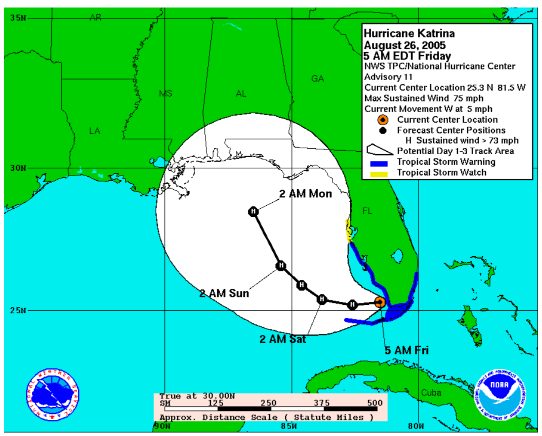
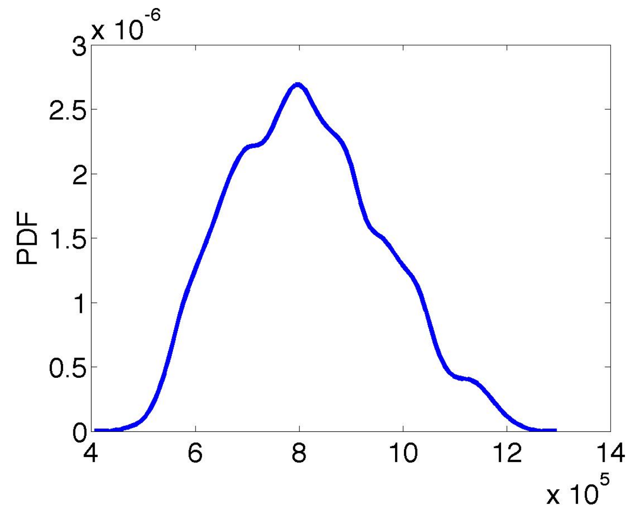
Uncertainty Quantification for Biological Models

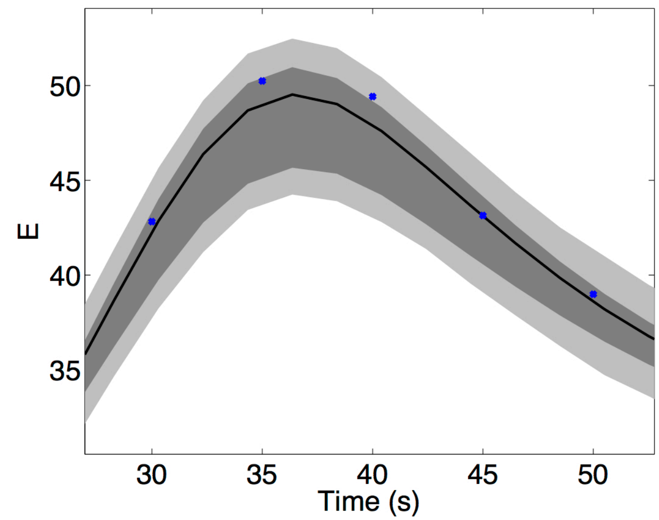
Ralph C. Smith

Department of Mathematics

North Carolina State University







Support: Air Force grant AFOSR FA9550-15-0299

NSF CMMI-1306290

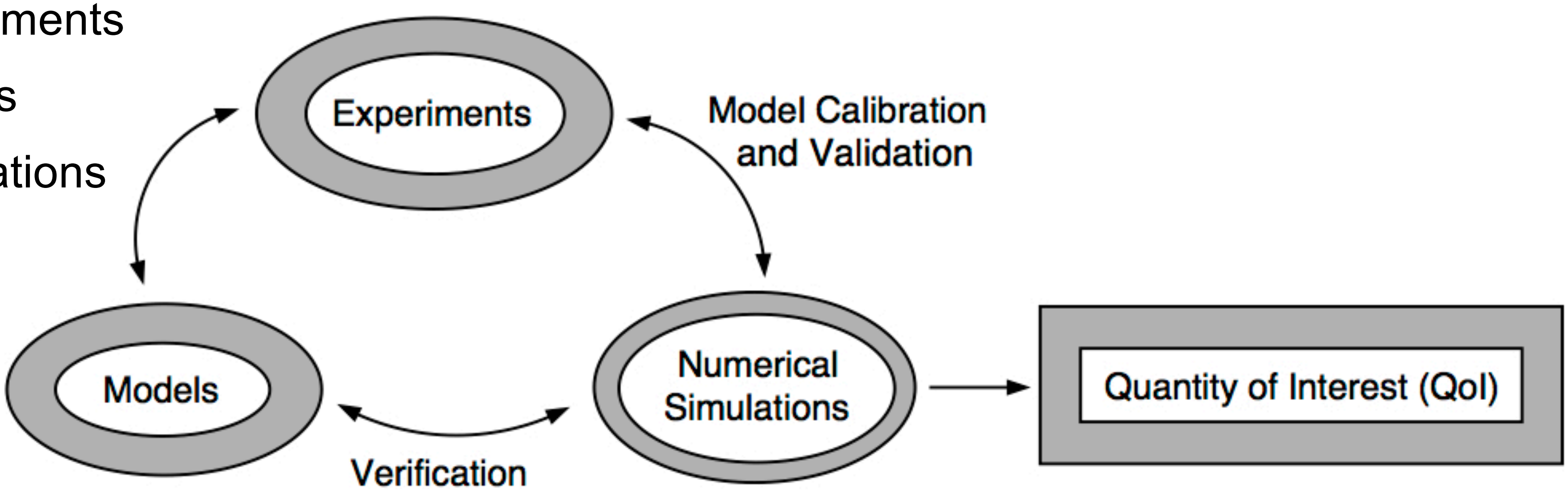
DOE Consortium for Advanced Simulation of LWR (CASL)

NNSA Consortium for Nonproliferation Enabling Capabilities (CNEC)

Predictive Science

Components: All involve uncertainty

- Experiments
- Models
- Simulations



- *Experimental results are believed by everyone, except for the person who ran the experiment*, source anonymous, quoted by Max Gunzburger, Florida State University.
- *Essentially, all models are wrong, but some are useful,* George E.P. Box, Industrial Statistician.
- *Computational results are believed by no one, except the person who wrote the code*, source anonymous, quoted by Max Gunzburger, Florida State University.
- *I have always done uncertainty quantification. The difference now is that it is capitalized.* Bill Browning, Applied Mathematics Incorporated.

Example 1: Weather Models

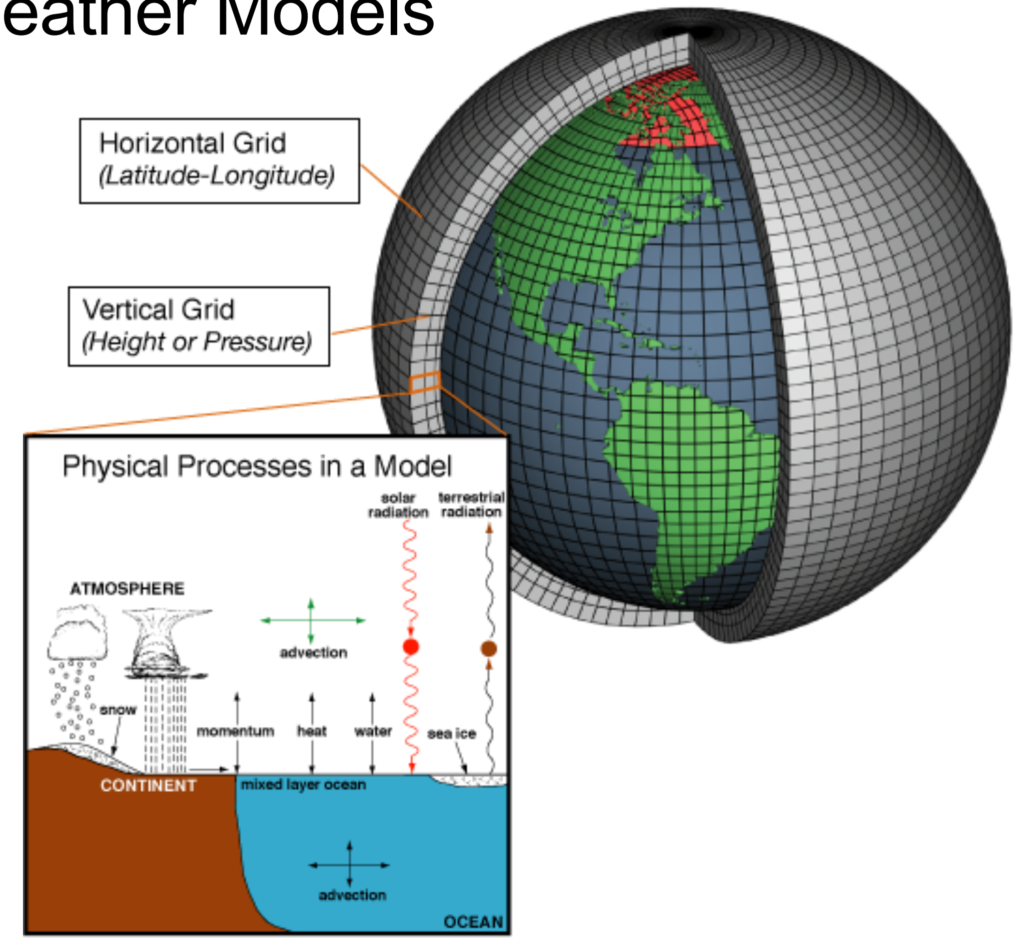
Challenges:

• Coupling between temperature, pressure gradients, wind, precipitation, aerosol species, etc.;

• Models and inputs contain uncertainties;

• Numerical grids are necessarily larger than many phenomena; e.g., clouds

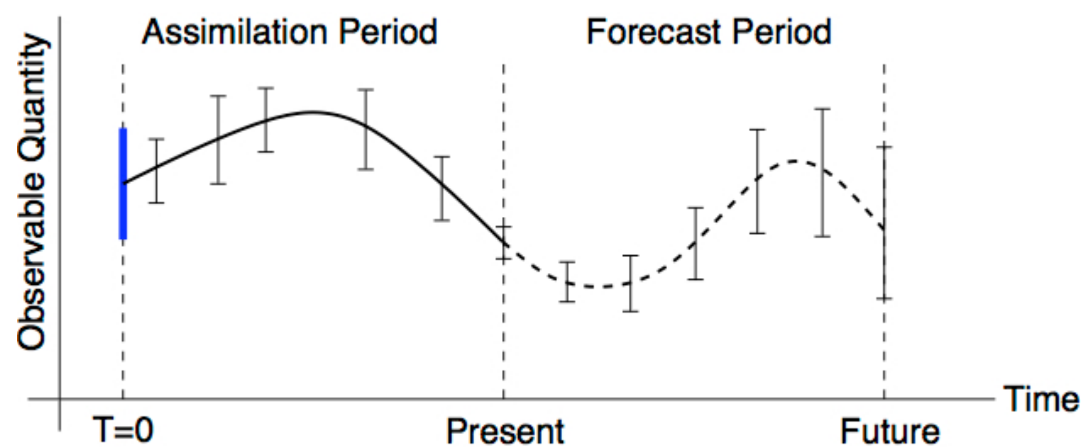
• Sensors have limited accuracy and positions may be uncertainty; e.g., weather balloons, ocean buoys.



Goal:

• Assimilate data to quantify uncertain initial conditions and parameters;

• Make predictions with quantified uncertainties.



Equations of Atmospheric Physics

Conservation Relations:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0$$

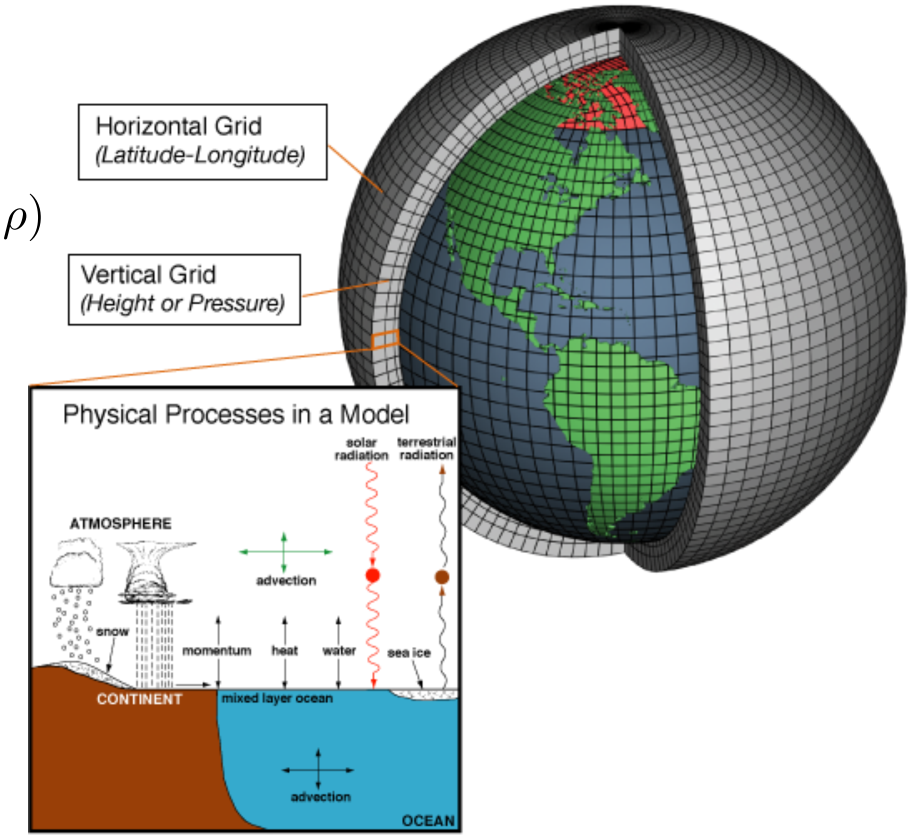
$$\frac{\partial v}{\partial t} = -v \cdot \nabla v - \frac{1}{\rho}\nabla p - g\hat{k} - 2\Omega \times v$$

$$\rho c_V \frac{\partial T}{\partial t} + p\nabla \cdot v = -\nabla \cdot F + \nabla \cdot (k\nabla T) + \rho \dot{q}(T, p, \rho)$$

$$p = \rho R T$$

$$\frac{\partial m_j}{\partial t} = -v \cdot \nabla m_j + S_{m_j}(T, m_j, \chi_j, \rho)\ ,\ j = 1, 2, 3,$$

$$\frac{\partial \chi_j}{\partial t} = -v \cdot \nabla \chi_j + S_{\chi_j}(T, \chi_j, \rho)\ ,\ j = 1, \cdots, J,$$



Constitutive Closure Relations: e.g.,

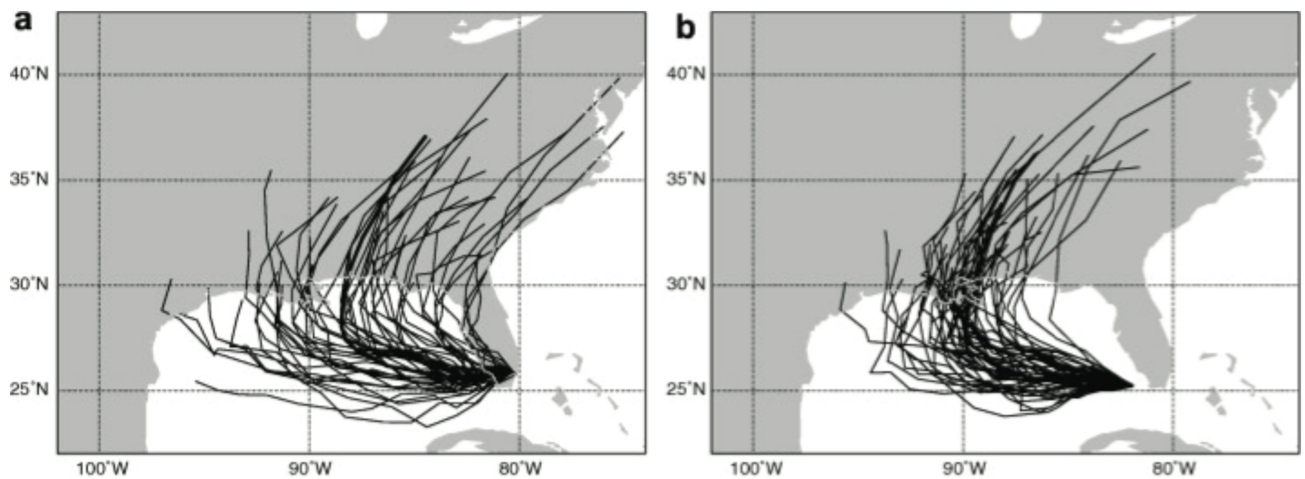
$$S_{m_2} = S_1 + S_2 + S_3 - S_4$$

where

$$S_1 = \bar{\rho}(m_2 - m_2^*)^2 \left[\underline{1.2 \times 10^{-4}} + \left(\underline{1.569 \times 10^{-12}} \frac{n_r}{d_0(m_2 - m_2^*)}\right)\right]^{-1}$$

Ensemble Predictions

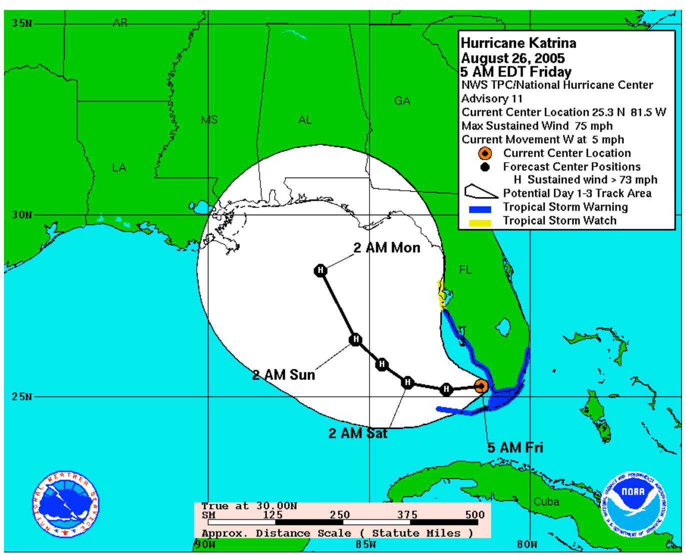
Ensemble Predictions:



00 UTC on August 26, 2005

12 UTC on August 26, 2005

Cone of Uncertainty:



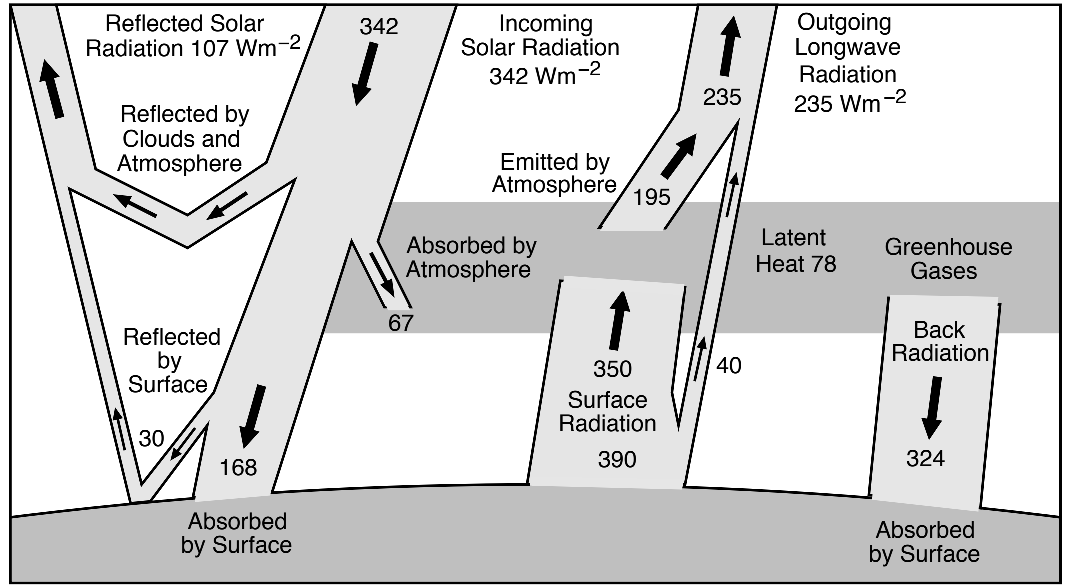
General Questions:

- What is expected rainfall on July 29?
- What are high and low temperatures?

