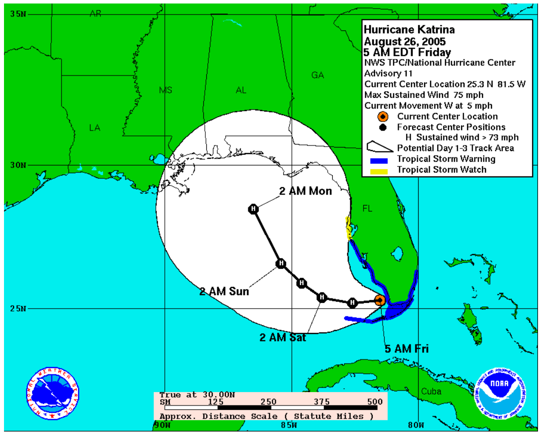
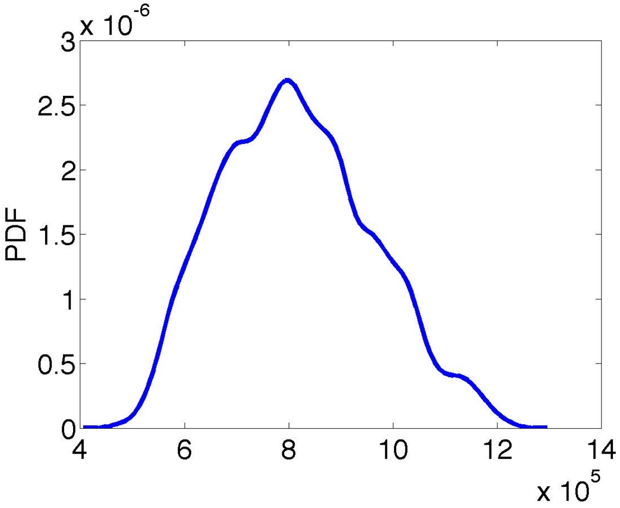
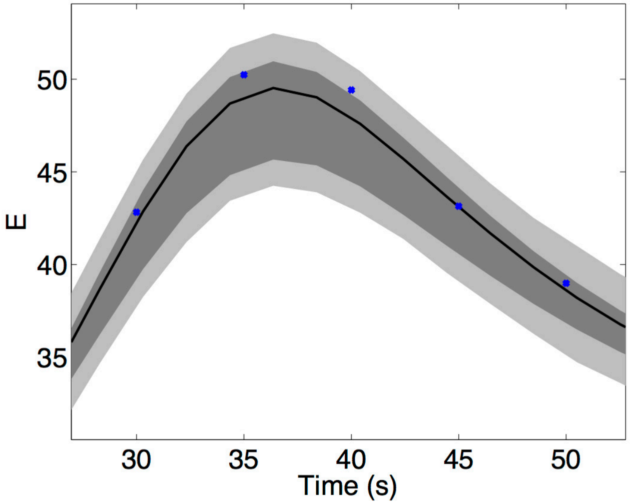
Roles of Sensitivity Analysis and Uncertainty Quantification for Science and Engineering Models

Ralph C. Smith
Department of Mathematics
North Carolina State University







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

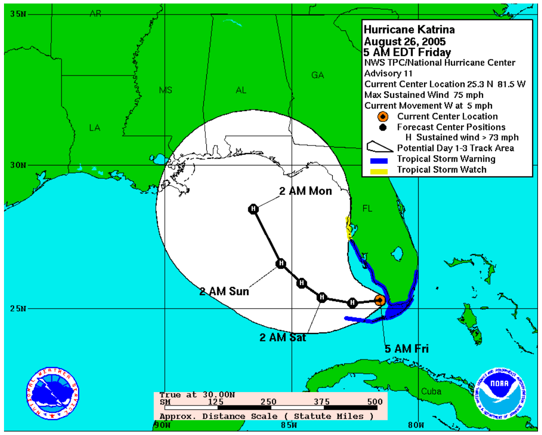
Support: DOE Consortium for Advanced Simulation of LWR (CASL)
NNSA Consortium for Nonproliferation Enabling Capabilities (CNEC)
National Science Foundation (NSF)
Air Force Office of Scientific Research (AFOSR)

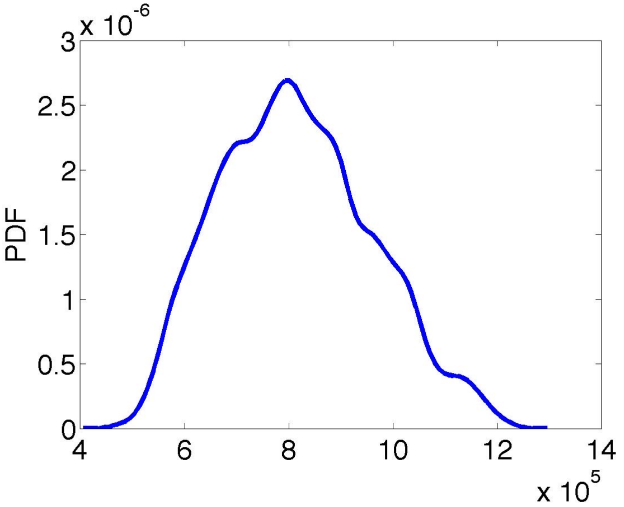
Roles of Sensitivity Analysis and Uncertainty Quantification for Science and Engineering Models

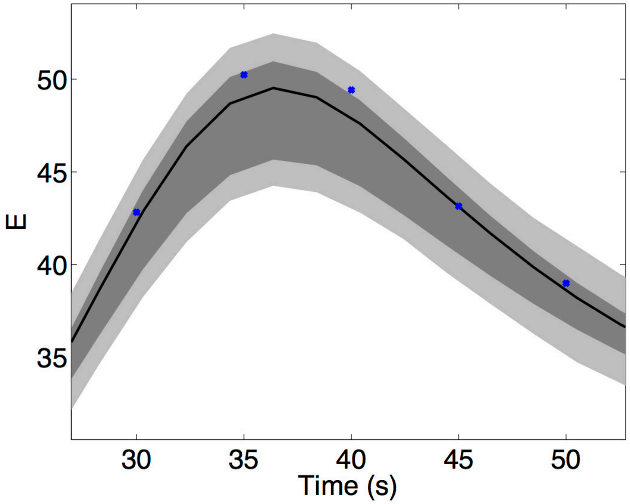
Ralph C. Smith

Department of Mathematics

North Carolina State University







”We”:

Kayla Coleman, Allison Lewis, Lider Leon, Paul Miles, Isaac Michaud (NCSU), Billy Oates (Florida State University)

Brian Williams (LANL), Russell Hooper, Brian Adams, Vince Mousseau (Sandia)

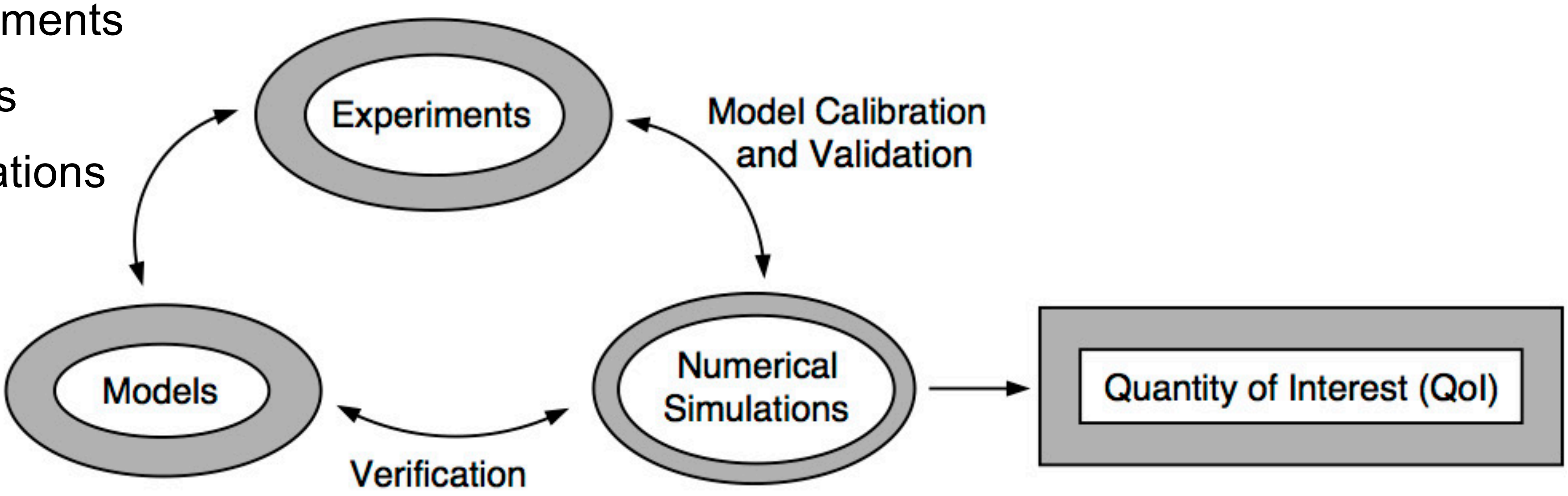
Emre Tatli and Yixing Sung (Westinghouse)

Max Morris (Iowa State University)

Predictive Science

Components: All involve uncertainty

- Experiments
- Models
- Simulations

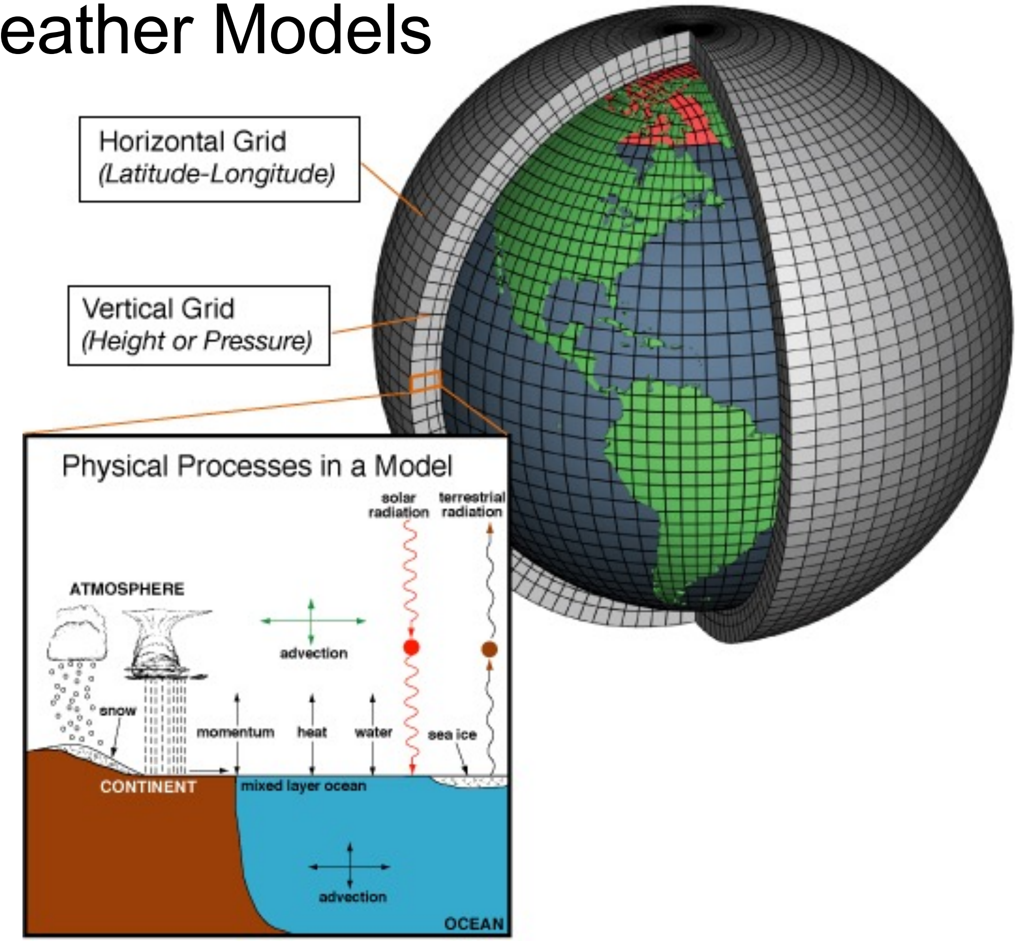


- *Essentially, all models are wrong, but some are useful,* George E.P. Box, Industrial Statistician.
- *Computational results are believed by no one, except the person who wrote the code*, source anonymous, quoted by Max Gunzburger, Florida State University.
- *Experimental results are believed by everyone, except for the person who ran the experiment,* 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

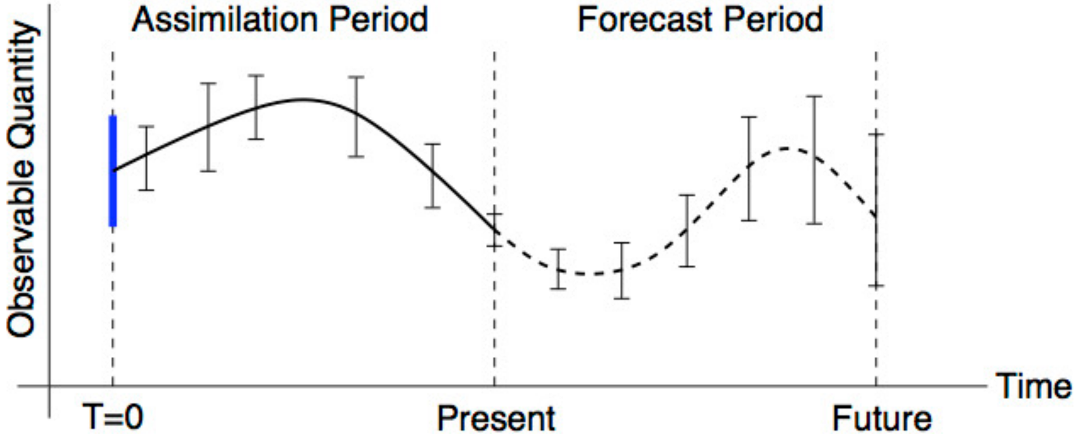
Challenges:

- Coupling between temperature, pressure gradients, precipitation, aerosol, etc.;
- Models and inputs contain uncertainties;
- Numerical grids necessarily larger than many phenomena; e.g., clouds
- Sensors positions may be uncertain; 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:

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

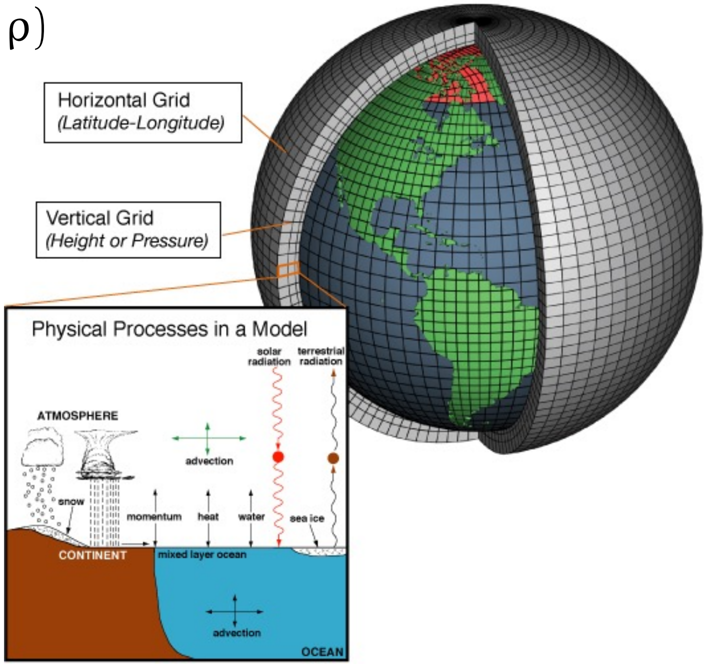
Momentum
$$\frac{\partial v}{\partial t} = -v \cdot \nabla v - \frac{1}{\rho}\nabla p - g\hat{k} - 2\Omega \times v$$

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

$$p = \rho R T$$

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

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



Constitutive Closure Relations: e.g.,

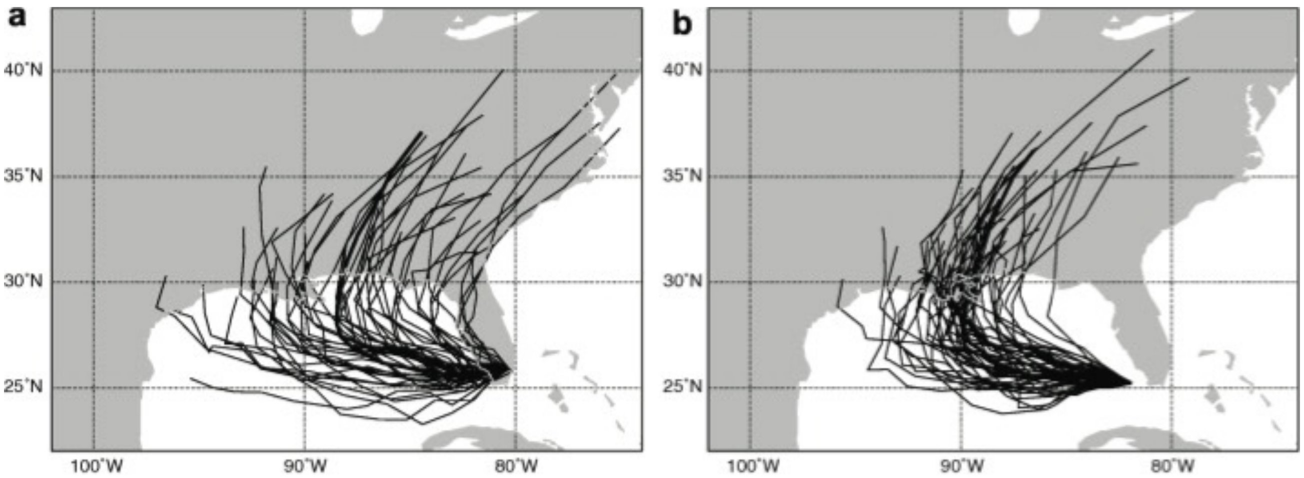
$$S_{m_2} = S_1 + S_2 + S_3 - S_4$$

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

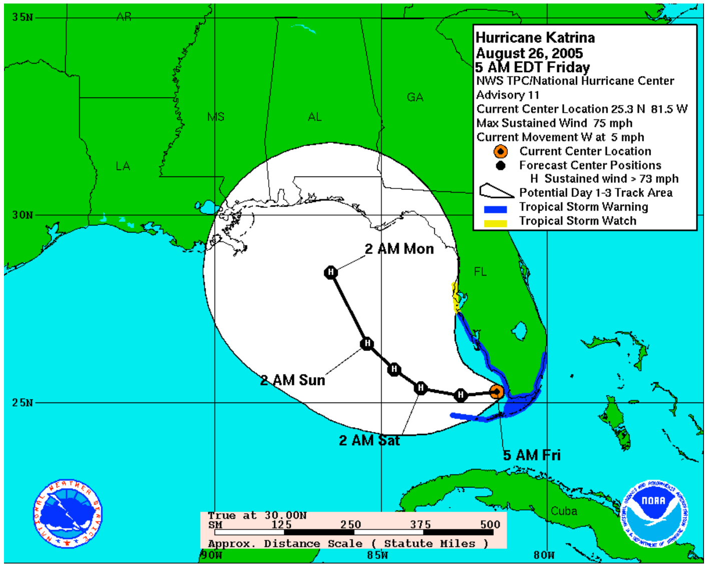
Ensemble Predictions:



00 UTC on August 26, 2005

12 UTC on August 26, 2005

Cone of Uncertainty:



General Questions:

- What is expected rainfall in Trieste on February 26?
- What are average high and low temperatures?
- Note: Quantities are statistical in nature.

