An introduction to Bayesian statistics and model calibration
… and a host of related topics

Derek Bingham
Statistics and Actuarial Science
Simon Fraser University

Department of Statistics and Actuarial Science
SFU SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Cast of thousands have participated in the experiments discussed

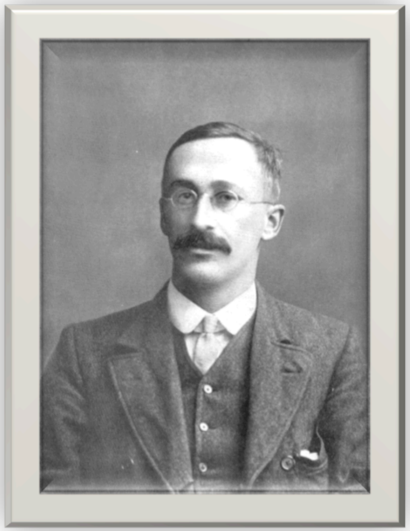
- Mike Grosskopf
- Dave Higdon
- Center for Radiative Shock Hydrodynamics
- Center for Exascale Radiation Transport

Outline

- Computer model examples
- Bayesian inference
- Bayesian approach to inverse problem
- Markov chain Monte Carlo (MCMC)
- Model calibration with limited model evaluations
  … and Gaussian process emulation
- Model calibration with discrepancy

Before we get started…

- Afternoons
- … and this afternoon



A New Holiday

Statistics and Beer Day is holiday that celebrates all of the ways in which the field of statistics has improved the world by focusing on how statistics has improved beer (which has also improved the world). It occurs each year on June 13, the birth date of William Sealy Gosset. Gosset worked for the Guinness Brewery where he applied statistical methods to improve the quality of the beer. Because his sample sizes were small, he needed new methods. His most well-known achievement was the t-distribution, which he published under the pseudonym Student. To celebrate, I recommend gathering with statisticians and other normal people and having a few pints. Of course, you should start out with a Guinness, but it's your choice after that.

statisticsandbeerday.wordpress.com

Many processes are investigated using computational models

- Many scientific applications use mathematical models to describe physical systems

- The computer models frequently:

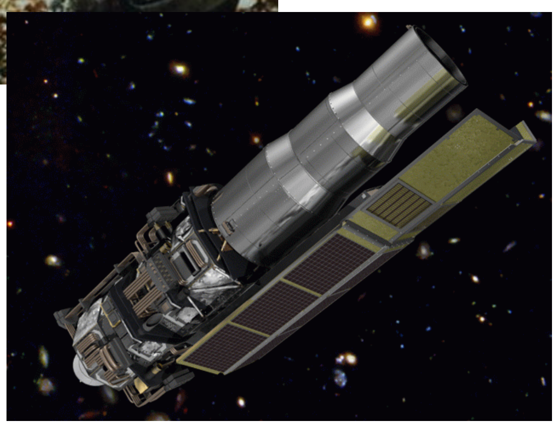
  1. require solutions to PDEs or use finite element analysis
  2. have high dimensional inputs
  3. have outputs which are complex functions of the input factors
  4. require a large amounts of computing time
  5. have features from some of the above

Glacier evolution

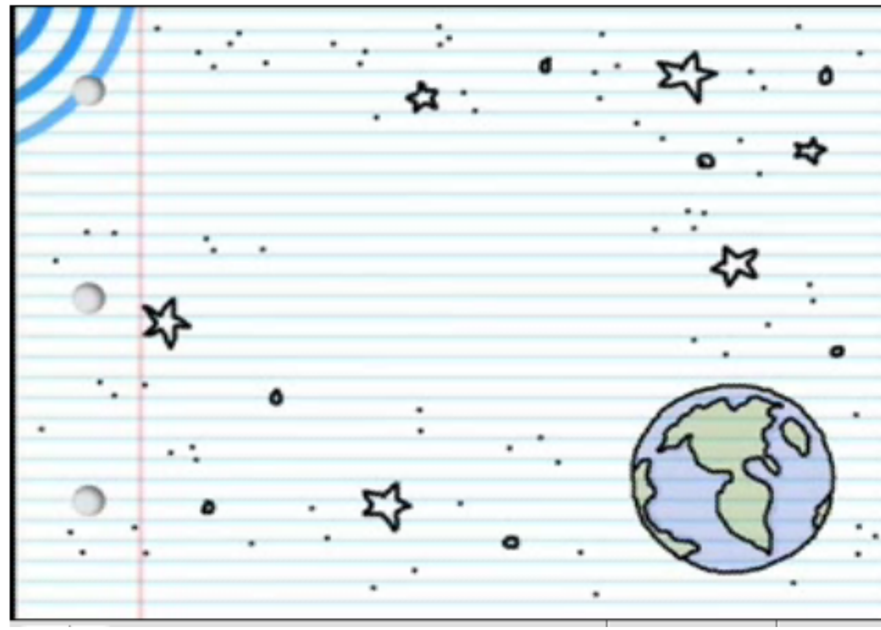
- Understanding the response of glaciers to climate is important globally for making accurate projections of sea level change
- Changes in glaciers can be the product of changes to the surface mass balance (accumulation and ablation)
- Have computer model to describe, say, ablation (output) given the season's weather trajectory (input)
- Have observations collected twice annually

SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Dark energy investigation



What Is a Redshift?

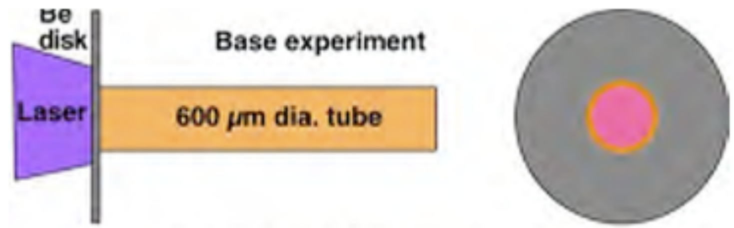


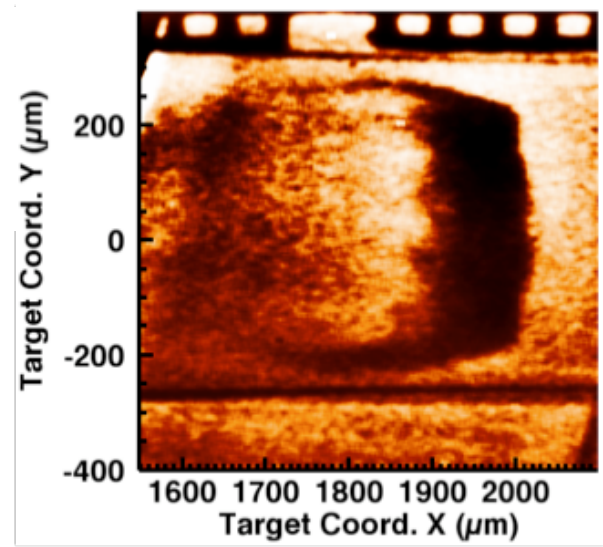
SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Center for Radiative shock Hydrodynamics (CRASH)

A conceptually simple experiment

– Launch a thin Be plasma down a shock tube at ~ 200 km/s

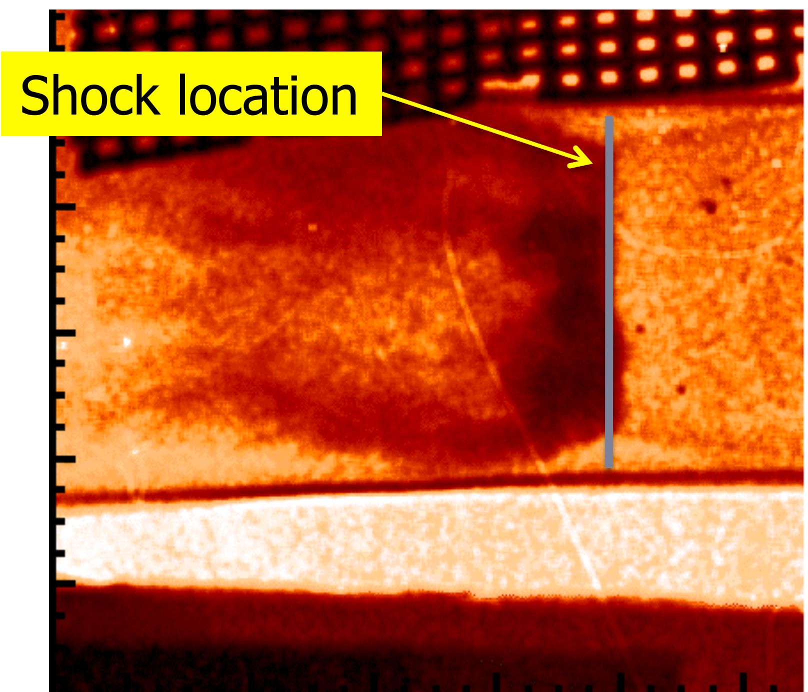




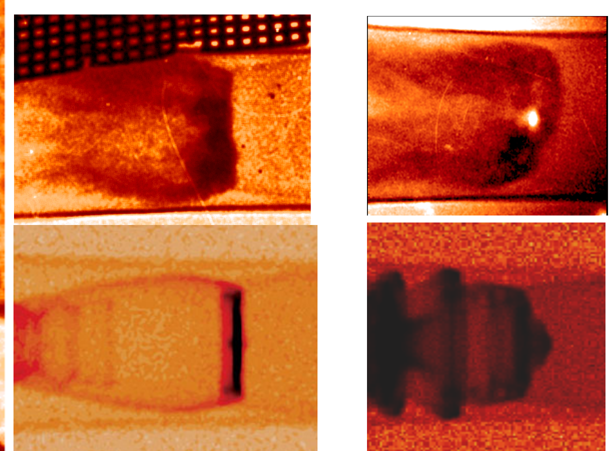
A radiative shock is a wave in which both hydrodynamic and radiation transport physics play a significant role in the shock's propagation and structure