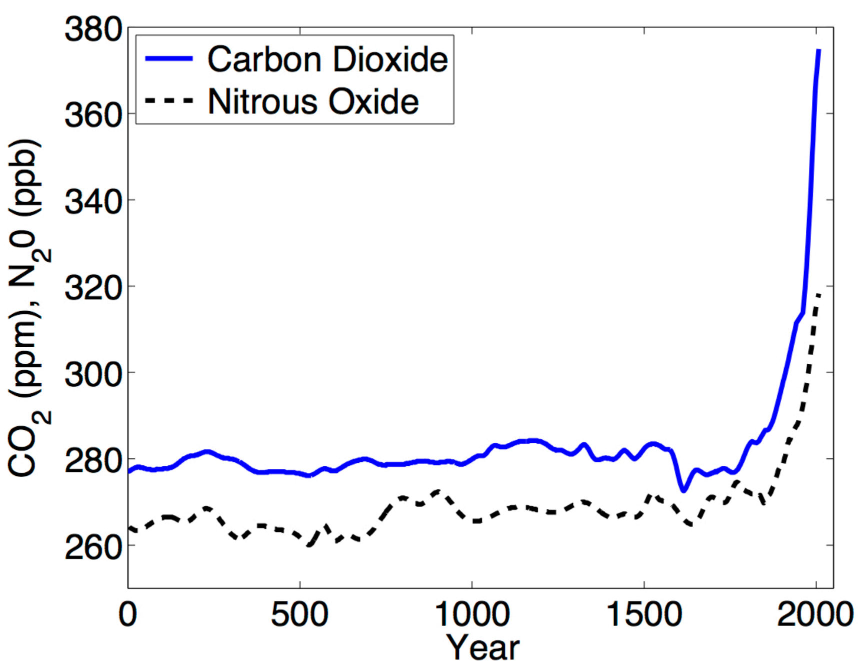
Example 2: Climate Models

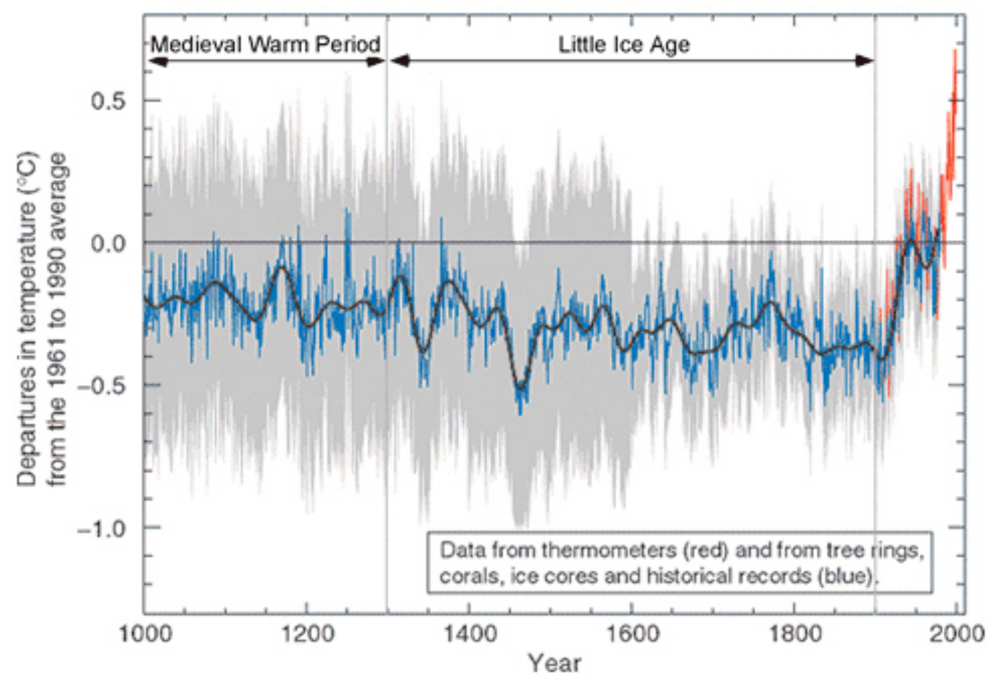
Strategy: Compute global energy balance

- Forced boundary value problem rather than initial value problem associated with weather models;
- This can reduce effects of chaotic dynamics.



Greenhouse Gases:





Climate Models – Uncertainty Quantification

Subset of Uncertainties listed by IPCC:

• Scenarios must be used for four ranges of demographic, economic, and technological growth.

• ``The magnitude of carbon dioxide emissions from land-use change and methane emissions from individual sources remain as key uncertainties.''

• ``Aerosol impacts on the magnitude of the temperature response, on clouds and on precipitation remain uncertain."

• ``Models differ considerably in their estimates of the strength of different feedbacks in the climate system, particularly cloud feedbacks, oceanic heat uptake and carbon cycle feedbacks, although progress has been made in these areas."

• ``Large-scale ocean circulation changes beyond the 21st century cannot be reliably assessed because of uncertainties in the meltwater supply from the Greenland ice sheet and model response to the warming.''

• ``Projections of climate change and its impacts beyond 2050 are strongly scenario- and model-dependent, and improved projections would require improved understanding of sources of uncertainty and enhancements in systematic observation networks."

Example 3: HIV Model for Characterization and Control Regimes

HIV Model:

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon) k_1 V T_1$$

$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon) k_2 V T_2$$

$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon) k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

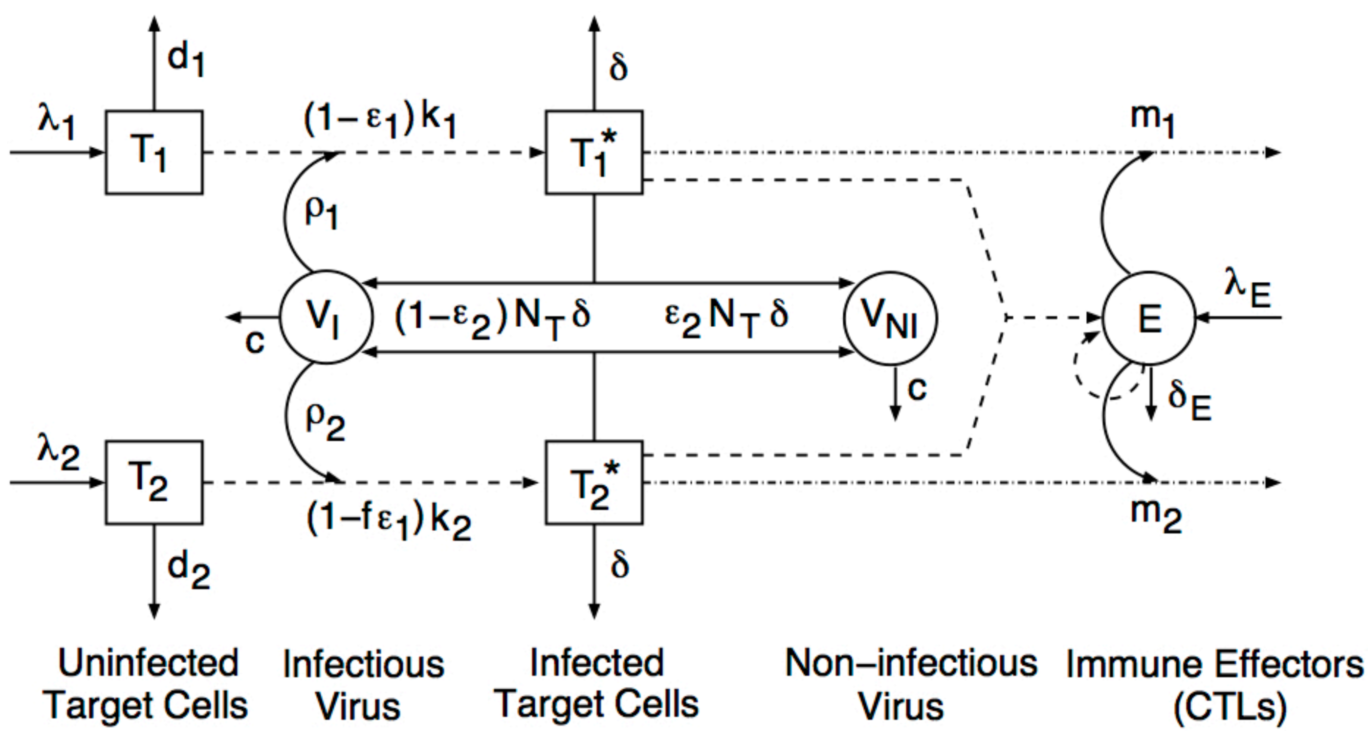
$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2] V$$

$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Notes: 21 parameters

[Adams, Banks et al., 2005, 2007]

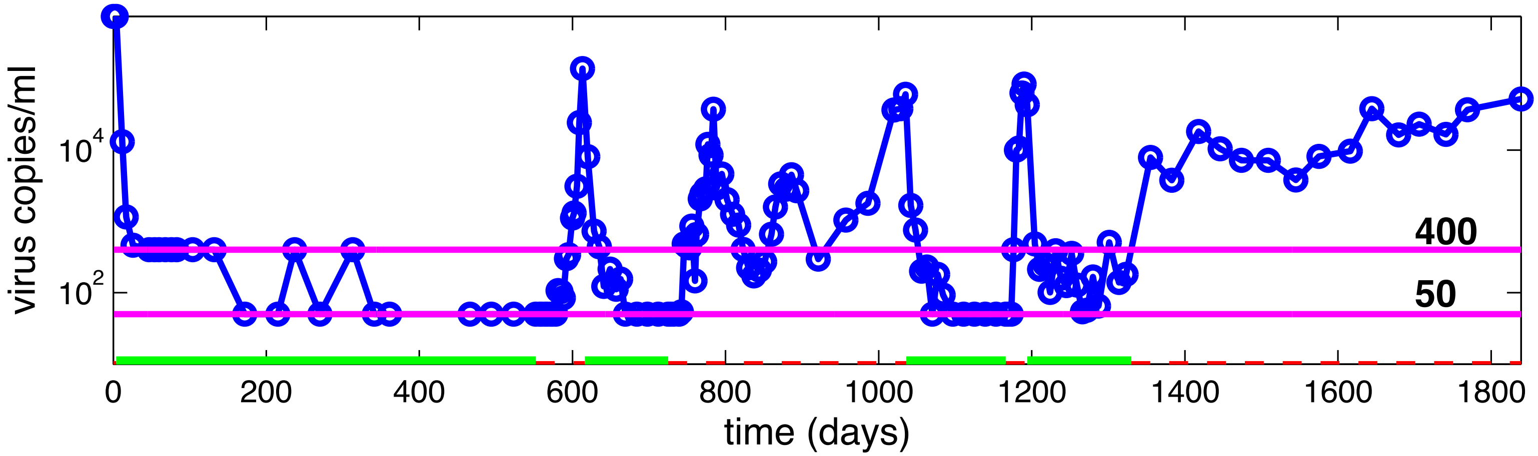
Compartments:



Example 3: HIV Model for Characterization and Control Regimes

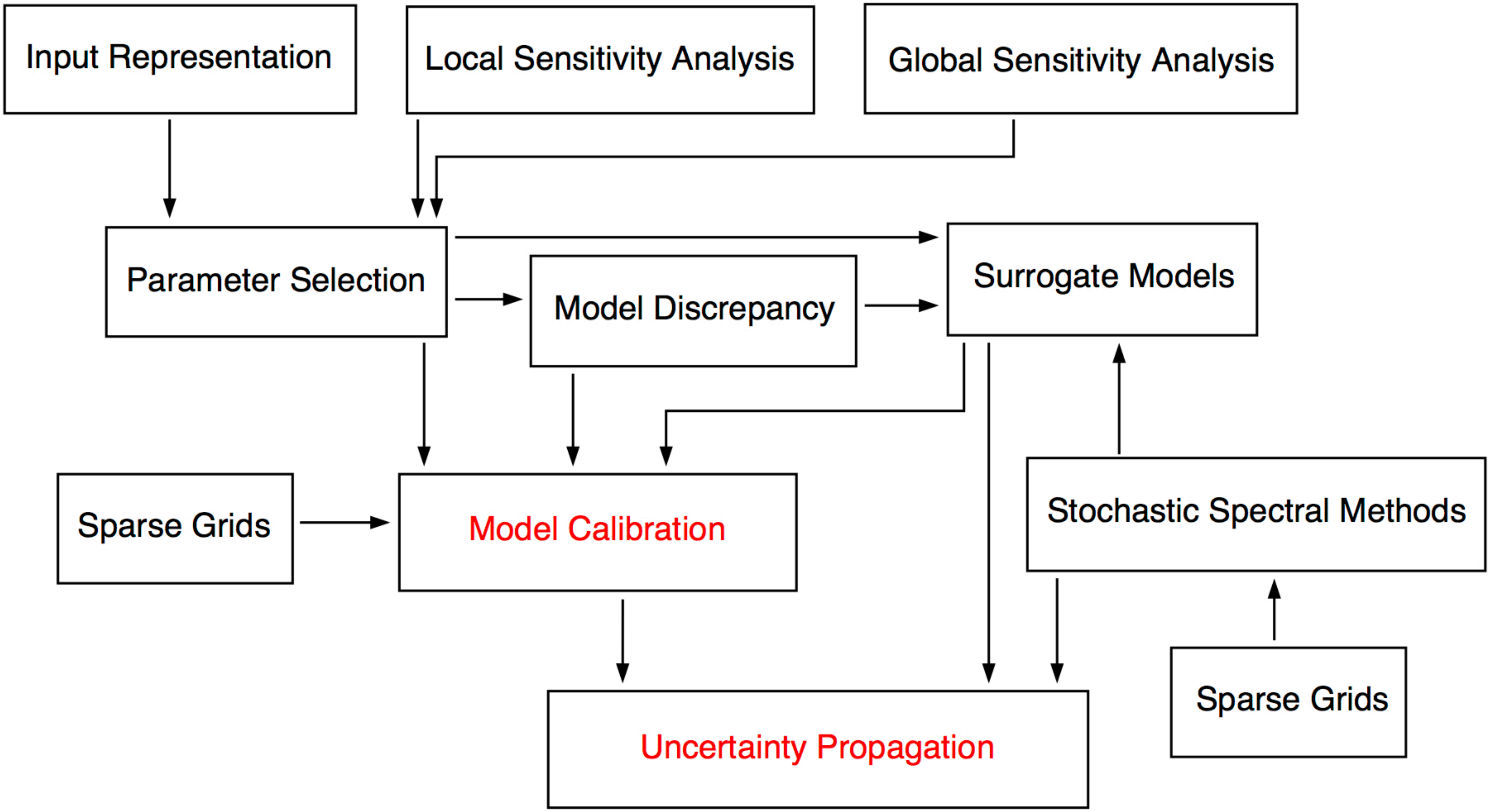
HIV Model: Several sources of uncertainty including viral measurement techniques

Example: Upper and lower limits to assay sensitivity

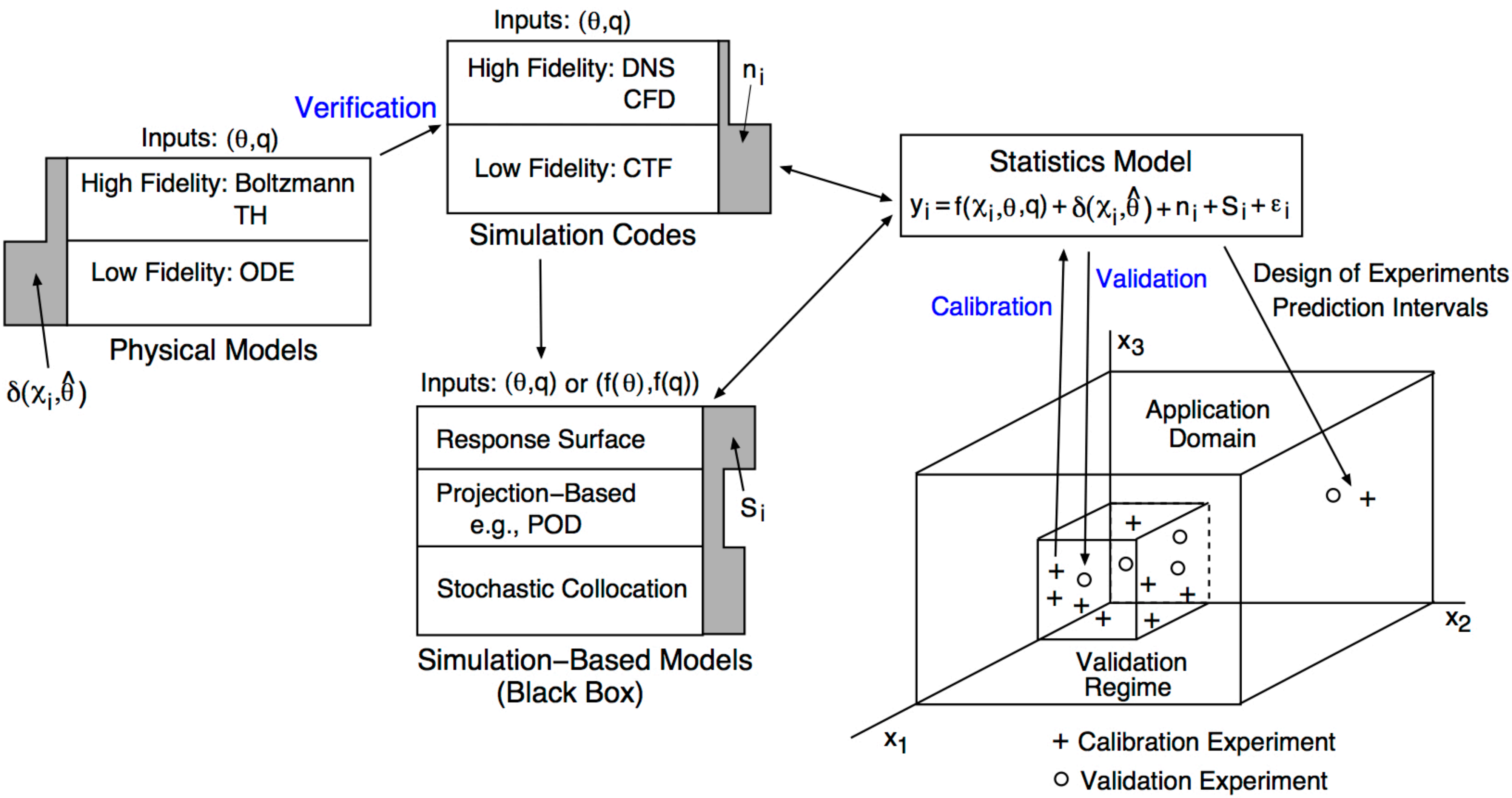


Steps in Uncertainty Quantification

Note: Uncertainty quantification requires synergy between statistics, mathematics and application area.



Relation Between UQ, Verification and Validation



First Challenge: Terminology and Notation

Terminology:

- Inputs: Parameters, initial conditions, boundary conditions, exogenous forces; e.g., Parameters in material models, initial conditions in weather models.

- Outputs or Responses: Quantities that we experimentally or numerically measure; e.g., Viral load,

- Quantities of Interest: Statistical quantity that we want to monitor; e.g., average temperature, peak nuclear power plant operating temperature.

Notation: Input notation can vary even within disciplines!