Later Application: Wetland Methane Model

Authors: Susiluoto, Raivonen, Backman, Laine, Makela, Peltola, Vesala, Aalto, *Geoscientific Model Development, 11, pp. 1199-1228, 2018.*

Objectives: Methane emissions from natural wetlands highly uncertain

- What is effect of climate change?
- How much uncertainty is there in model parameters that control physical processes?
- How do parameters and wetland behavior react to environmental changes?

Model: HelsinkI Model of MEthane buiLd-up and emIssion for peatlands (HIMMELI)

- Incorporates process descriptions for methane production from anaerobic respiration, oxygen consumption, and carbon dioxide production;
- Submodel in regional and biosphere models.

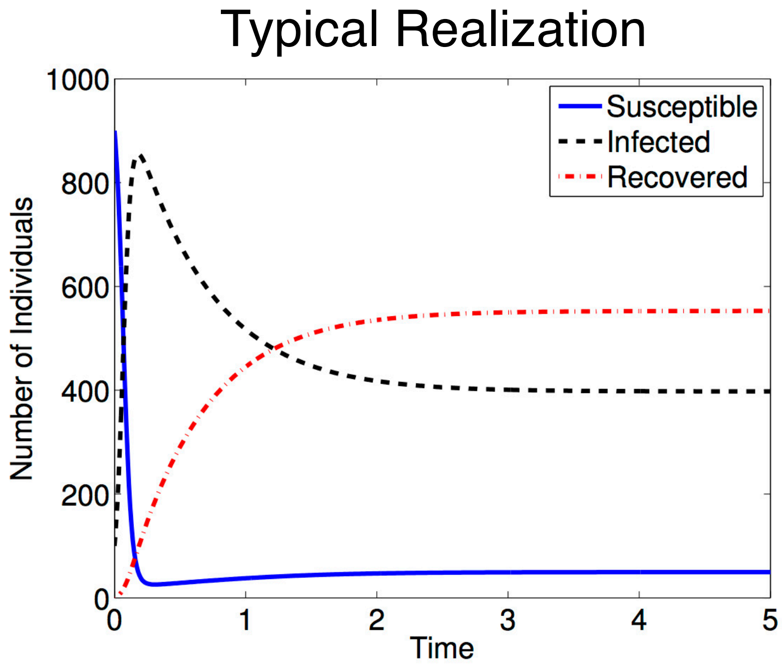
Example 2: SIR Model for Disease Dynamics

SIR Model: Population N

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

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

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


Typical Realization

Parameters:

- γ: Infection coefficient
- k: Interaction coefficient
- r: Recovery rate
- δ: Birth/death rate

Response:

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

Note: Parameters $q = [\gamma, k, r, \delta]$ not uniquely determined by data

SA Goal: Use sensitivity analysis to isolate subset of influential or identifiable parameters

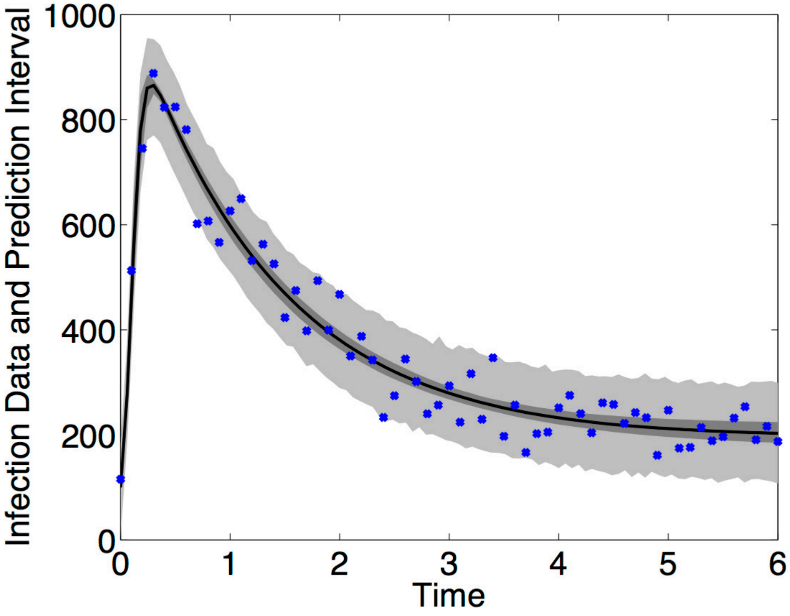
Example 2: SIR Model for Disease Dynamics

SIR Model: Population N

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

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

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



Parameters:

- γ: Infection coefficient
- k: Interaction coefficient
- r: Recovery rate
- δ: Birth/death rate

Response:

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

UQ Goal: Predict I(t) with uncertainty intervals

Example 3: HIV Model for Characterization and Treatment

HIV Model:

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

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

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

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

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

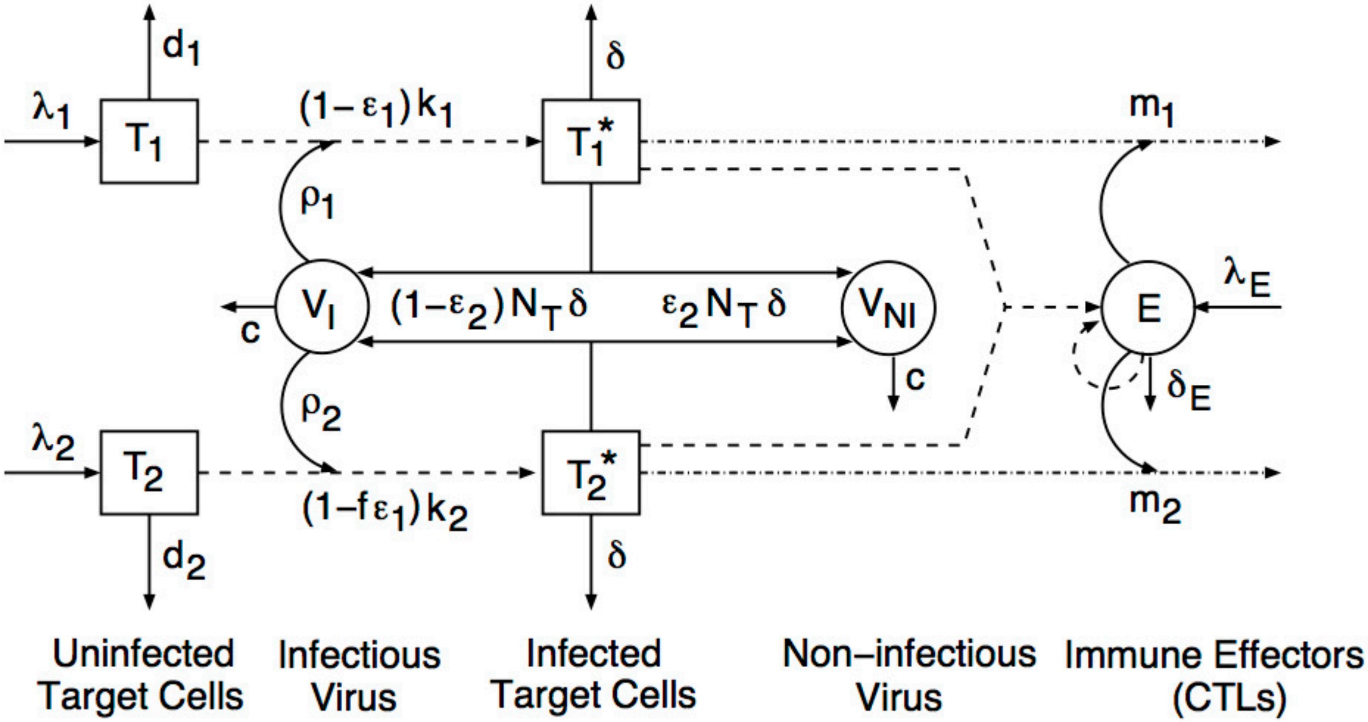
$$\dot{E} = \lambda_E + \frac{b_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_b} E - \frac{d_E (T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Notes: 21 parameters

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

Notation: $\dot{E} \equiv \frac{dE}{dt}$

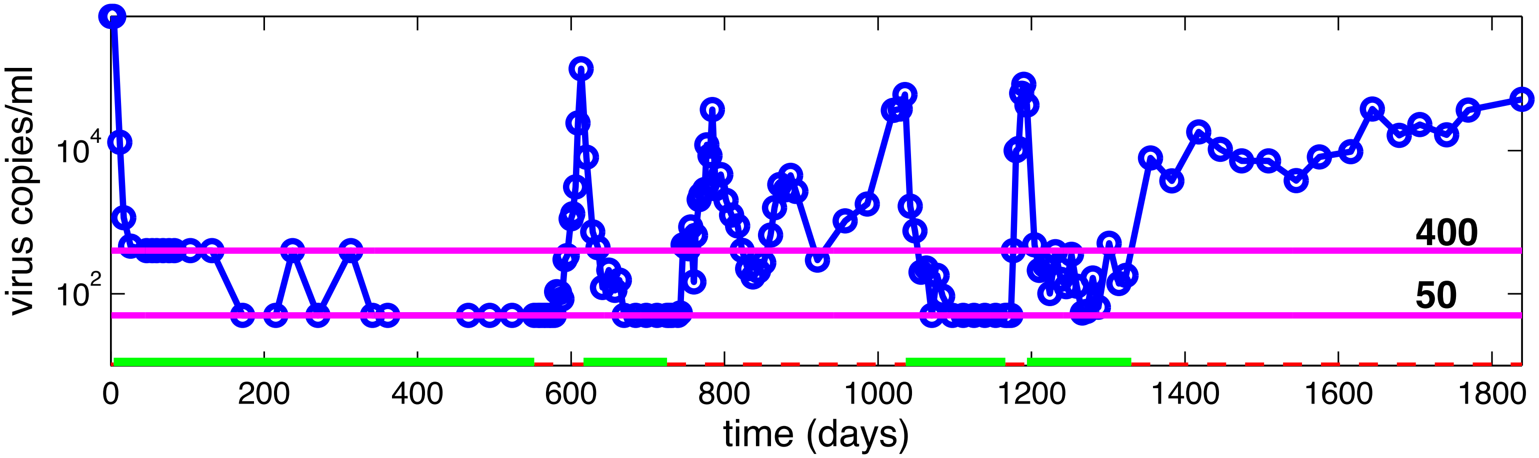
Compartments:



HIV Model for Characterization and Treatment Regimes

HIV Model: Several sources of uncertainty including viral measurement techniques

Example: Upper and lower limits to assay sensitivity

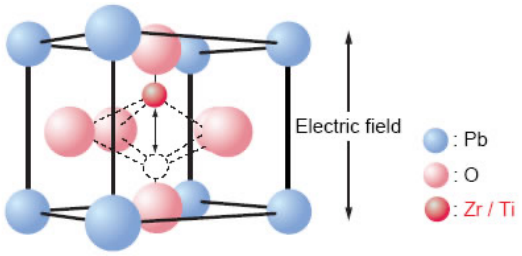


UQ Questions:

- What are the uncertainties in parameters that cannot be directly measured?
- What is optimal treatment regime that is “safe” for patient?
- What is expected viral load? Issue: very often requires high-dimensional integration!
- e.g.,

$$\mathbb{E}[V(t)] = \int_{\mathbb{R}^{21}} V(t,q)\rho(q)dq$$

Example 4: Quantum-Informed Continuum Models

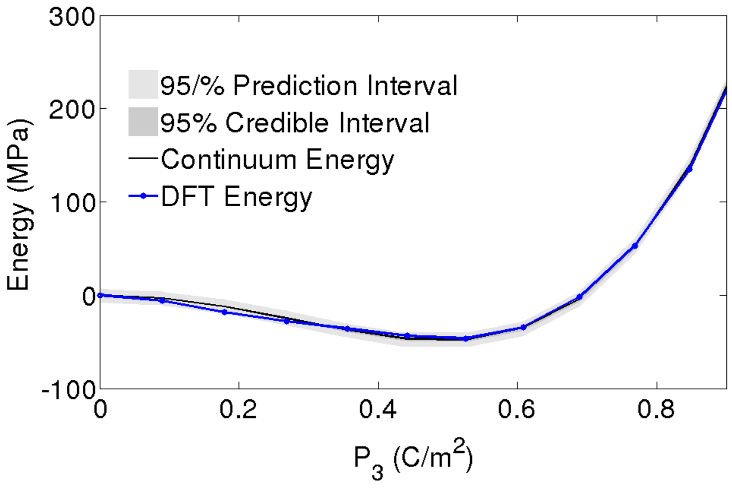


Lead Titanate Zirconate (PZT)

Objectives:

- Employ density function theory (DFT) to construct energy relations;
  - e.g., Helmholtz energy

$$\psi(P) = \alpha_1 P^2 + \alpha_{11} P^4 + \alpha_{111} P^6$$



Helmholtz Energy

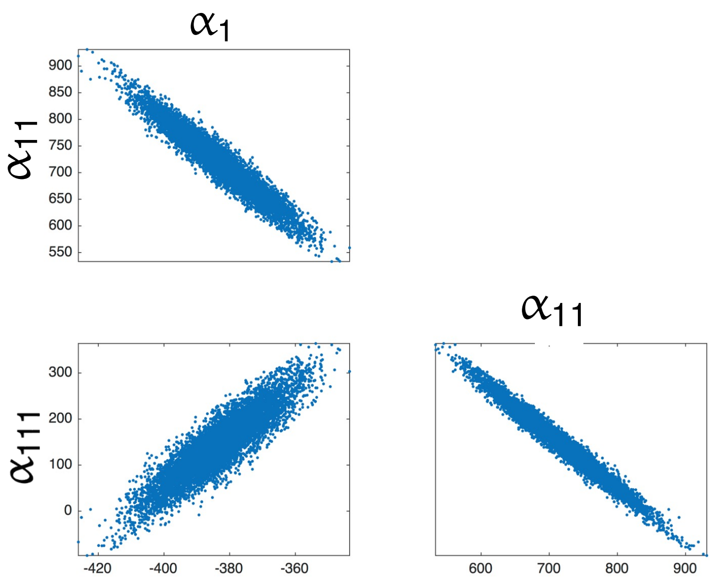
UQ and SA Issues:

- Is 6th order term required to accurately characterize material behavior?

Broad Objective:

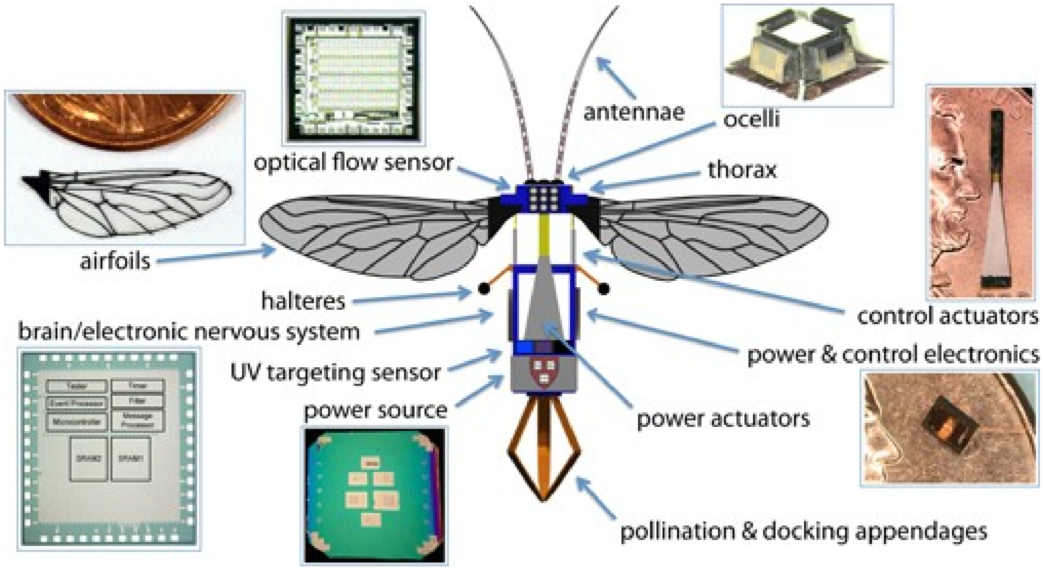
- Use UQ/SA to help bridge scales from quantum to system

Note: Correlated parameters



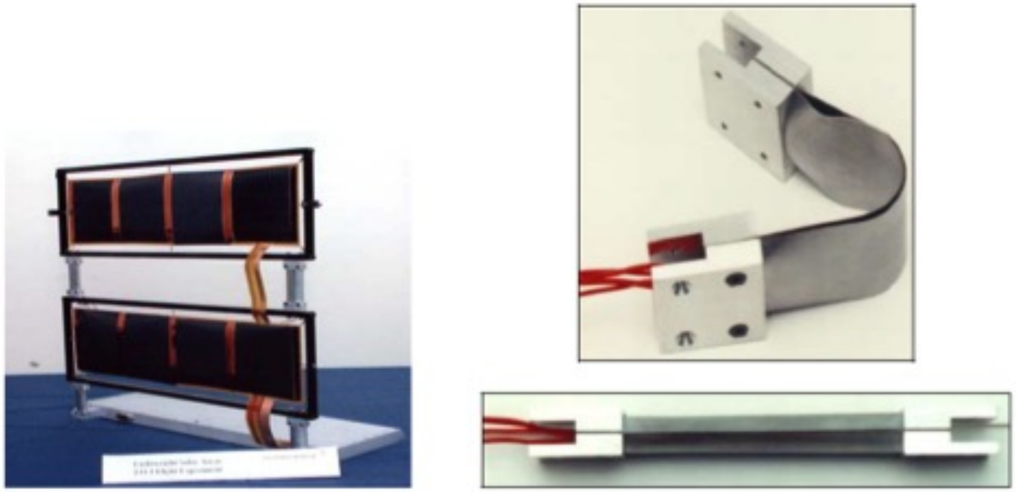
Smart Material Applications

PZT: Robobee -- Rob Wood, Harvard University

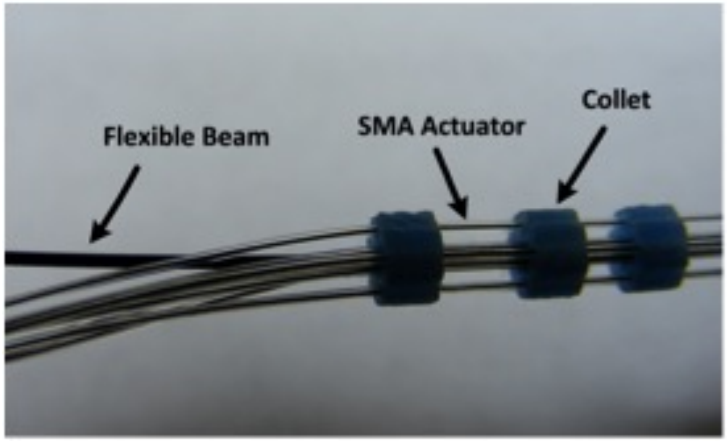


Shape Memory Alloys (SMA):