Have several outputs & inputs

- Outputs (y)
  - Shock location
  - Shock breakout time
  - Wall shock location
  - Axial centroid of Xe
  - Area of dense Xe
- Inputs (X)
  - Observation time
  - Laser energy
  - Be disk thickness
  - Xe gas pressure
- Calibration parameters (θ)
  - Electron flux limiter
  - Laser scale factor



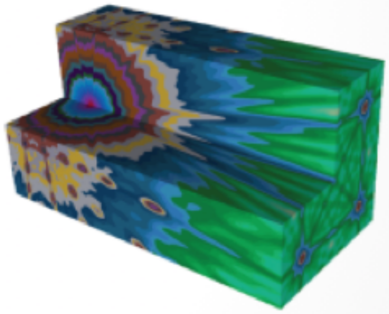
We measured and computed



Shocks at 13 ns

Goal:

1. build a predicative model using observations and physics model
2. constrain calibration parameters



Center for Exascale Radiation Transport (CERT)

- Focus on development of computational techniques for efficiently simulating thermal radiation propagation using exascale computers

- Development of predictive science techniques to quantify uncertainty in simulated results

- Radiation propagation plays a major role in high-energy density laboratory physics experiments

- Have experiments and have computational model

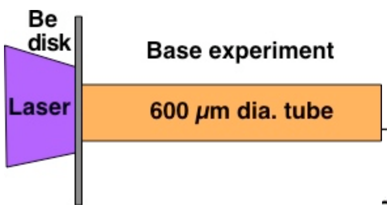
Have a really good computational model, but …

- For this application, the Boltzmann equation (linearized) should exactly represent the mean of the physical process if the impurities in the graphite are known
- Problem is that it is too computationally intensive to do this for the system at a level of fidelity that they would like
- Have observations from CERT model and experiments
- Would like to characterize brick impurities and "validate" model

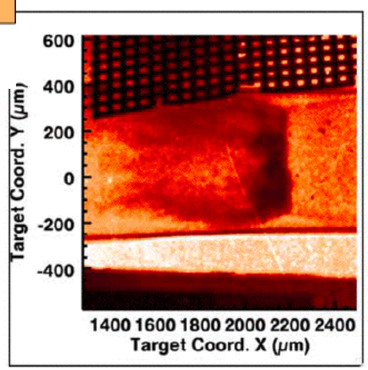
graphite "brick"

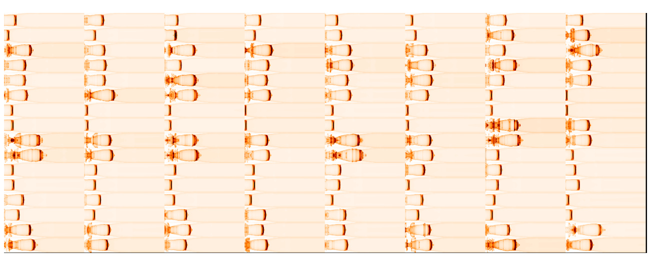
Common thread: combining simulations, field observations for prediction, calibration and uncertainty quantification

1 ns, 3.8 kJ laser irradiates Be disk

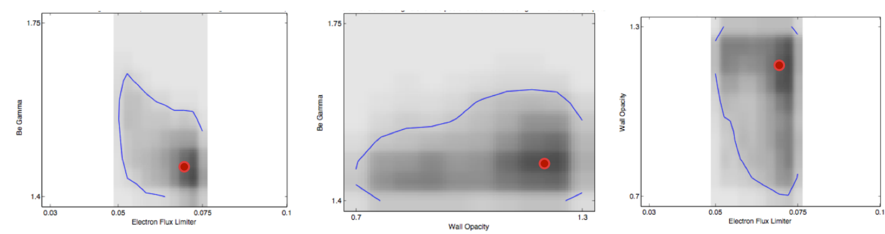


Shot 52665 TIM3

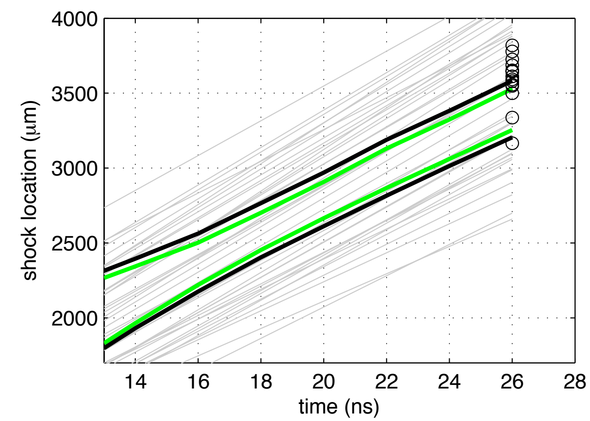




Calibration: finding input parameter settings consistent with observations



Prediction of new observations with uncertainty

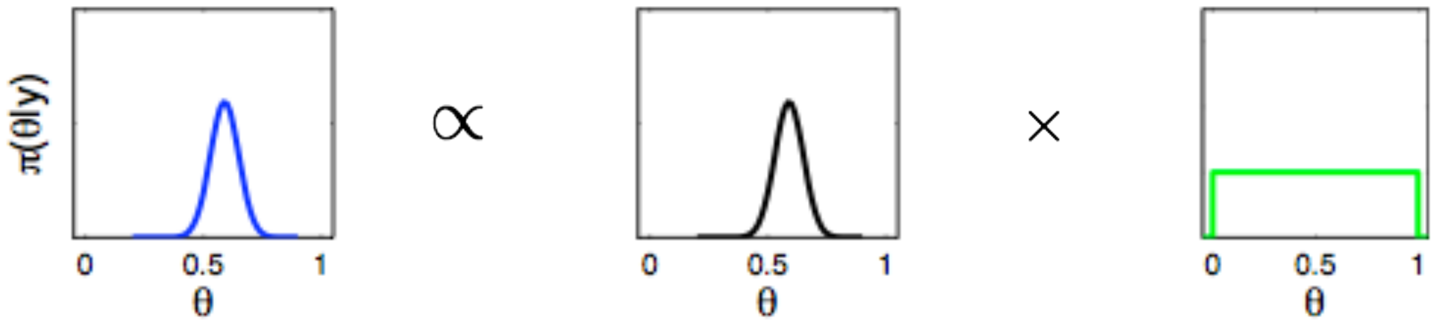


SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Bayesian inference

- The goal is to learn about the unknowns (say, θ) from the observables
- We have some idea about the unknowns captured in our prior distributions ($\pi\left(\theta\right)$)
- We also have some idea about how the observables depend on the unknowns given by the data generating mechanism (likelihood), $f\left(y|\theta\right)$
- Inference is done via the posterior distribution: $\pi\left(\theta|y\right) = \dfrac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta}$
- Note the denominator is a constant, so $\pi\left(\theta|y\right) \propto f\left(y|\theta\right) \times \pi(\theta)$

Very general look at Bayes rule



$$\pi(\theta|y) \propto f(y|\theta) \times \pi(\theta)$$

- posterior pdf $\pi(\theta|y)$ describes uncertainty in θ given data
- prior pdf for θ is required
- inference proceeds through the samples from posterior distribution

Example: Mean of a Gaussian with known variance

- Model: $y = \theta + \epsilon$
- $y|\theta \sim N(\theta, \sigma^2)$ ← Likelihood
- $\theta \sim N(\mu, \delta^2)$ ← Prior distribution
- Observations: $y_1, y_2, \ldots, y_n$
- $\pi(\theta|y) \propto exp\left(-\frac{1}{2\delta^2}(\theta-\mu)^2\right) exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right)$ ← Posterior distribution (without the normalizing constant)

$$\longrightarrow \quad \theta|y \sim N(\nu, \gamma^2)$$

$$\nu = \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\delta^2}\mu}{\frac{n}{\sigma^2} + \frac{1}{\delta^2}} \qquad \gamma = \left(\frac{n}{\sigma^2} + \frac{1}{\delta^2}\right)^{-1}$$

SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Basic inverse problem (calibration)

(Will take a Bayesian approach)

- Have a computer model that is a function of inputs, t and x
- The computer model represents the mean of a physical system
- Have noisy observations from the field that have been observed at $t = \theta$ and, possibly different values of x
- Unfortunately, θ is not known to the experimenters