- Math Control Community: $q = [q_1, \ldots, q_p]$
- Math Reduced-Order Community: $p = [p_1, \ldots, p_q]$
- Statistics: $\theta = [\theta_1, \ldots, \theta_d]$
- Nuclear Engineering: $\alpha = [\alpha_1, \ldots, \alpha_k]$

Notation: Same variability for outputs and quantities of interest

- Math Control Community: ν or y
- Nuclear Engineering: R

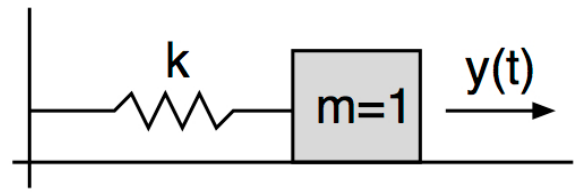
First Challenge: Terminology and Notation

Terminology:

- Linearly parameterized problems: e.g., portfolio model $y = c_1 q_1 + c_2 q_2$
    - Rarely occur in applications except image processing
- Nonlinearly parameterized problems: typical case
    - Differs from linear or nonlinear in state; e.g., spring model

$$\frac{d^2 y(t)}{dt^2} + ky(t) = 0$$

$$y(0) = y_0 \ , \ \frac{dy}{dt}(0) = 0$$



Notation: $\dot{y} \equiv \frac{dy}{dt} \ , \ \ddot{y} \equiv \frac{d^2 y}{dt^2}$

Inputs: $q = [k, y_0]$

Response: Displacement $y(t) = y_0 \cos(\sqrt{k} \cdot t)$

$$\Rightarrow \quad \begin{array}{l} \ddot{y}(t) + ky(t) = 0 \\ y(0) = y_0 \ , \ \frac{dy}{dt}(0) = 0 \end{array}$$

Note:

- Linear state dependence
- Nonlinear parameter dependence

Model Calibration and Uncertainty Propagation

Sources of Uncertainty:

- Model
- Parameters
- Sensor measurements
- Initial/boundary conditions

Strategy:

- Quantify uncertainty in parameters
- Propagate uncertainty through model

Example: HIV model

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon) k_1 V T_1$$

$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon) k_2 V T_2$$

$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon) k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2] V$$

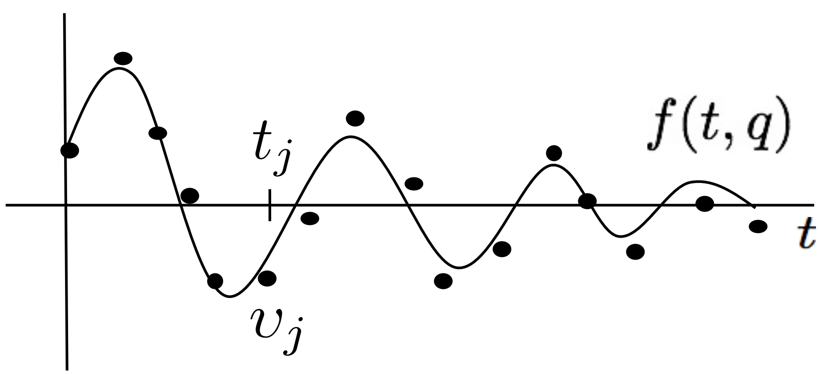
$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Parameters: Reduced set

$$q = [b_E, \delta, d_1, k_2, \lambda_1, K_d]$$

Point Estimates: Ordinary least squares – Kevin Flores and Alun Lloyd yesterday

$$q^0 = \arg\min_q \frac{1}{2} \sum_{j=1}^{N} [\upsilon_j - f(t_j, q)]^2$$



Note: Scaling critical since parameter values vary by 8 orders of magnitude.

Bayesian Model Calibration

Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayesian Model Calibration:

- Parameters assumed to be random variables

$$\pi(q|\upsilon) = \frac{\pi(\upsilon|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon|q)\pi_0(q)dq}$$

Example: Coin Flip

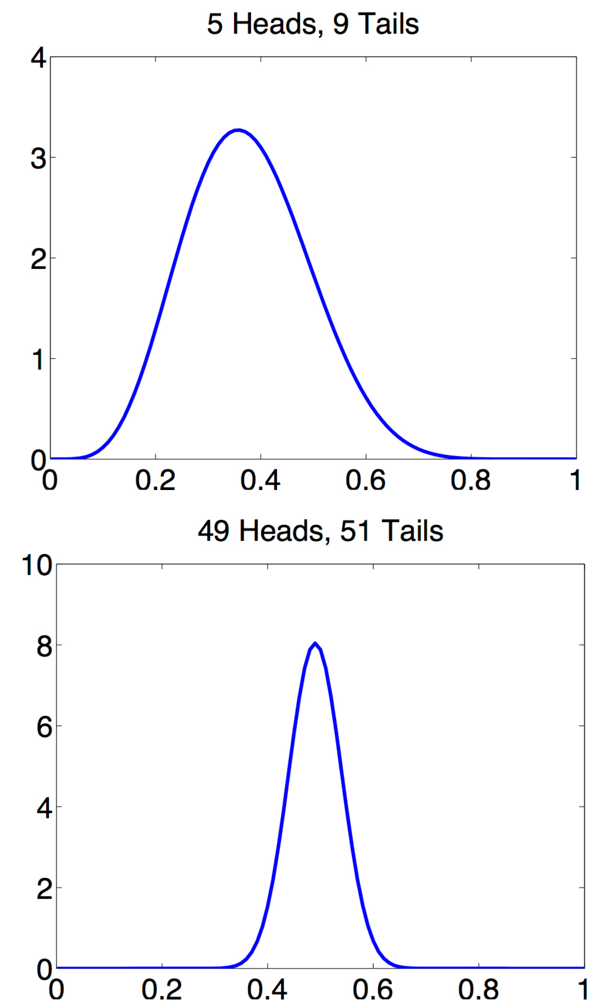
$$\Upsilon_i(\omega) = \begin{cases} 0 & , \quad \omega = T \\ 1 & , \quad \omega = H \end{cases}$$

Likelihood:

$$\begin{aligned} \pi(\upsilon|q) &= \prod_{i=1}^{N} q^{\upsilon_i}(1-q)^{1-\upsilon_i} \\ &= q^{\sum \upsilon_i}(1-q)^{N-\sum \upsilon_i} \\ &= q^{N_1}(1-q)^{N_0} \end{aligned}$$

Posterior with Noninformative Prior: $\pi_0(q) = 1$

$$\pi(q|\upsilon) = \frac{q^{N_1}(1-q)^{N_0}}{\int_0^1 q^{N_1}(1-q)^{N_0}dq} = \frac{(N+1)!}{N_0!N_1!}q^{N_1}(1-q)^{N_0}$$

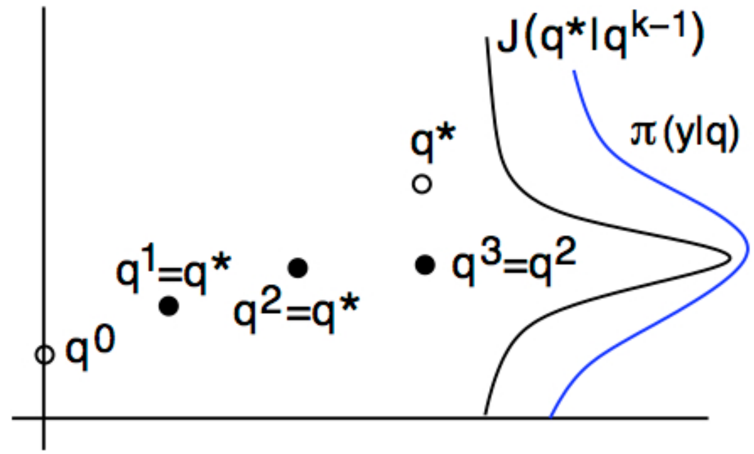


Bayesian Model Calibration

Bayesian Model Calibration:

• Parameters considered to be random variables with associated densities.

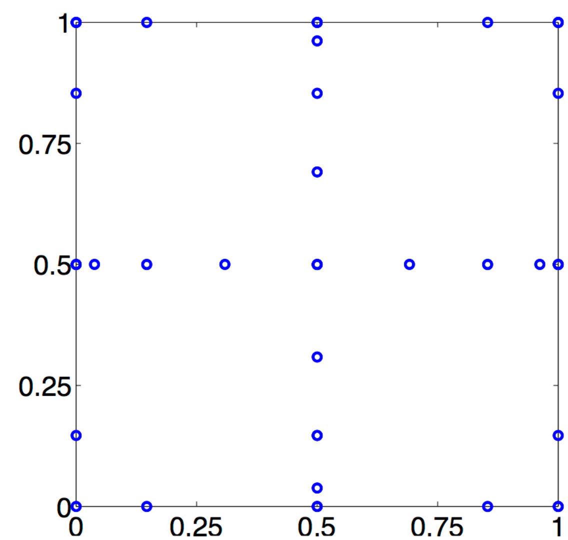
$$\pi(q|\upsilon) = \frac{\pi(\upsilon|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon|q)\pi_0(q)dq}$$



Problem:

• Often requires high dimensional integration;

- ○ e.g., p = 6-23 for HIV model
- ○ p = hundreds to thousands for some models



Strategies:

• Sampling methods

• Sparse grid quadrature techniques

• Details to come in Saturday presentations by Daniela Calvetti, Andrea Arnold, Franz Hamilton.

Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006]

1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$

2. Construct covariance estimate V

3. For $k = 1, \cdots, M$

   (a) Construct candidate
   $$q^* \sim N(q^{k-1}, V)$$

   (b) Compute
   $$SS_{q^*} = \sum_{i=1}^{N} [\upsilon_i - f(t_i, q^*)]^2$$
   $$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

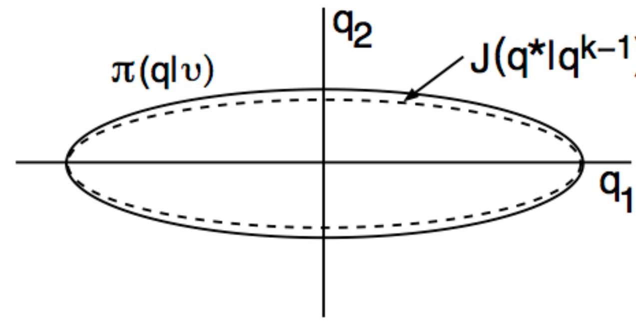
   (c) Compute
   $$\alpha(q^*|q^{k-1}) = \min\left(1, e^{-[SS_{q^*} - SS_{q^{k-1}}]/2\sigma^2}\right)$$

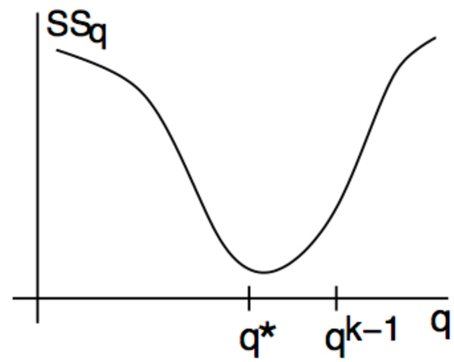
   (f) Accept q^* with probability α

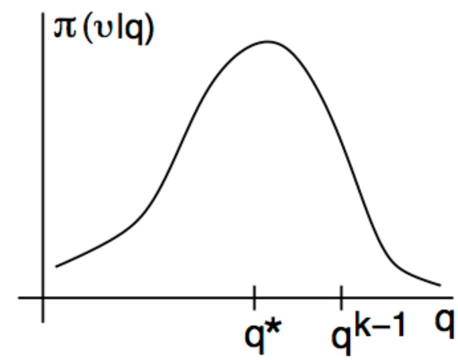
4. Update covariance as samples accepted

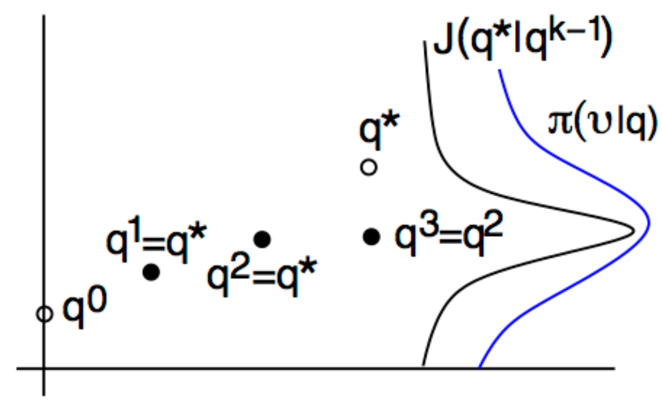
Proposal Distribution

$$J(q^*|q^{k-1}) = N(q^{k-1}, V)$$









Bayesian Model Calibration – HIV Example

HIV Model:

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon) k_1 V T_1$$

$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon) k_2 V T_2$$

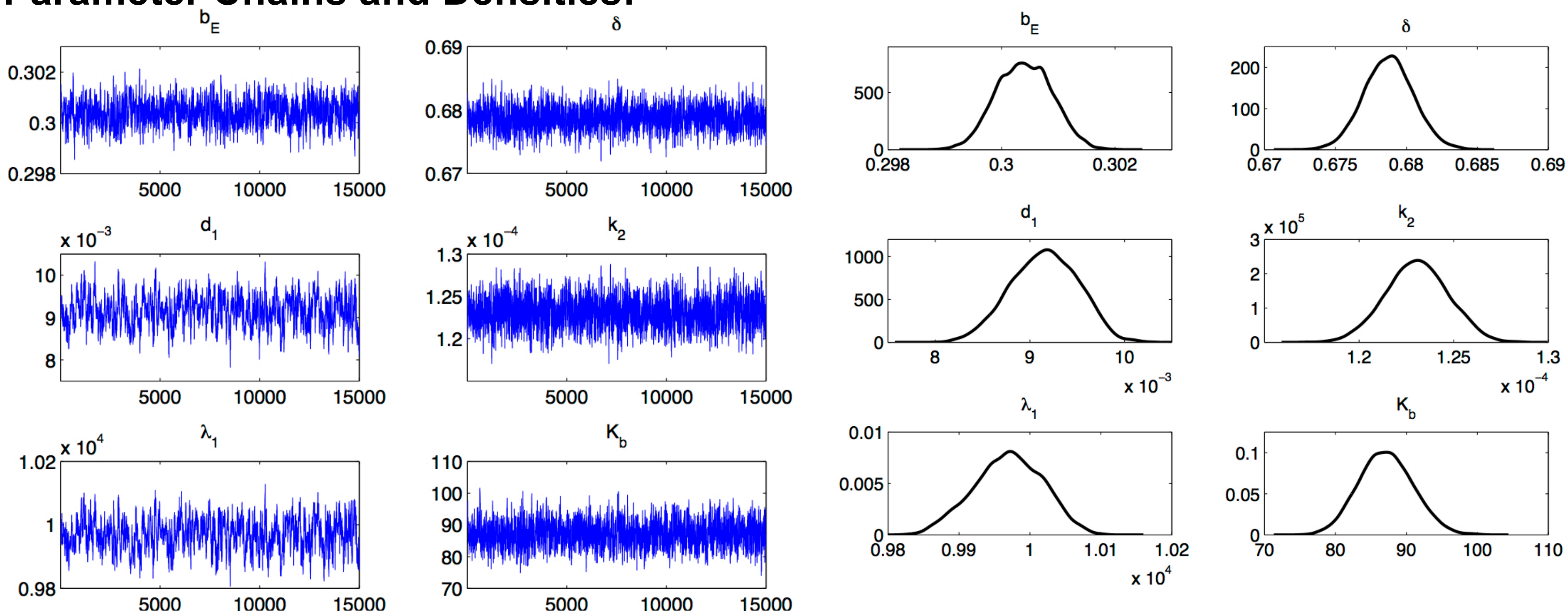
$$\dot{T}_1^* = (1-\varepsilon) k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon) k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2] V$$

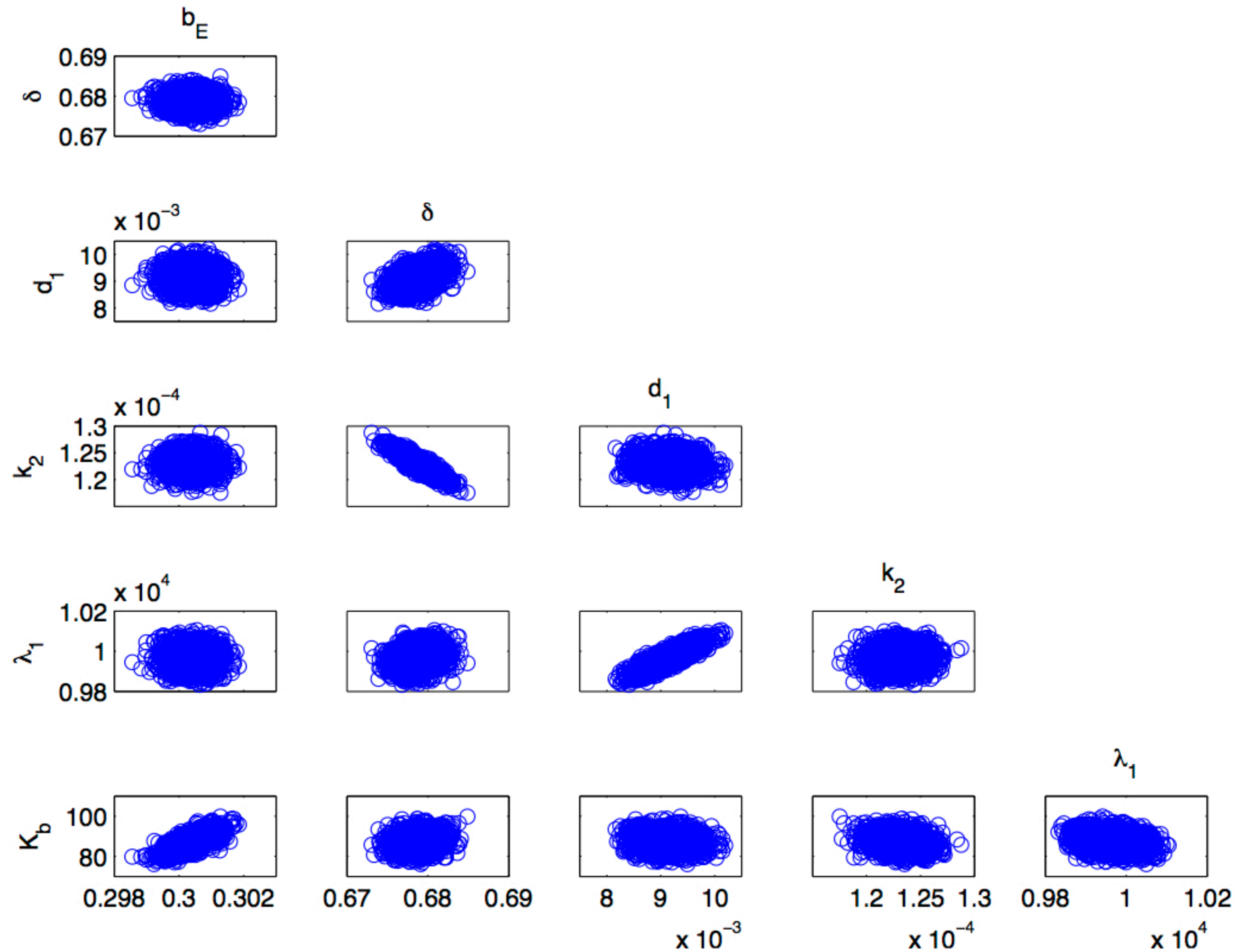
$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Parameter Chains and Densities:



Bayesian Model Calibration – HIV Example

Joint Sample Points:



Issues:

- Model parameters often correlated.
- Single-valued joint plots can indicate non-identifiable parameters (see talks by Marisa Eisenberg and Chris Durden).
- Bayesian methods feasible for non-identifiable parameters if prior is tight.