Chevrons for Noise Reduction/Fuel efficiency



SMA Hinges for Solar Arrays

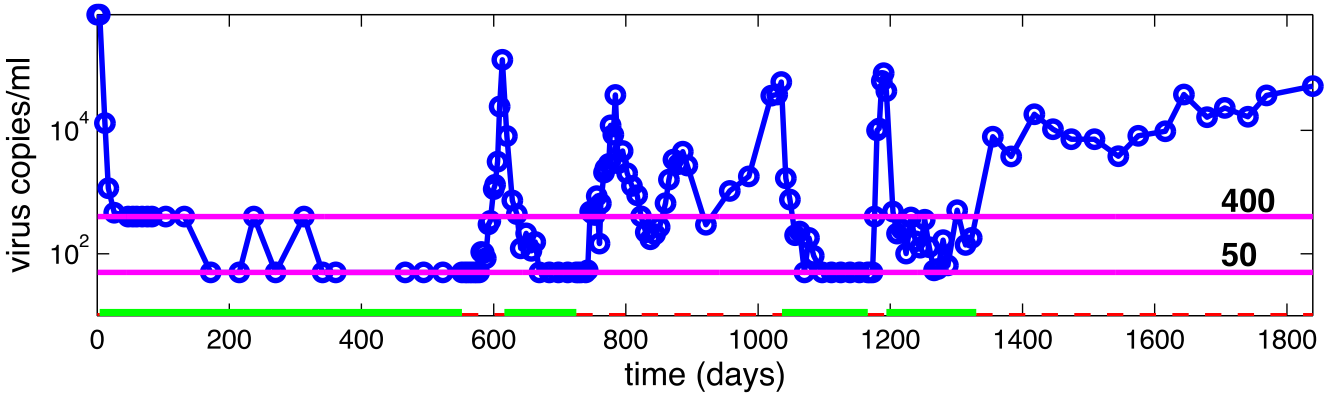


Catheters for Laser Ablation

Experimental Uncertainties

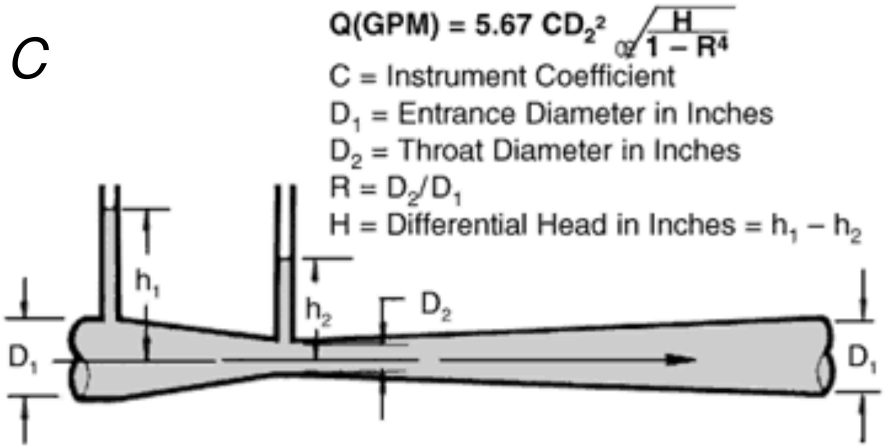
Experimental results are believed by everyone, except for the person who ran the experiment, Max Gunzburger, Florida State University.

Example: Upper and lower limits to assay sensitivity



Note: Many sensors rely on calibrated physical relations; e.g., Pitot tube to measure airspeed using Bernoulli's principle.

$$\frac{u^2}{2} + gz + \frac{p}{\rho} = C$$



$$Q(GPM) = 5.67\, C D_2^2 \sqrt{\frac{H}{1 - R^4}}$$

C = Instrument Coefficient

D_1 = Entrance Diameter in Inches

D_2 = Throat Diameter in Inches

$R = D_2/D_1$

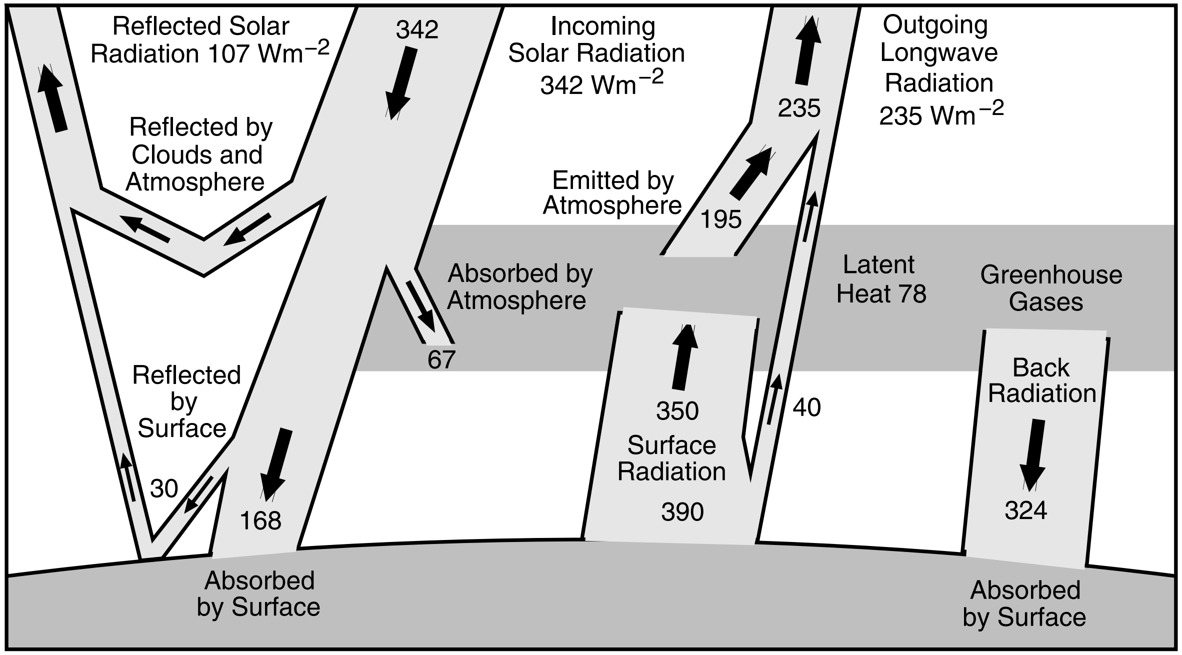
H = Differential Head in Inches = $h_1 - h_2$

Model-Form Uncertainties

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

- Numerous components of weather and climate models --- e.g., aerosol-induced cloud formation --- are highly complex and not well-understood.

- Model-form uncertainty constitutes a significant UQ challenge. ”Hawkmoth Effect” versus “Butterfly Effect,” associated with initial conditions, discussed by Frigg, Bradley, Du and Smith, Philosophy of Science, 2014.



Parameter Uncertainties

Example:

- Closure relations for equations of atmospheric physics

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

where

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

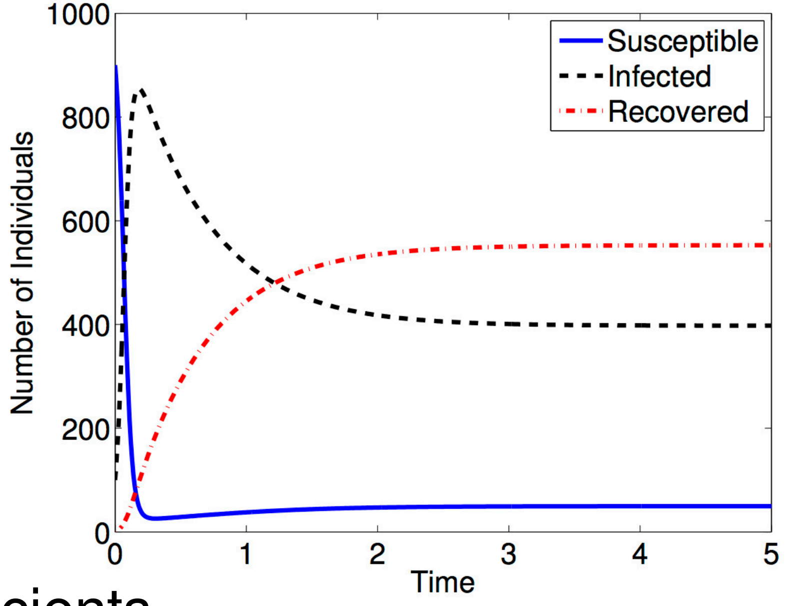
Example: SIR Model

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

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

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

Typical Realization

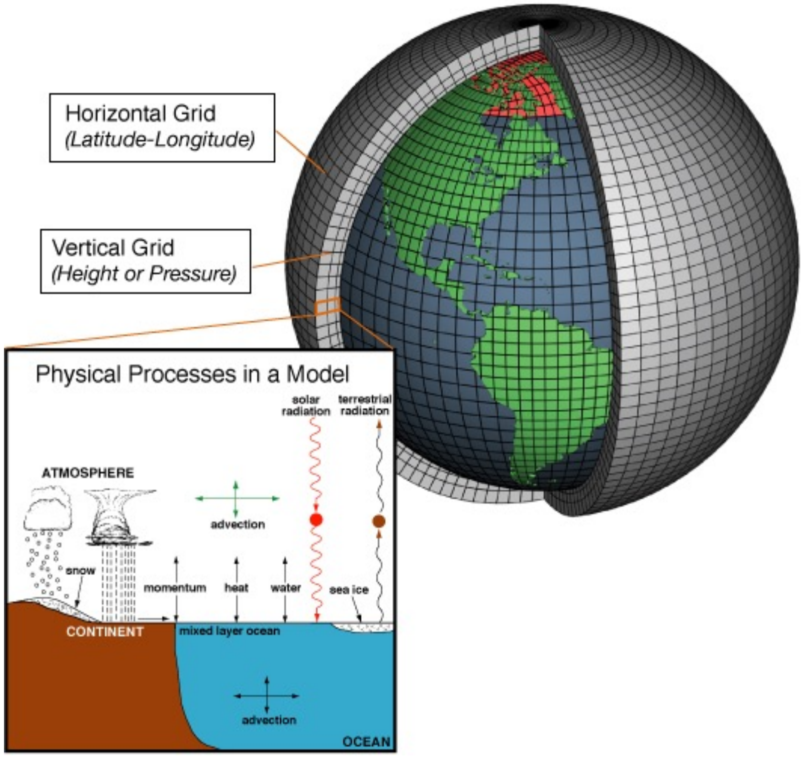


E.g.; γ and k: Infection and interaction coefficients

Numerical Errors

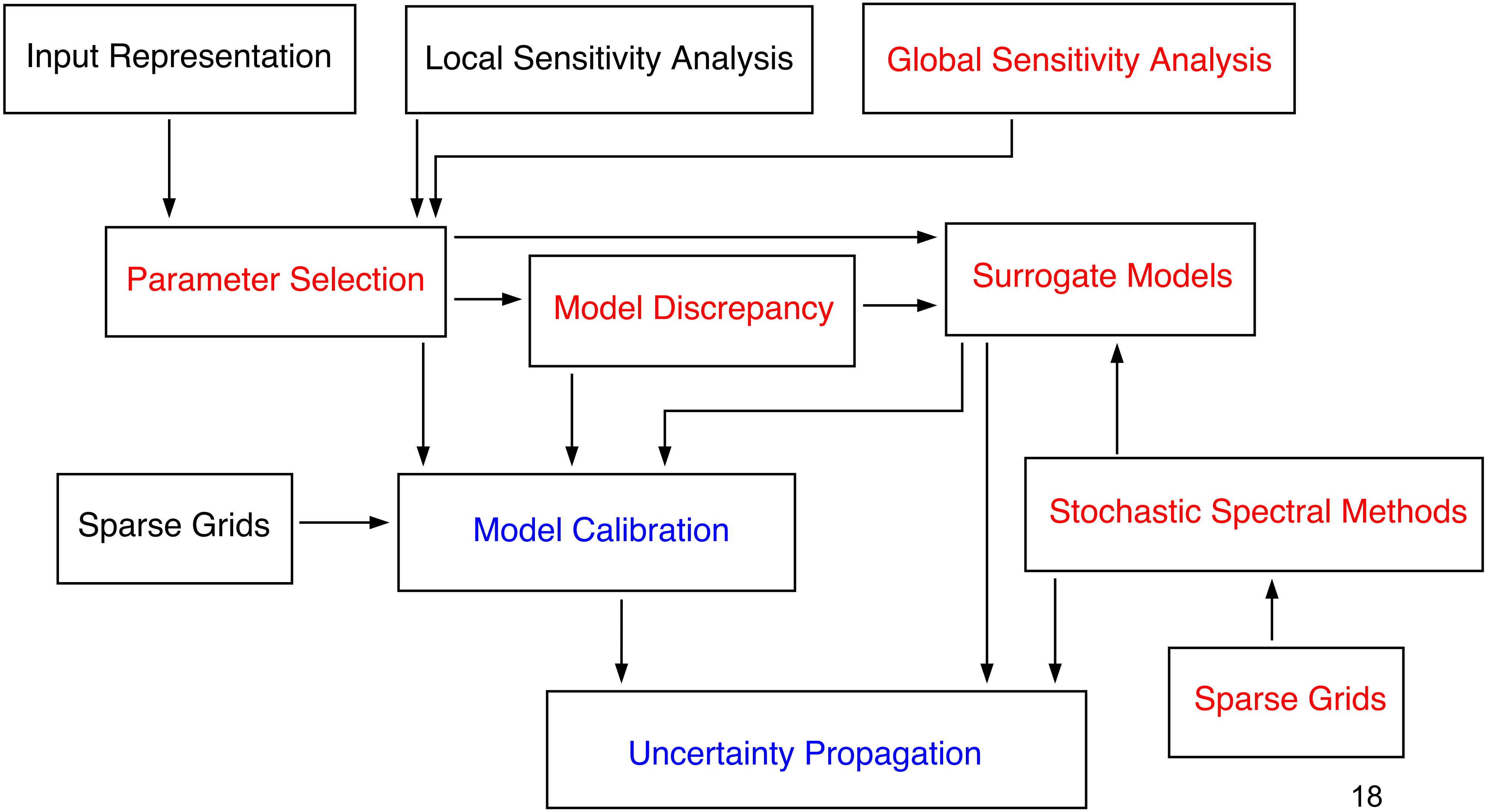
Computational results are believed by no one, except the person who wrote the code, Max Gunzburger, Florida State University.

- Roundoff, discretization or approximation errors;
- Bugs or coding errors;
- Bit-flipping, hardware failures and uncertainty associated with future exascale and quantum computing;
- Grids required for applications such as weather (e.g., 50~km) are much larger than the scale of physics being modeled.



Steps in Uncertainty Quantification

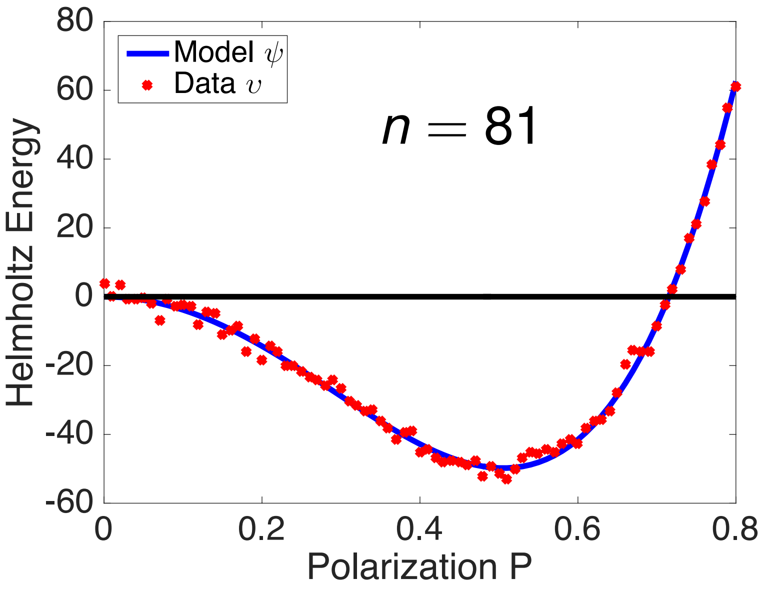
Note: Uncertainty quantification requires synergy between statistics, applied mathematics and domain sciences.