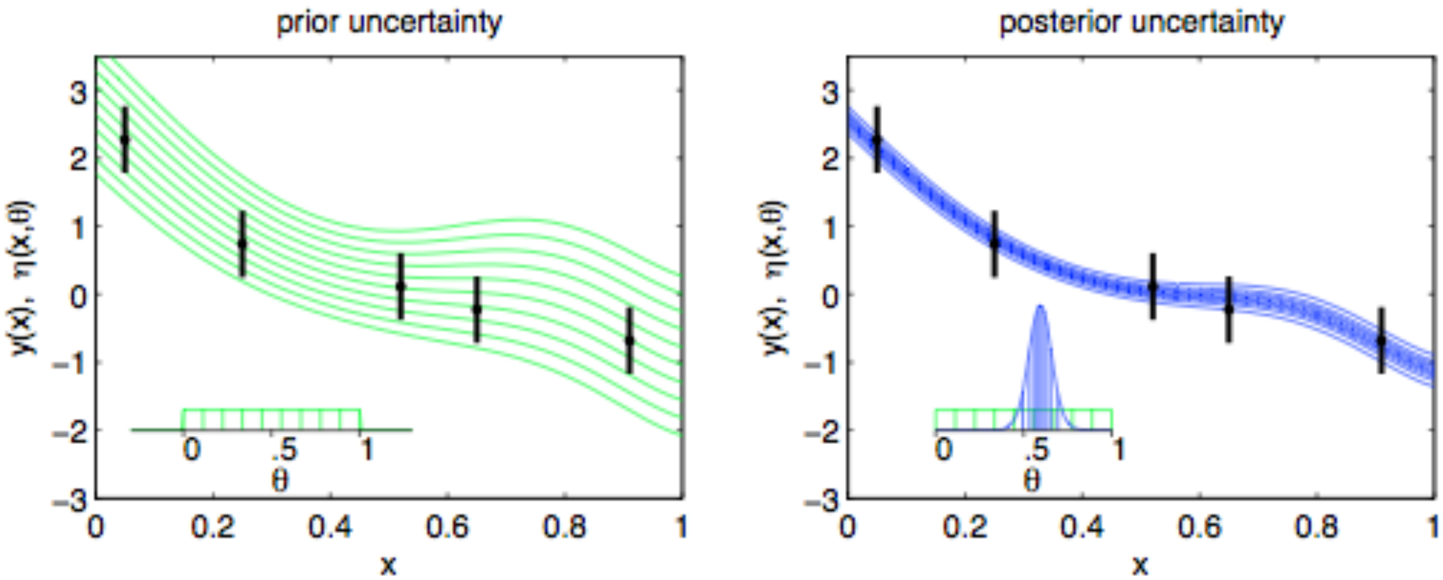
Basic inverse problem

- Two types of inputs:
  - **Design inputs (*x*)**: inputs that can be measured or adjusted in the field
  - **Calibration inputs**: inputs that are needed to run the model, may govern the behaviour of the physical system, but whose value is not known in the experiments

- **Goal:** estimate the calibration parameters, θ, using field observations and computational model

Statistical formulation

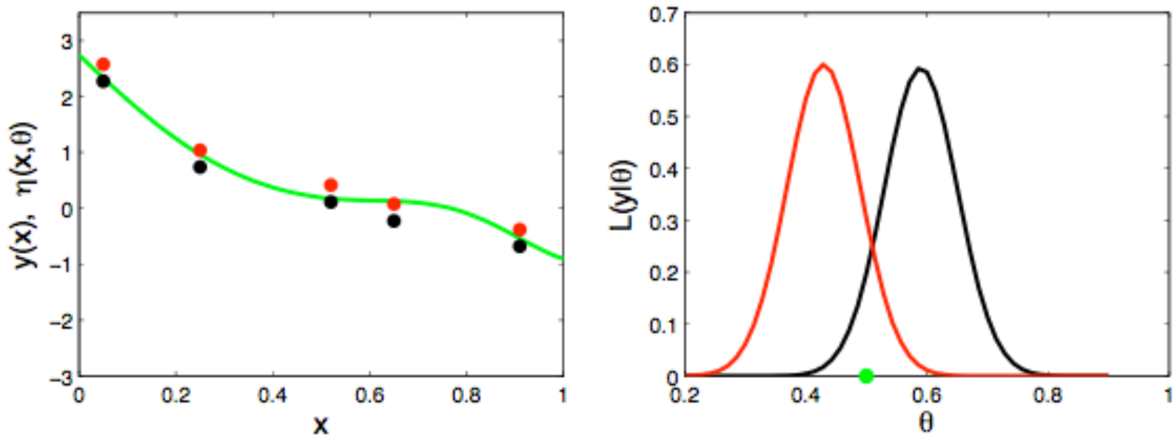
- Model for field observations: $y_f(x_i) = \eta(x_i, \theta) + \epsilon_i$



Likelihood

- Probability distribution of the data given the model parameters (assume the variance is known for this example)

$$f(y_f|x,\theta) \propto exp\left(\frac{-1}{2\sigma^2}(y_f - \eta(x,\theta))^T(y_f - \eta(x,\theta))\right)$$



SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

General Bayesian approach

- Collect data from a process with probability density *f* ( ) (likelihood)… joint distribution of the obseervations given unknown parameters

- Summarize our belief about parameters with a distribution (prior distribution)

- Update our belief in parameters using Bayes' rule (posterior distribution)

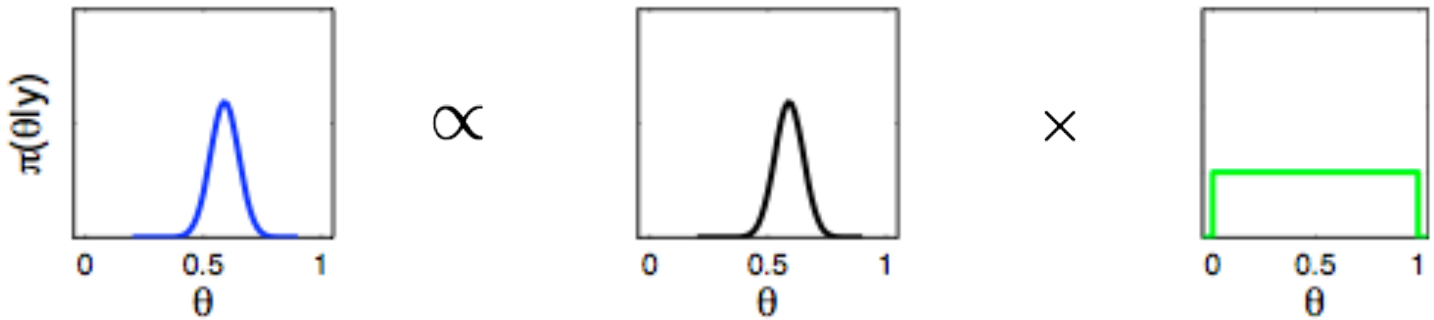
- The Bayesian approach uses probability to describe what is known about random variables and also parameters

- $\pi\left(\theta|y_f, x\right) = \dfrac{f(y_f|x,\theta)\pi(\theta)}{\int f(y_f|x,\theta)\pi(\theta)d\theta}$

Denominator (normalizing constant) usually hard to find

- $\pi\left(\theta|y_f, x\right) \propto \; f\left(y_f|x,\theta\right) \times \; \pi\left(\theta\right)$

Bayes rule for this problem



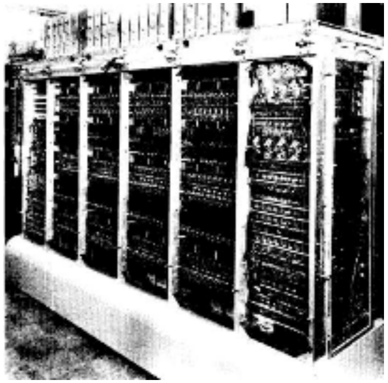
$$\pi\left(\theta \middle| y_f, x\right) \propto f\left(y_f \middle| x, \theta\right) \times \pi\left(\theta\right)$$

- very general approach for inference
- posterior pdf $\pi\left(\theta \middle| y_f, x\right)$ describes uncertainty in θ given data
- prior pdf for θ is required
- inference proceeds through the samples from posterior distribution

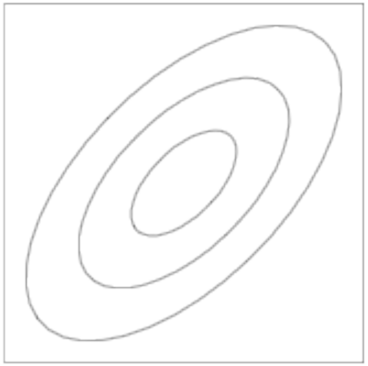
Exploring the posterior distribution

Nick Metropolis

MANIAC I



2-d pdf

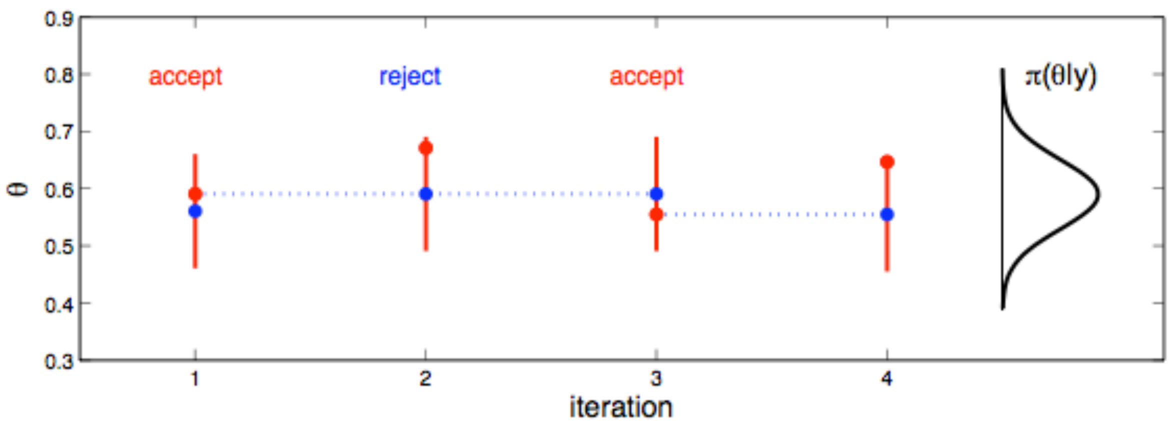


MCMC sample

- Use Markov chain Monte Carlo (MCMC) to build a Markov chain with limit distribution
- Realizations are a draw from posterior distribution
- Need not be normalized

SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Metropolis-Hastings recipe for MCMC



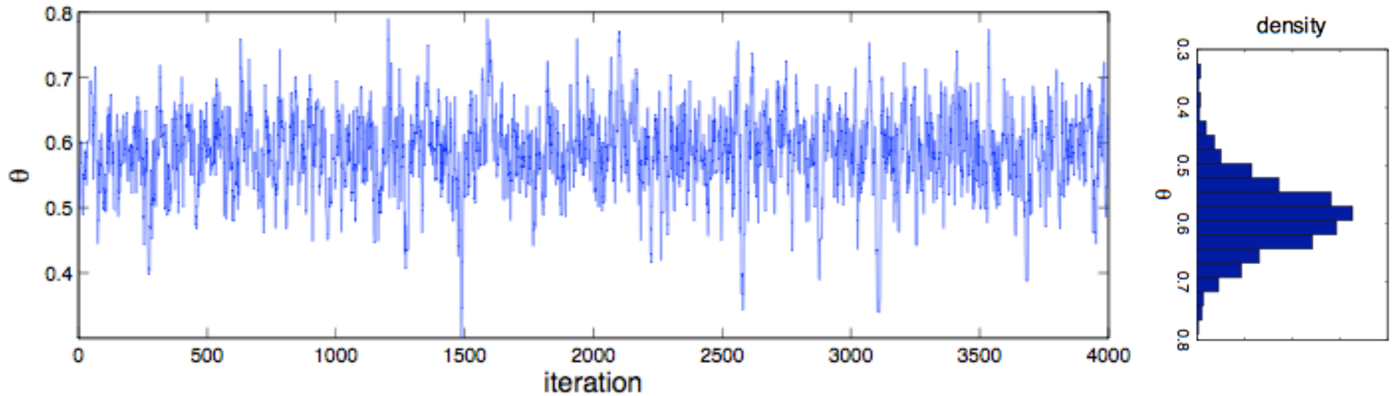
0. Initialize chain at θ^0
1. Given current realization of θ^t, generate θ^* from a symmetric distribution
2. Compute the acceptance probability $\alpha = min(1, \frac{\pi(\theta^*|y)}{\pi(\theta^t|y)})$
3. Set $\theta^{t+1} = \theta^*$ with probability α, otherwise $\theta^{t+1} = \theta^t$
4. Iterate steps 1-3

Notice that the normalizing constants cancel

Comment on Metropolis-Hastings

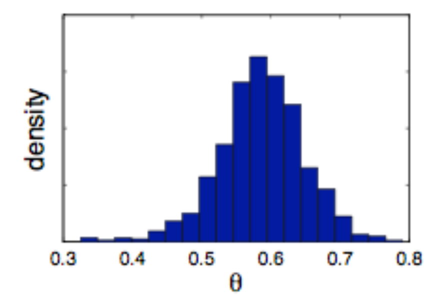
- Metropolis-Hastings proposal distribution in step 1?
  - Symmetric distribution centered at last accepted value
    e.g., $\theta^* \sim U(\theta^t - c, \theta^t + c)$ or $\theta^* \sim N(\theta^t, \phi^2)$
  - Wide variance: explores the parameter space but rarely accepts proposals
  - Small variance: accepts almost every proposed value (about as good as the last accepted value) but is takes a long time to explore the posterior distribution
- Normal random walk proposal distribution and Normal-ish target density, the optimal proposal width has acceptance rate of 44% for 1-d
- About 25% for larger than 4-d

Metropolis sampling for the inverse problem



- Chain $\theta^0, \theta^1, \ldots, \theta^{4000}$ is a draw from $\pi\left(\theta | y_f, x\right)$

- By the law of large numbers the distribution of our sample will converge to the target posterior distribution eventually



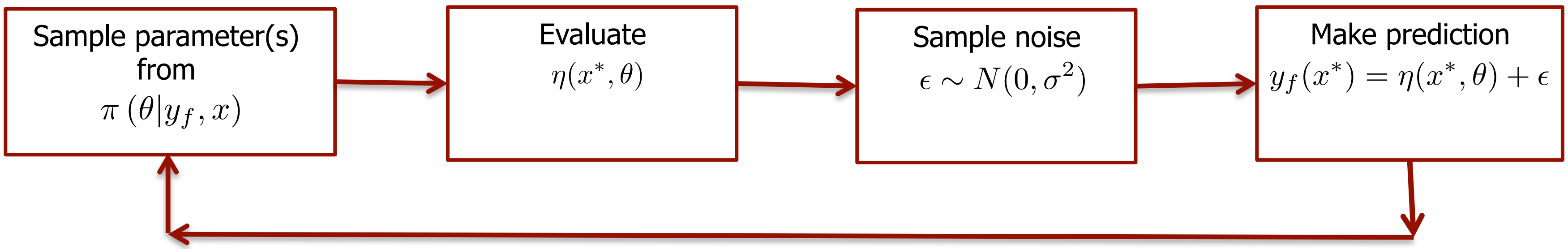
- Use Monte Carlo realizations to estimate expectations, variances, prediction uncertainty, etc.

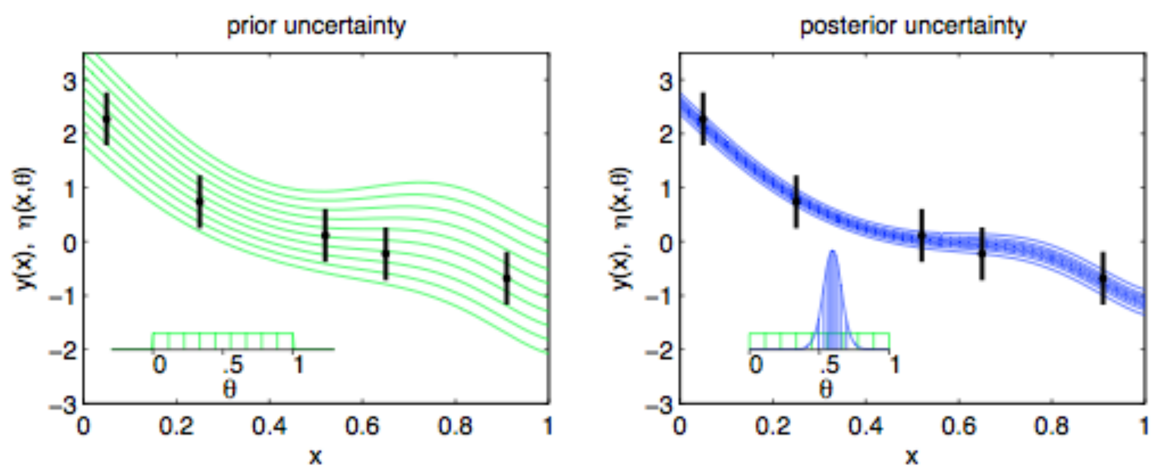
Can we trust our answer?

- MCMC uses similar diagnostic principles as optimization
- Write out a list of potential things you need to try to make sure optimization worked
  - Run multiple chains and see if they converge to the same distribution
  - Find and discard burn-in, keep only the recurrent part of the Markov Chain
  - Tuning MCMC proposal distribution for better success
- Myriad of diagnostics to assess chains

Prediction strategy at new input, x^*

- Can use predictive posterior distribution to get point estimates and prediction intervals





Limited simulator evaluations

- Frequently, the simulator is too slow to embed in the MCMC
- Instead, will use a limited number of runs of the computer model (*computer experiment*) to build a surrogate model
- Will model $\eta(\ )$ and θ simultaneously
- Model $\eta(\ )$ using a Gaussian process model (GP)

Basic inverse problem – Hard to evaluate computer model

$$y_s(x,t) = \eta(x,t)$$

Process mean was embedded in the previous solution

$$y_f(x,\theta) = \eta(x,\theta) + \epsilon$$

Kennedy and O'Hagan, 2001
Higdon et al., 2008

- Where,
  - y_f system response
  - y_s simulator response at input t
  - θ calibration parameter(s)
  - ϵ random error

Have data from 2 separate sources – field observations and computer model outputs

Problem is to estimate the *calibration* parameters

Perhaps make predictions

Aside: Use Gaussian processes for emulating computer model output

- GP' s have proven effective for emulating computer model output (Sacks et al., 1989)

- Emulating computer model output
  - output varies smoothly with input changes
  - output is noise free
  - passes through the observed response
  - GP's outperform other modeling approaches in this arena

Aside: Why a statistical emulator?

