as a 'black box'

Example – *Simple Reaction Sequence*

$A \xrightarrow{k_1} pB + (1-p)D \qquad p \sim U[0.9 - 1.0],\ [A(0)] \sim N[100,10]$

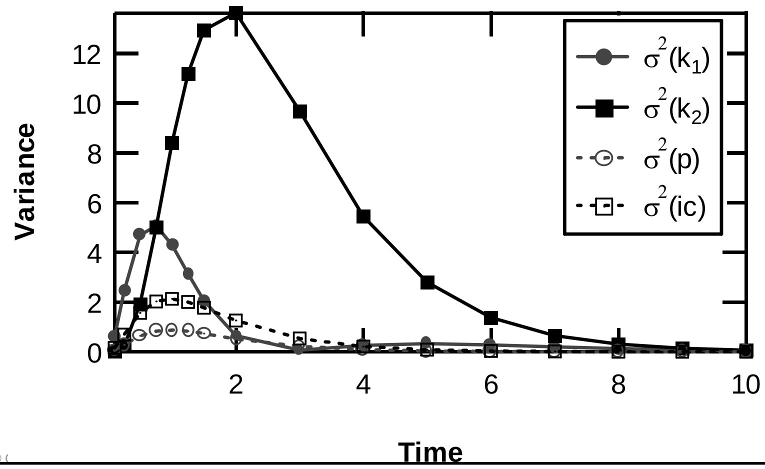
$B \xrightarrow{k_2} C \qquad k_1 \sim N[0.5,0.1], k_2 \sim N[2.0,0.5]$



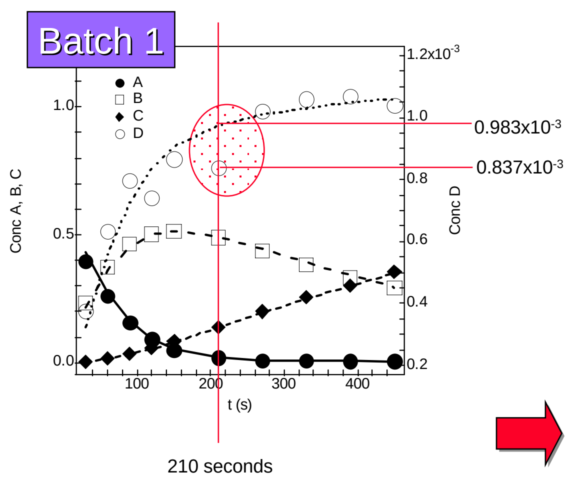
Parameter Uncertainty on Variance of B

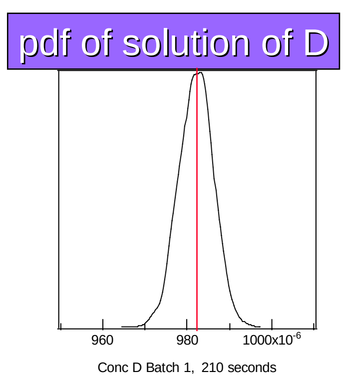
$$A \xrightarrow{k_1} pB + (1-p)D \qquad p \sim U[0.9\text{-}1.0],\ [A(0)] \sim N[100,10]$$

$$B \xrightarrow{k_2} C \qquad k_1 \sim N[0.5,0.1],\ k_2 \sim N[2.0,0.5]$$



Outlier Evaluation / Detection





Probability of experimental measurement at 210 s being a good fit is very small

Conclusions

- Critical need for uncertainty estimates in data bases
- New uncertainty algorithms can help shape design, analysis and assimilation of experimental data

"... While there are always lots of uncertainties, the key challenge in engineering is to find those problem components that contribute most to uncertainties in ***outcomes****..."*

MIT
Chemical Engineering
Course 10

Contact Address:

Department of Chemical Engineering
Room 66-372
Massachusetts Institute of Technology
Cambridge, MA 02139
(617) 253 6564
(617) 258 0546 (fax)
mcrae@mit.edu (email)
http://www.mit.edu