Deterministic Model Calibration

Example: Helmholtz energy $\psi(P,q) = \underline{\alpha_1} P^2 + \underline{\alpha_{11}} P^4 + \underline{\alpha_{111}} P^6$

$$q = [\alpha_1, \alpha_{11}, \alpha_{111}]$$



Statistical Model: Describes observation process

$$\upsilon_i = \psi(P_i, q) + \varepsilon_i \quad , \; i = 1, \ldots, n$$

Point Estimates: Ordinary least squares

$$q^0 = \arg\min_q \frac{1}{2} \sum_{j=1}^{n} [\upsilon_i - \psi(P_i, q)]^2$$

PZT: Robobee

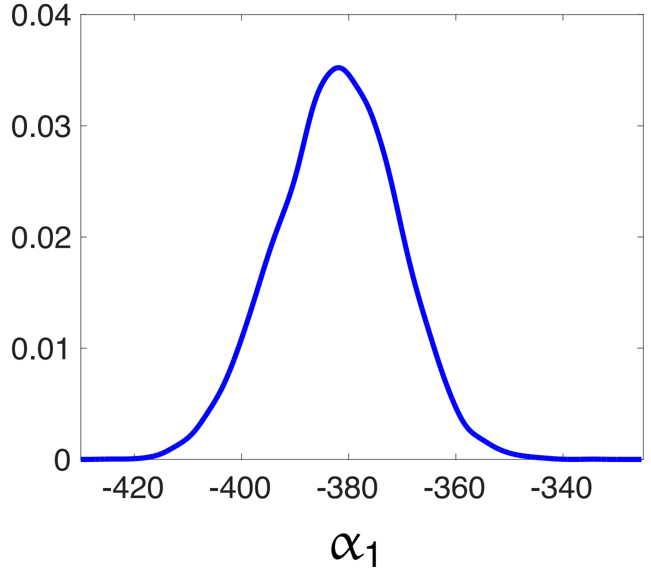
Note: Provides point estimates; Does not quantification of uncertainty in:

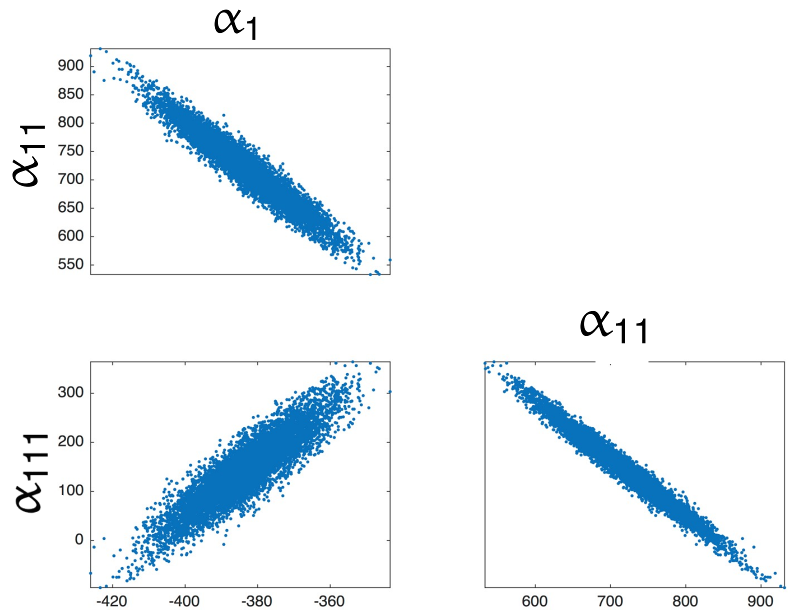
- Model
- Parameters
- Data

Objectives for Uncertainty Quantification

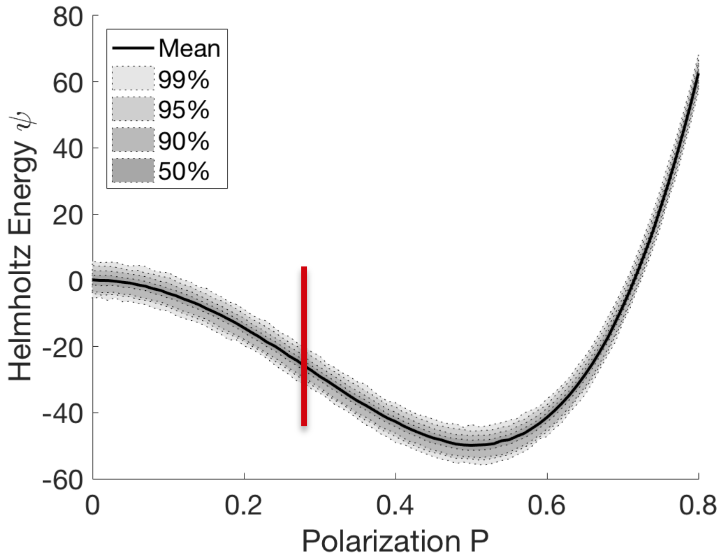
Goal: Replace point estimates with distributions or credible intervals

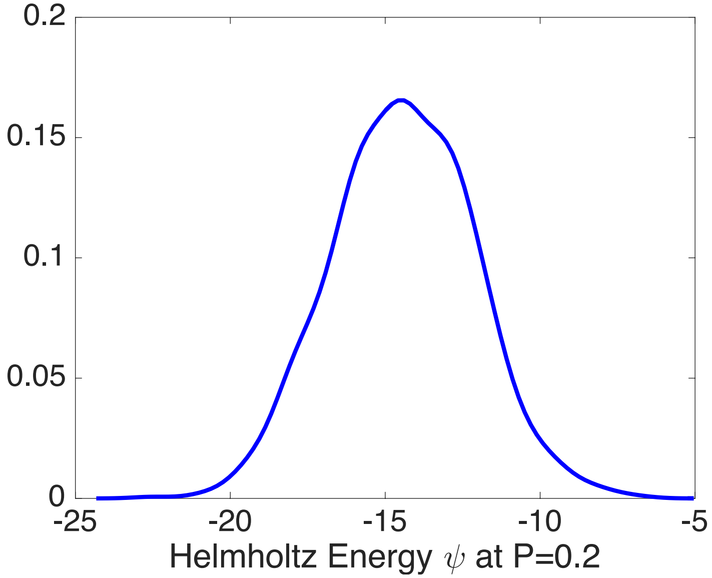
E.g., Parameter Densities



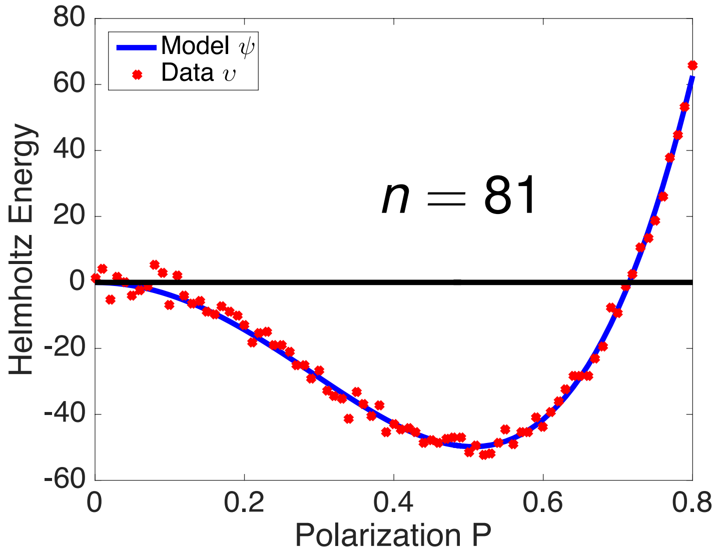


E.g., Response Intervals





Frequentist Inference



Statistical Model: For i = 1, …, n

$$\upsilon_i = \psi(P_i, q) + \varepsilon_i \longleftarrow \varepsilon_i \sim N(0, \underline{\sigma^2})$$

$$= \alpha_1 P_i^2 + \alpha_{11} P_i^4 + \varepsilon_i$$

$$\Rightarrow \begin{bmatrix} \upsilon_i \end{bmatrix} = \begin{bmatrix} P_i^2 & P_i^4 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_{11} \end{bmatrix} + \begin{bmatrix} \varepsilon_i \end{bmatrix}$$

$$\Rightarrow \upsilon = Xq + \varepsilon$$

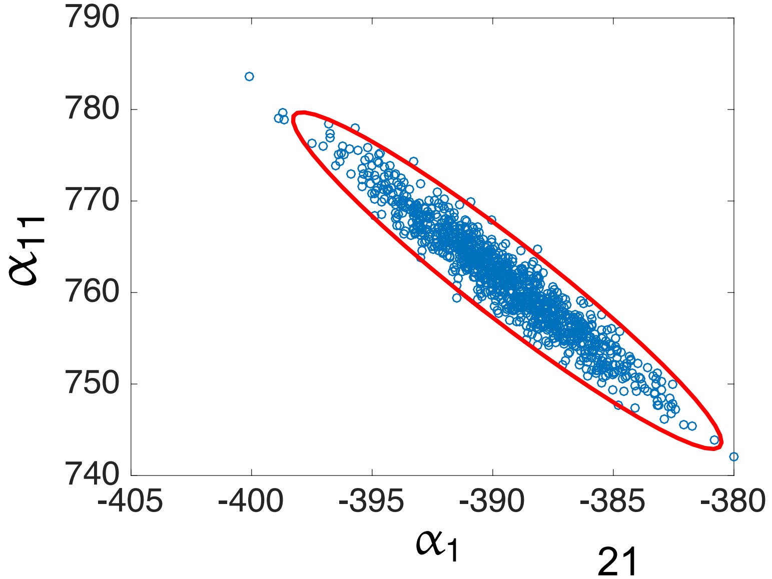
Statistical Quantities:

$$q = (X^T X)^{-1} X^T \upsilon$$

$$\mathrm{var}(\alpha_1)$$

$$V = \underline{\sigma^2}(X^T X)^{-1} = \begin{bmatrix} 8.8 & -17.4 \\ -17.4 & 37.6 \end{bmatrix}$$

$$\mathrm{cov}(\alpha_1, \alpha_{11}) \qquad \mathrm{var}(\alpha_{11})$$



Note: Covariance matrix incorporates "geometry"

Goal: Employ Bayesian inference for UQ

Statistical Inference

Goal: Goal in statistical inference is to make conclusions about a phenomenon based on observed data.

Frequentist: Observations made in the past analyzed to determine confidence about state of modeled phenomenon.

- Probabilities defined as frequencies for which an event occurs if experiment is repeated several times.
- Parameters may be unknown but are fixed and deterministic.

Bayesian: Interpretation of probability is subjective and can be updated with new data.

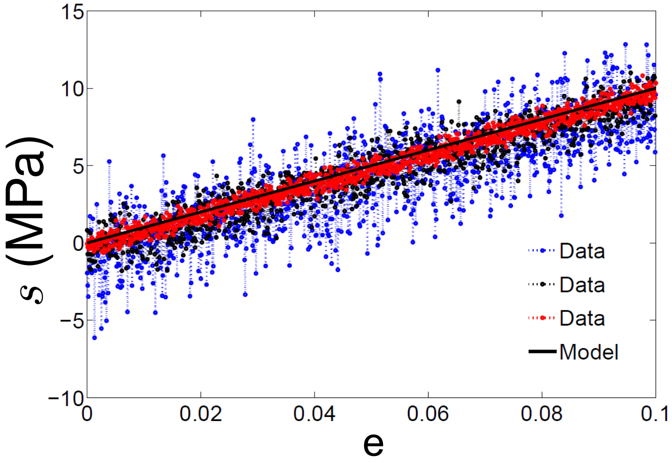
- Parameters considered to be random variables having associated densities.

Bayesian Inference: Motivation

Example: Displacement-force relation (Hooke's Law)

$$s_i = Ee_i + \varepsilon_i \;,\; i = 1, \ldots, N$$

$$\overset{\nwarrow}{\varepsilon_i \sim N(0, \sigma^2)}$$



Parameter: Stiffness E

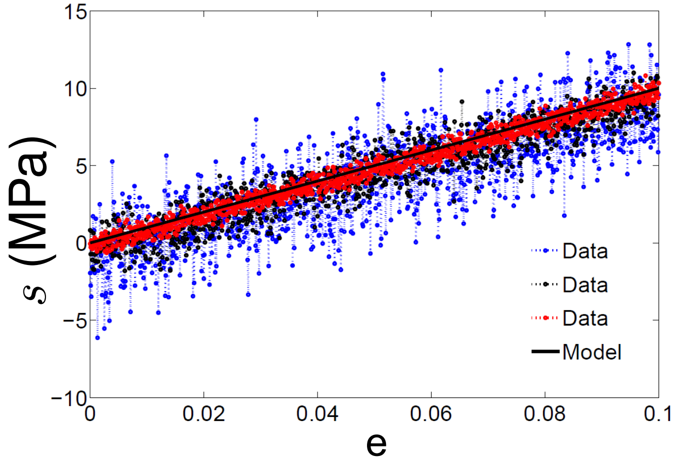
Strategy: Use model fit to data to update prior information

Bayesian Inference: Motivation

Example: Displacement-force relation (Hooke's Law)

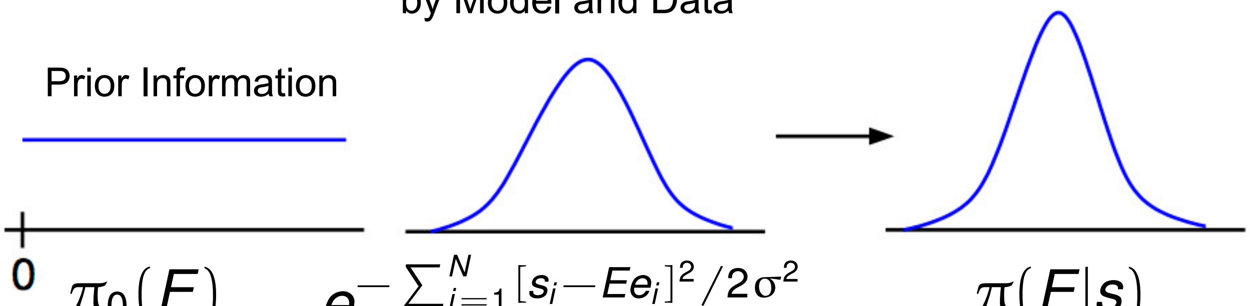
$$s_i = Ee_i + \varepsilon_i \ , \ i = 1, \ldots, N$$

$$\varepsilon_i \sim N(0, \sigma^2)$$



Parameter: Stiffness E

Strategy: Use model fit to data to update prior information



Prior Information

Information Provided by Model and Data

Updated Information

$$\pi_0(E) \qquad e^{-\sum_{i=1}^N [s_i - Ee_i]^2 / 2\sigma^2} \qquad \pi(E|s)$$

Data

Model

Non-normalized Bayes' Relation:

$$\pi(E|s) = e^{-\sum_{i=1}^N [s_i - Ee_i]^2 / 2\sigma^2} \pi_0(E)$$

Bayesian Inference

Bayes' Relation: Specifies posterior in terms of likelihood and prior

Likelihood: $e^{-\sum_{i=1}^{N}[s_i - Ee_i]^2/2\sigma^2}$, $q = E$

$\upsilon = [s_1, \ldots, s_N]$

Posterior Distribution

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

Prior Distribution

Normalization Constant

Problem: Can require high-dimensional integration

- e.g., HIV Model: p = 21!
- Solution: Sampling-based Markov Chain Monte Carlo (MCMC) algorithms.
- Metropolis algorithms first used by nuclear physicists during Manhattan Project in 1940's to understand particle movement underlying first atomic bomb.
- Current applications: Delayed Rejection Adaptive Metropolis (DRAM) [Haario et al., 2006] – MATLAB, Python, R, Dakota

Delayed Rejection Adaptive Metropolis (DRAM)

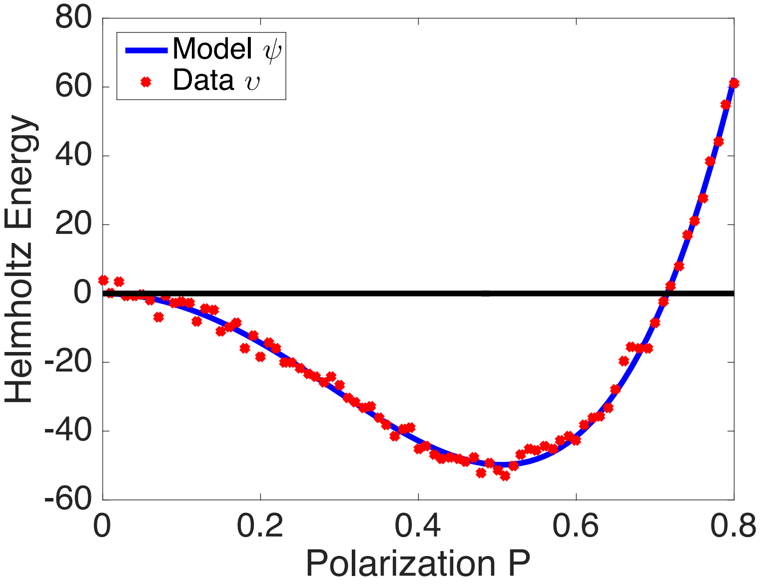
Algorithm: [Haario et al., 2006] – MATLAB, Python, R

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

Example: Helmholtz energy

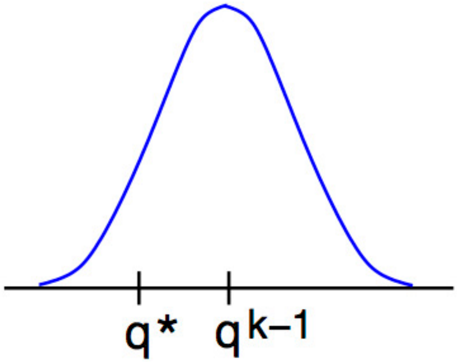
$$\upsilon_i = \psi(P_i, q) + \varepsilon_i \longleftarrow \varepsilon_i \sim N(0, \sigma^2)$$

$$= \alpha_1 P_i^2 + \alpha_{11} P_i^4 + \varepsilon_i$$



Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R



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

2. For $k = 1, \ldots, M$

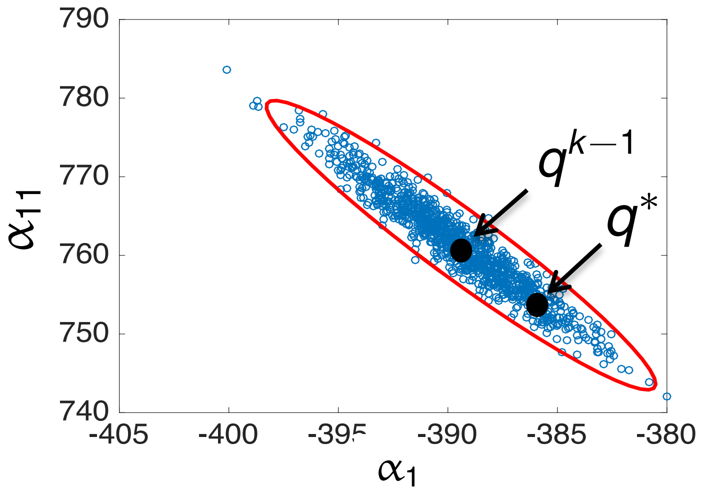
   (a) Construct candidate $q^* \sim N(q^{k-1}, \underline{V})$

Example: Helmholtz energy

$$\upsilon_i = \psi(P_i, q) + \varepsilon_i \longleftarrow \varepsilon_i \sim N(0, \sigma^2)$$

$$= \alpha_1 P_i^2 + \alpha_{11} P_i^4 + \varepsilon_i$$

Recall: Covariance V incorporates geometry



Delayed Rejection Adaptive Metropolis (DRAM)