- Can only run the code a limited number of times
  - where to run the code
  - how many times do you run the code
- Want to **predict** output with **uncertainty** at un-observed inputs… need foundation for statistical inference

Aside: Statistical formulation for GP emulation of computer models

- Computer model: $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$ Usually scale inputs to unit cube

- Function is expensive so, get to observe a sample of *n* runs from the computer model

- Specify a set of inputs where we will run the code $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$

- Run the code and get the outputs $\mathbf{y}^T = (y_1, y_2, \ldots, y_n)$

Aside: Statistical formulation for GP emulation of computer models

- Will view the computer code as a single realization of a Gaussian process :

$$y(\mathbf{x}) = \mu + z(\mathbf{x})$$

where,

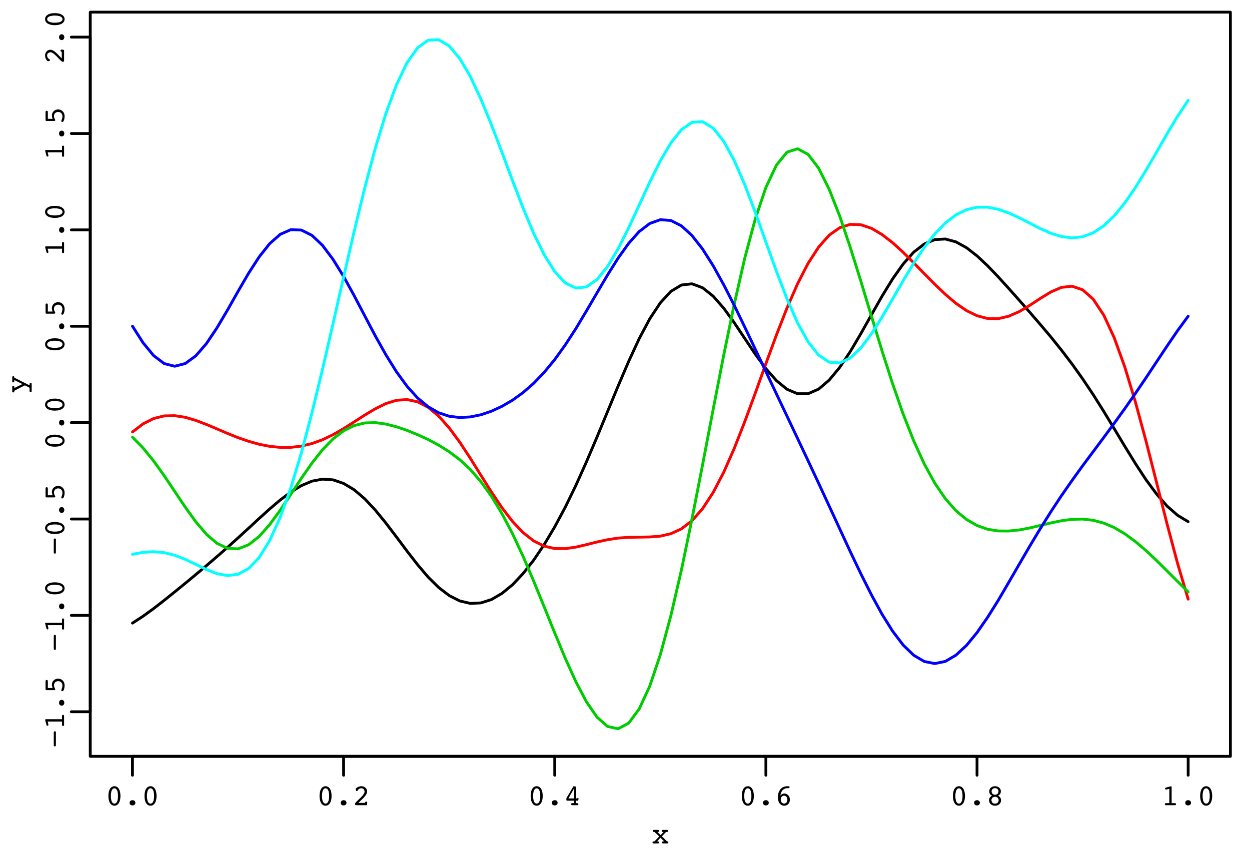
$$E\left(z(\mathbf{x})\right) = 0$$
$$Var\left(z(\mathbf{x})\right) = \sigma^2$$
$$z(\mathbf{x}) \sim N(0, \sigma^2)$$
$$Corr\left(z(\mathbf{x}), z(\mathbf{x}')\right) = \prod_{i=1}^{d} e^{-\phi_i (x_i - x_i')^2}$$

- For *n* data points, will have the covariance matrix, $\Sigma = \sigma^2 R$

$$\mathbf{y} \sim N(\mu 1_n, \Sigma)$$

Aside: Realizations of a GP for a fixed model

$\mu = 0 \; ; \; \sigma^2 = 25 \; ; \; \phi = 52$



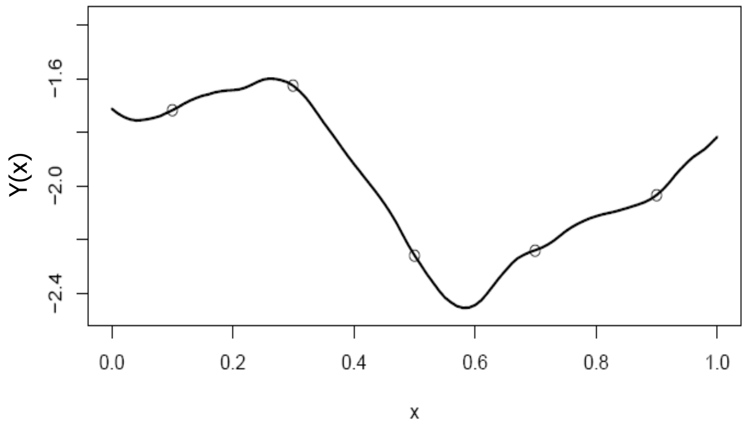
SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Aside: The parameters have meaning

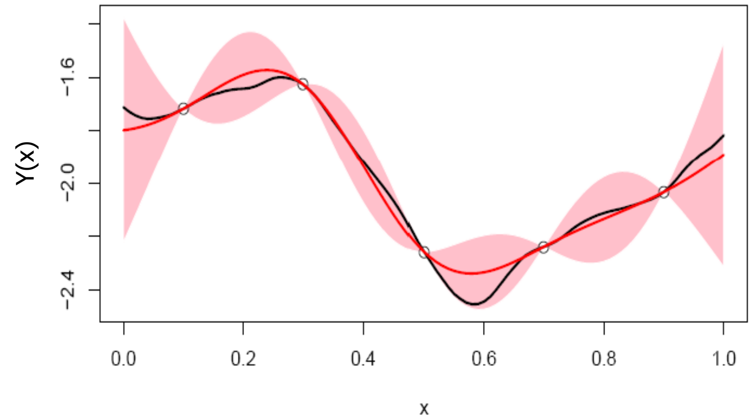
- The mean, μ, is the mean over all realizations
- Making the variance, σ^2, larger re-scales the vertical axis
- If $\phi_i = 0$, the function does not vary with respect to this input
- When ϕ_i is big, the function will be wigglier (a technical term?)
- Response where the inputs are close together will be more highly correlated that inputs that are far apart

Aside: Can emulate computer model with uncertainty

True function and observations



True function, emulated mean function and 95% prediction intervals



Basic inverse problem – Hard to evaluate computer model

Process mean was embedded in the previous solution

$$y_s(x,t) = \eta(x,t)$$

$$y_f(x,\theta) = \eta(x,\theta) + \epsilon$$

Kennedy and O'Hagan, 2001
Higdon et al., 2008

- Where,
  - y_f system response
  - y_s simulator response at input t
  - θ calibration parameter(s)
  - ϵ random error

Have data from 2 separate sources – field observations and computer model outputs

Problem is to estimate the *calibration* parameters

Perhaps make predictions

Hierarchical model is used to combine simulations and observations

- View computational model as a draw of a random process
- Denote vectors of simulation trials as and field measurements as $\mathbf{y_s}$ and $\mathbf{y_f}$ respectively
- Suppose that these are *n* and *m* -vectors respectively

- Can combine sources of information using a single GP

$$y_s(x,t) = \eta(x,t)$$

$$y_f(x,\theta) = \eta(x,\theta) + \epsilon$$

$$\begin{pmatrix} \mathbf{y_s} \\ \mathbf{y_f} \end{pmatrix} \sim MVN\left(\mu, \Sigma_\eta + \Sigma_\epsilon\right)$$

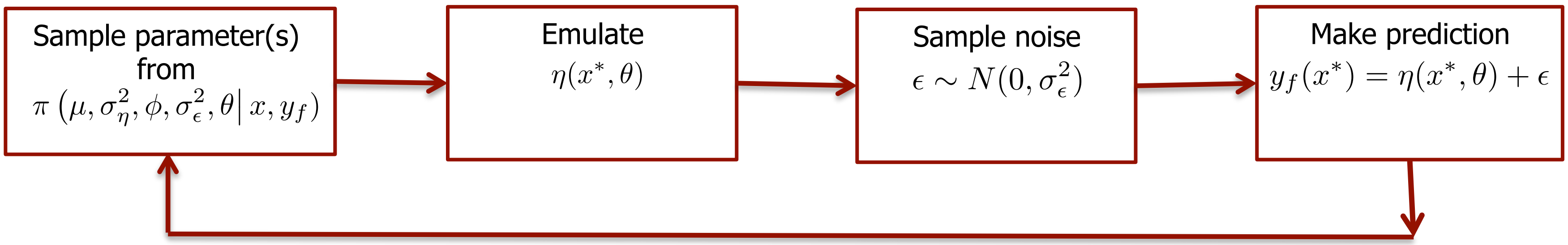
Independent Gaussian process models

$$\Sigma_\eta + \Sigma_\epsilon = \Sigma_\eta + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_\epsilon^2 I_n \end{pmatrix}$$

