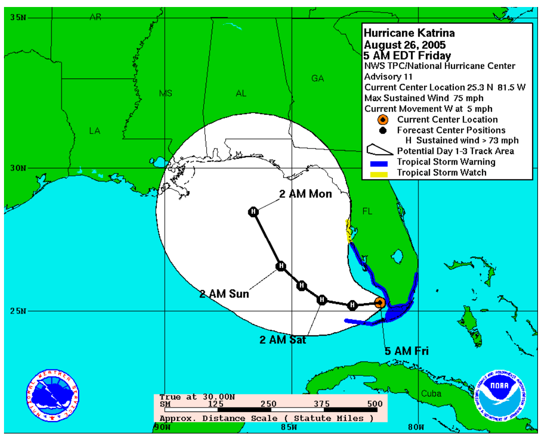
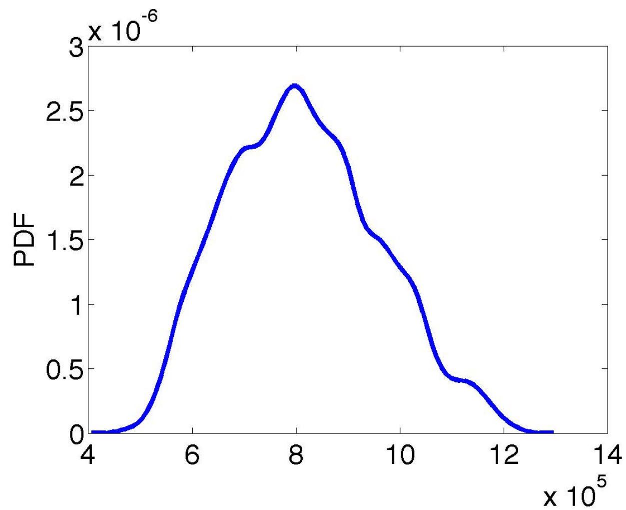
Bayesian Inference and Uncertainty Propagation for Physical and Biological Models

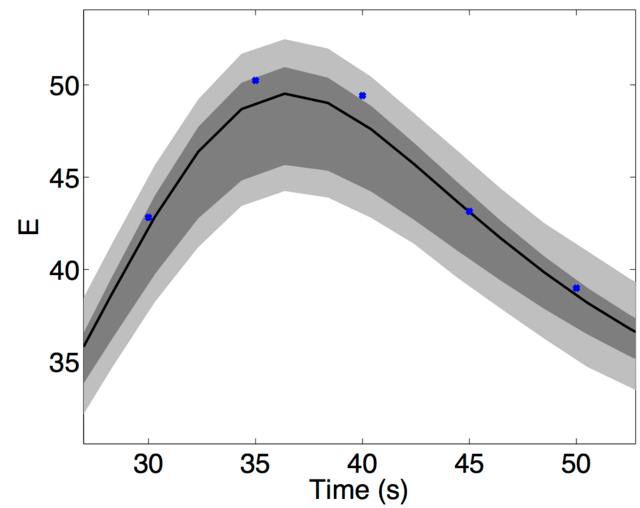
Ralph C. Smith

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)

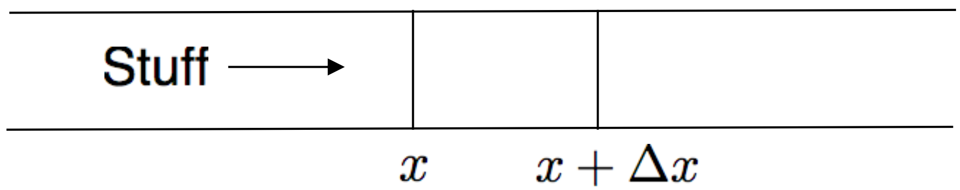
NNSA Consortium for Nonproliferation Enabling Capabilities (CNEC)

National Science Foundation (NSF)

Air Force Office of Scientific Research (AFOSR)