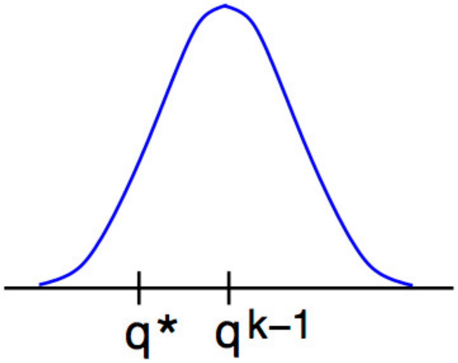
Algorithm: [Haario et al., 2006] – MATLAB, Python, R

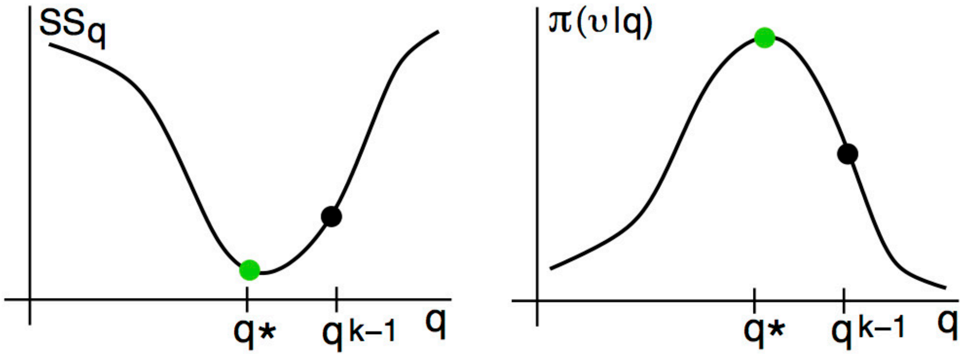
1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \dots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

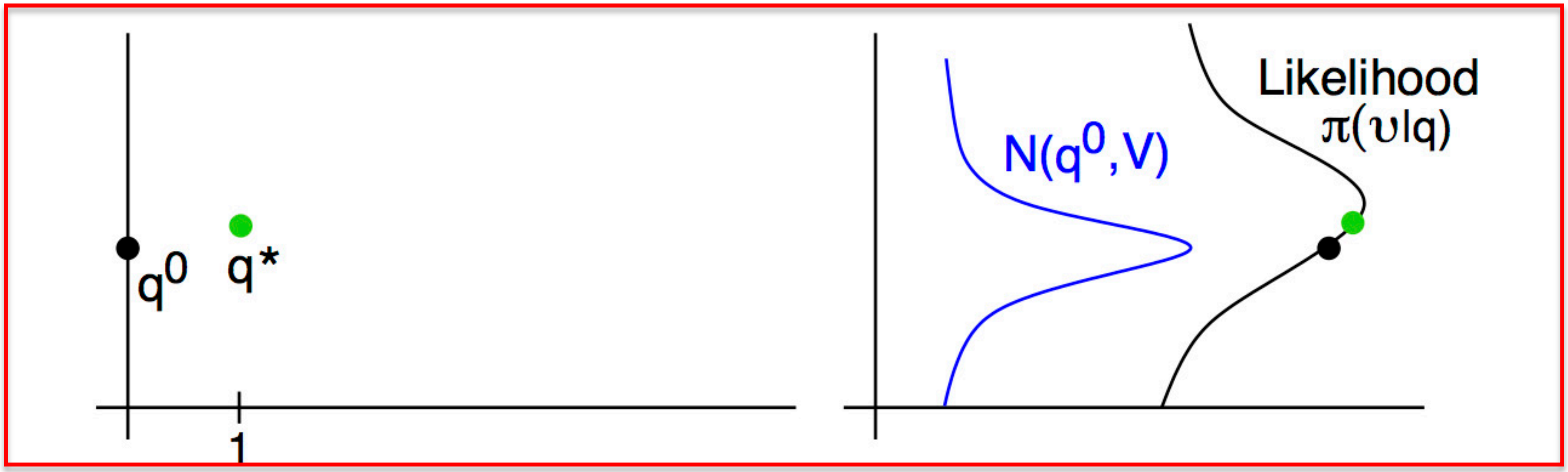
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$

   (c) Accept q^* with probability dictated by likelihood

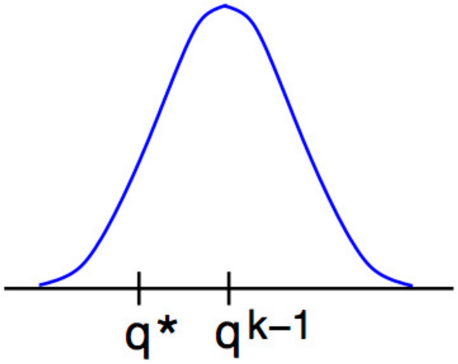






Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R



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

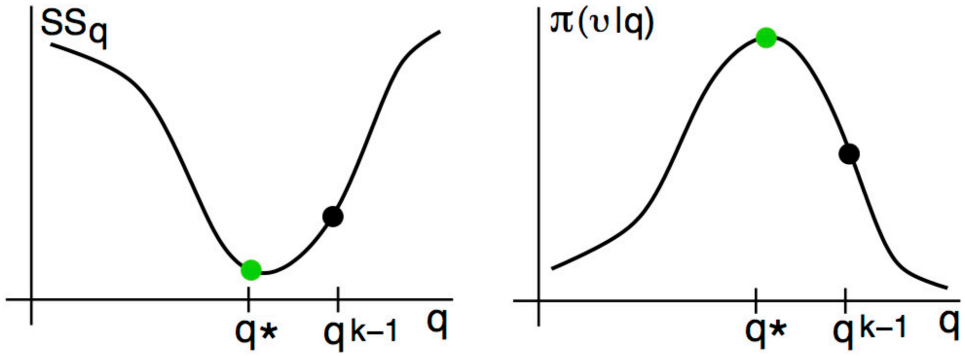
2. For $k = 1, \ldots, M$

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

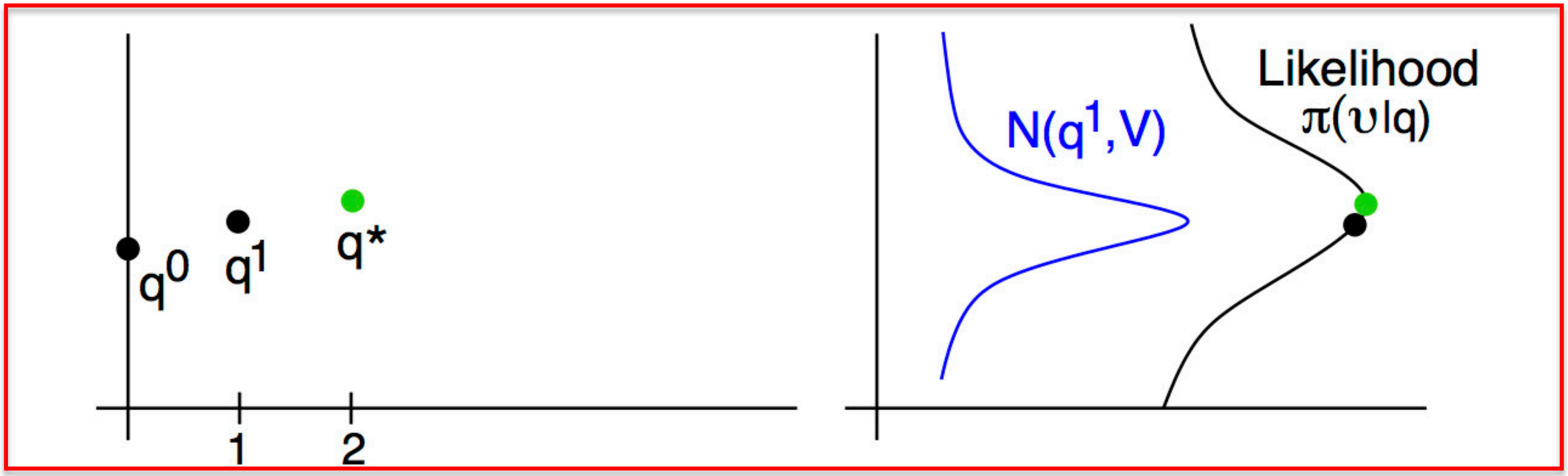
   (b) Compute likelihood

$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$



   (c) Accept q^* with probability dictated by likelihood



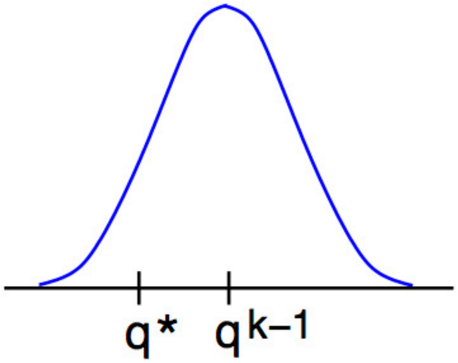
Delayed Rejection Adaptive Metropolis (DRAM)

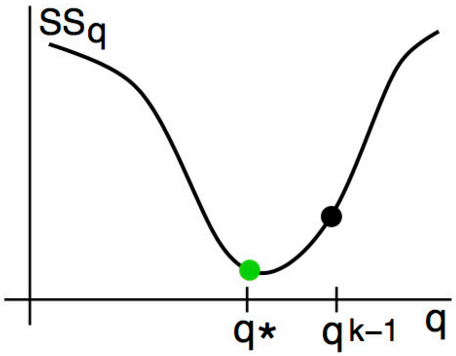
Algorithm: [Haario et al., 2006] – MATLAB, Python, R

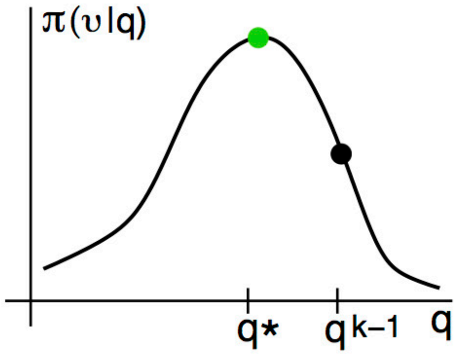
1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
    (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
    (b) Compute likelihood

$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

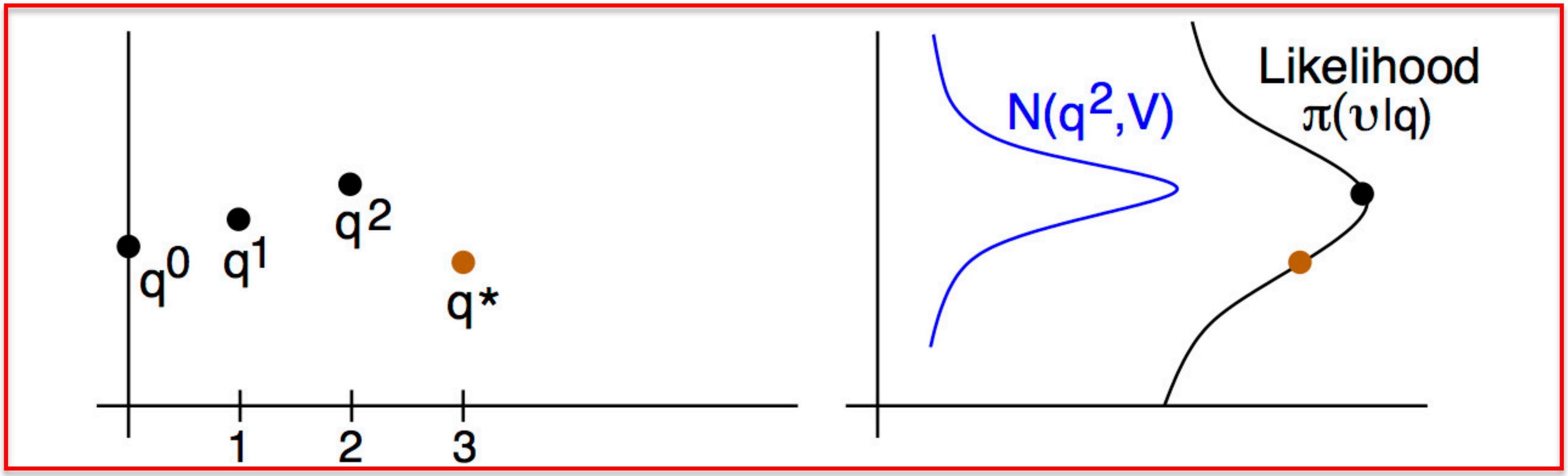
$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$





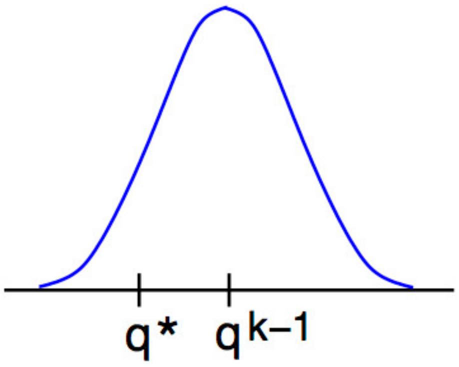


    (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

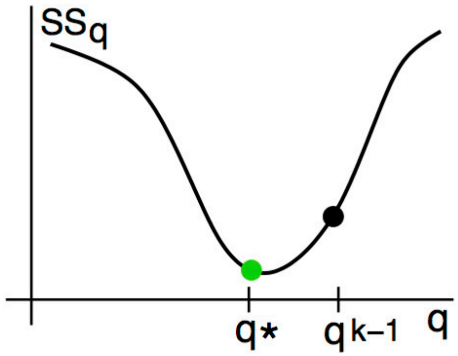
Algorithm: [Haario et al., 2006] – MATLAB, Python, R

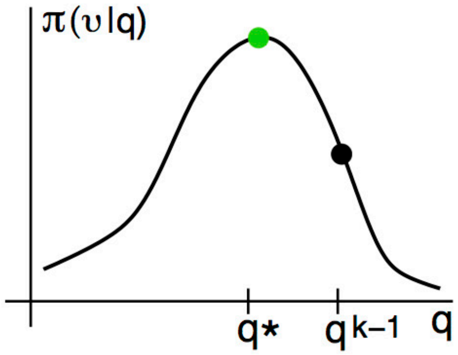


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
   - (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   - (b) Compute likelihood

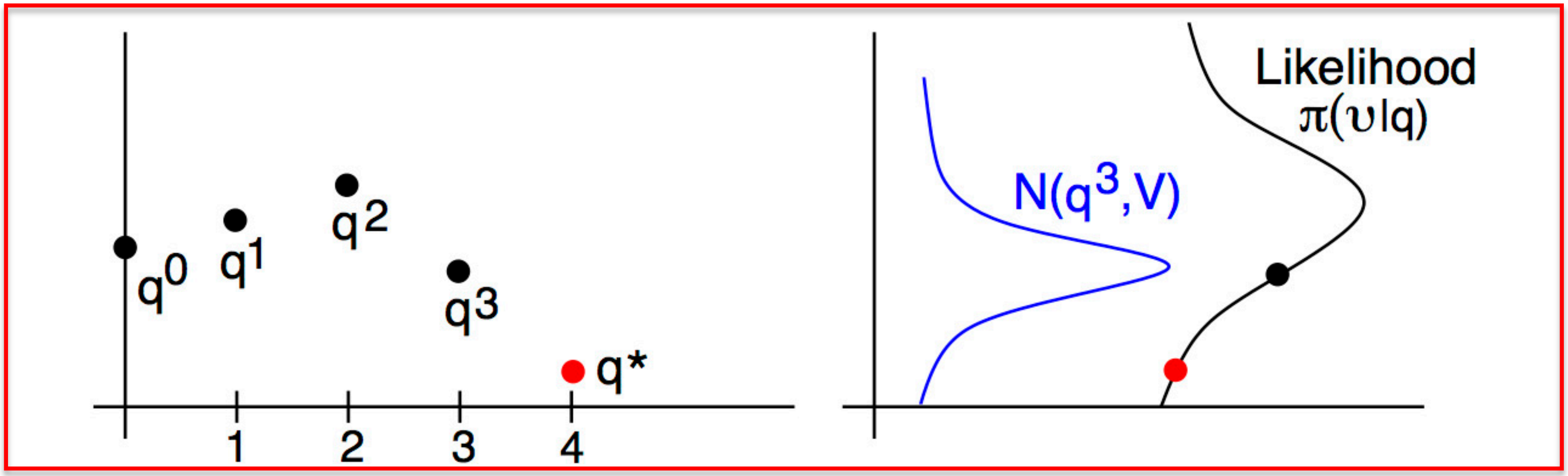
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$





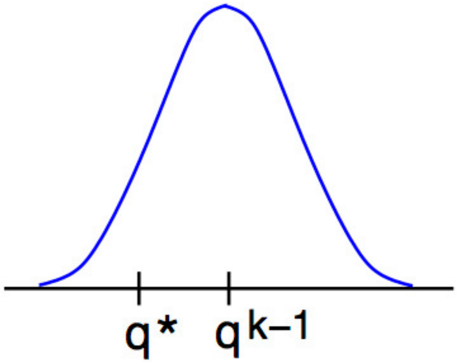
   - (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

Algorithm: [Haario et al., 2006] – MATLAB, Python, R

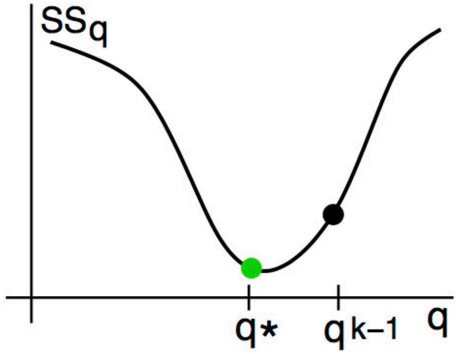
1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - \psi(P_i, q)]^2]$
2. For $k = 1, \ldots, M$
    (a) Construct candidate $q^* \sim N(q^{k-1}, V)$