Use a Bayesian specification for model parameters

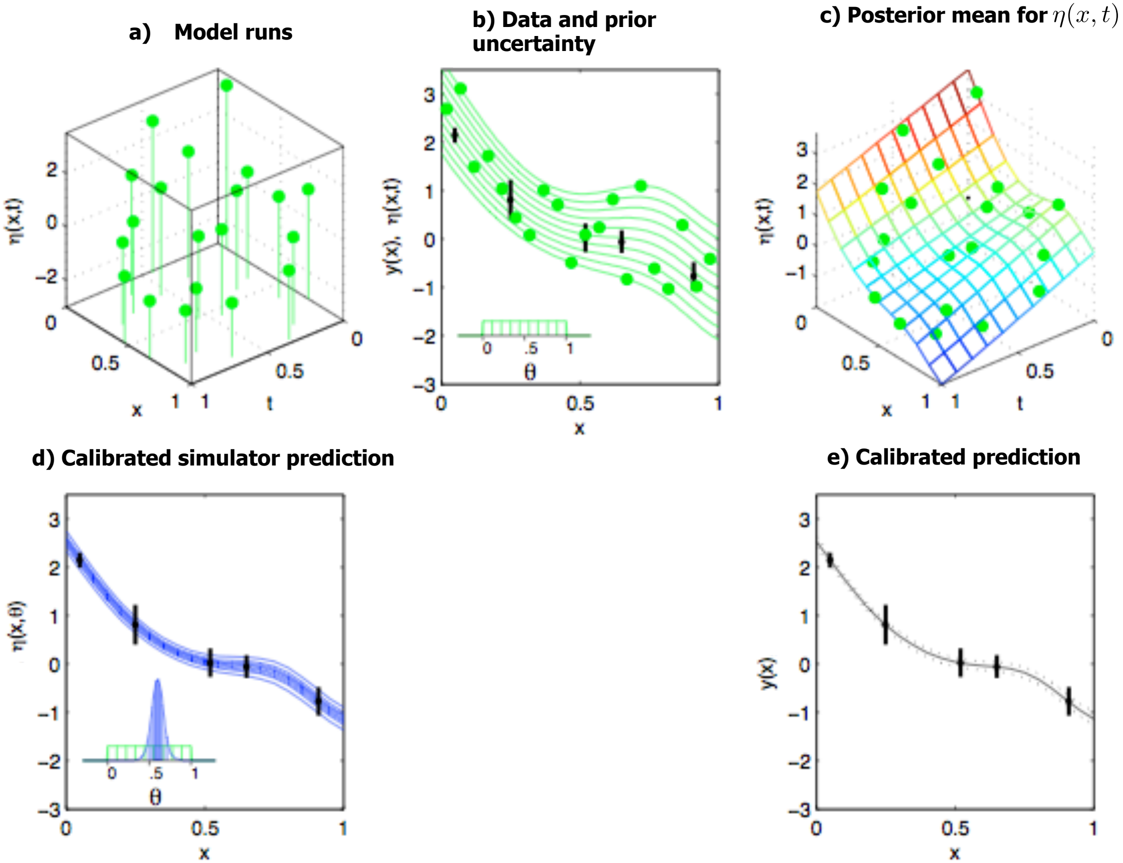
- Parameters to be estimated: μ, σ_ϵ^2, ϕ, σ_η^2 and θ
- Need to specify prior distributions for each
- Usually inverse gamma priors for variances and expert knowledge for calibrations parameters
- Often use exponential priors for correlation parameters
- Use MCMC to draw samples from joint posterior distribution

Prediction strategy

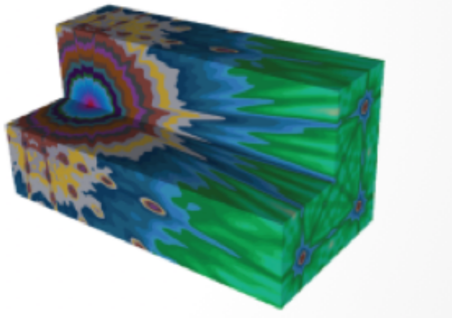


- Can use posterior predictive distribution to get point estimates and prediction intervals
- Attempts to get full accounting of uncertainty

Model calibration – Idea



Example
Center for Exascale Radiation Transport (CERT)



- Have observations from CERT model and experiments

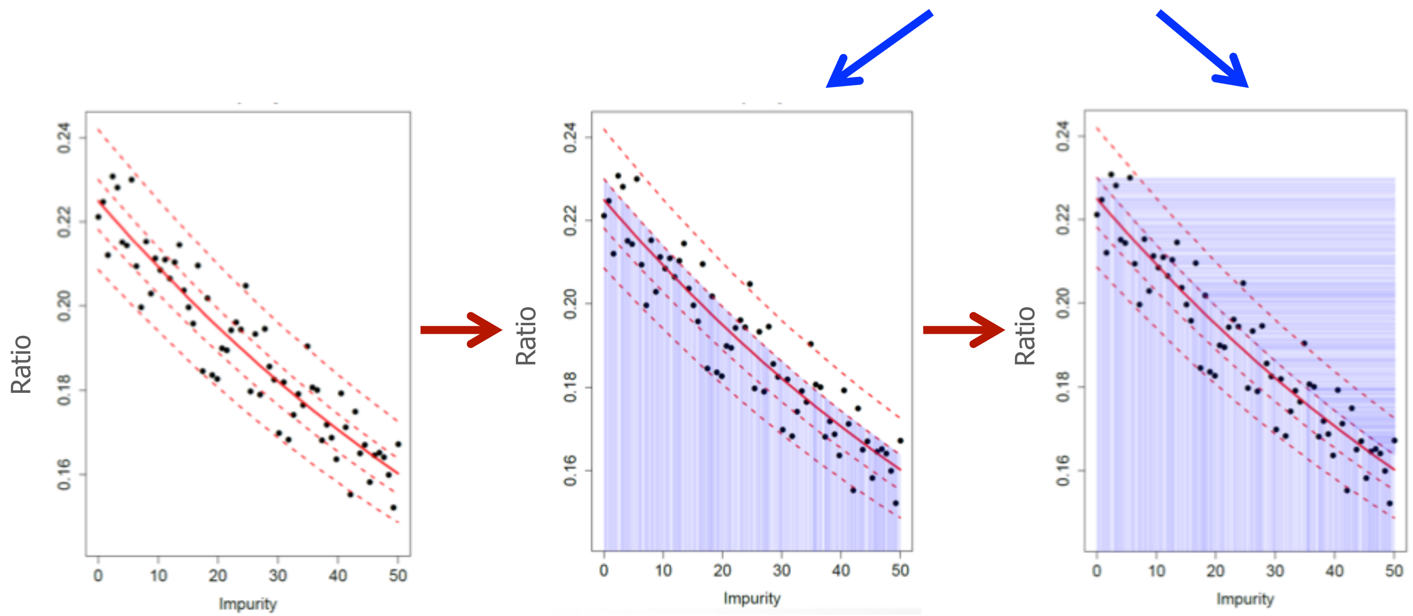
- Experiment data:
  - 6 experiments (different bricks)
  - Responses: Ratio of counts with/counts without graphite

- CERT Model runs
  - 64 simulations of the computational model with inputs chosen using maximin Latin-hypercube design
  - Inputs: Axial position and Impurity Concentration
  - Response: Ratio of count rates (with/without) graphite

- **Goal:** Characterize impurities in the graphite… what does this mean?