Propagation of Uncertainty in Models – HIV Example

HIV Example:

$$\dot{T}_1 = \lambda_1 - d_1 T_1 - (1-\varepsilon)k_1 V T_1$$

$$\dot{T}_2 = \lambda_2 - d_2 T_2 - (1-f\varepsilon)k_2 V T_2$$

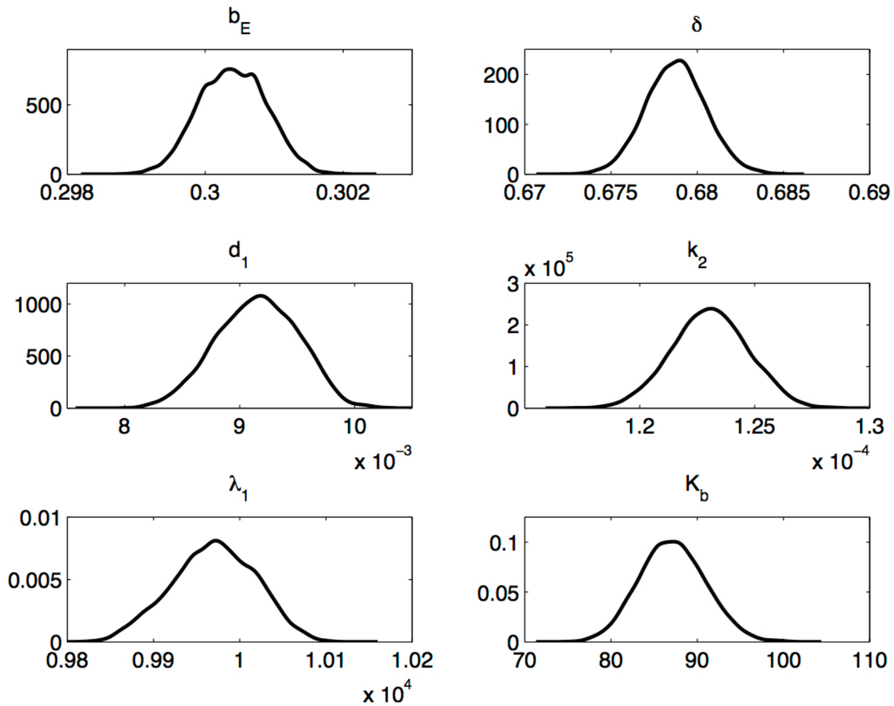
$$\dot{T}_1^* = (1-\varepsilon)k_1 V T_1 - \delta T_1^* - m_1 E T_1^*$$

$$\dot{T}_2^* = (1-f\varepsilon)k_2 V T_2 - \delta T_2^* - m_2 E T_2^*$$

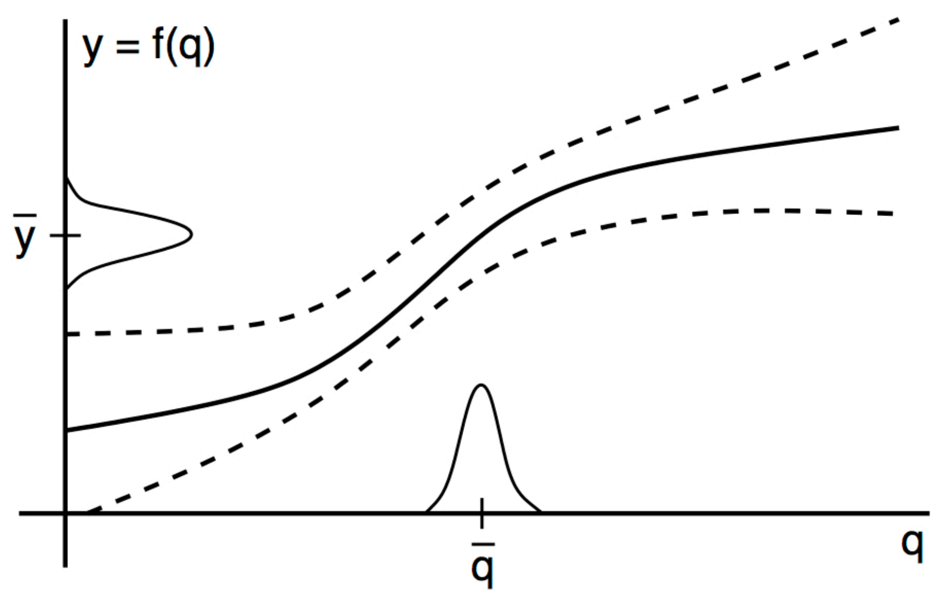
$$\dot{V} = N_T \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2]V$$

$$\dot{E} = \lambda_E + \frac{b_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Parameter Densities:



General Strategy



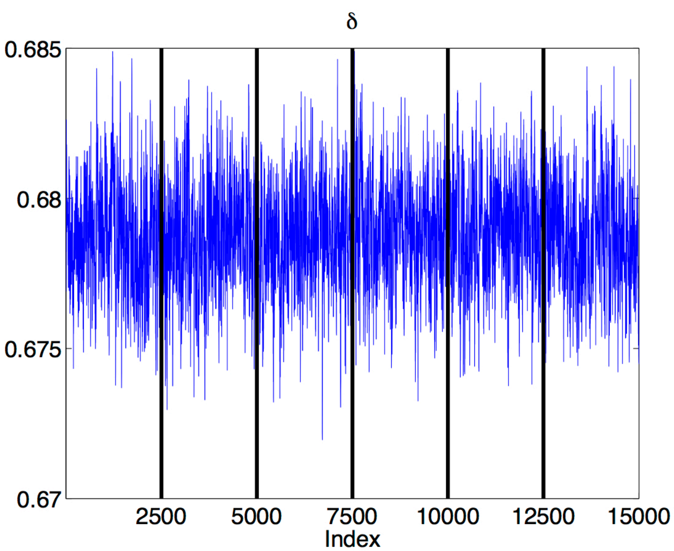
DRAM to Construct Prediction Intervals

Advantages:

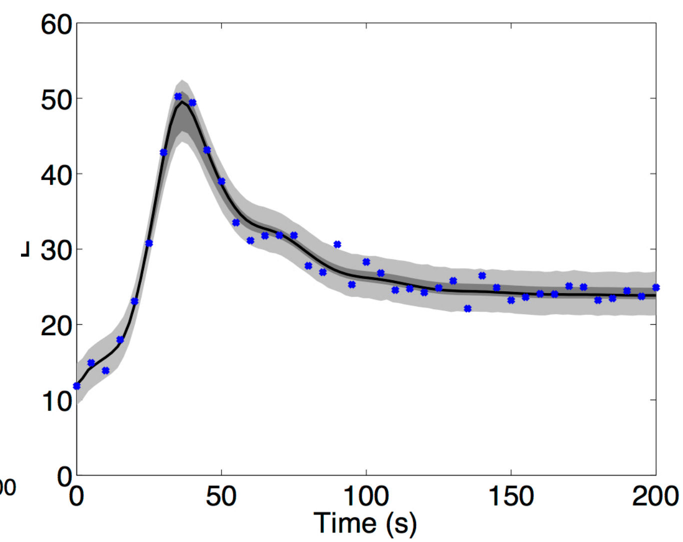
- No additional cost for DRAM if interpolating;
- One of few techniques for correlated parameters; e.g., Polynomial chaos requires mutually independent parameters or joint density.
- Does not require Gaussian densities;
- Incorporates both parameter and measurement uncertainties.

Disadvantages:

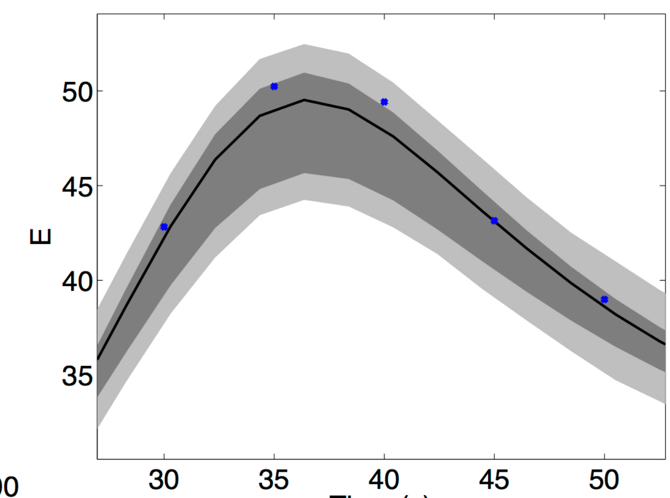
- Slow convergence rate $\mathcal{O}(1/\sqrt{M})$
- 100-fold more evaluations required to gain additional place of accuracy.
- Significant numerical analysis used to efficiently propagate densities.
- May require surrogate models.



Samples from Chain

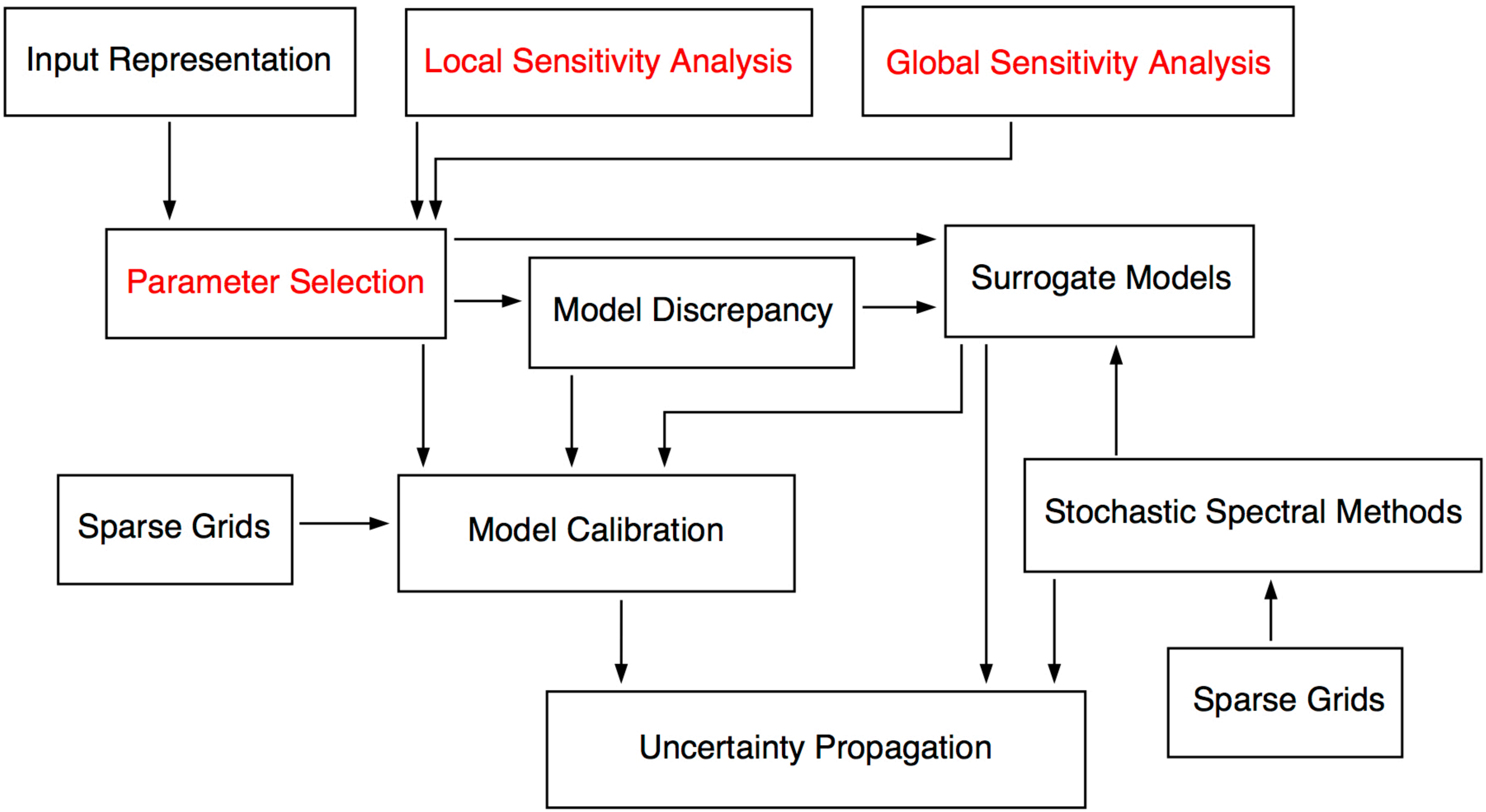


Data, Credible Intervals and Prediction Intervals



Non-Gaussian Credible and Prediction Intervals

Steps in Uncertainty Quantification



Parameter Selection: Required for models with unidentifiable or noninfluential inputs

- e.g., Nuclear neutron transport codes can have 100,000 inputs

Parameter Selection Techniques

First Issue: Parameters often not *identifiable* in the sense that they are uniquely determined by the data (See talks by Marisa Eisenberg and Chris Durdent).

Example: Simple harmonic oscillator

$$m\frac{d^2z}{dt^2} + c\frac{dz}{dt} + kz = f_0\cos(\omega_F t)$$

$$z(0) = z_0 \; , \; \frac{dz}{dt}(0) = z_1$$

Note: Parameter sets $q = [m, c, k, f_0]$ and $q = \left[1, \frac{c}{m}, \frac{k}{m}, \frac{f_0}{m}\right]$ yield same states

Solution: Reformulate problem as

$$\frac{d^2z}{dt^2} + C\frac{dz}{dt} + Kz = F_0\cos(\omega_F t)$$

$$z(0) = z_0 \; , \; \frac{dz}{dt}(0) = z_1$$

where $C = \frac{c}{m}, K = \frac{k}{m}$ and $F_0 = \frac{f_0}{m}$

Techniques:

• Linear algebra analysis;

○ e.g., SVD or QR algorithms

• Local sensitivity analysis

• Global sensitivity analysis

Second Issue: Nuclear neutronics problems can have 100,000 parameters but only 25-50 are influential.

Global Sensitivity Analysis

Example: Portfolio model

$$Y = c_1 Q_1 + c_2 Q_2$$

Note:

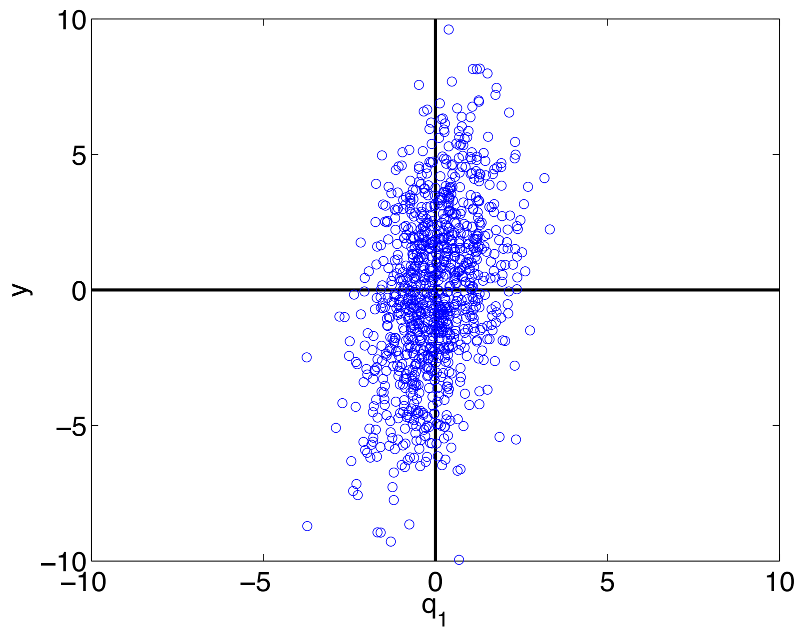
- Q_1 and Q_2 represent hedged porfolios
- c_1 and c_2 amounts invested in each portfolio

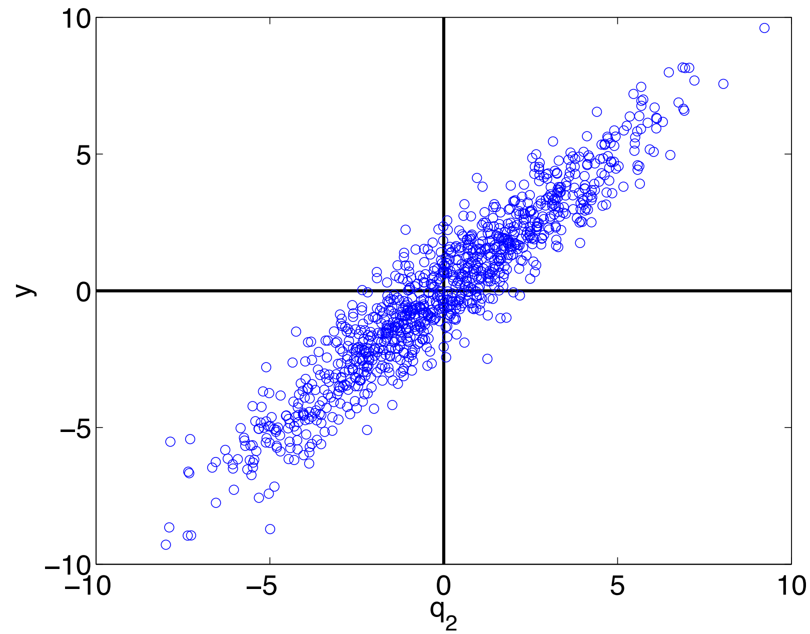
Take

$$c_1 = 2 \ , \ c_2 = 1$$

$Q_1 \sim N(0, \sigma_1^2)$ with $\sigma_1 = 1$

$Q_2 \sim N(0, \sigma_2^2)$ with $\sigma_2 = 3$





Local Sensitivities: Alun Lloyd

$$s_i \equiv \frac{\partial Y}{\partial Q_i} \Rightarrow s_1 = 2 > s_2 = 1$$

Solutions:

- Response correlation
- Variance methods
- Random sampling of local sensitivities

Variance-Based Methods

Sobol Representation: For now, take $Q_i \sim \mathcal{U}(0,1)$ and $\Gamma = [0,1]^p$

Take

$$f(q) = f_0 + \sum_{i=1}^{p} f_i(q_i) + \sum_{1 \le i < j \le p} f_{ij}(q_i, q_j)$$

Analogy: Taylor or Fourier series

subject to

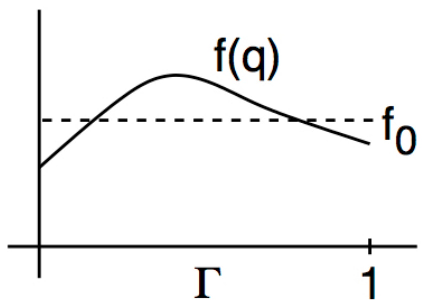
$$\int_0^1 f_i(q_i)dq_i = \int_0^1 f_{ij}(q_i,q_j)dq_i = \int_0^1 f_{ij}(q_i,q_j)dq_j = 0$$

Analogy:

- Derivatives for Taylor
- Orthogonality of sines and cosines for Fourier

Then

$$f_0 = \int_\Gamma f(q)dq$$



$$f_i(q_i) = \int_{\Gamma^{p-1}} f(q)dq_{\sim i} - f_0$$

Notation: $q_{\sim i} = [q_1, \cdots, q_{i-1}, q_{i+1}, \cdots, q_p]$

$$f_{ij}(q_i,q_j) = \int_{\Gamma^{p-2}} f(q)dq_{\sim\{ij\}} - f_i(q_i) - f_j(q_j) - f_0$$

Variance-Based Methods

Variances:

$$D_i = \int_0^1 f_i^2(q_i)dq_i$$

$$D_{ij} = \int_0^1 \int_0^1 f_{ij}^2(q_i, q_j)dq_i dq_j.$$

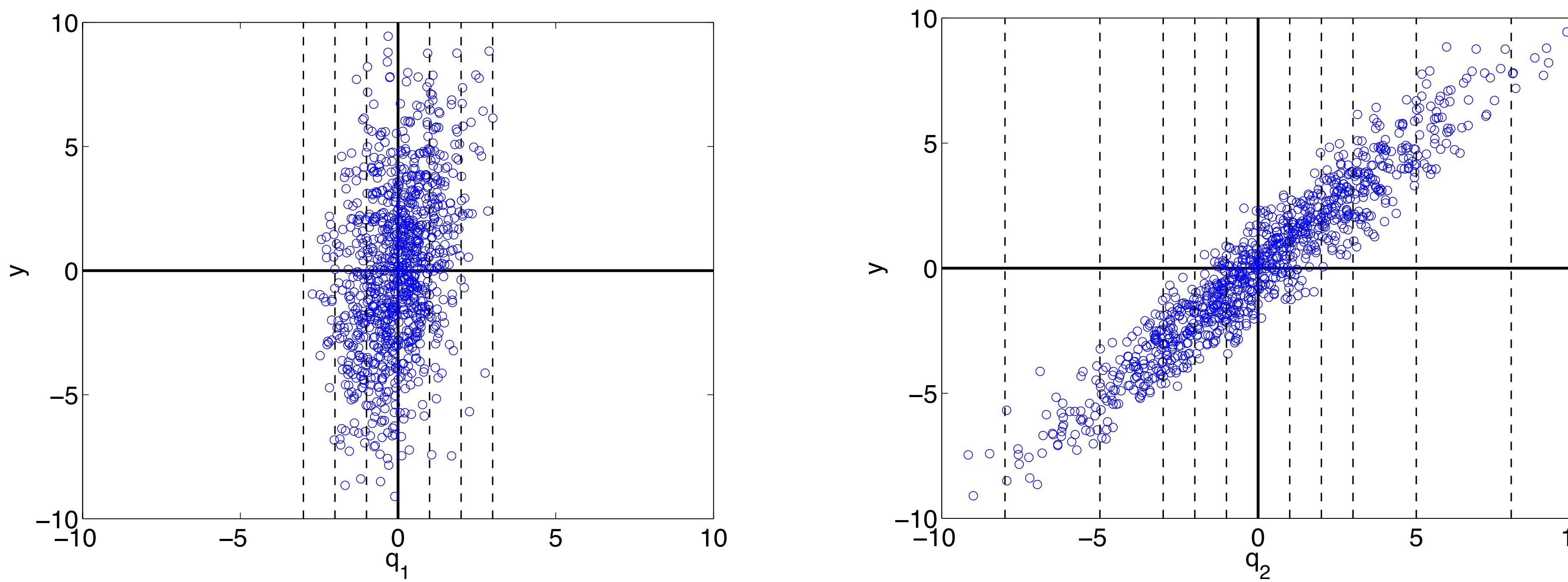
$$D = \mathrm{var}(Y) = \int_\Gamma f^2(q)dq - f_0^2$$

Sobol Indices:

$$S_i = \frac{D_i}{D} \quad , \; S_{ij} = \frac{D_{ij}}{D} \quad , \; i, j = 1, \cdots, p$$

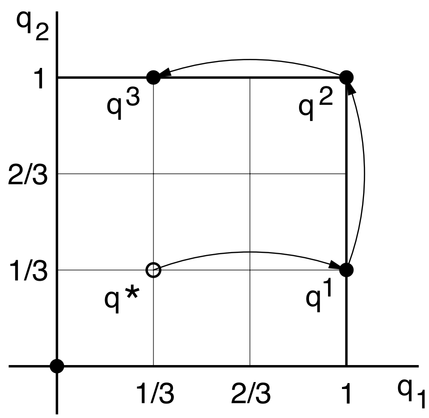
$$S_{T_i} = S_i + \sum_{j=1}^{p} S_{ij}$$