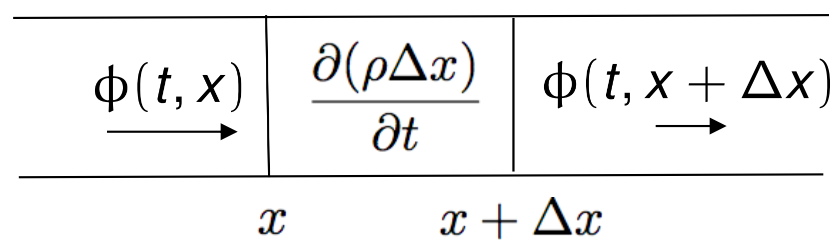
Modeling Strategy

General Strategy: Conservation of stuff



$$\frac{dStuff}{dt} = \text{Stuff in - Stuff out + Stuff created - Stuff destroyed}$$

Continuity Equation:



$$\frac{\partial(\rho \Delta x)}{\partial t} = \phi(t, x) - \phi(t, x + \Delta x)$$

$$\Rightarrow \lim_{\Delta x \to 0} \frac{\partial \rho}{\partial t} = \lim_{\Delta x \to 0} \frac{\phi(t, x) - \phi(t, x + \Delta x)}{\Delta x}$$

$$\boxed{\Rightarrow \frac{\partial \rho}{\partial t} + \frac{\partial \phi}{\partial x} = 0}$$

Density: $\rho(t, x)$ - Stuff per unit length or volume

Rate of Flow: $\phi(t, x)$ - Stuff per second

More Generally:

$$\boxed{\Rightarrow \frac{\partial \rho}{\partial t} + \frac{\partial \phi}{\partial x} = \text{Sources - Sinks}}$$

Example 1: Weather Models

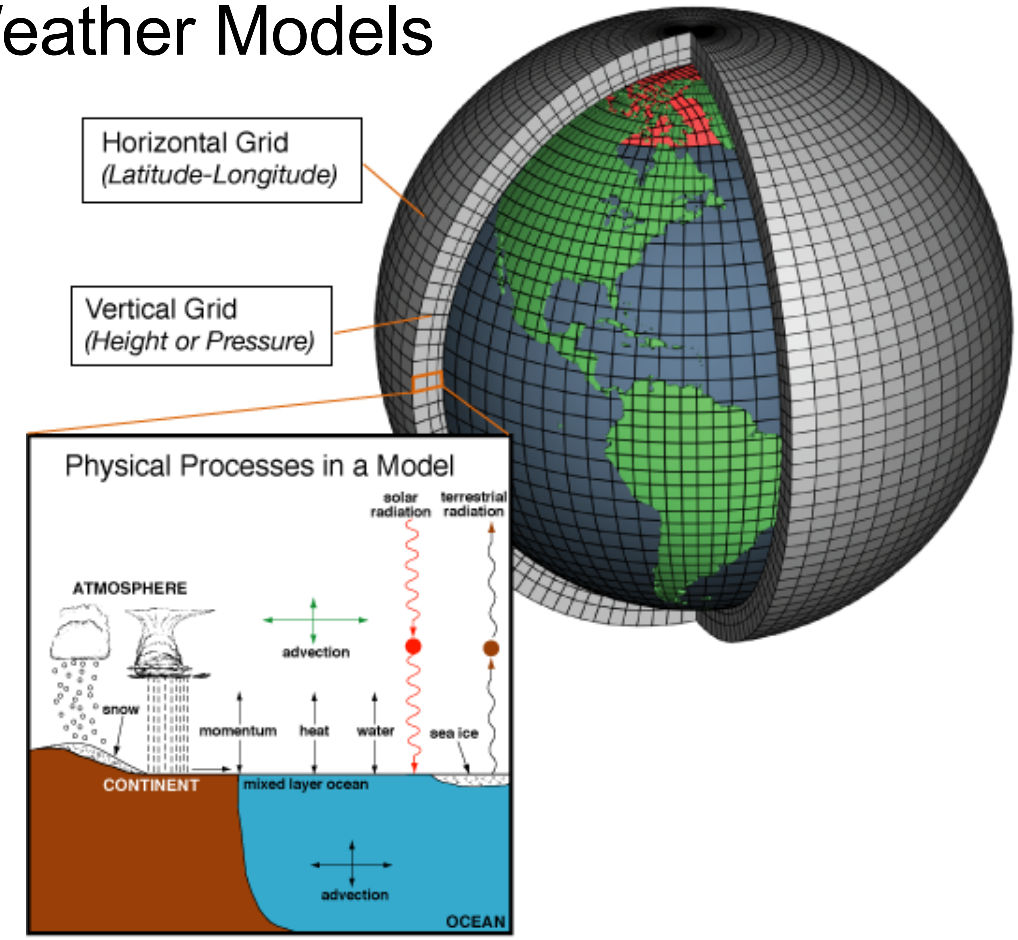
Challenges:

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

• Models and inputs contain uncertainties;

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

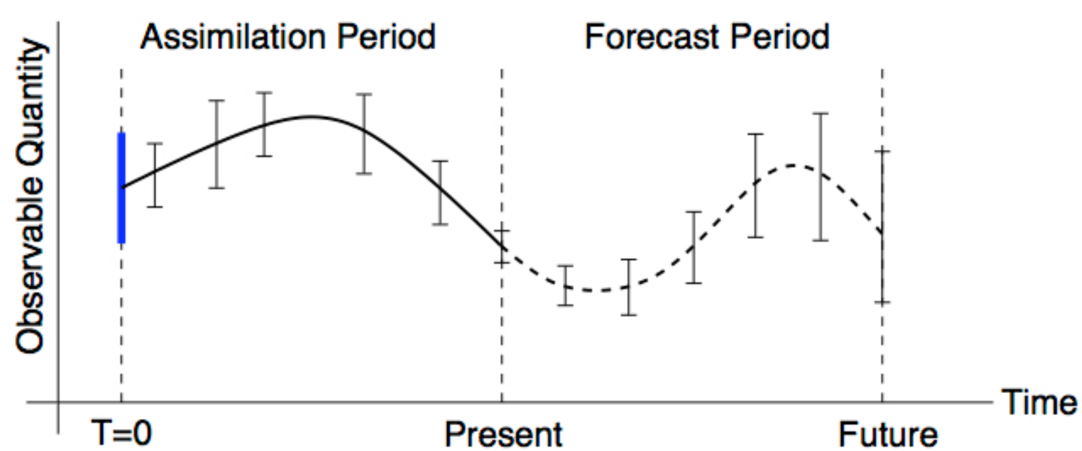
• Sensors positions may be uncertain; e.g., weather balloons, ocean buoys.



Goal:

• Assimilate data to quantify uncertain initial conditions and parameters;

• Make predictions with quantified uncertainties.



Equations of Atmospheric Physics

Conservation Relations:

$$\boxed{\frac{\partial \rho}{\partial t} + \frac{\partial \phi}{\partial x} = \text{Sources - Sinks}}$$

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

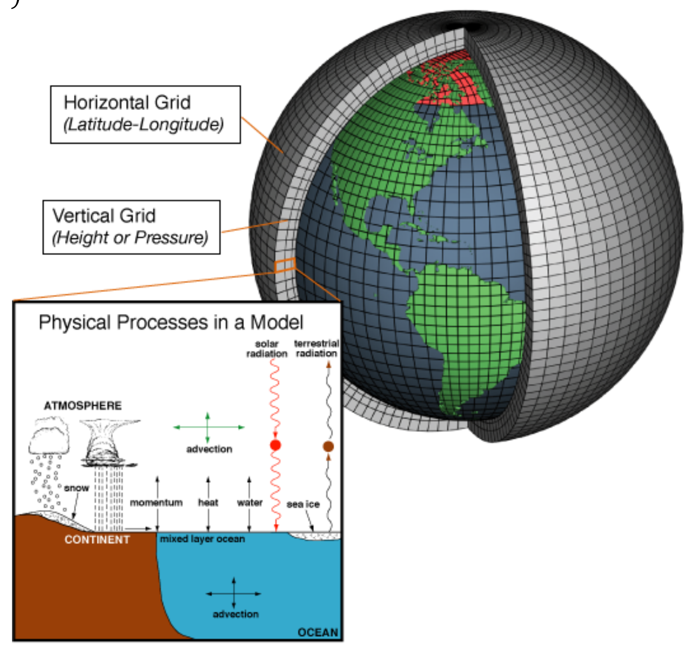
Momentum $$\frac{\partial v}{\partial t} = -v \cdot \nabla v - \frac{1}{\rho}\nabla p - g\hat{k} - 2\Omega \times v$$