Idea: without data

Possible impurities correspond to possible ratios

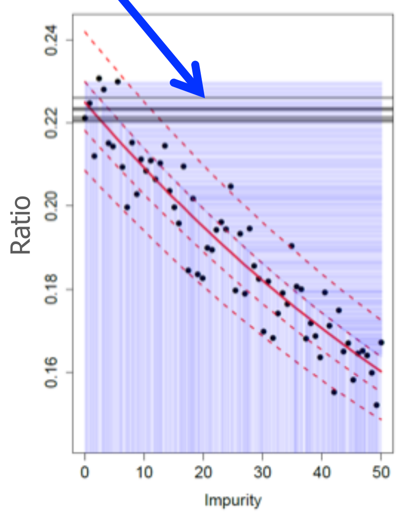


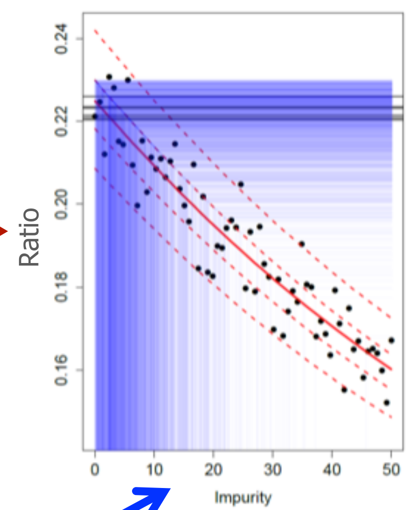
Courtesy of Mike Grosskopf

SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Idea: with data

Field data

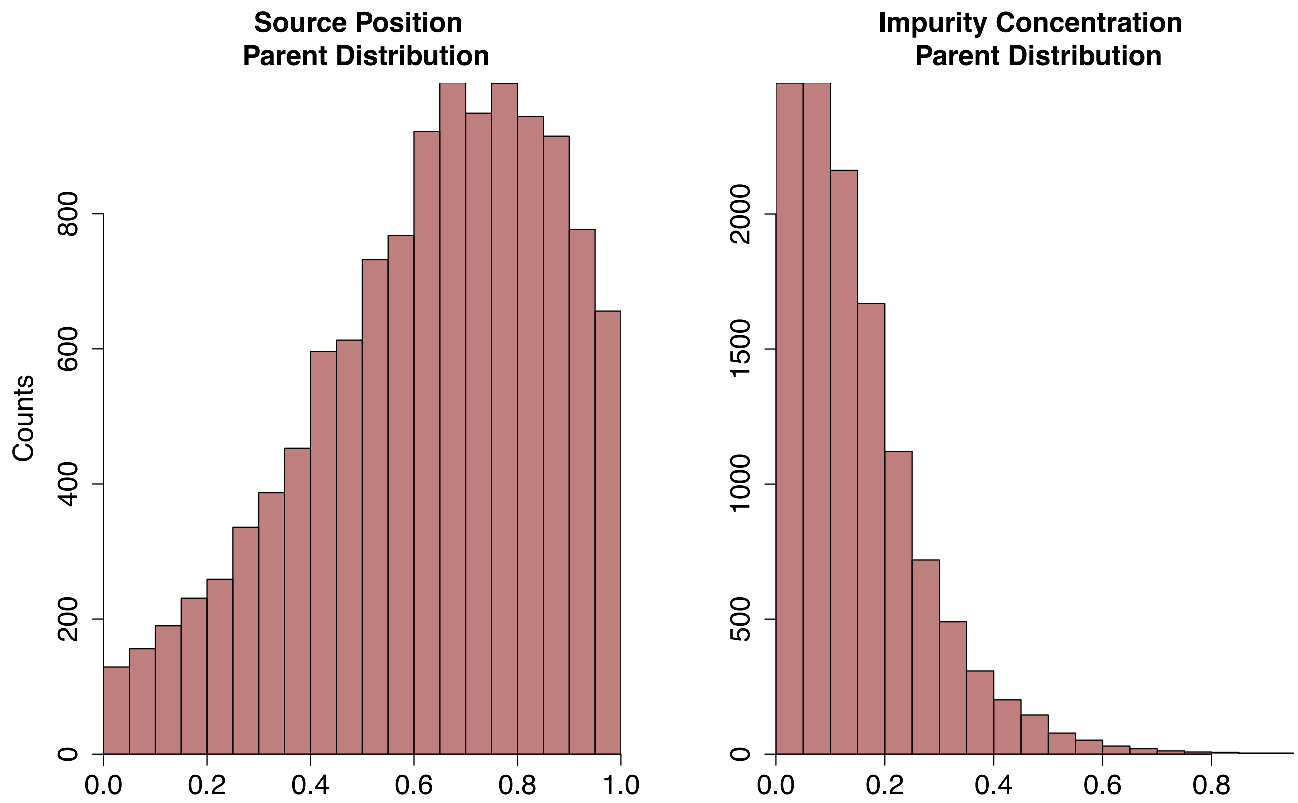




Field data help constrain impurity concentration

SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Our interest was in posterior distribution for the calibration parameters



Mathematical models are a computer implementation of what we know

- So, we have mathematical models to represent physical systems
- What are they?
  - They are an implementation of our knowledge about the system at hand
- **Problem:** We may not know everything about the system
- **Problem:** Even if we know about everything, we not be able to compute what we want to compute

What happens if the model is not quite right

- The computer model gives us insight into the system
- Does it help us predict the system response?
- If not, how can a statistical formulation help us?

Model calibration – Statistical formulation

$$y_s(x,t) = \eta(x,t)$$
$$y_f(x,\theta) = \eta(x,\theta) + \delta(x) + \epsilon$$

Kennedy and O'Hagan, 2001
Higdon et al., 2008

- Where,
  - x model or system inputs;
  - y_f system response
  - y_s simulator response
  - θ calibration parameters
  - ϵ random error

Have data from 2 separate sources – field observations and computer model outputs

Also have a model for systematic discrepancy, $\delta(x)$

Model calibration – Statistical formulation

$$y_s(x,t) = \eta(x,t)$$

$$y_f(x,\theta) = \eta(x,\theta) + \delta(x) + \epsilon$$

Gaussian process models

Have Gaussian process model for the discrepancy

- GP's are the same for the simulator and error terms as specified previously

- For the discrepancy, will also specify a GP (though this is not always a great idea)

$$E\left(\delta(\mathbf{x})\right) = 0$$

$$Var\left(\delta(\mathbf{x})\right) = \sigma_\delta^2$$

$$Corr\left(\delta(\mathbf{x}), \delta(\mathbf{x}')\right) = \prod_{i=1}^{d} e^{-\gamma_i (x_i - x_i')^2}$$

- For all the field observations: $\delta(\mathbf{X}) \sim N(0, \sigma_\delta^2 R_\delta) \equiv N(0, \Sigma_\delta)$

Hierarchical model is used to combine simulations and observations with discrepancy

- View computational model as a draw of a random process (again!!)

- Denote vectors of simulation trials as and field measurements as $\mathbf{y_s}$ and $\mathbf{y_f}$ respectively

- Suppose that these are *n* and *m* -vectors respectively

→

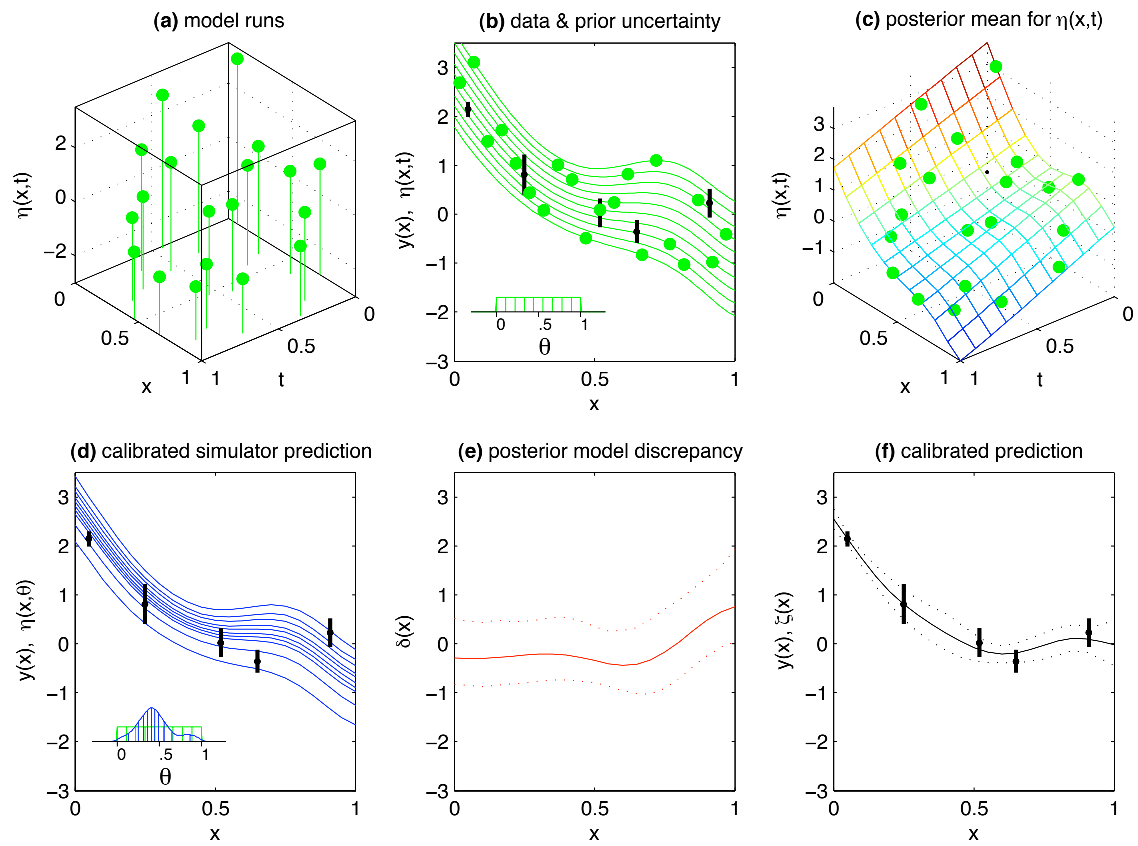
- Can combine sources of information using a single GP

$$y_s(x,t) = \eta(x,t)$$

$$y_f(x,\theta) = \eta(x,\theta) + \delta(x) + \epsilon$$

$$\begin{pmatrix} \mathbf{y_s} \\ \mathbf{y_f} \end{pmatrix} \sim MVN\left(\mu, \Sigma_\eta + \Sigma_\delta + \Sigma_\epsilon\right)$$

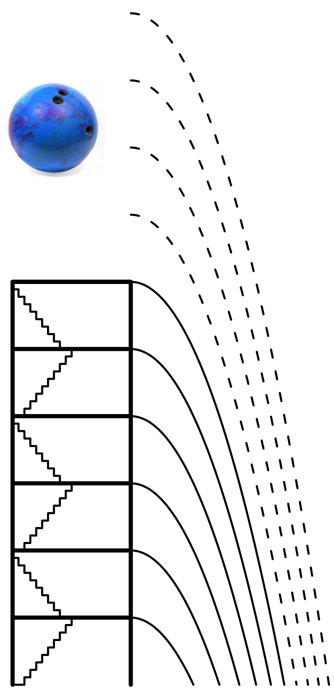
Calibration idea: Discrepancy model

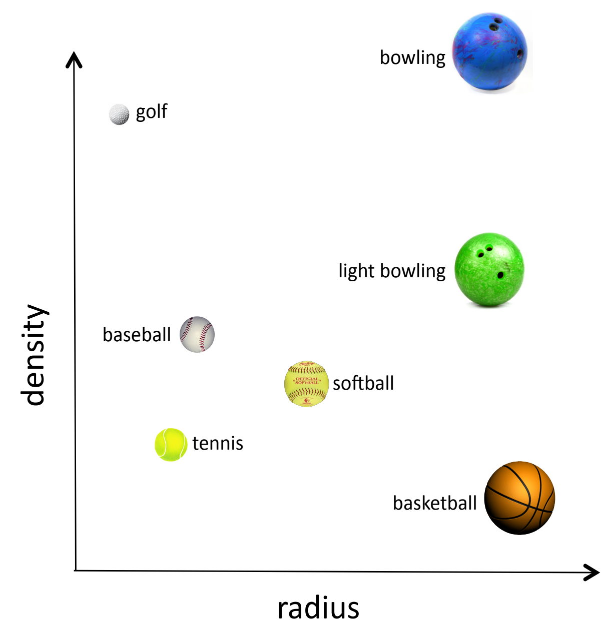


SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Basic ball drop example

$$h = \text{vertical distance} = \frac{1}{2} \times g \times t^2 \longrightarrow \frac{d^2 h}{dt^2} = g$$