Statistical Interpretation: $D_i = \mathrm{var}[\mathbb{E}(Y|q_i)] \Rightarrow S_i = \frac{\mathrm{var}[\mathbb{E}(Y|q_i)]}{\mathrm{var}(Y)}$

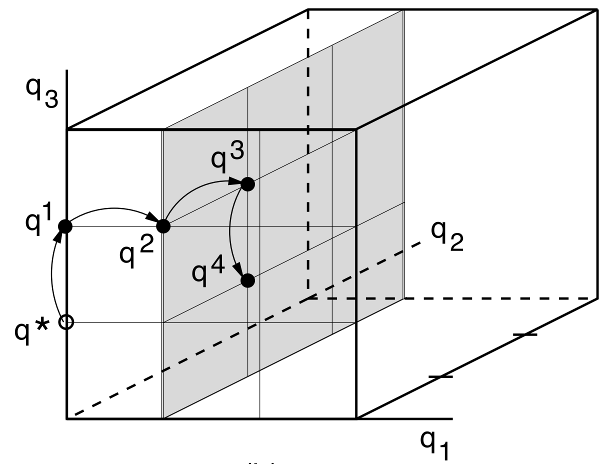


Morris Screening

Example: Consider uniformly distributed parameters on $\Gamma = [0,1]^p$



(a)



(b)

Elementary Effect:

$$d_i^j = \frac{f(q^j + \Delta e_i) - f(q^j)}{\Delta} \quad i^{th} \text{ parameter, } j^{th} \text{ sample}$$

$$\Delta \in \left\{ \frac{1}{\ell - 1}, \cdots, 1 - \frac{1}{\ell - 1} \right\} \quad \ell \text{ is level}$$

Global Sensitivity Measures: r samples

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^{r} |d_i^j(q)|$$

$$\sigma_i^2 = \frac{1}{r-1} \sum_{j=1}^{r} \left(d_i^j(q) - \mu_i \right)^2 \ , \ \mu_i = \frac{1}{r} \sum_{j=1}^{r} d_i^j(q)$$

SIR Disease Example

SIR Model:

$$\frac{dS}{dt} = \delta N - \delta S - \gamma k I S \quad , \ S(0) = S_0 \qquad \text{Susceptible}$$

$$\frac{dI}{dt} = \gamma k I S - (r + \delta) I \quad , \ I(0) = I_0 \qquad \text{Infectious}$$

$$\frac{dR}{dt} = rI - \delta R \quad , \ R(0) = R_0 \qquad \text{Recovered}$$

Note: Parameter set $q = [\gamma, k, r, \delta]$ is not identifiable

Assumed Parameter Distribution:

$$\gamma \sim \mathcal{U}(0,1) \ , \ k \sim \mathrm{Beta}(\alpha, \beta) \ , \ r \sim \mathcal{U}(0,1) \ , \ \delta \sim \mathcal{U}(0,1)$$

Infection Coefficient	Interaction Coefficient	Recovery Rate	Birth/death Rate

Response:

$$y = \int_0^5 R(t,q)\,dt$$

SIR Disease Example

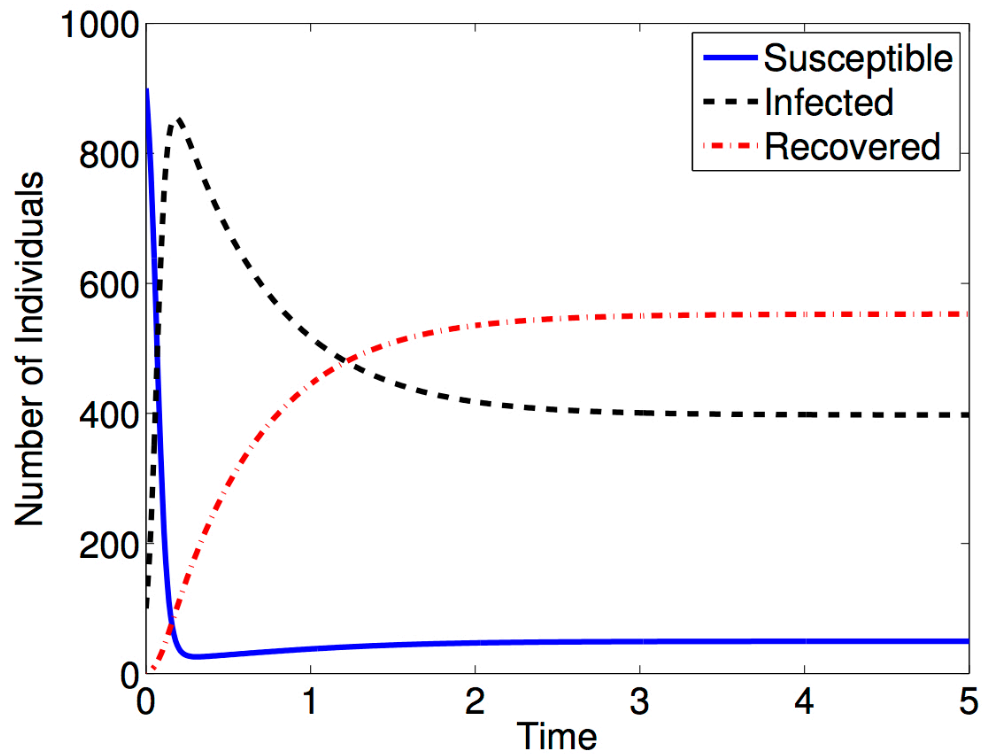
SIR Model:

$$\frac{dS}{dt} = \delta N - \delta S - \gamma k I S \quad , \; S(0) = S_0$$ Susceptible

$$\frac{dI}{dt} = \gamma k I S - (r + \delta) I \quad , \; I(0) = I_0$$ Infectious

$$\frac{dR}{dt} = rI - \delta R \quad , \; R(0) = R_0$$ Recovered

Typical Realization:

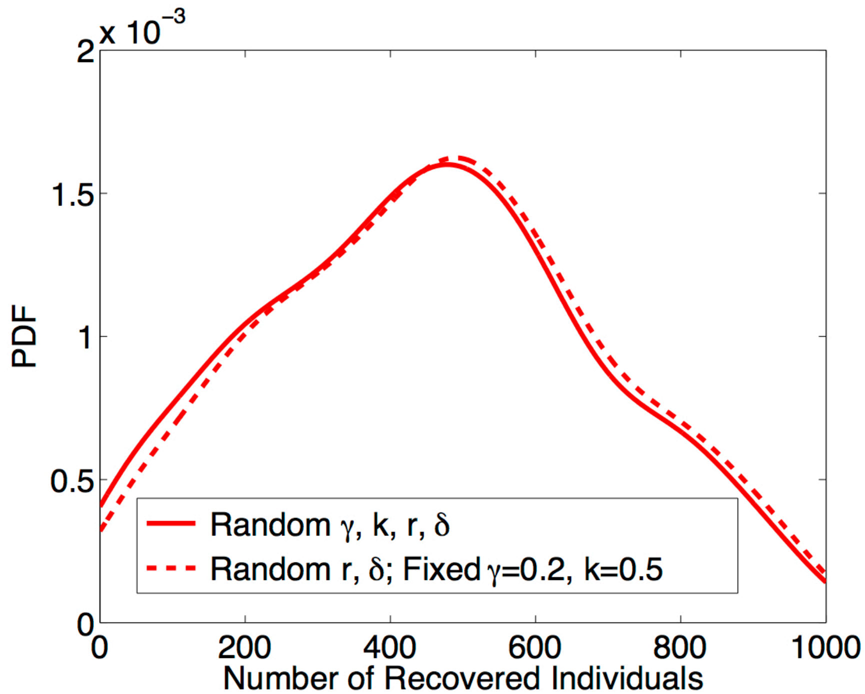


SIR Disease Example

Global Sensitivity Measures:

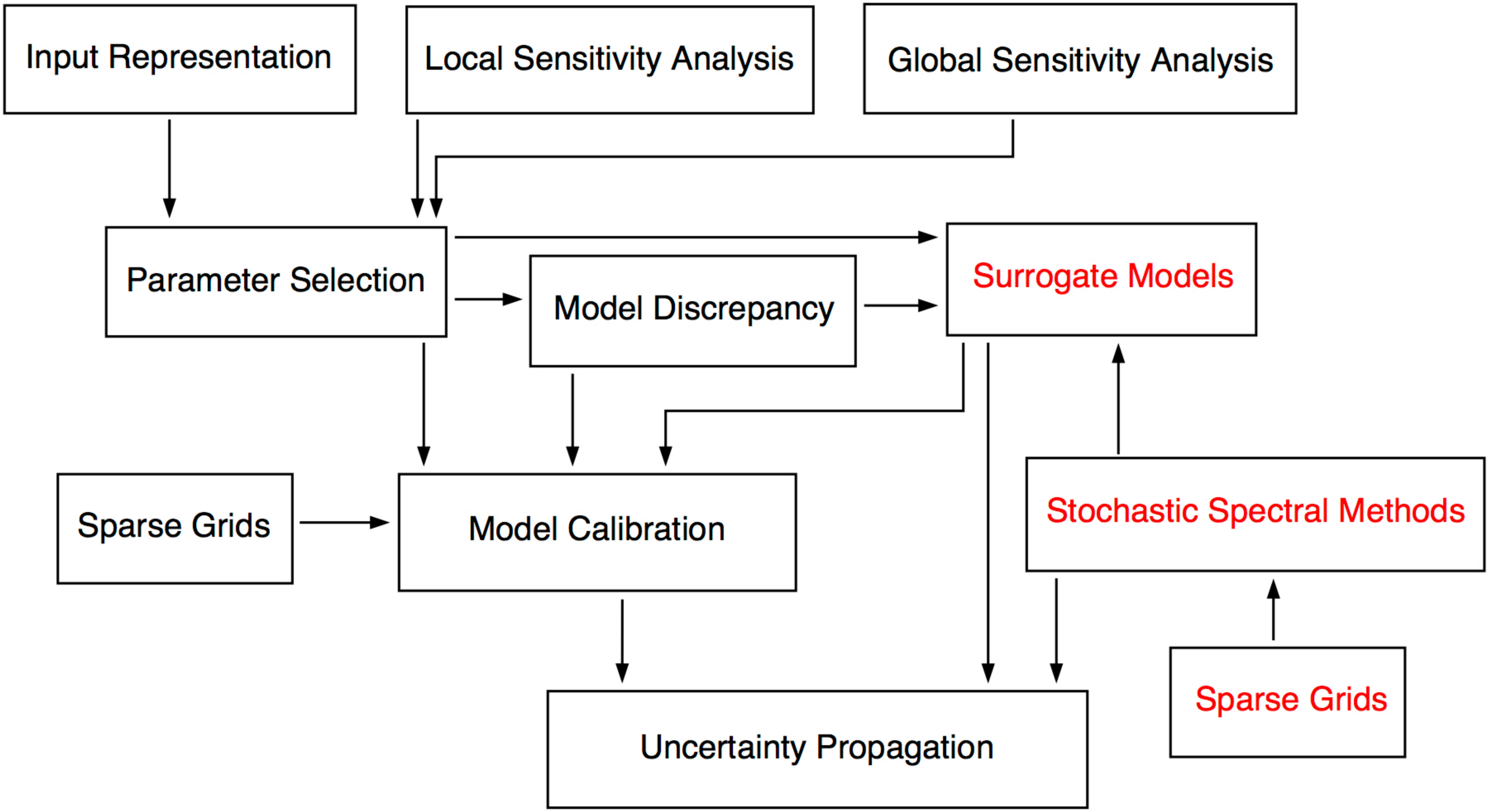
		γ	k	r	δ
	S_i	0.0997	0.0312	0.7901	0.1750
Sobol	S_{T_i}	-0.0637	-0.0541	0.5634	0.2029
	μ_i^* $(\times 10^3)$	0.2532	0.2812	2.0184	1.2328
Morris	σ_i $(\times 10^3)$	0.9539	1.6245	6.6748	3.9886

Result: Densities for $R(t_f)$ at $t_f = 5$



Note: Can fix non-influential parameters

Steps in Uncertainty Quantification



Surrogate Models: Similar goals and strategies as used for control implementation

- e.g., POD, Koopman representations

Surrogate Models

Problem: Difficult to obtain sufficient number of realizations of discretized PDE models for Bayesian model calibration and uncertainty propagation.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0$$

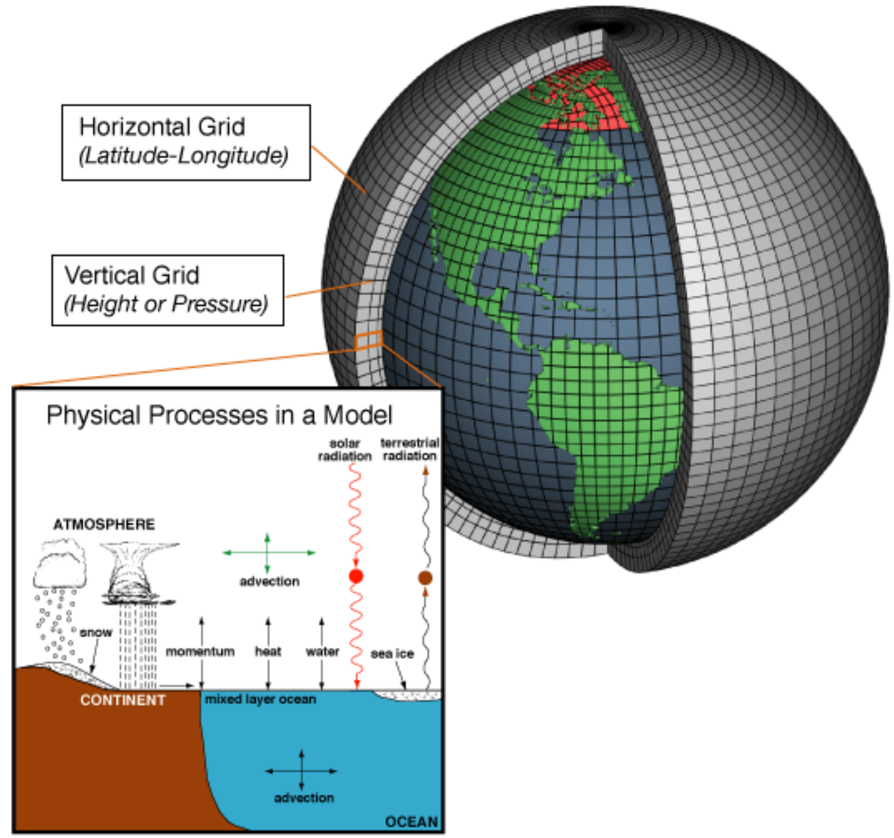
$$\frac{\partial v}{\partial t} = -v \cdot \nabla v - \frac{1}{\rho}\nabla p - g\hat{k} - 2\Omega \times v$$

$$\rho c_V \frac{\partial T}{\partial t} + p\nabla \cdot v = -\nabla \cdot F + \nabla \cdot (k\nabla T) + \rho \dot{q}(T, p, \rho)$$

$$p = \rho R T$$

$$\frac{\partial m_j}{\partial t} = -v \cdot \nabla m_j + S_{m_j}(T, m_j, \chi_j, \rho)\ ,\ j = 1, 2, 3,$$

$$\frac{\partial \chi_j}{\partial t} = -v \cdot \nabla \chi_j + S_{\chi_j}(T, \chi_j, \rho)\ ,\ j = 1, \cdots, J,$$



Solution: Construct surrogate models

- Also termed data-fit models, response surface models, emulators, meta-models
- Projection-based models often called reduced-order models

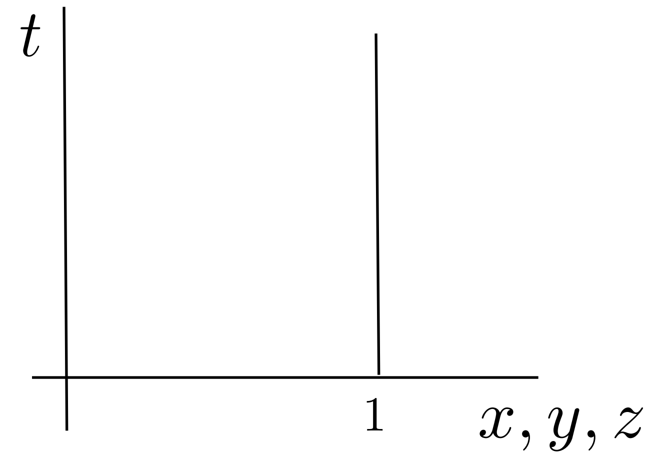
Surrogate Models: Motivation

Example: Consider the model

$$\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = f(q)$$

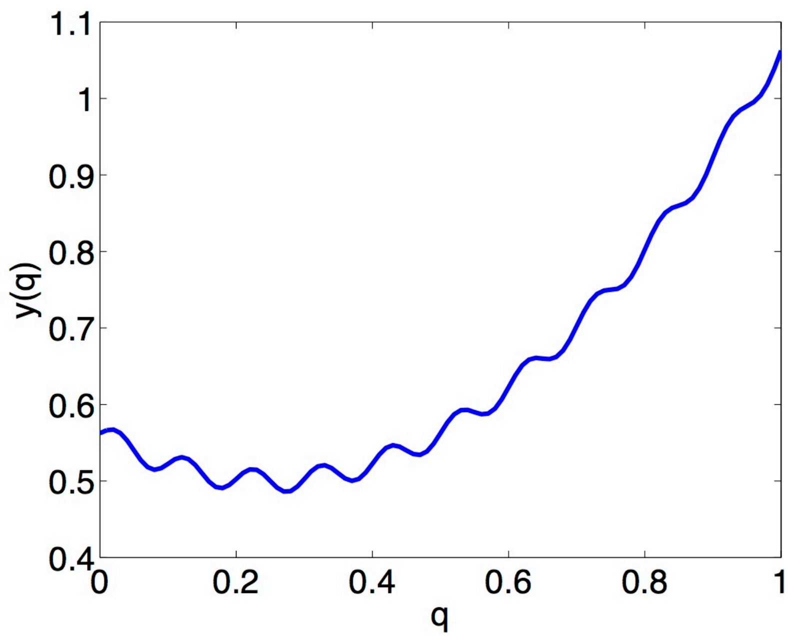
$$\text{Boundary Conditions}$$

$$\text{Initial Conditions}$$



with the response

$$y(q) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 u(t, x, y, z; q)\, dx dy dz dt$$



Notes:

- Requires approximation of PDE in 3-D
- What would be a simple surrogate?