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

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.,

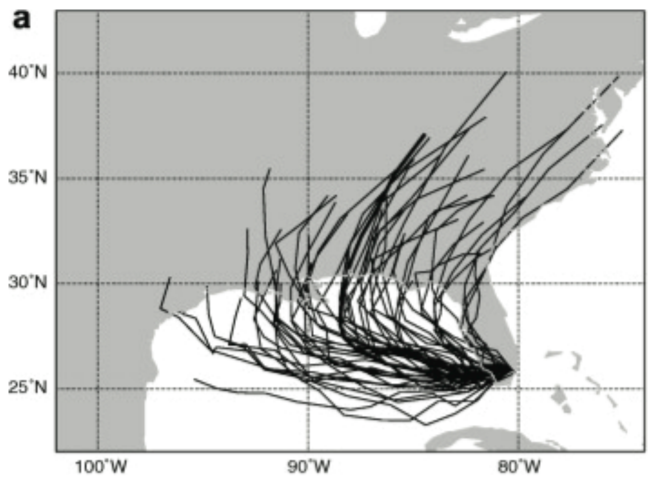
$$S_{m_2} = S_1 + S_2 + S_3 - S_4$$

where

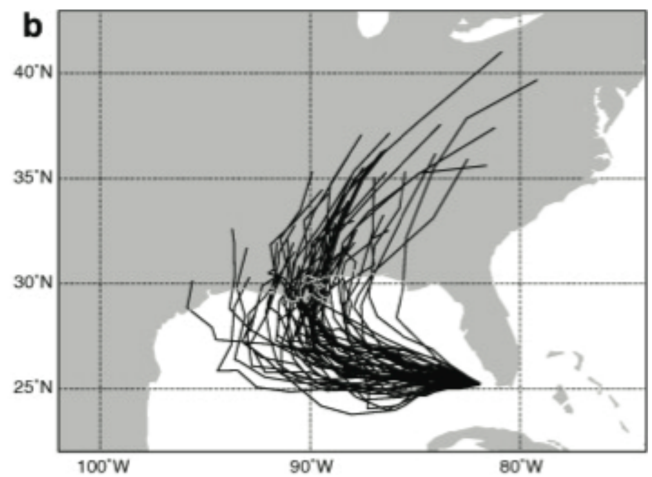
$$S_1 = \bar{\rho}(m_2 - m_2^*)^2 \left[\underline{1.2 \times 10^{-4}} + \left(\underline{1.569 \times 10^{-12}} \frac{n_r}{d_0(m_2 - m_2^*)} \right) \right]^{-1}$$

Ensemble Predictions

Ensemble Predictions:

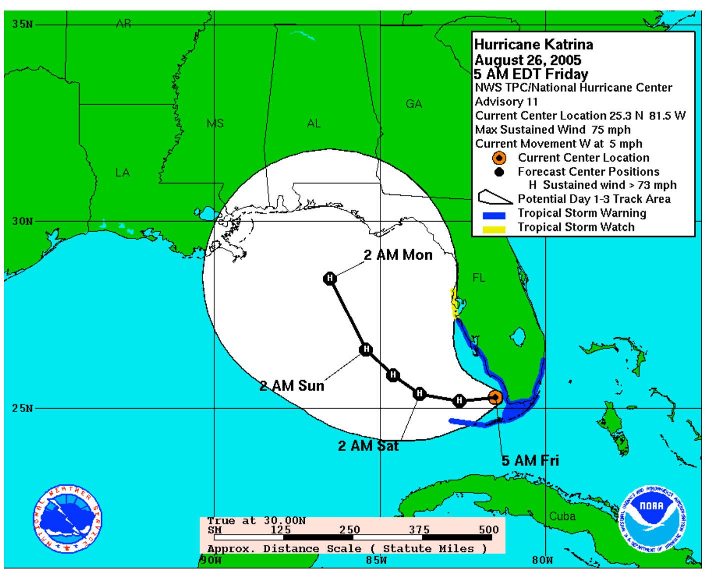


00 UTC on August 26, 2005



12 UTC on August 26, 2005

Cone of Uncertainty:



General Questions:

- What is expected rainfall on May 24?
- 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, *Geosci. Model Dev. 11, 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: HIV Model for Characterization and Control Regimes

HIV Model:

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

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

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

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

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

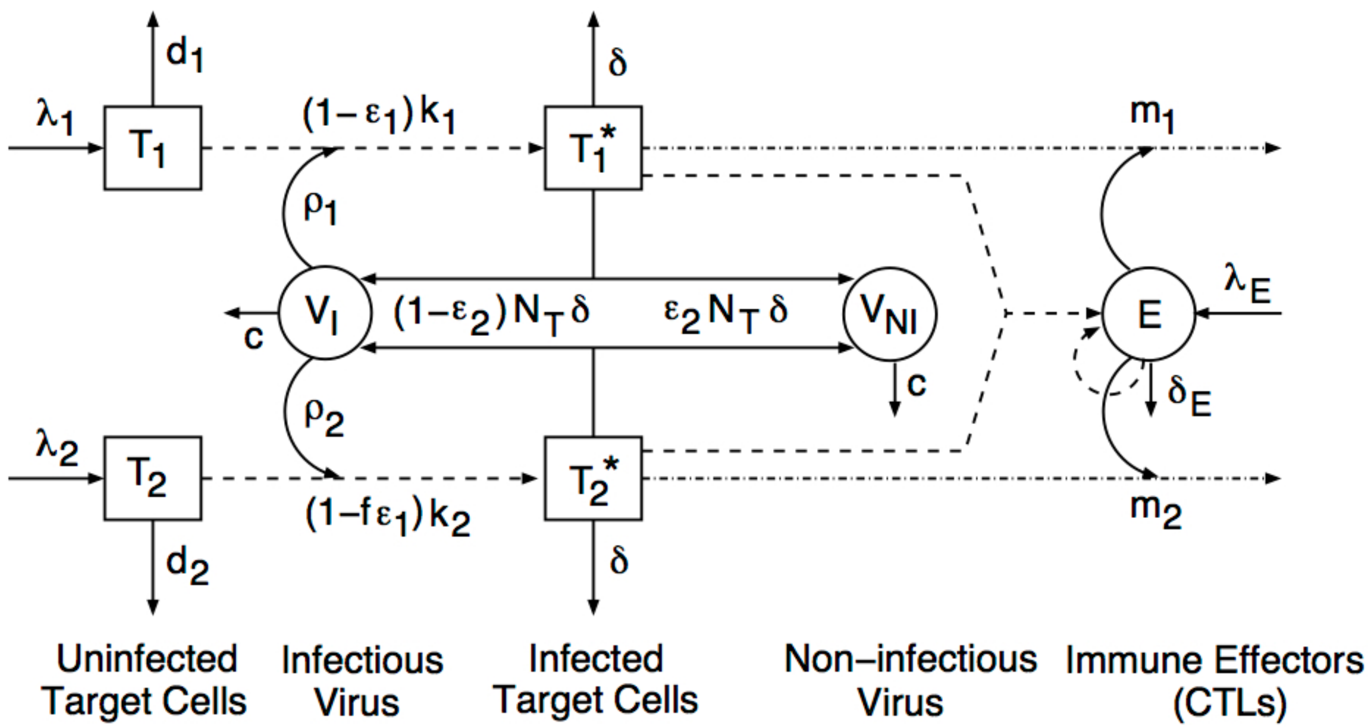
$$\dot{E} = \lambda_E + \frac{b_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_b}E - \frac{d_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_d}E - \delta_E E$$