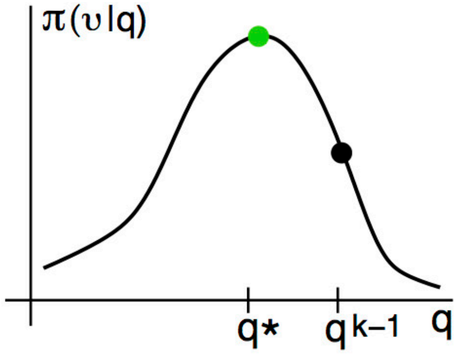


    (b) Compute likelihood

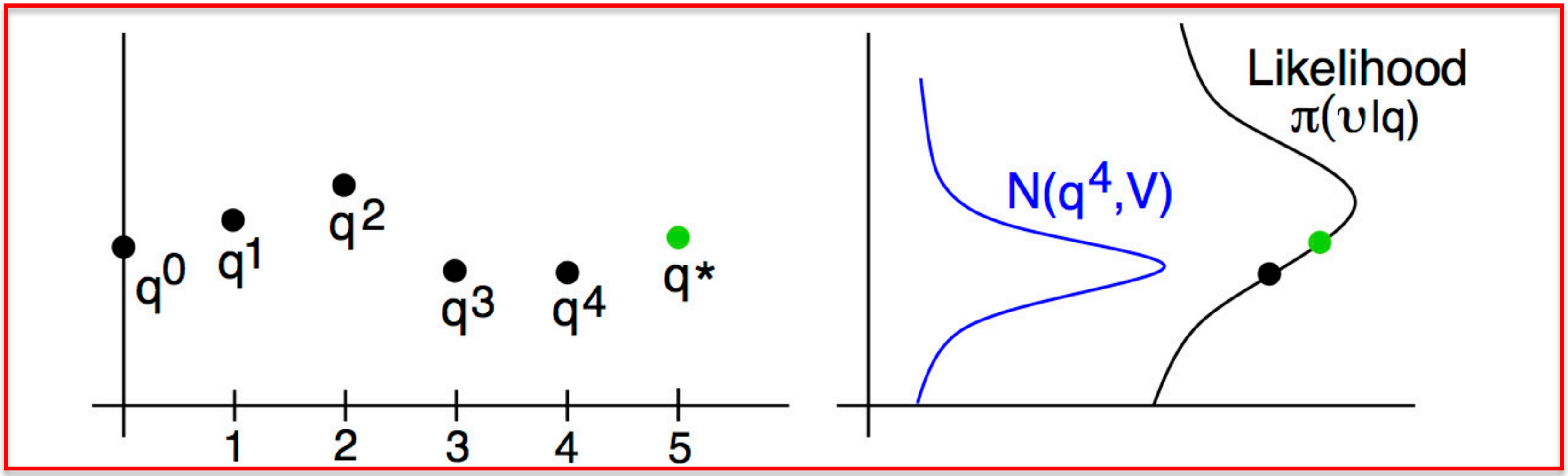
$$SS_{q^*} = \sum_{i=1}^{N} \upsilon_i - \psi(P_i, q^*)]^2$$

$$\pi(\upsilon|q) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}$$





    (c) Accept q^* with probability dictated by likelihood



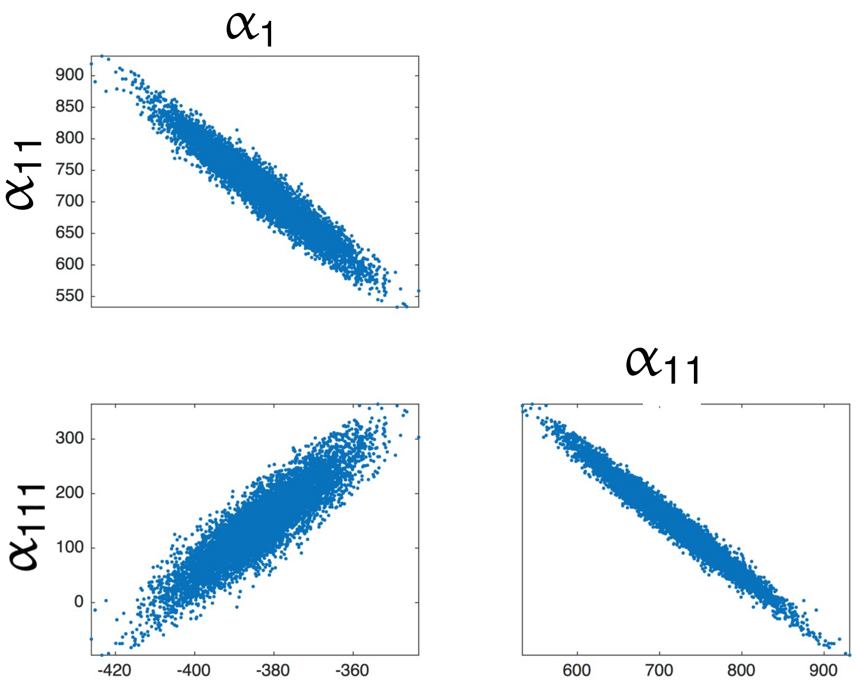
Delayed Rejection Adaptive Metropolis (DRAM)

Example: Helmholtz energy with 3 parameters

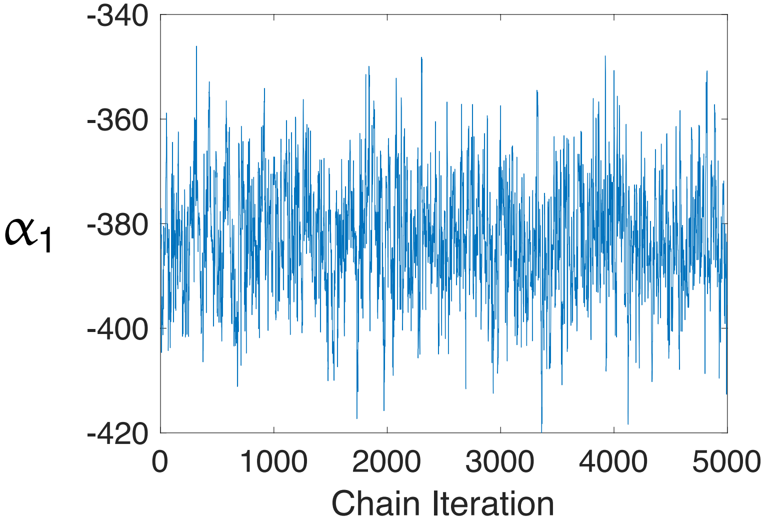
$$\psi(P,q) = \underline{\alpha_1} P^2 + \underline{\alpha_{11}} P^4 + \underline{\alpha_{111}} P^6$$

Note: Similar results for α_{11} and α_{111}

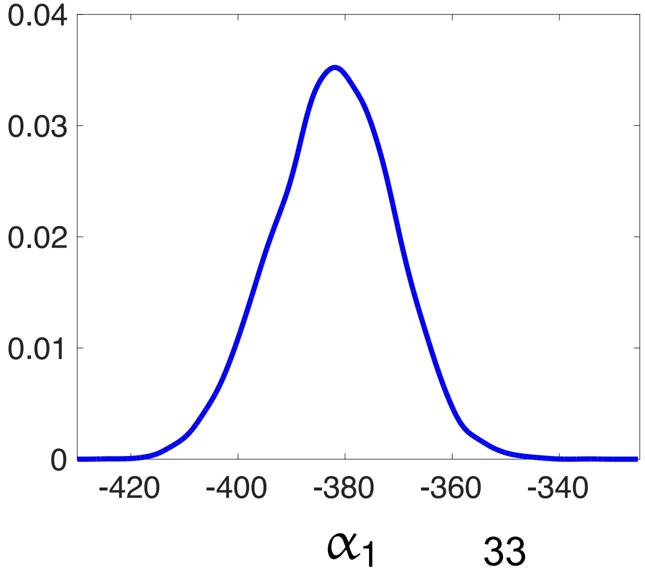
Pairwise Plots: Quantify correlation



Chain for α_1 with 5000 samples



Marginal density for α_1



Bayesian Model Calibration – HIV Example

Model:

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

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

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

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

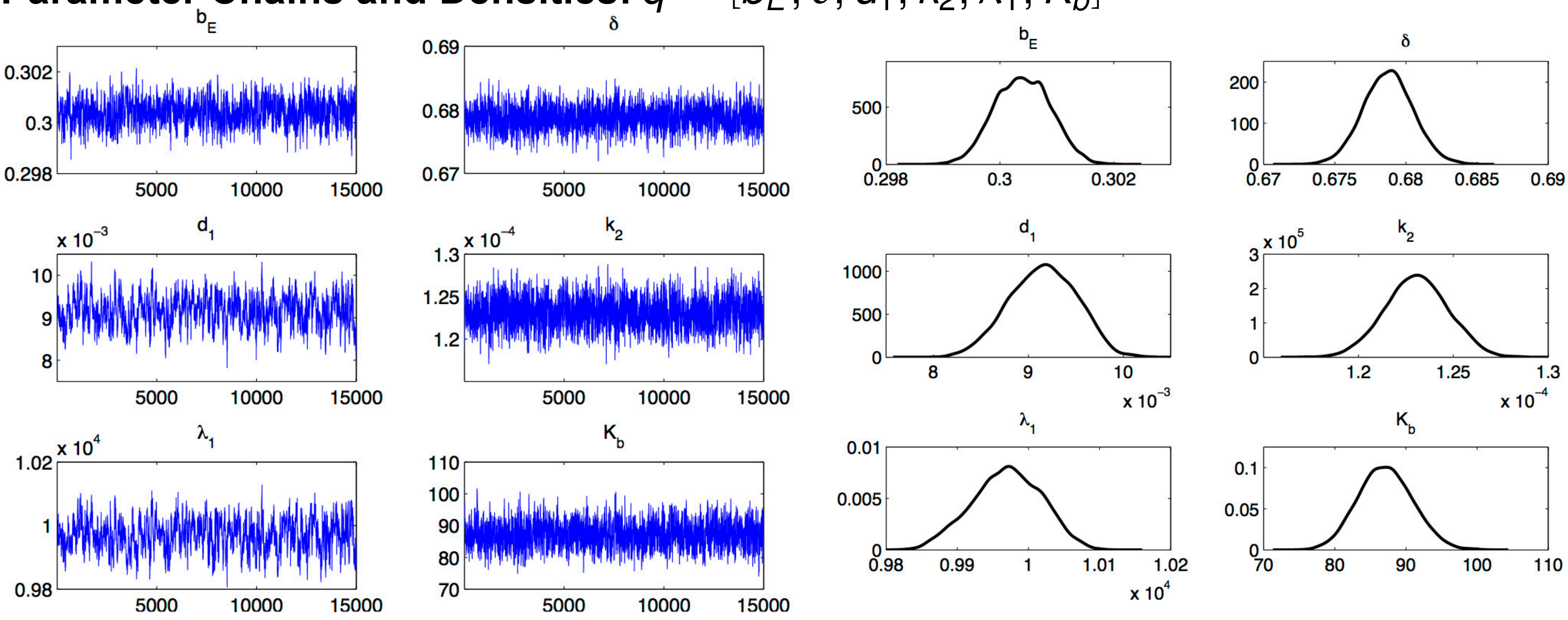
$$\dot{V} = N_T \underline{\delta}(T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1-f\varepsilon)\rho_2 k_2 T_2]V$$

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

Verification: Why do we trust results??

- Compare results from different algorithms; e.g., DRAM and Gibbs

Parameter Chains and Densities: $q = [b_E, \delta, d_1, k_2, \lambda_1, K_b]$



Example: Wetland Methane Model

Authors: Susiluoto, Raivonen, Backman, Laine, Makela, Peltola, Vesala, Aalto, *Geoscientific Model Development, 11, pp. 1199-1228, 2018.*

Objectives: Methane emissions from natural wetlands highly uncertainty

- What is effect of climate change?
- How much uncertainty is there in model parameters that control physical processes?
- How do parameters and wetland behavior react to environmental changes?

Model: HelsinkI Model of MEthane buiLd-up and emIssion for peatlands (HIMMELI)

- Incorporates process descriptions for methane production from anaerobic respiration, oxygen consumption, and carbon dioxide production;
- Submodel in regional and biosphere models.

Initial Calibration Parameters: 14

Wetland Methane Model

Subset of Governing Relations:

$$T_X(t,z) = Q_X^{diff} + Q_X^{plant} + Q_X^{ebu}$$

$$\frac{\partial[CH_4]}{\partial t}(t,z) = -T_{CH_4} + R_{CH_4}^{exu} + R_{CH_4}^{peat} - R_{CH_4}^{oxid}$$

$$\frac{\partial[O_2]}{\partial t}(t,z) = -T_{O_2} - R_{aerob}^{peat} - R_{CO_2}^{exu} - 2R_{CH_4}^{oxid}$$

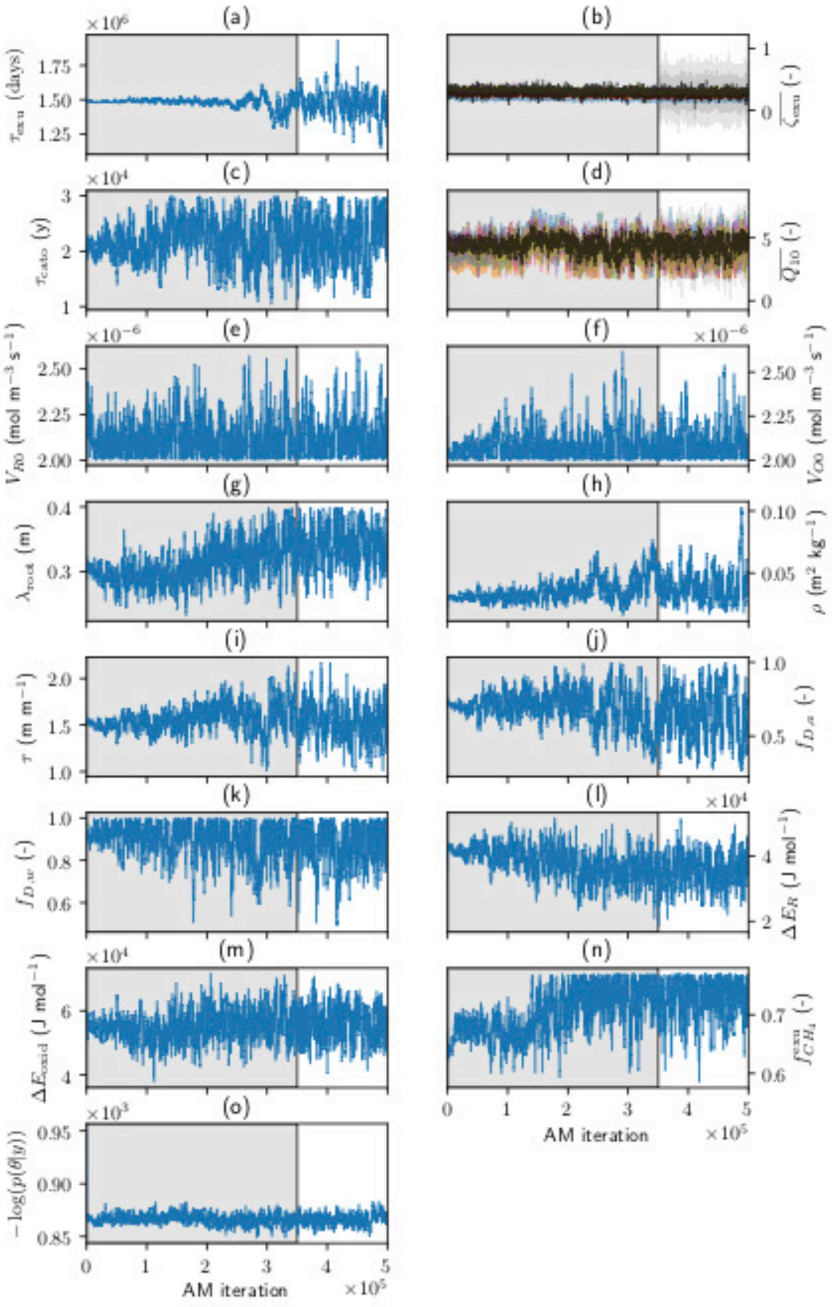
$$\frac{\partial[CO_2]}{\partial t}(t,z) = -T_{CO_2} + R_{CO_2}^{exu} + R_{CO_2}^{peat} + R_{CH_4}^{oxid} + R_{aerob}^{peat}$$

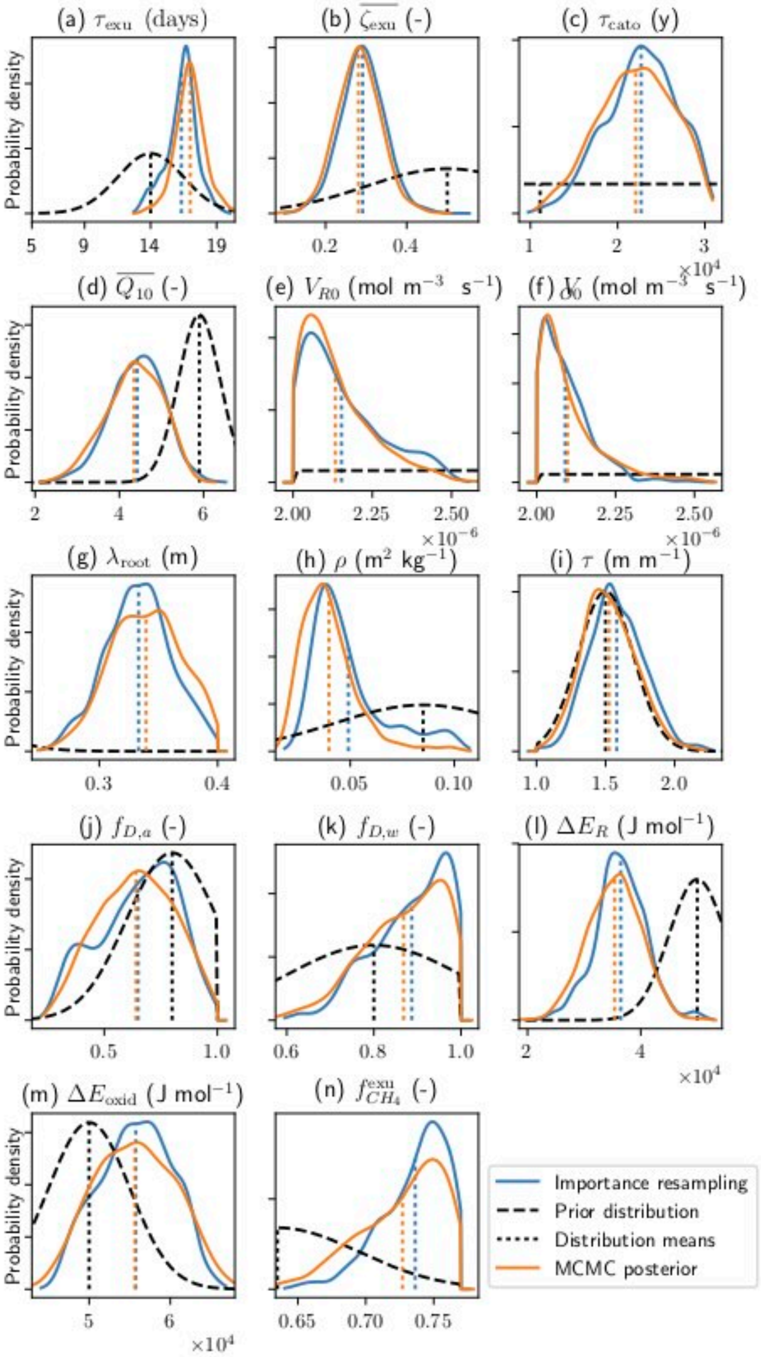
Initial Calibration Parameters: 14

MCMC Techniques:

- Metropolis-within-Gibbs for sampling hierarchical parameters
- Modified DRAM used to sample model parameters
- Model and algorithm parallelization and tuning reduced computation times from months to days

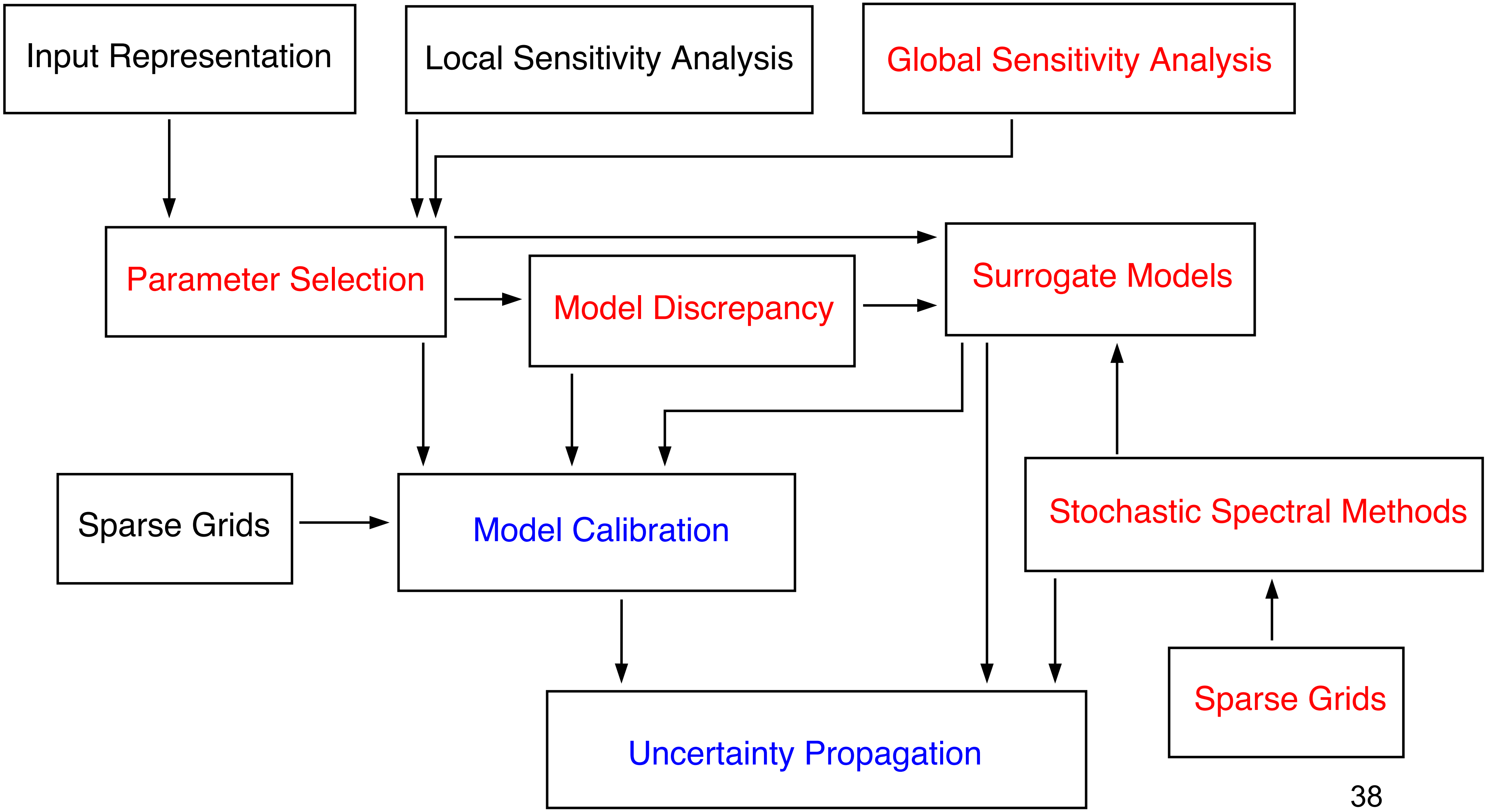
Wetland Methane Model: Chains and Marginal Distributions





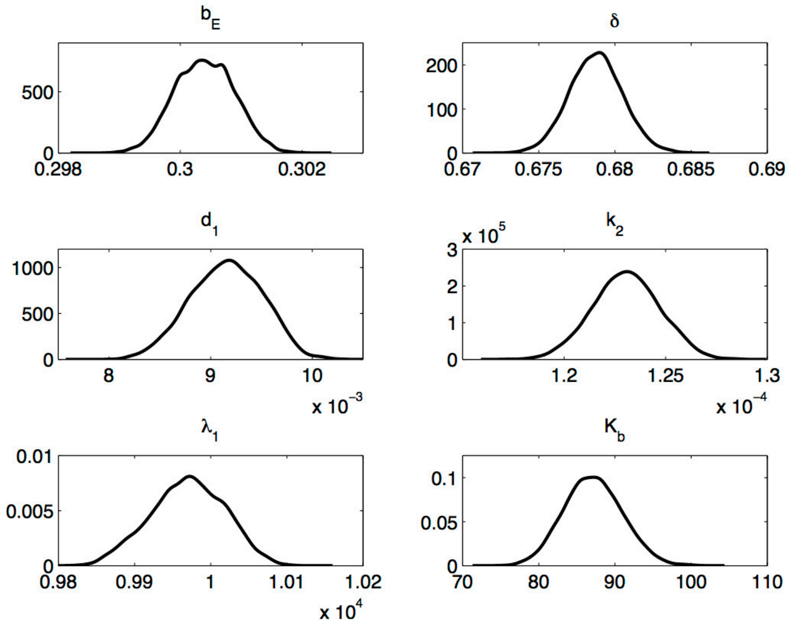
Steps in Uncertainty Quantification

Note: Uncertainty quantification requires synergy between statistics, applied mathematics and domain sciences.



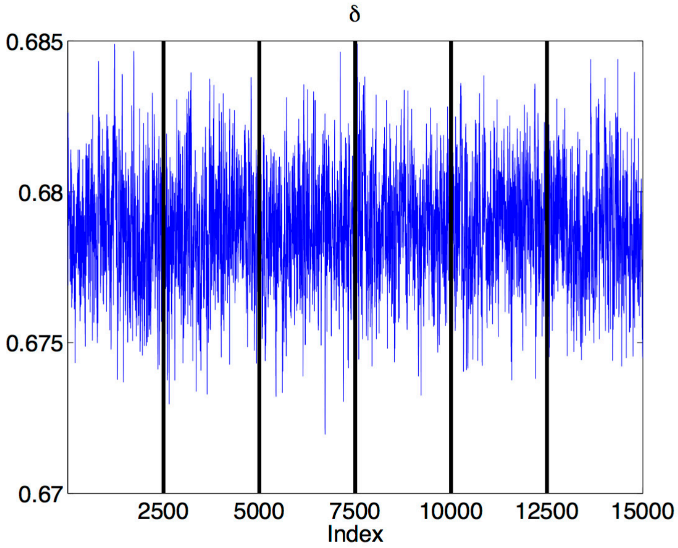
Propagation of Uncertainty in Models – HIV Example

Parameter Densities:

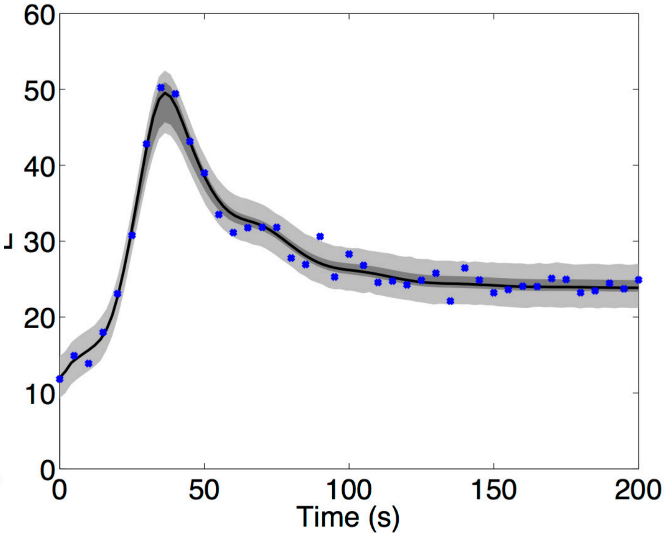


Techniques:

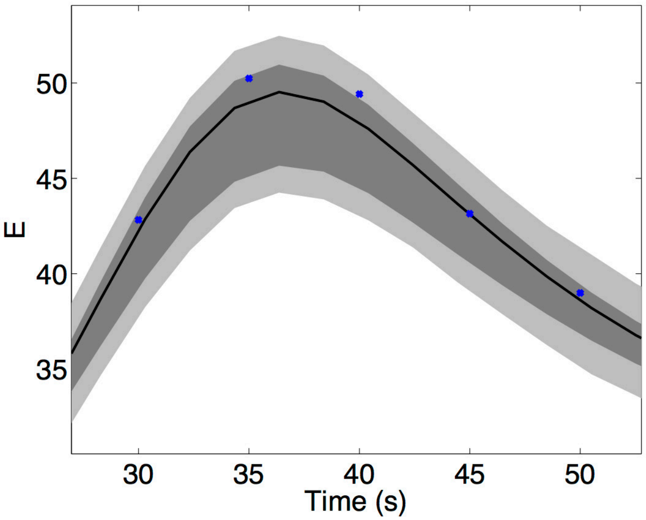
- Sample from parameter densities to construct prediction intervals for QoI.
- Slow convergence rate $\mathcal{O}(1/\sqrt{M})$
- 100-fold more evaluations required to gain additional place of accuracy.
- Significant numerical analysis used to efficiently propagate densities.



Samples from Chain



Data, Credible Intervals and Prediction Intervals

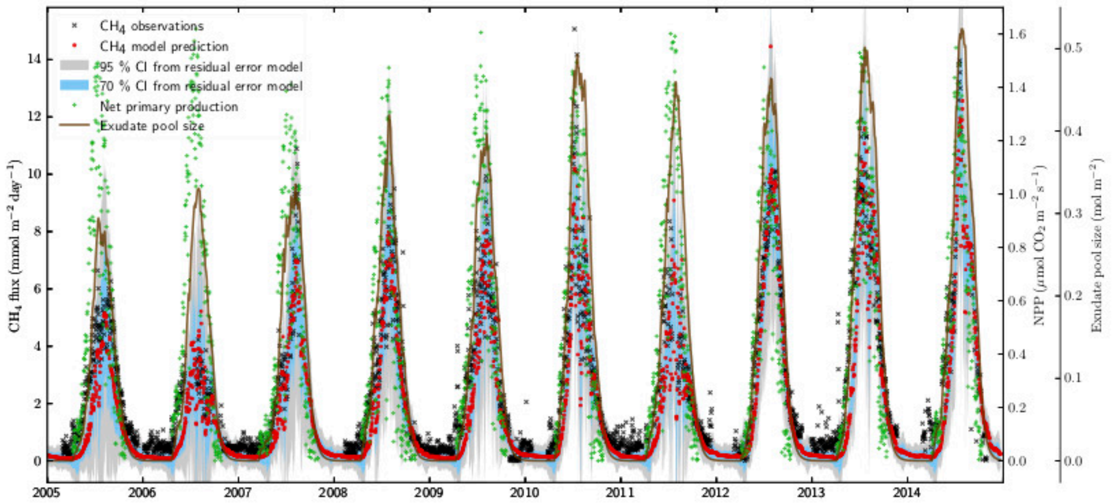


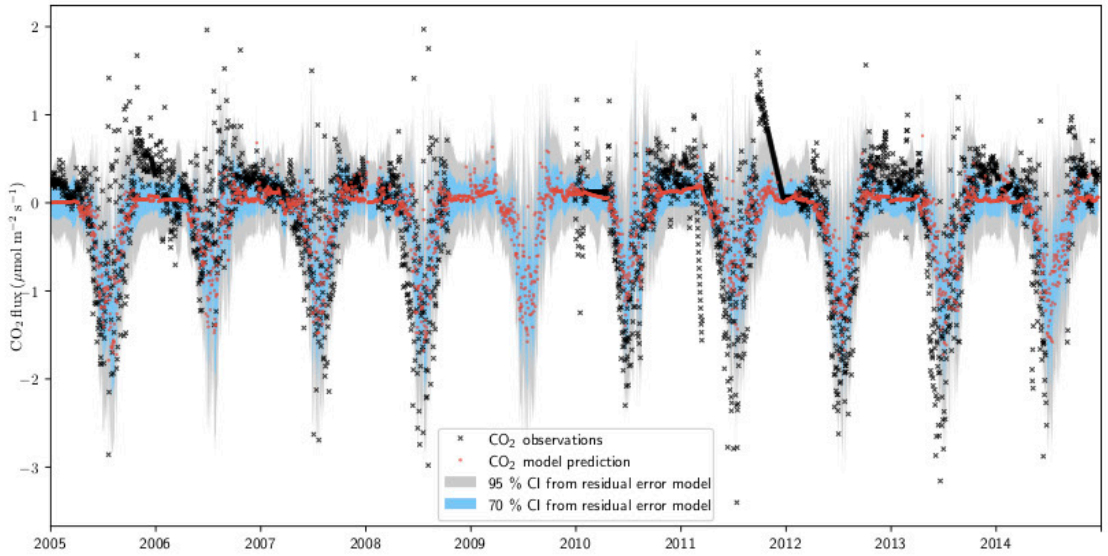
Non-Gaussian Credible and Prediction Intervals

Wetland Methane Model

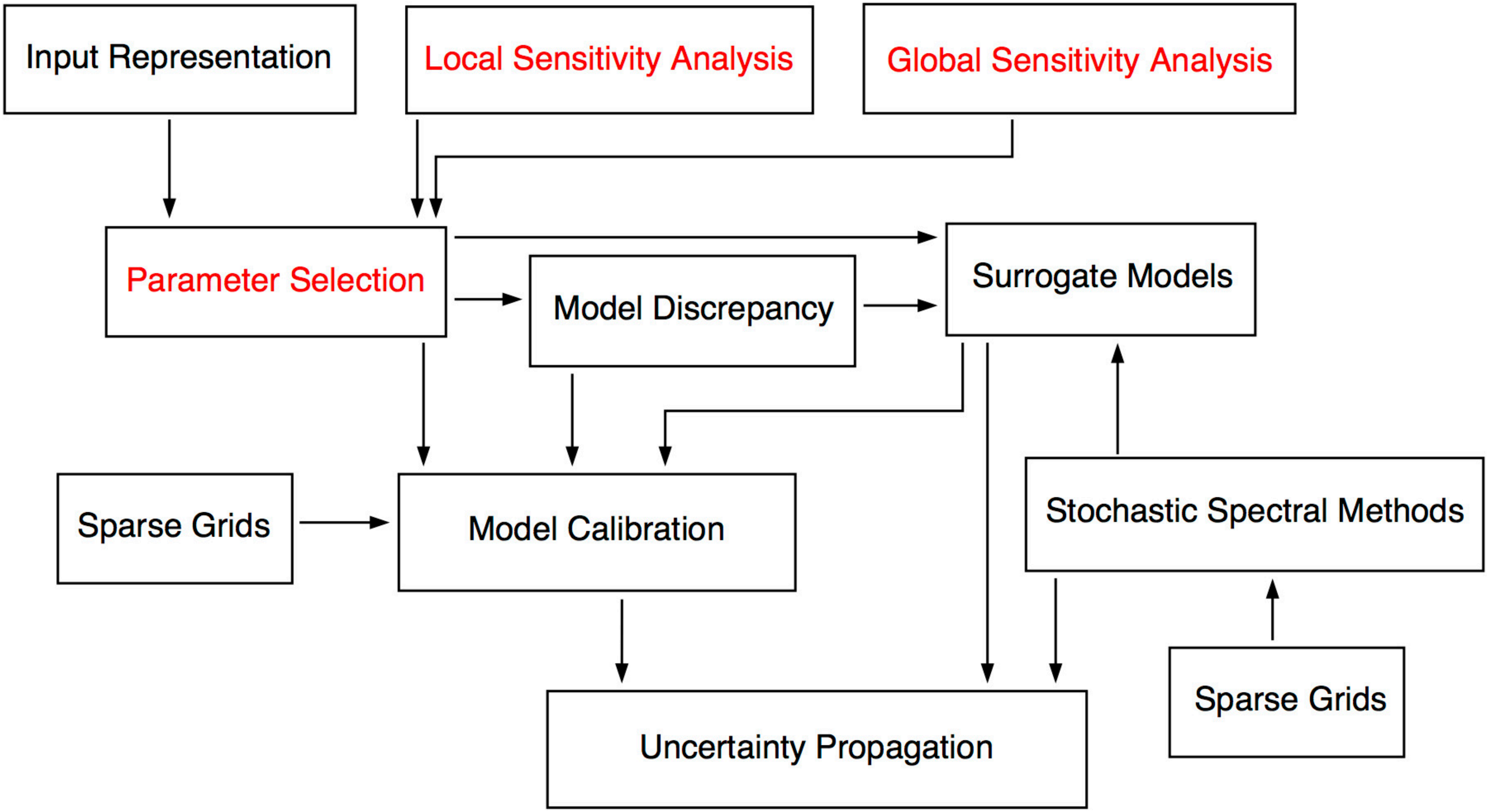
Authors: Susiluoto, Raivonen, Backman, Laine, Makela, Peltola, Vesala, Aalto, *Geoscientific Model Development, 11, pp. 1199-1228, 2018.*

Observation: Credible intervals consistent for methane and carbon dioxide





Steps in Uncertainty Quantification



Input Representation

Local Sensitivity Analysis

Global Sensitivity Analysis

Parameter Selection

Model Discrepancy

Surrogate Models

Sparse Grids

Model Calibration

Stochastic Spectral Methods

Uncertainty Propagation

Sparse Grids

Parameter Selection Techniques: Required for models with high-dimensional input spaces and/or unidentifiable or noninfluential inputs.

Parameter Selection Techniques

First Issue: Parameters often *not identifiable* in the sense that they are not uniquely determined by the data.