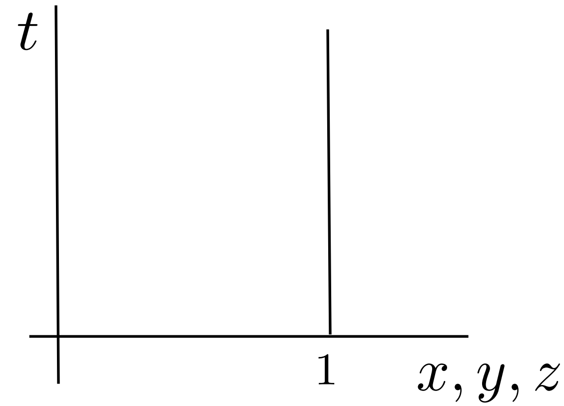
Surrogate Models

Example: Consider the model

$$\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = f(q)$$

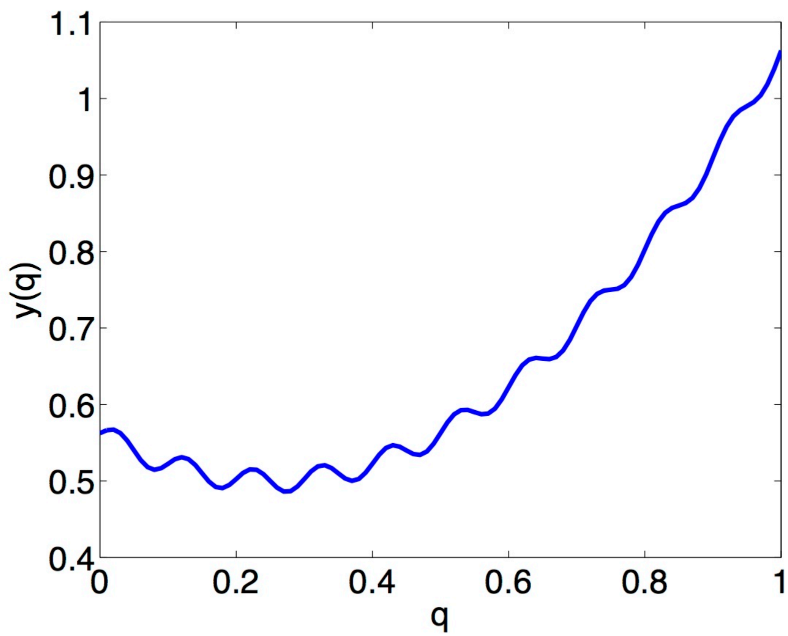
Boundary Conditions

Initial Conditions



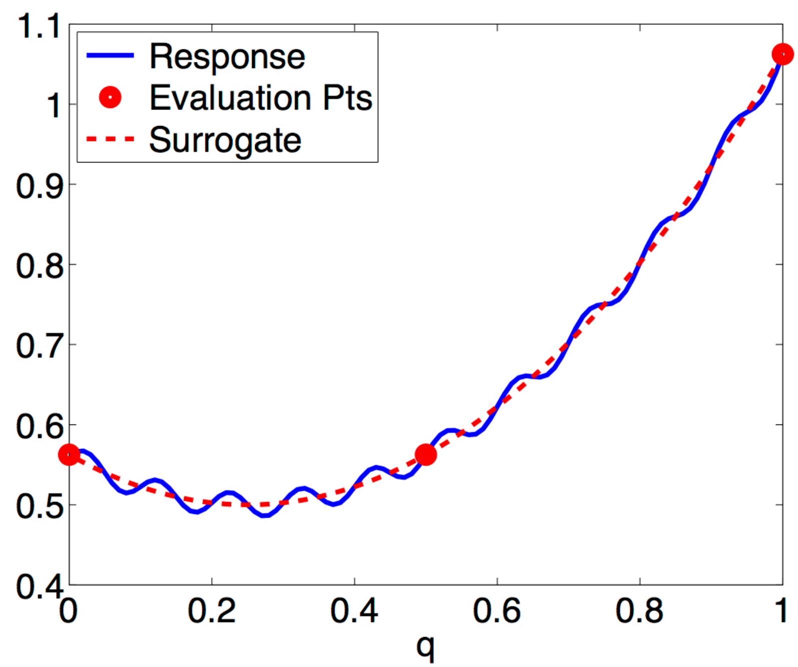
with the response

$$y(q) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 u(t,x,y,z;q)\,dx\,dy\,dz\,dt$$



Surrogate: Quadratic

$$y_s(q) = (q - 0.25)^2 + 0.5$$



Data-Fit Models

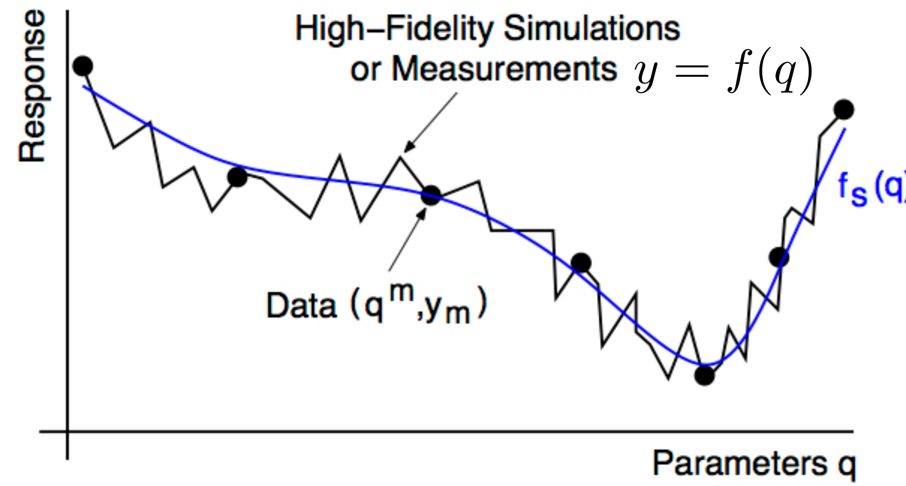
Notes:

• Often termed response surface models, emulators, meta-models;

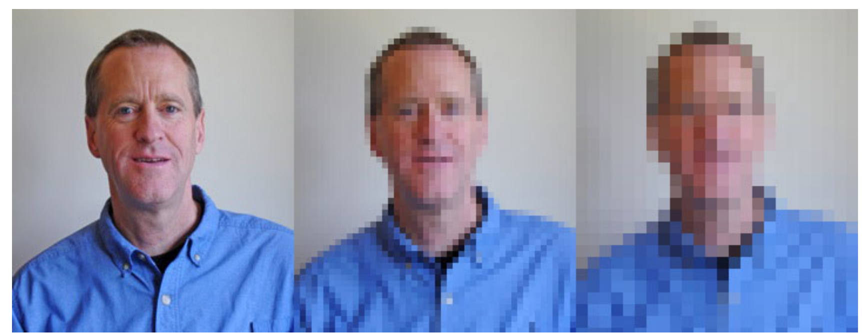
• Rely on interpolation or regression;

• Data can consist of high-fidelity simulations or experiments.

• Common techniques: polynomial models, kriging (Gaussian process regression), stochastic collocation.



Example:



Strategy: Consider high fidelity model

$$y = f(q)$$

with M model evaluations

$$y_m = f(q^m)\ ,\ m = 1, \cdots, M$$

Statistical Model: $f_s(q)$: Emulator for $f(q)$

$$y_m = f_s(q^m) + \varepsilon_m\ ,\ m = 1, \cdots, M$$

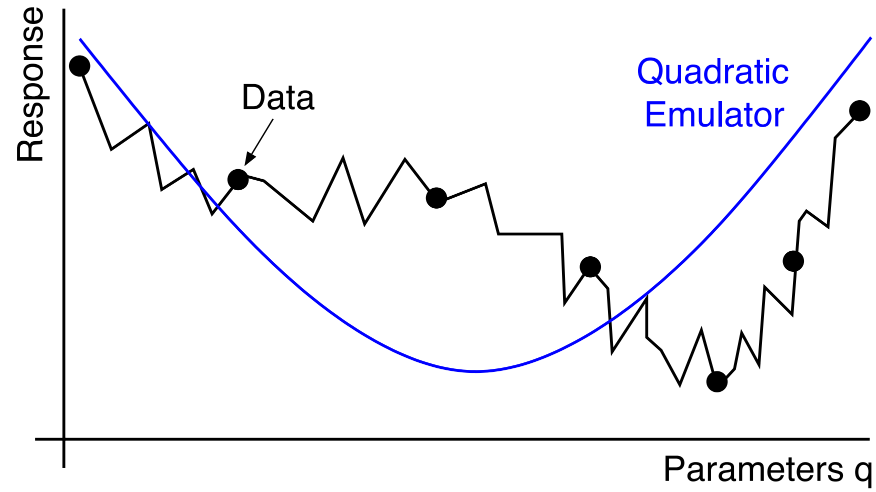
Data-Fit Models – Polynomial Emulator

Quadratic Emulator: Regression

$$f_s(q;\beta) = \beta_0 + \beta_1 q + \beta_2 q^2$$

Deterministic System: $y_{obs} = X\beta$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} 1 & q^1 & (q^1)^2 \\ \vdots & & \vdots \\ 1 & q^M & (q^M)^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$



Least Squares Estimate:

$$\beta = [X^T X]^{-1} X^T y_{obs}$$

Notes:

- Good choice for optimization;
- Accurate approximation may require high-order polynomials;
- Does not provide uncertainty bounds for uncertainty quantification.

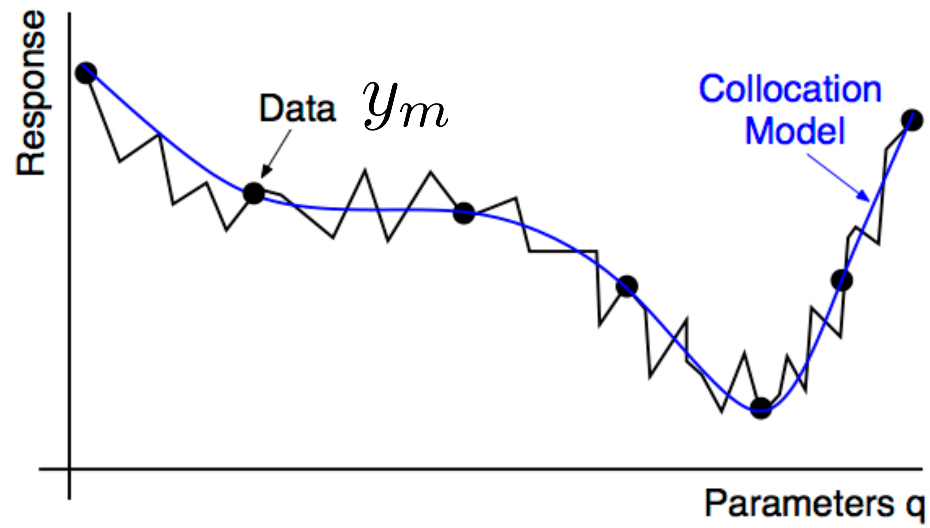
Data-Fit Models – Stochastic Collocation

Strategy: Consider high fidelity model

$$y = f(q)$$

with M model evaluations

$$y_m = f(q^m) \ , \ m = 1, \cdots, M$$



Collocation Surrogate:

$$y_s(q) = f_s(q) = \sum_{m=1}^{M} y_m L_m(q)$$

where $L_m(q)$ is a Lagrange polynomial, which in 1-D is represented by

$$L_m(q) = \prod_{\substack{j=0 \\ j \neq m}}^{M} \frac{q - q^j}{q^m - q^j} = \frac{(q - q^1) \cdots (q - q^{m-1})(q - q^{m+1}) \cdots (q - q^M)}{(q^m - q^1) \cdots (q^m - q^{m-1})(q^m - q^{m+1}) \cdots (q^m - q^M)}$$

Note:

$$L_m(q^j) = \delta_{jm} = \begin{cases} 0 & , \quad j \neq m \\ 1 & , \quad j = m \end{cases}$$

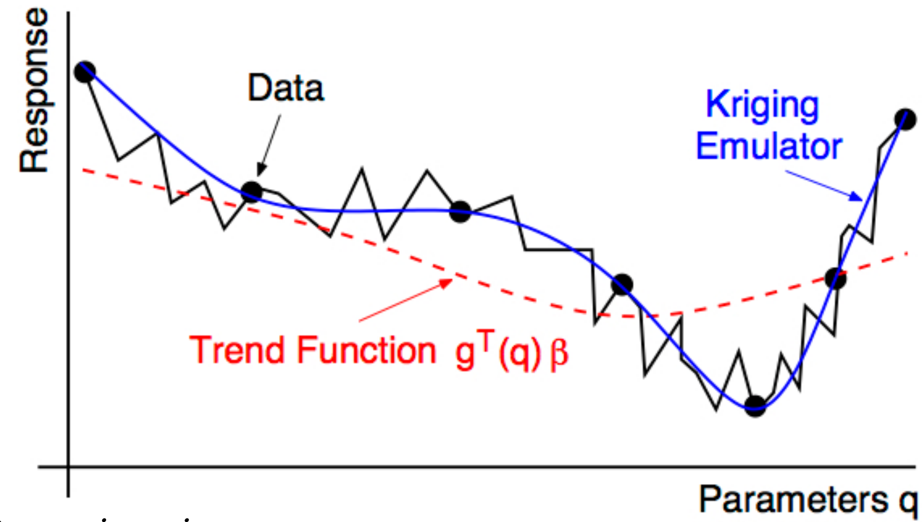
Result: $y_s(q^m) = f(q^m)$

Note: Method is nonintrusive and treats code as blackbox.

Data-Fit Models – Gaussian Process Emulator

Kriging (Gaussian Process):

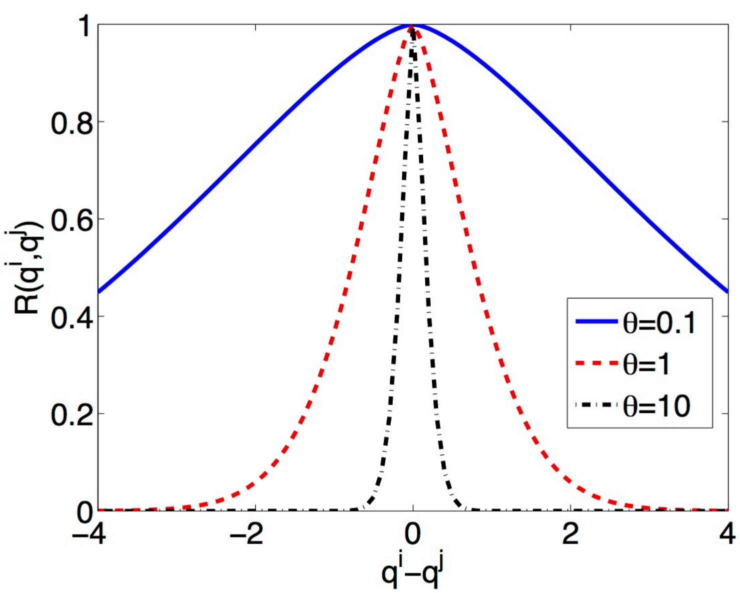
$$f_s(q;\beta) = g^T(q)\beta + Z(q)$$



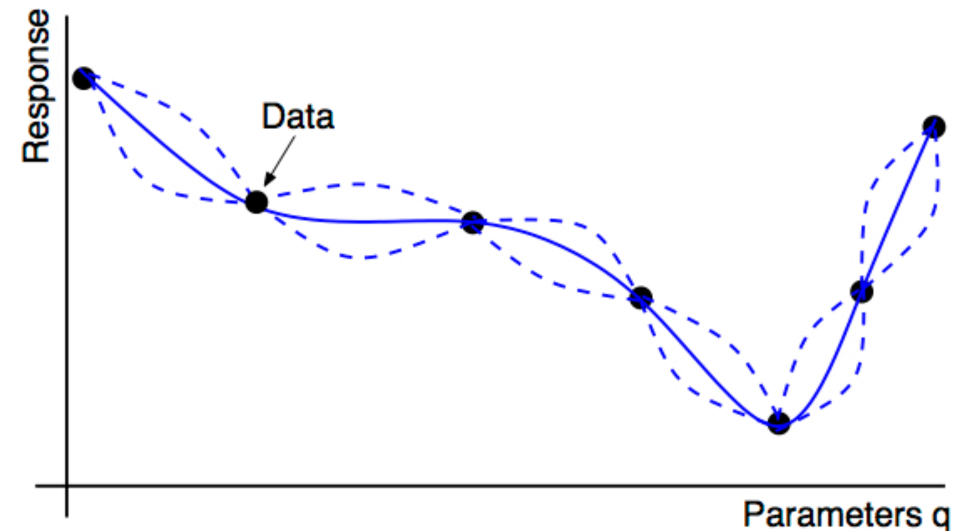
- $g^T(q)\beta$: Trend function
- $Z(q)$: Gaussian process with

$$\mathrm{cov}(Z(q^i), Z(q^j)] = \sigma^2 R(q^i,q^j) + \sigma_0^2\delta(q^i,q^j)$$

$$R(q^i,q^j) = \exp\left(-\sum_{k=1}^{p} |\theta_k(q_k^i - q_k^j)|^{\gamma_k}\right)$$



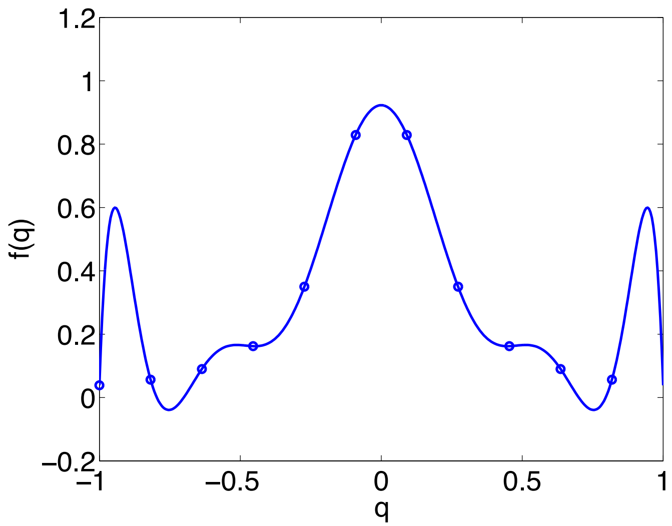
Error Bounds:



Surrogate Models – Grid Choice

Example: Consider the Runge function $f(q) = \dfrac{1}{1+25q^2}$ with points

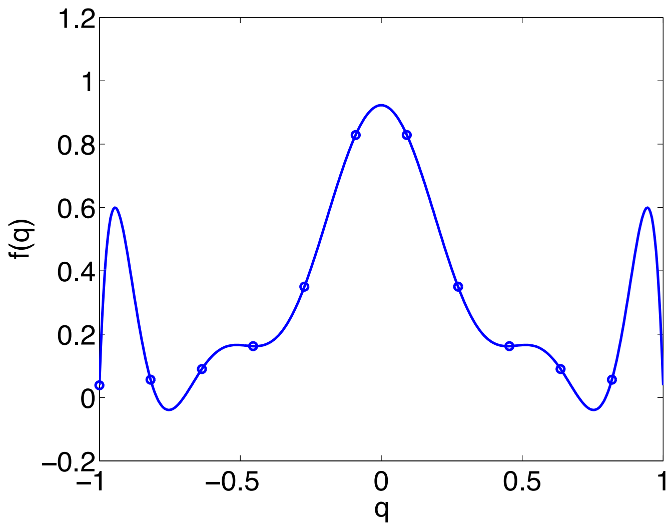
$$q^j = -1 + (j-1)\frac{2}{M_1}\ ,\ j = 1, \cdots, M_1 + 1$$

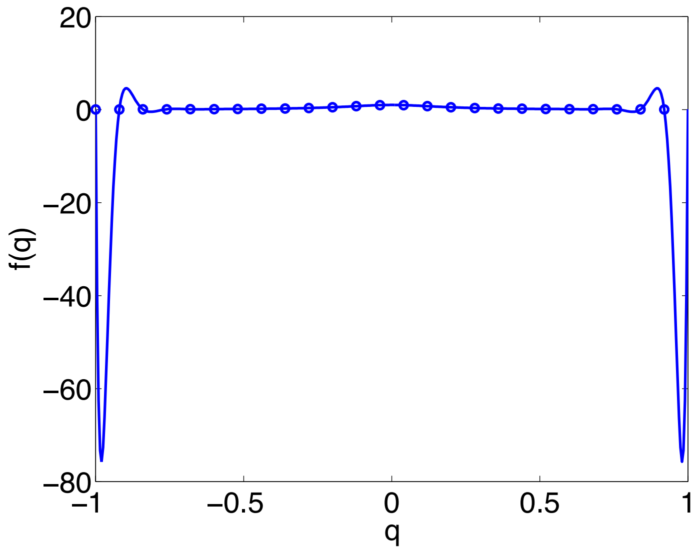


Surrogate Models – Grid Choice

Example: Consider the Runge function $f(q) = \dfrac{1}{1+25q^2}$ with points

$$q^j = -1 + (j-1)\frac{2}{M_1}\ ,\ j = 1, \cdots, M_1 + 1$$

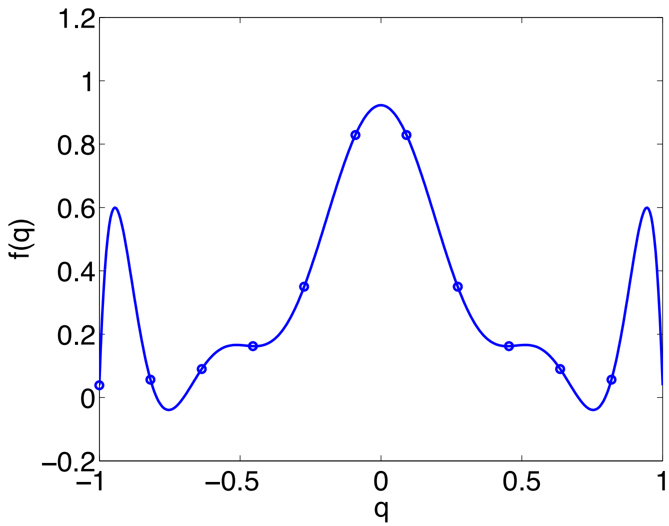


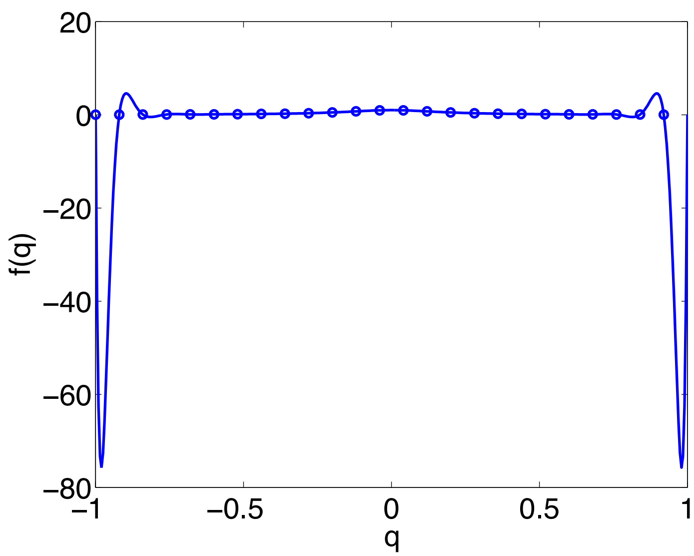


Surrogate Models – Grid Choice

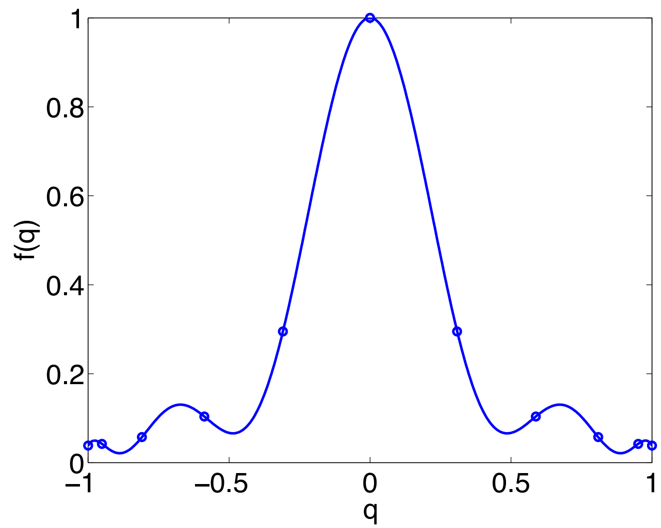
Example: Consider the Runge function $f(q) = \dfrac{1}{1+25q^2}$ with points