Notes: 21 parameters

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

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

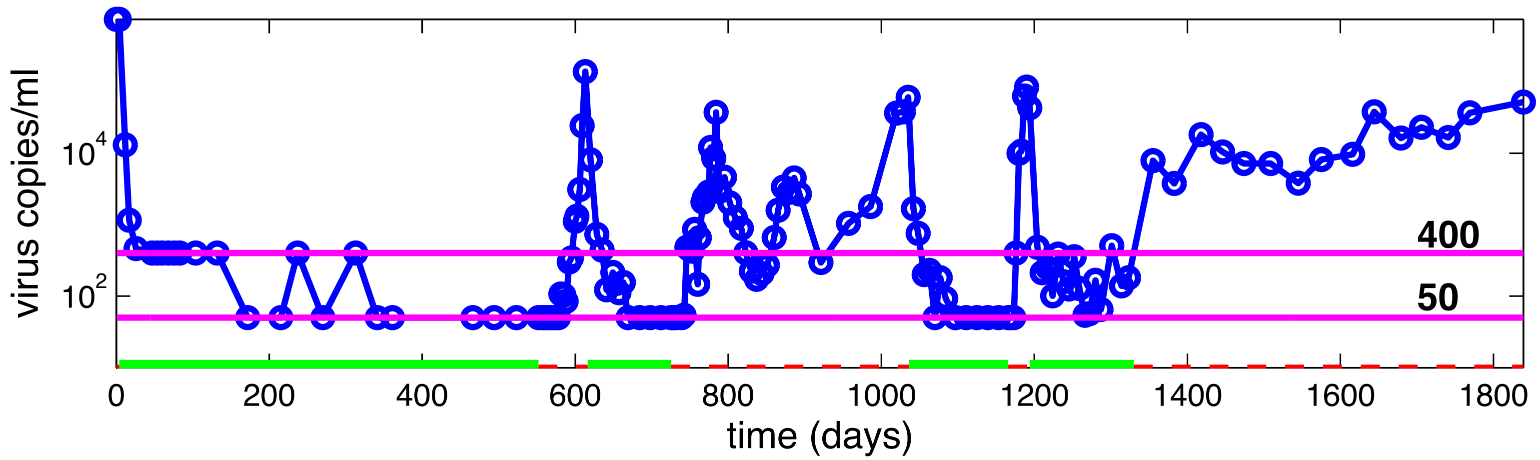
Compartments:



Example: 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

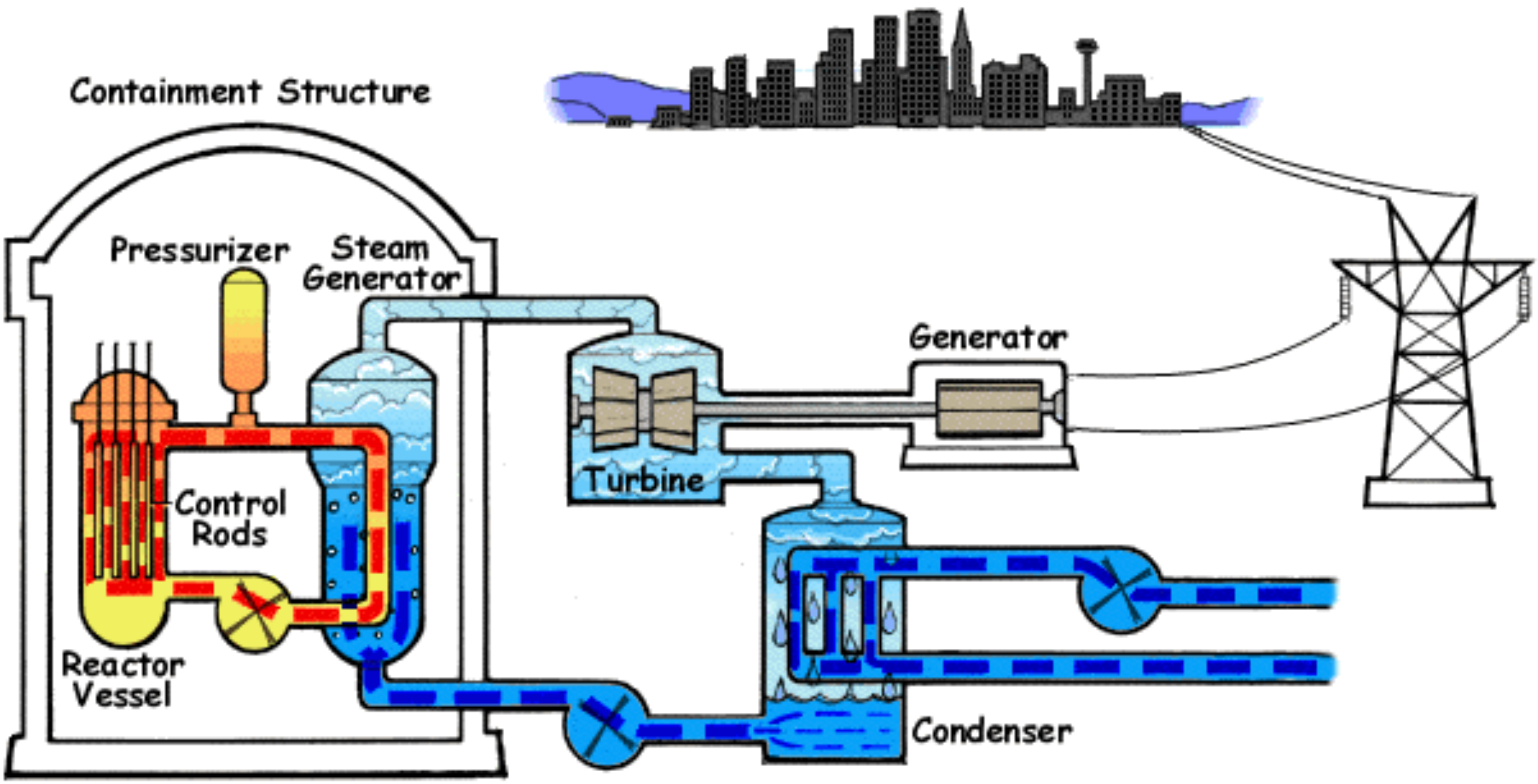


UQ Questions:

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

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

Example 3: Pressurized Water Reactors (PWR)



Models:

- Involve neutron transport, thermal-hydraulics, chemistry, fuels
- Inherently multi-scale, multi-physics -- Must be incorporated in surrogate models

Objective: Develop Virtual Environment for Reactor Applications (VERA)

Example: Pressurized Water Reactors (PWR)

3-D Neutron Transport Equations:

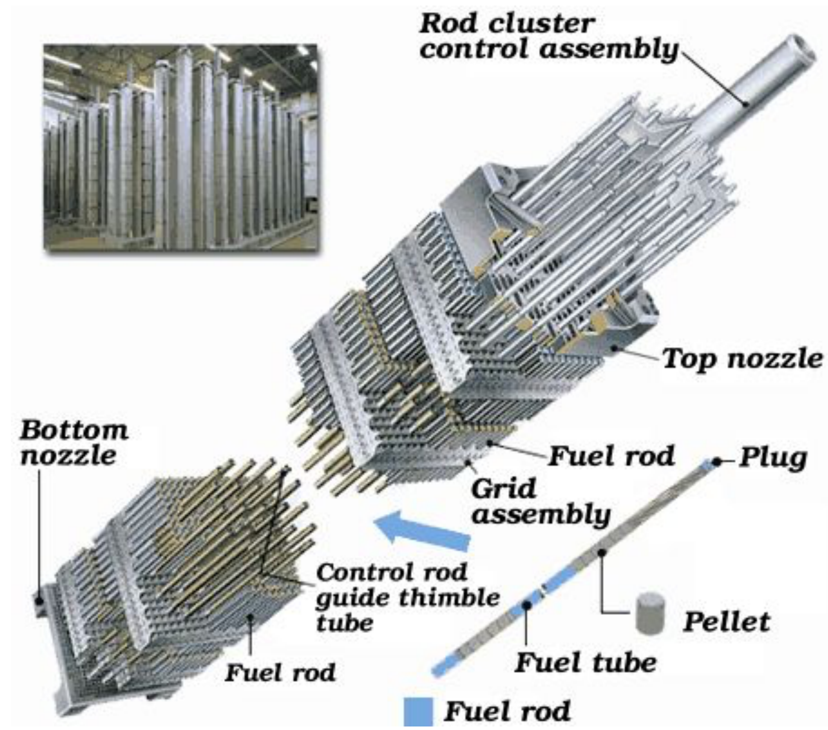
$$\frac{1}{|v|}\frac{\partial\varphi}{\partial t} + \Omega\cdot\nabla\varphi + \Sigma_t(r,E)\varphi(r,E,\Omega,t)$$

$$= \int_{4\pi} d\Omega' \int_0^\infty dE' \Sigma_s(E' \to E, \Omega' \to \Omega)\varphi(r,E',\Omega',t)$$

$$+ \underline{\frac{\chi(E)}{4\pi}} \int_{4\pi} d\Omega' \int_0^\infty dE' \underline{\nu(E')}\,\underline{\Sigma_f(E')}\varphi(r,E',\Omega',t)$$