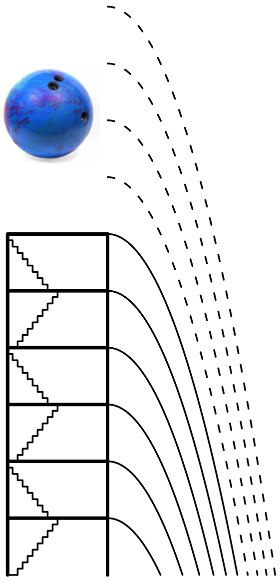


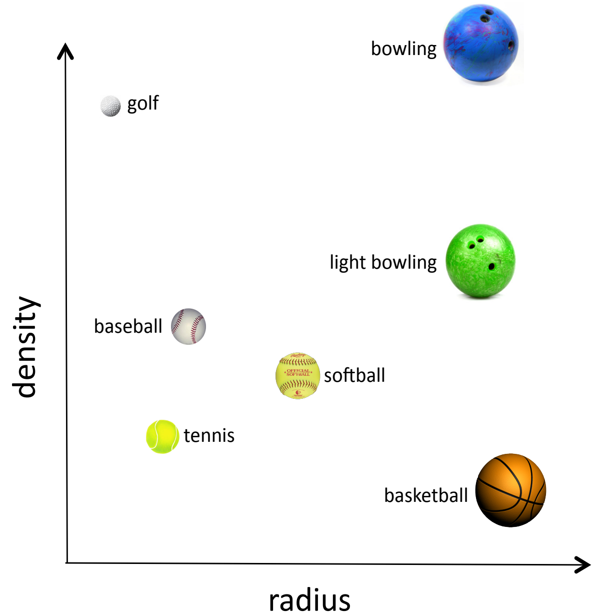


SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

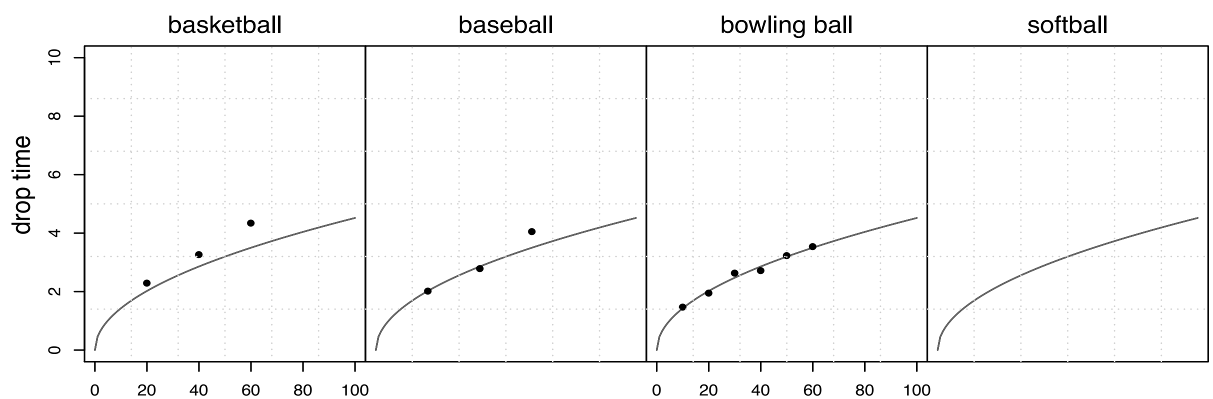
SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Basic ball drop example



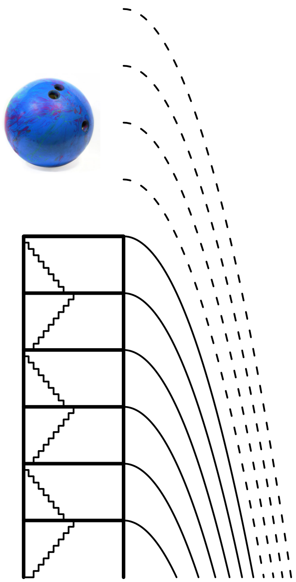


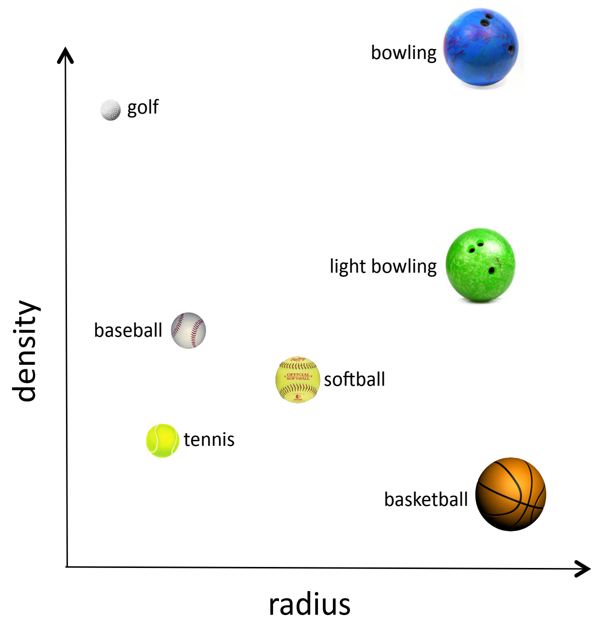
$$\frac{d^2h}{dt^2} = g \qquad \text{use} \quad \hat{g} = 9.8\,\frac{m}{s^2}$$



SFU
Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

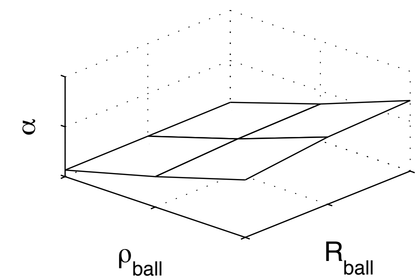
Basic ball drop example with discrepancy

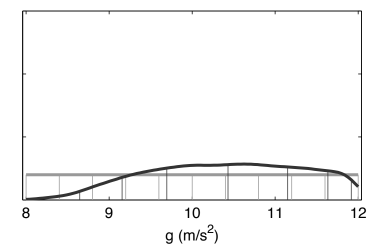


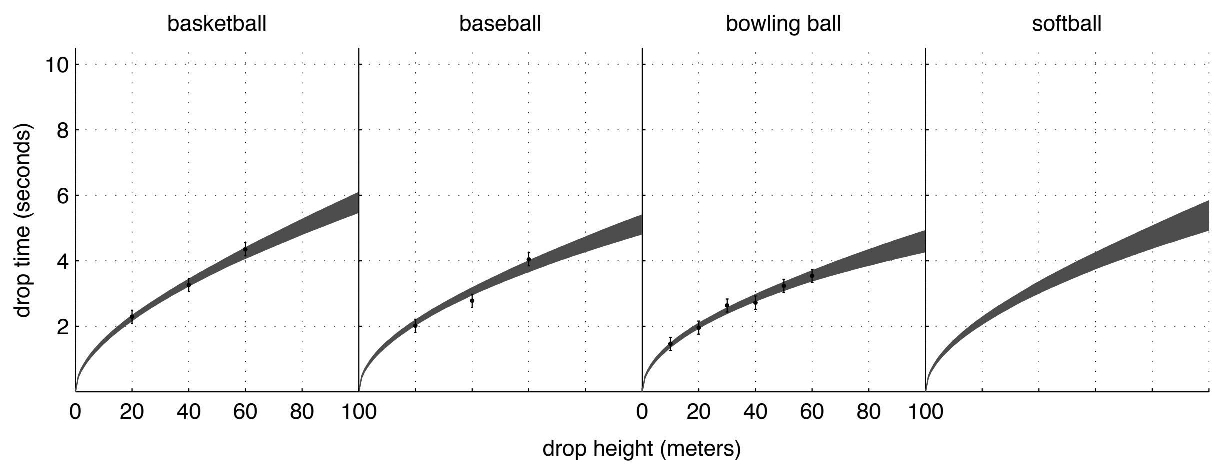


$$\frac{d^2h}{dt^2} = g$$

drop time = simulated drop time + α(ρ,R) × drop height







Department of Statistics and Actuarial Science
SIMON FRASER UNIVERSITY
ENGAGING THE WORLD

Thanks for your time

References

- Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008). "Computer model calibration using high dimensional output", *Journal of the American Statistical Association*, 103, 570–583.
- Kennedy, M. and O'Hagan, A. (2001), "Bayesian calibration of computer models (with discussion)". *Journal of the Royal Statistical Society, Series B*. 63, 425-464.
- Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P (1989), "Design and analysis of computer experiments," *Statistical Science*, 4, 409-435.