$$q^j = -1 + (j-1)\frac{2}{M_1} \ , \ j = 1, \cdots, M_1 + 1$$





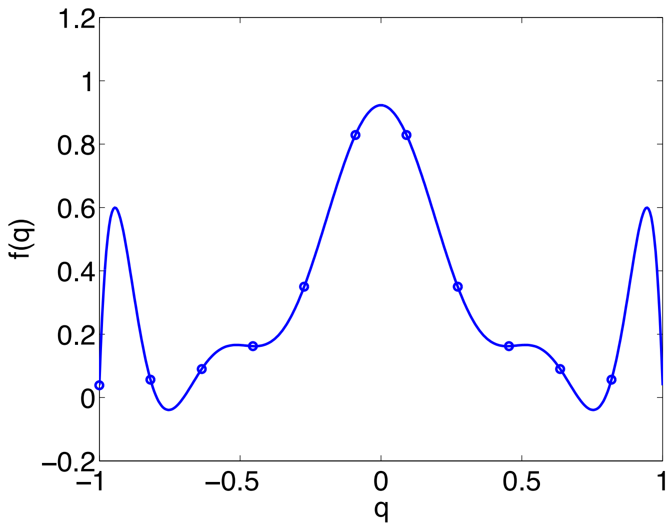
$$q^j = -\cos\frac{\pi(j-1)}{M_1 - 1} \ , \ j = 1, \cdots, M_1$$

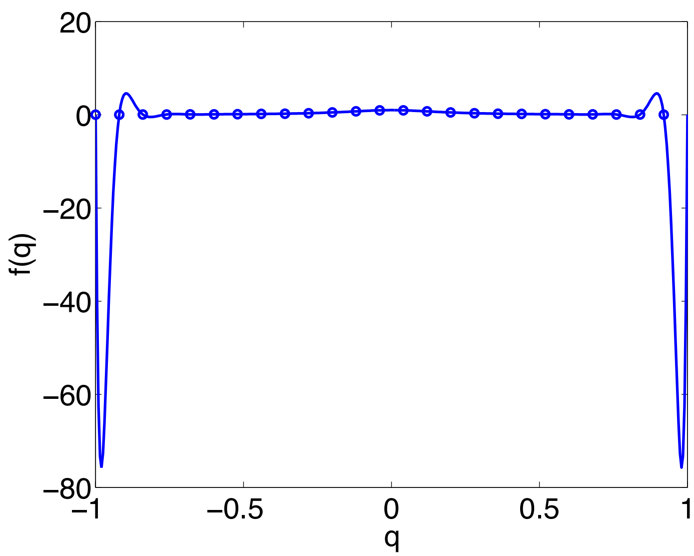


Surrogate Models – Grid Choice

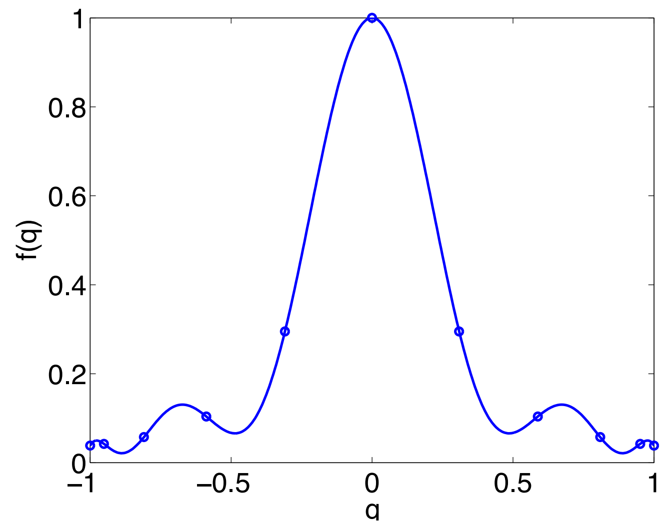
Example: Consider the Runge function $f(q) = \dfrac{1}{1+25q^2}$ with points

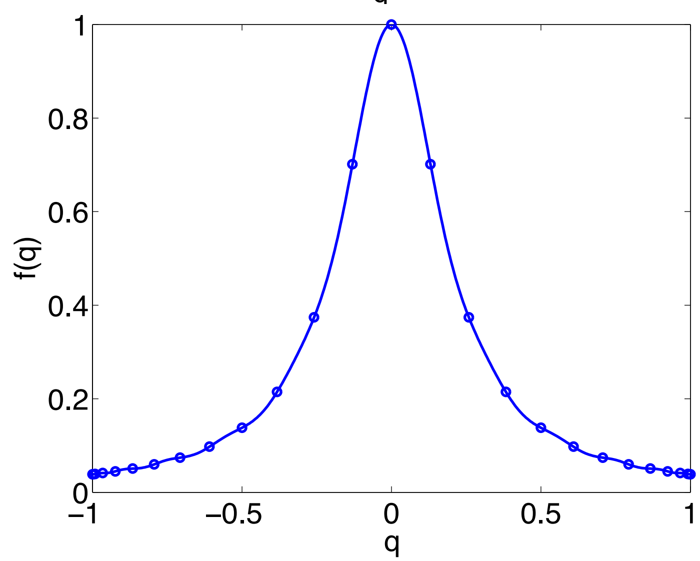
$$q^j = -1 + (j-1)\frac{2}{M_1} \ , \ j = 1, \cdots, M_1 + 1$$





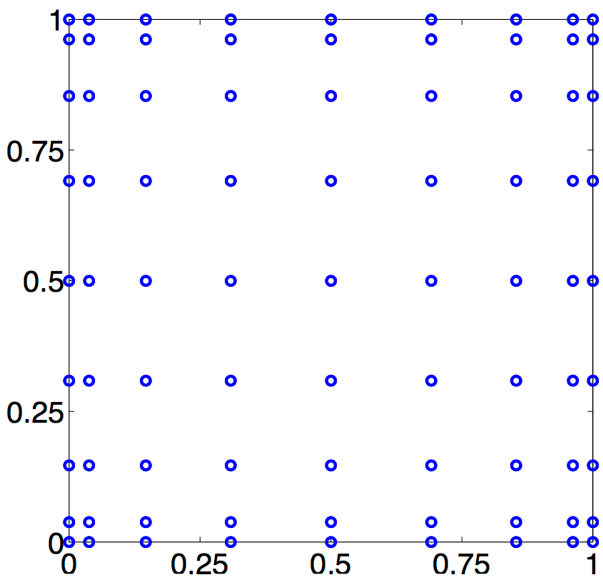
$$q^j = -\cos\frac{\pi(j-1)}{M_1 - 1} \ , \ j = 1, \cdots, M_1$$



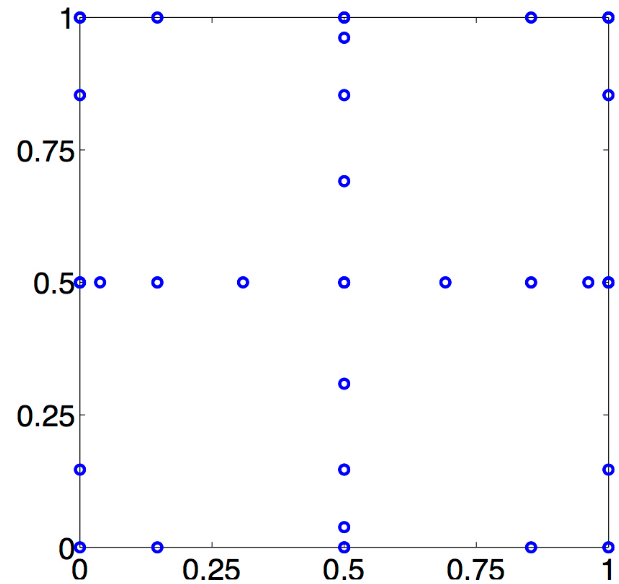


Sparse Grid Techniques

Tensored Grids: Exponential growth

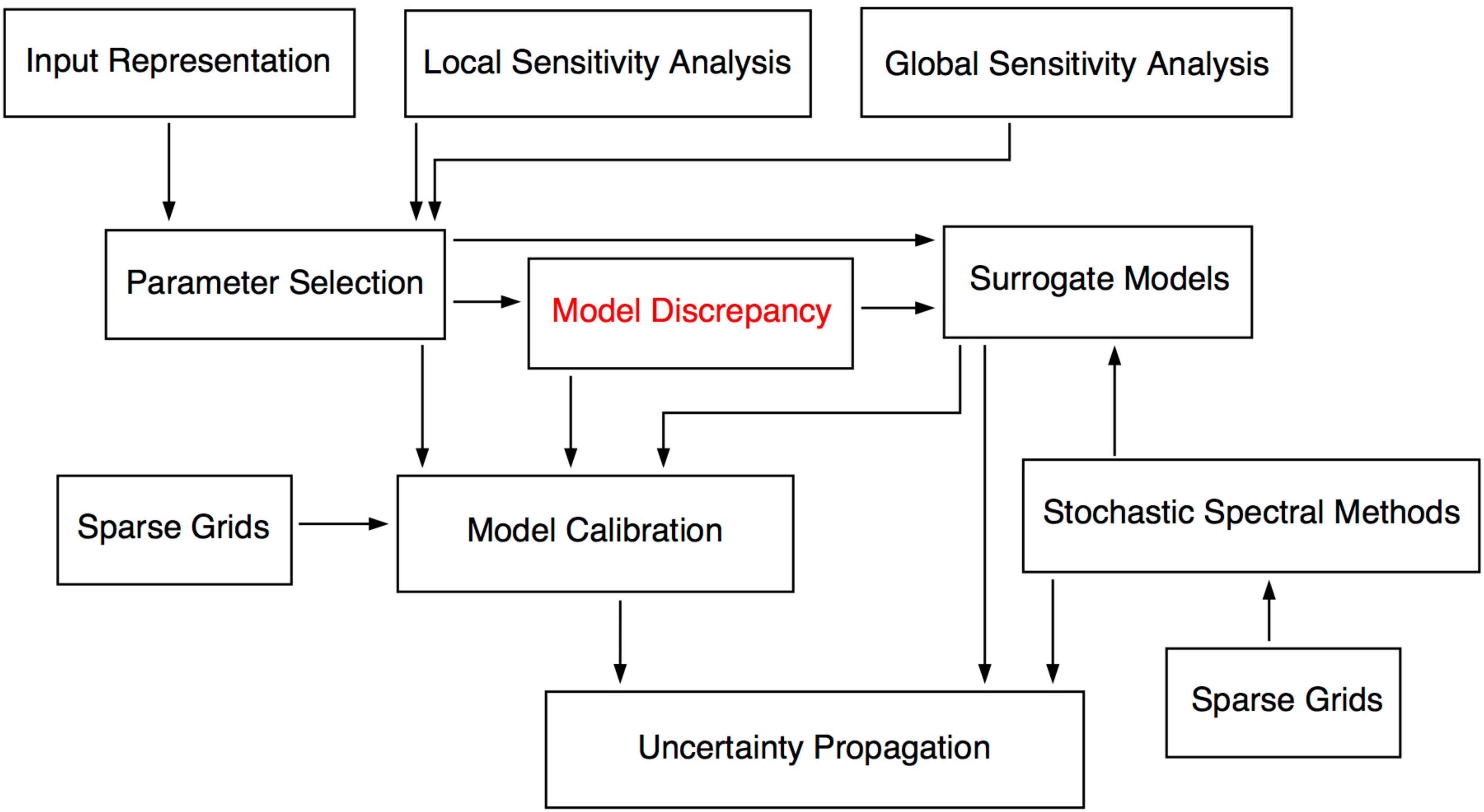


Sparse Grids: Same accuracy



p	R_ℓ	Sparse Grid $\mathcal{R}$	Tensored Grid $R = (R_\ell)^p$
2	9	29	81
5	9	241	59,049
10	9	1581	$> 3 \times 10^9$
50	9	171,901	$> 5 \times 10^{47}$
100	9	1,353,801	$> 2 \times 10^{95}$

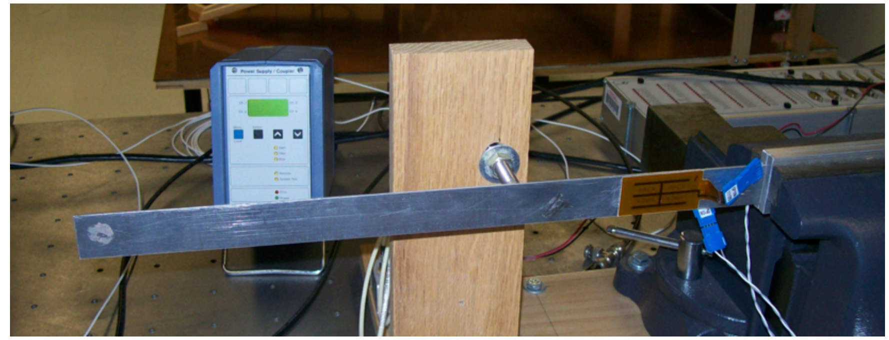
Steps in Uncertainty Quantification



Quantification of Model Discrepancy – Thin Beam

"Essentially all models are wrong, but some are useful" George E.P. Box

Example: Thin beam driven by PZT patches



Euler-Bernoulli Model: For all $\phi \in V$

$$\int_0^L \left[\rho(x)\frac{\partial^2 w}{\partial t^2} + \gamma\frac{\partial w}{\partial t}\right]\phi dx + \int_0^L \left[YI(x)\frac{\partial^2 w}{\partial x^2} + cI(x)\frac{\partial^3 w}{\partial x^2 \partial t}\right]\phi'' dx$$

$$= k_p V(t)\int_{x_1}^{x_2} \phi'' dx$$

with

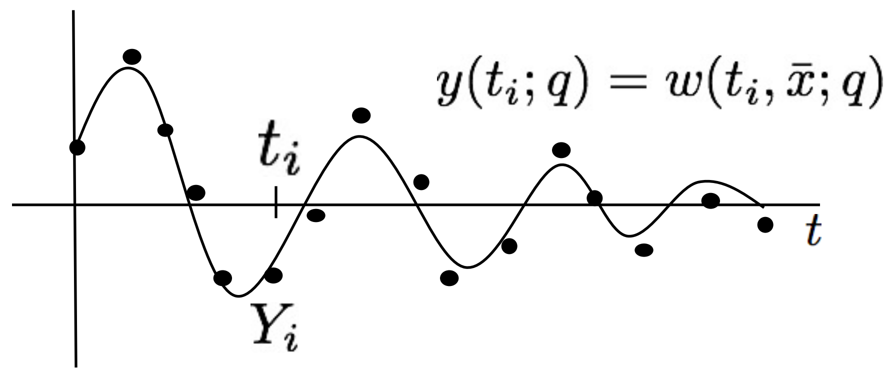
$$\rho(x) = \rho hb + \rho_p h_p b_p \chi_p(x)\ ,\ YI(x) = YI + Y_p I_p \chi_p(x)$$

$$cI(x) = cI + c_p I_p \chi_p(x)$$

Note: 7 parameters, 32 states

Statistical Model:

$$Y_i = y(t_i;q) + \delta(t_i) + \varepsilon_i$$

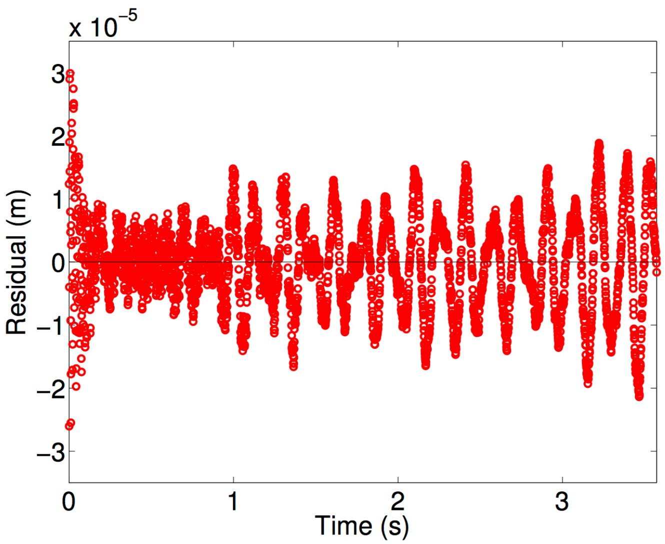


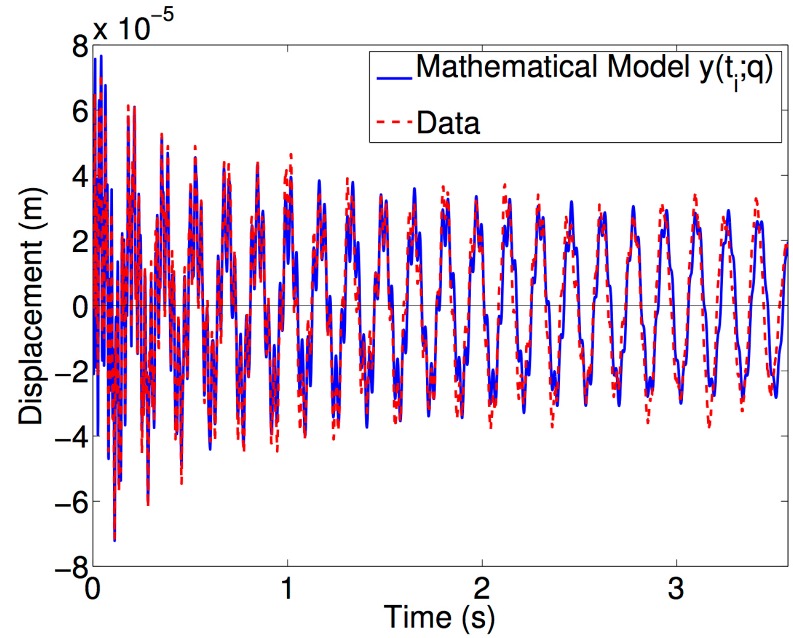
Quantification of Model Discrepancy – Thin Beam

Example: Good model fit

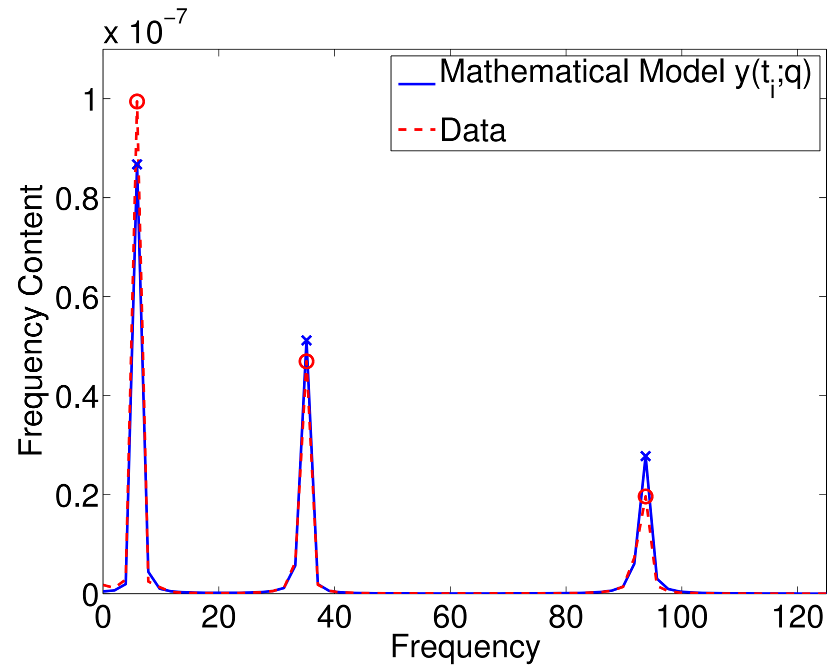
$$Y_i = y(t_i;q) + \delta(t_i) + \varepsilon_i$$