Example 1: Spring model

$$\underline{m}\frac{dy}{dt^2} + \underline{k}y = 0$$

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

Solution: $y(t,q) = y_0 \cos\left(\underline{\sqrt{k/m}} \cdot t\right)$

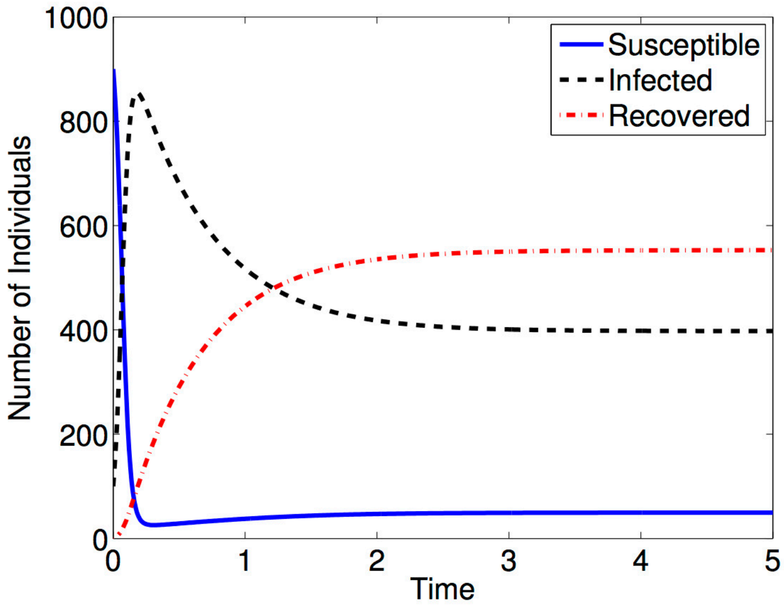
Note: q = [k,m] not jointly identifiable

Example 2: SIR Model

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

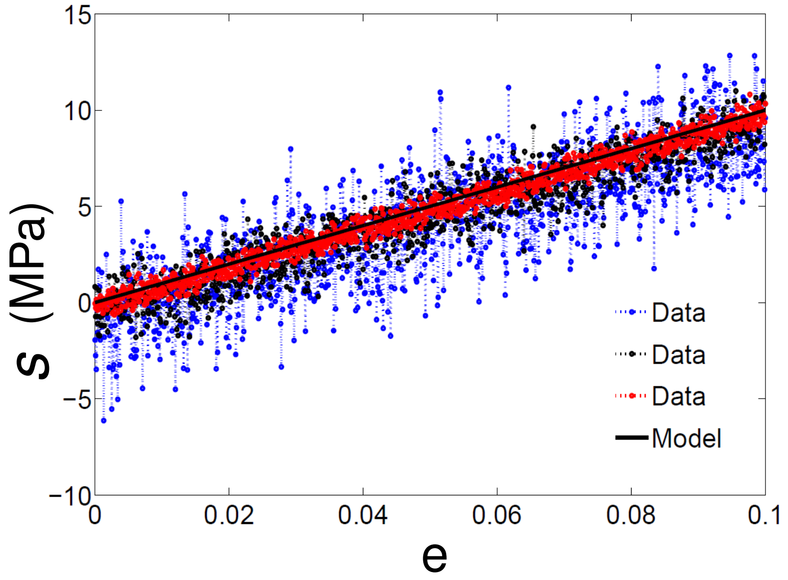
$$\frac{dI}{dt} = \underline{\gamma k}IS - (r+\delta)I \quad , \ I(0) = I_0$$

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



Note: Parameters $q = [\gamma, k, r, \delta]$ not uniquely determined by data

Sensitivity Analysis: Motivation



Example: Linear constitutive relation

$$s = Ee + c\frac{de}{dt}$$

Nominal Values: $E = 100$, $c = 0.1$

$$e = 0.001 \; , \; \frac{de}{dt} = 0.1$$

Question: To which parameter E or c is stress s most sensitive?

Local Sensitivity Analysis:

$$\frac{\partial s}{\partial E} = e = 0.001$$

$$\boxed{\frac{\partial s}{\partial c} = \frac{de}{dt} = 0.1}$$

Conclusion: Model most sensitive to damping parameter c

Limitations:

- Does not accommodate potential uncertainty in parameters.
- Does not accommodate potential correlation between parameters.
- Sensitive to units and magnitudes of parameters.
- Local but can be made quasi-global through Morris sampling.

Global Sensitivity Analysis: Motivation

Example: Linearly parameterized model

$$s = Ee + c\frac{de}{dt}$$

Nominal Values: $E = 100, \; c = 0.1$

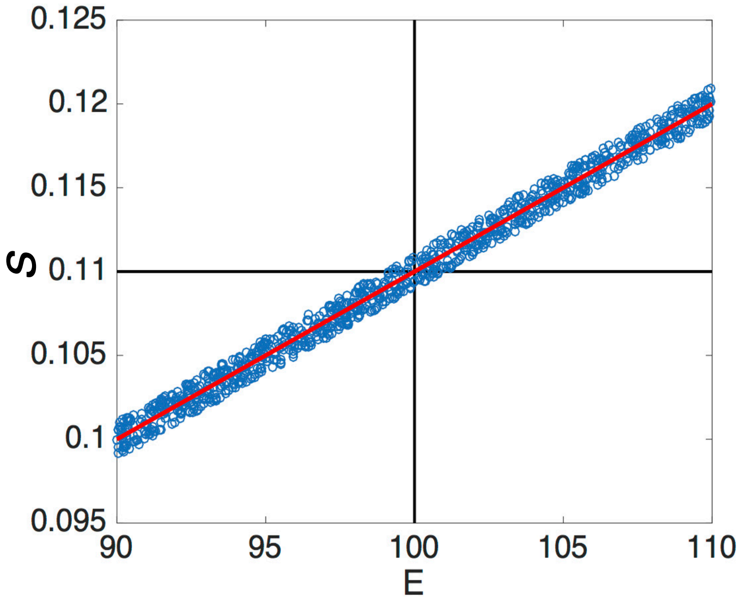
Uncertainty: 10% of nominal values

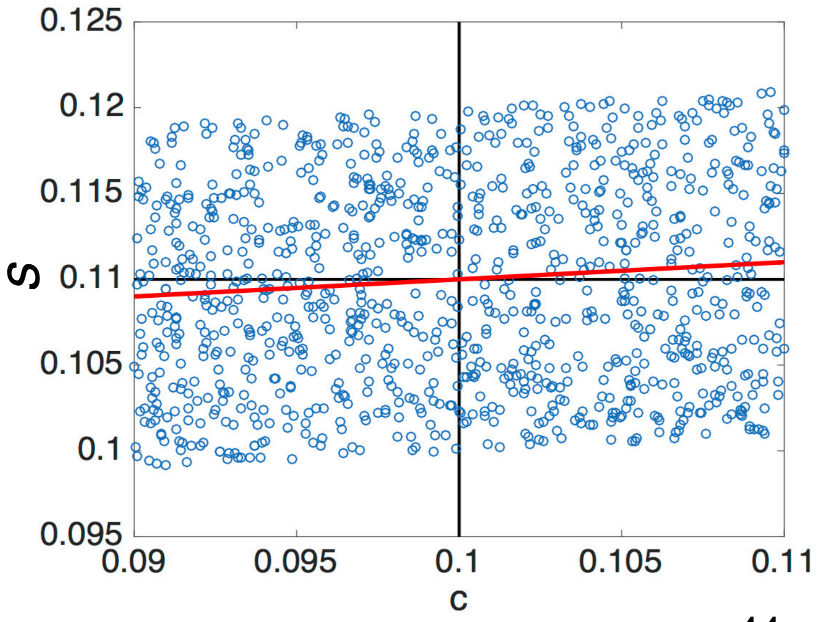
$$E \sim \mathcal{U}(90, 110) \; , \; c \sim \mathcal{U}(0.09, 0.11)$$

Local Sensitivities:

$$\frac{\partial s}{\partial E} = e = 0.001$$

$$\frac{\partial s}{\partial c} = \frac{de}{dt} = 0.1$$





Global Sensitivity: E is more influential

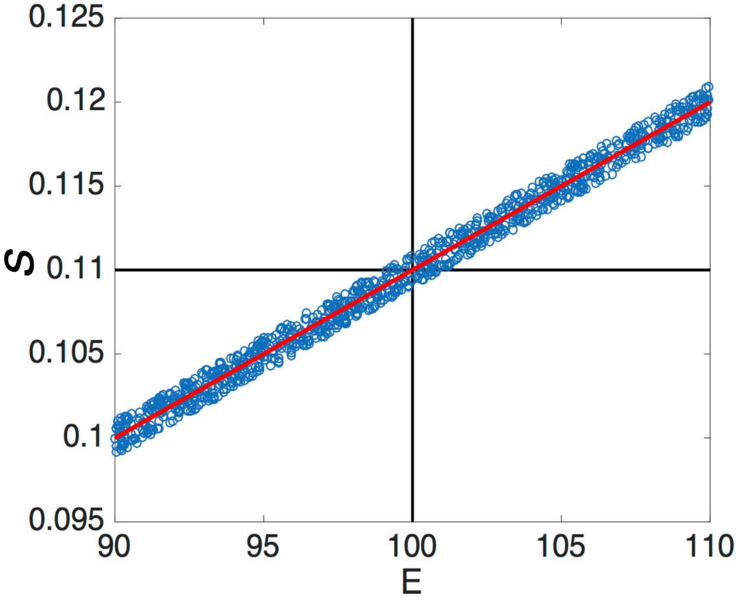
Variance-Based Methods

Sobol Representation: Standard assumption: independent parameters

Take

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

Analogy: Taylor or Fourier series



With appropriate assumptions,

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

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

Variances:

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

$$D = \mathrm{var}(Y)$$

Sobol Indices: $S_i = \frac{D_i}{D}$

Statistical Interpretation:

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

Global Sensitivity Analysis: Motivation

Example: Linearly model

$$s = Ee + c\frac{de}{dt}$$

Nominal Values: $E = 100, \;\; c = 0.1$

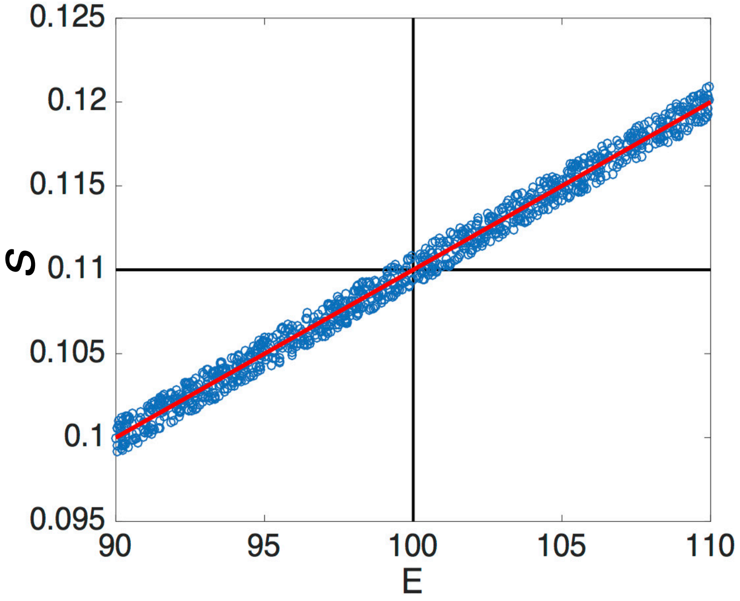
Uncertainty: 10% of nominal values

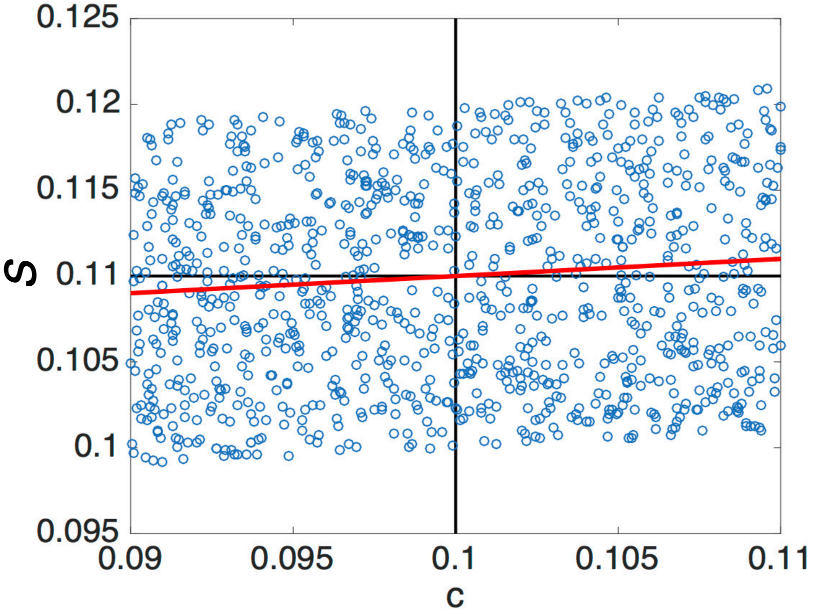
$$E \sim \mathcal{U}(90, 110) \; , \;\; c \sim \mathcal{U}(0.09, 0.11)$$

Statistical Interpretation:

$$D_i = \text{var}[\mathbb{E}(Y|q_i)]$$

$$S_i = \frac{\text{var}[\mathbb{E}(Y|q_i)]}{\text{var}(Y)}$$

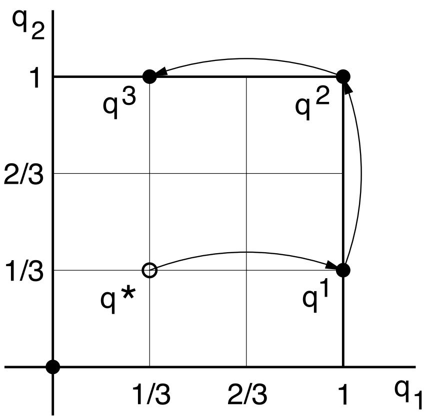




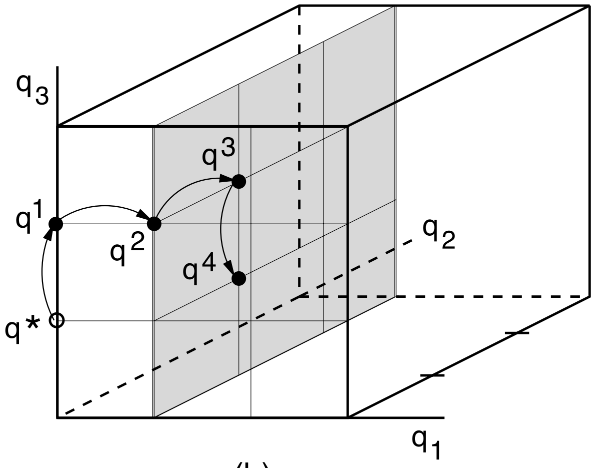
Global Sensitivity: E is more influential

Quasi-Global Sensitivity Analysis: Morris Screening

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



(a)



(b)

Elementary Effect:

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

Global Sensitivity Measures: r samples

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

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

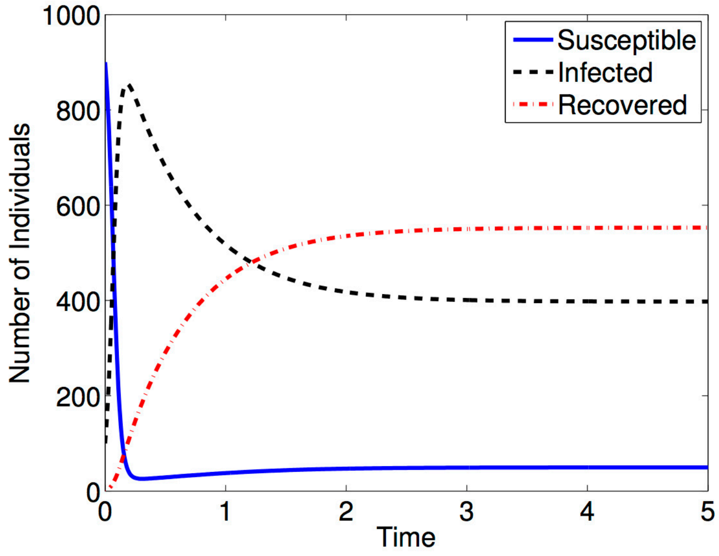
Global Sensitivity Analysis: SIR Disease Model

SIR Model:

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

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

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



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

Assumed Parameter Distribution:

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

Infection Coefficient	Interaction Coefficient	Recovery Rate	Birth/death Rate

Response:

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

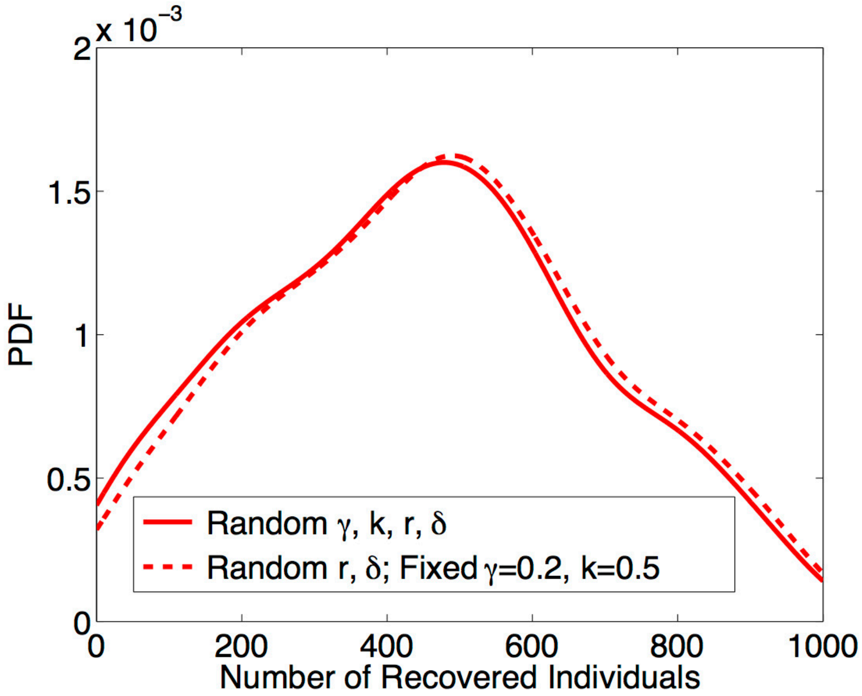
SIR Disease Example

Global Sensitivity Measures:

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

Influential Parameters

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



Note: Can fix non-influential parameters γ, k

Concluding Remarks

Notes:

- UQ requires a synergy between statistics, applied mathematics, engineering and science.
- Model calibration, model selection, uncertainty propagation and experimental design natural in a Bayesian framework.
- Goal: Predict model responses with quantified and reduced uncertainties.
- Parameter selection critical to isolate identifiable and influential parameters.
- Surrogate models critical for computationally intensive simulation codes.
- Codes and packages: MATLAB, Python, R, nanoHUB, Sandia Dakota.
- *Prediction is very difficult, especially if it's about the future*. Niels Bohr.