Problem: Measurement errors not iid





Model Fit to Data

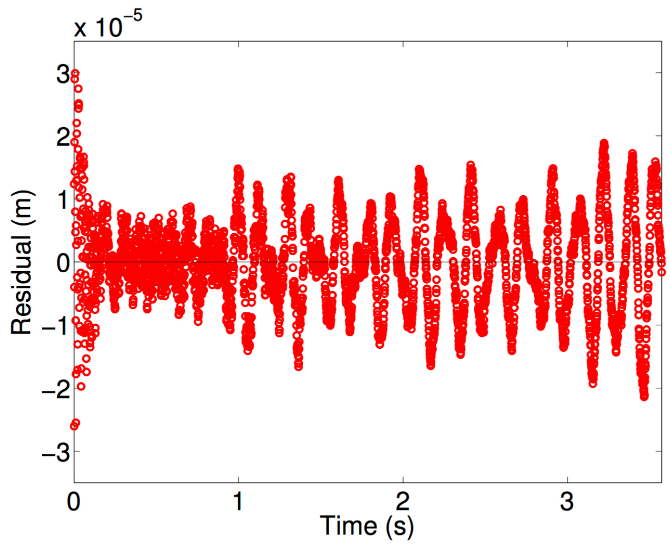


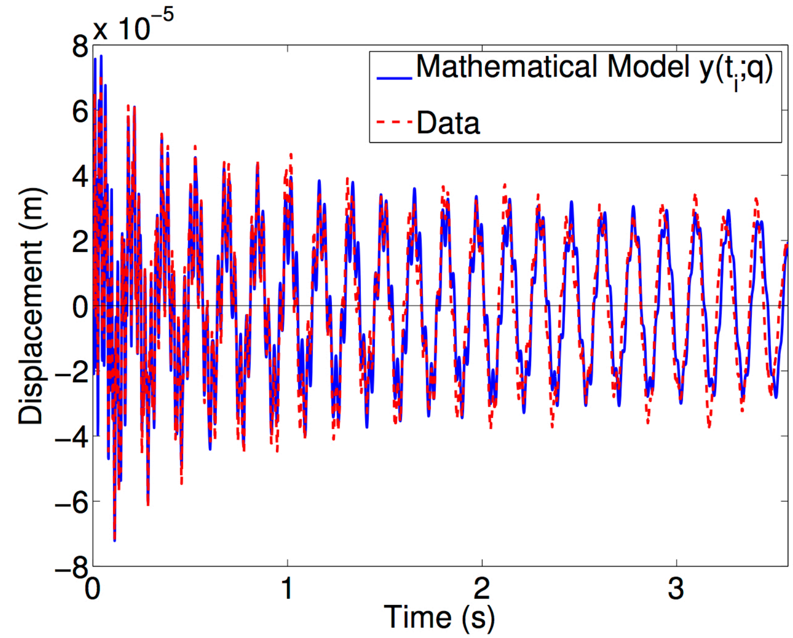
Quantification of Model Discrepancy – Thin Beam

Example: Good model fit

$$Y_i = y(t_i;q) + \delta(t_i) + \varepsilon_i$$

Problem: Measurement errors not iid



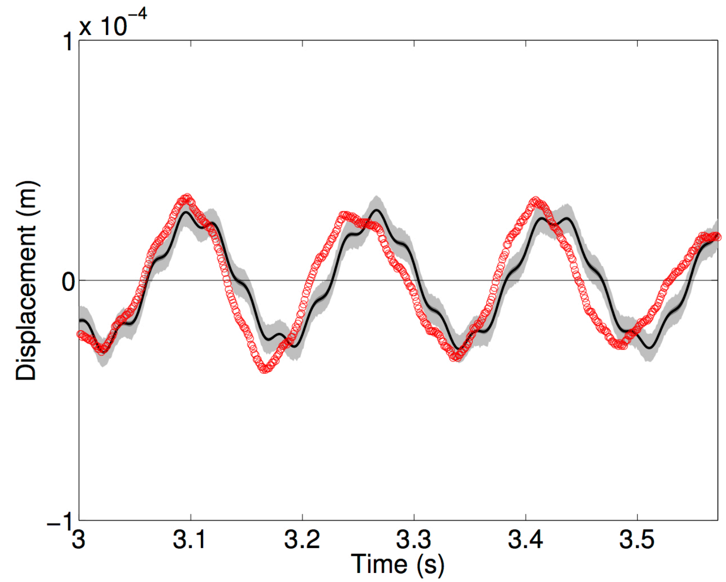


Model Fit to Data

Result: Prediction intervals wrong

Approaches:

- GP Model: Inaccurate for extrapolation
- Control-based approaches
- Illustrate first for heat example
- Return to beam in a bit



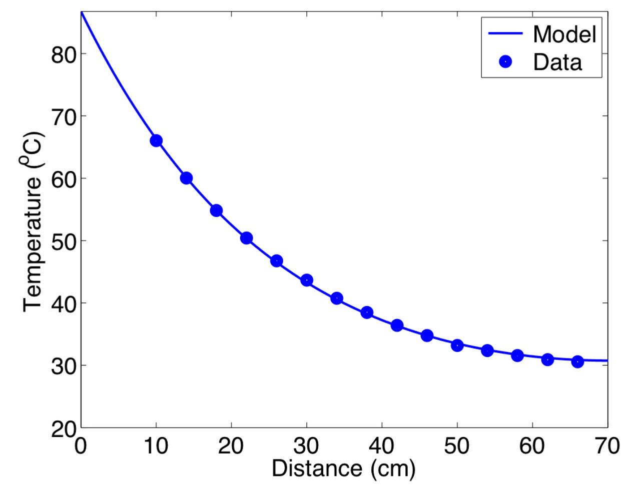
Prediction Intervals and Data

Quantification of Model Discrepancy – Heat Equation

Example: Steady state heat model

$$\frac{d^2T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left[T_s(x) - T_{amb}\right]$$

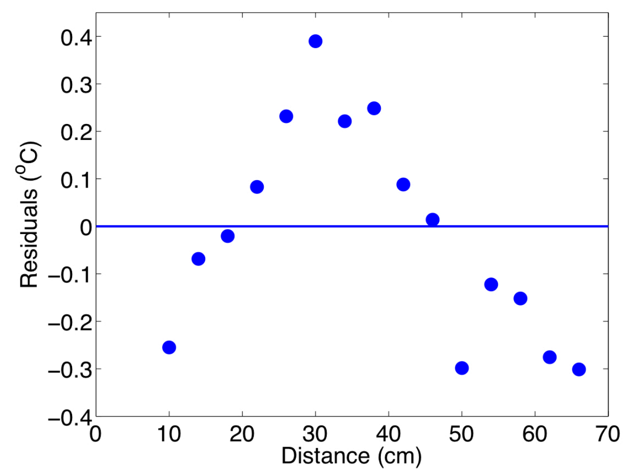
$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k} \quad , \quad \frac{dT_s}{dx}(L) = \frac{h}{k}[T_{amb} - T_s(L)]$$



Problem: Correlated residuals

Solution: Consider statistical model

$$Y_i = T_s(x_i;q) + \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$

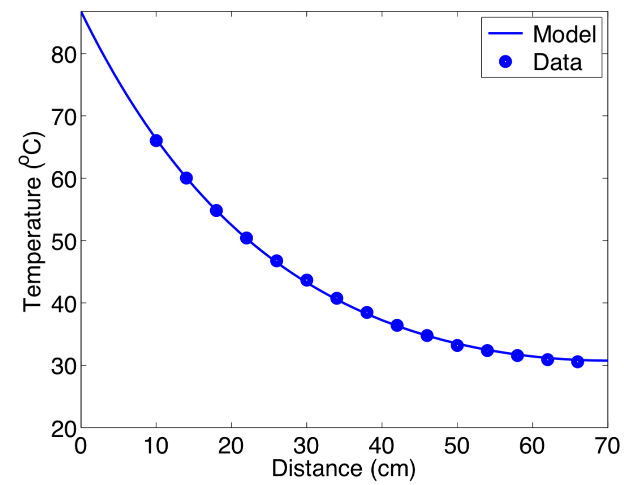


Quantification of Model Discrepancy – Heat Equation

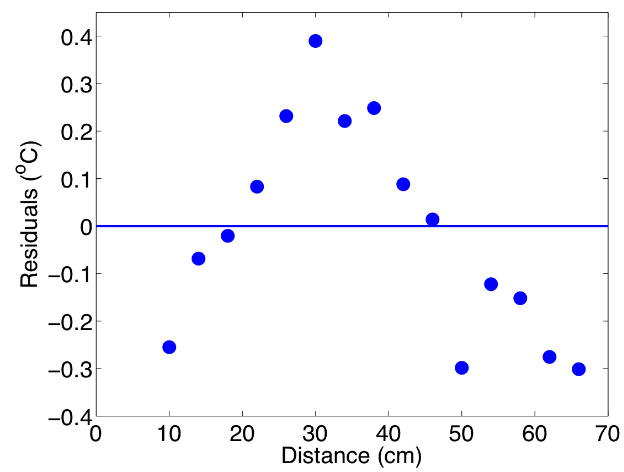
Example: Steady state heat model

$$\frac{d^2T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left[T_s(x) - T_{amb}\right]$$

$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k} \quad , \quad \frac{dT_s}{dx}(L) = \frac{h}{k}\left[T_{amb} - T_s(L)\right]$$

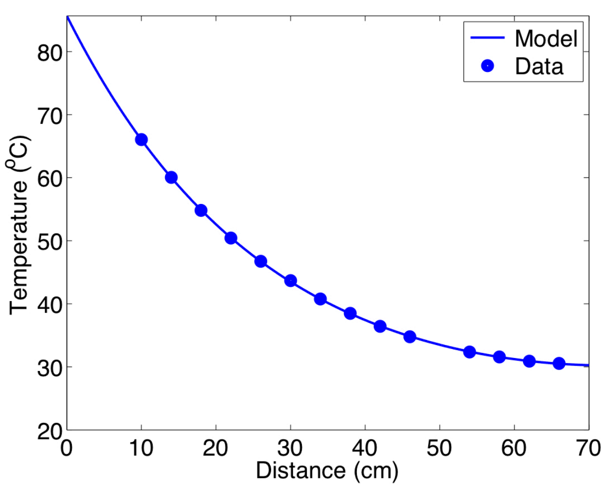


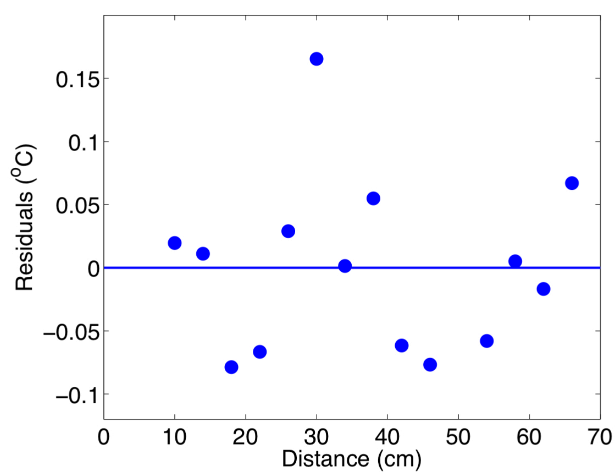
Problem: Correlated residuals



Solution: Consider statistical model

$$Y_i = T_s(x_i;q) + \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$$



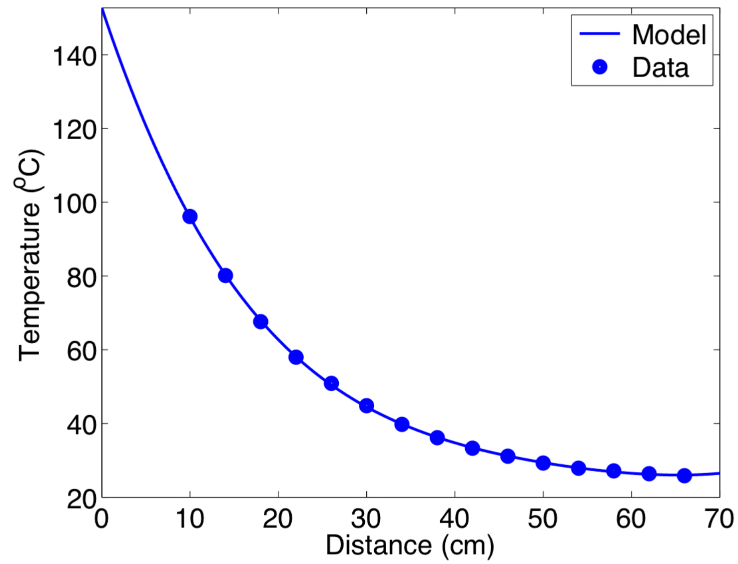


Quantification of Model Discrepancy – Heat Equation

Example: Steady state heat model

$$\frac{d^2T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left[T_s(x) - T_{amb}\right]$$

$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k} \quad , \quad \frac{dT_s}{dx}(L) = \frac{h}{k}[T_{amb} - T_s(L)]$$



Issue: Confounding between physical and phenomenological components

$$Y_i = T_s(x_i; q) + \delta(x_i) + \varepsilon_i$$

Example: Purely phenomenological

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \varepsilon_i$$

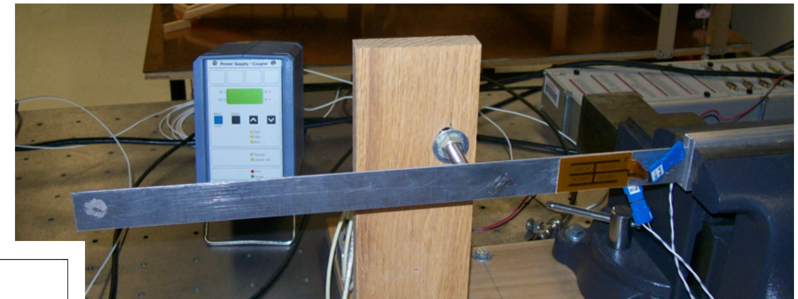
Result: Cannot provide extrapolatory predictions

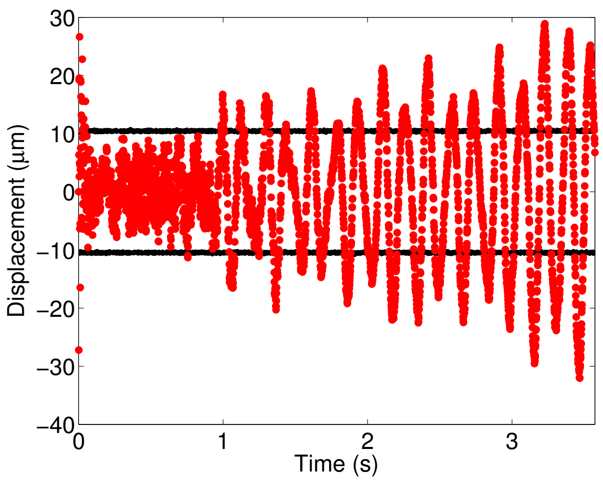
Conclusion: Must incorporate prior information about physical parameters and model discrepancy term.

Quantification of Model Discrepancy – Thin Beam

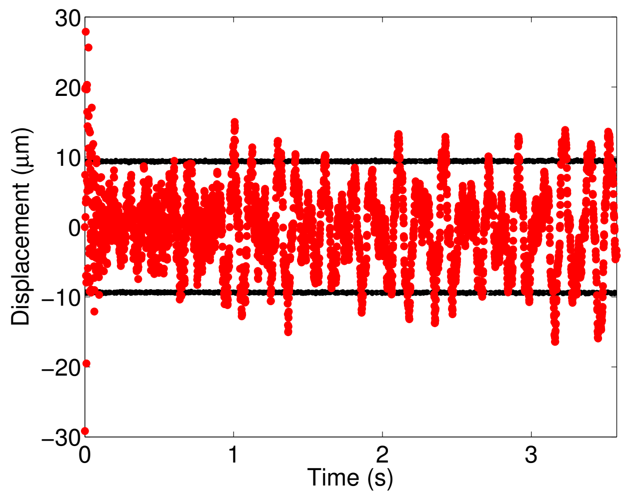
Partial Solution: "Optimize" calibration interval

•Use damping/frequency domain results to guide.



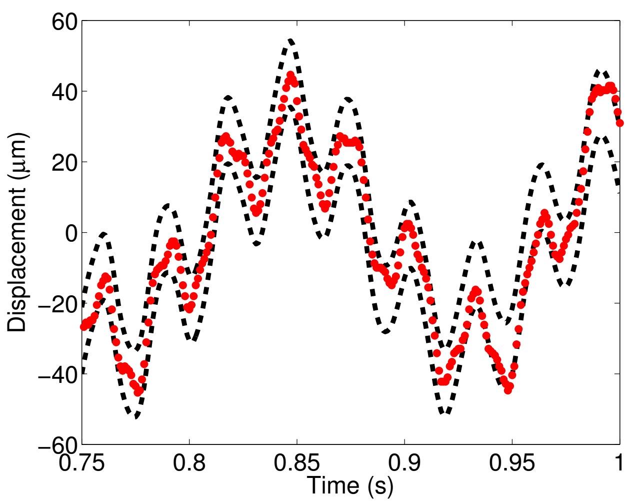


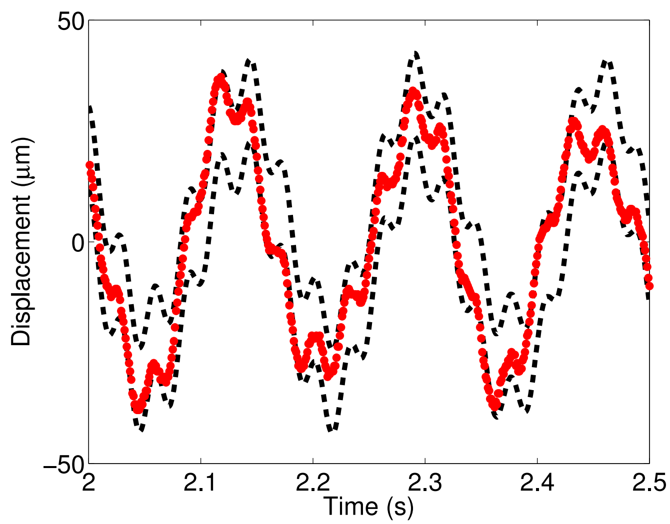
Calibrate on [0,1]

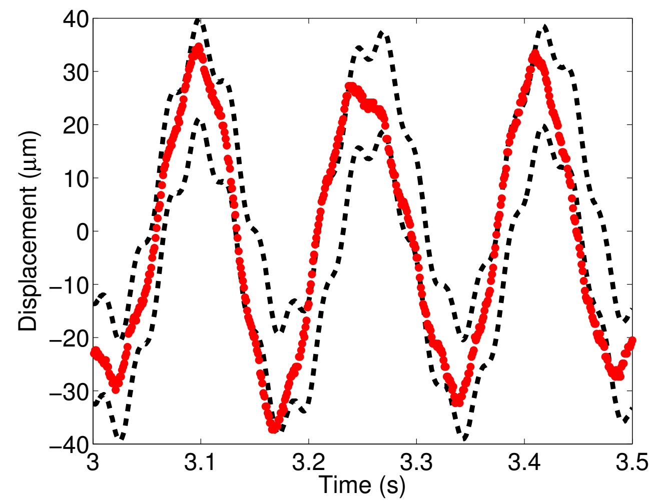


Calibrate on [0.25,1.25]

Note: We have substantially extended calibration regime.







Concluding Remarks

Notes:

• Uncertainty quantification critical for large-scale biological and physical models.

• UQ requires a synergy between applied mathematics, statistics, and domain sciences.

• Model calibration, model selection, uncertainty propagation and experimental design are natural in a Bayesian framework.

• Goal is to predict model responses with quantified and reduced uncertainties.

• Parameter selection and subspace construction is critical to isolate identifiable and influential parameters.

• Due to complexity of models, surrogate models are required for many applications.

• The quantification of model discrepancy constitutes an important research area.

• Algorithms and techniques are new and evolving.