Challenges:

- Very large number of inputs; e.g., 10,000.
- ORNL Code SCALE: Can take hours to run.
- Time-dependent surrogate models must accommodate PDE structure.
- Predicting future requires extrapolatory or out-of-data predictions; one must address model discrepancy to construct validation intervals.



Example: Pressurized Water Reactors (PWR)

Thermo-Hydraulic Equations: Mass, momentum and energy balance for fluid

$$\frac{\partial}{\partial t}(\alpha_f \rho_f) + \nabla \cdot (\underline{\alpha_f} \rho_f v_f) = -\Gamma$$

$$\alpha_f \rho_f \frac{\partial v_f}{\partial t} + \alpha_f \rho_f v_f \cdot \nabla v_f + \nabla \cdot \sigma_f^R + \alpha_f \nabla \cdot \sigma + \alpha_f \nabla p_f$$
$$= -F^R - F + \Gamma(v_f - v_g)/2 + \alpha_f \rho_f g$$

$$\frac{\partial}{\partial t}(\alpha_f \rho_f e_f) + \nabla \cdot (\underline{\alpha_f} \rho_f e_f v_f + Th) = (T_g - T_f)H + T_f \Delta_f$$
$$-T_g(H - \alpha_g \nabla \cdot h) + h \cdot \nabla T - \Gamma[e_f + T_f(s^* - s_f)]$$
$$-p_f \left(\frac{\partial \alpha_f}{\partial t} + \nabla \cdot (\alpha_f v_f) + \frac{\Gamma}{\rho_f} \right)$$

$$\boxed{\frac{\partial \rho}{\partial t} + \frac{\partial \phi}{\partial x} = \text{Sources - Sinks}}$$

Notes:

- Similar relations for gas and bubbly phases

Challenges:

- Codes can have 15-30 closure relations and up to 75 parameters.
- Codes and closure relations often ”borrowed” from other physical phenomena; e.g., single phase fluids, airflow over a car (CFD code STAR-CCM+)
- Calibration necessary and closure relations can conflict.

Example: Pressurized Water Reactors (PWR)

Thermo-Hydraulic Equations: Mass, momentum and energy balance for fluid

$$\frac{\partial}{\partial t}(\alpha_f \rho_f) + \nabla \cdot (\underline{\alpha_f} \rho_f v_f) = -\Gamma$$

$$\alpha_f \rho_f \frac{\partial v_f}{\partial t} + \alpha_f \rho_f v_f \cdot \nabla v_f + \nabla \cdot \sigma_f^R + \alpha_f \nabla \cdot \sigma + \alpha_f \nabla p_f$$
$$= -F^R - F + \Gamma(v_f - v_g)/2 + \alpha_f \rho_f g$$

$$\frac{\partial}{\partial t}(\alpha_f \rho_f e_f) + \nabla \cdot (\underline{\alpha_f \rho_f} e_f v_f + Th) = (T_g - T_f)H + T_f \Delta_f$$
$$-T_g(H - \alpha_g \nabla \cdot h) + h \cdot \nabla T - \Gamma[e_f + T_f(s^* - s_f)]$$
$$-p_f \left(\frac{\partial \alpha_f}{\partial t} + \nabla \cdot (\alpha_f v_f) + \frac{\Gamma}{\rho_f} \right)$$

$$\boxed{\frac{\partial \rho}{\partial t} + \frac{\partial \phi}{\partial x} = \text{Sources - Sinks}}$$

Notes:

- Similar relations for gas and bubbly phases

Challenges:

- Codes can have 15-30 closure relations and up to 75 parameters.

Example: Dittus—Boelter Relation

$$Nu = 0.023 Re^{0.8} Pr^{0.4}$$

Nu: Nusselt number
Re: Reynolds number
Pr: Prandtl number

Example: Pressurized Water Reactors (PWR)

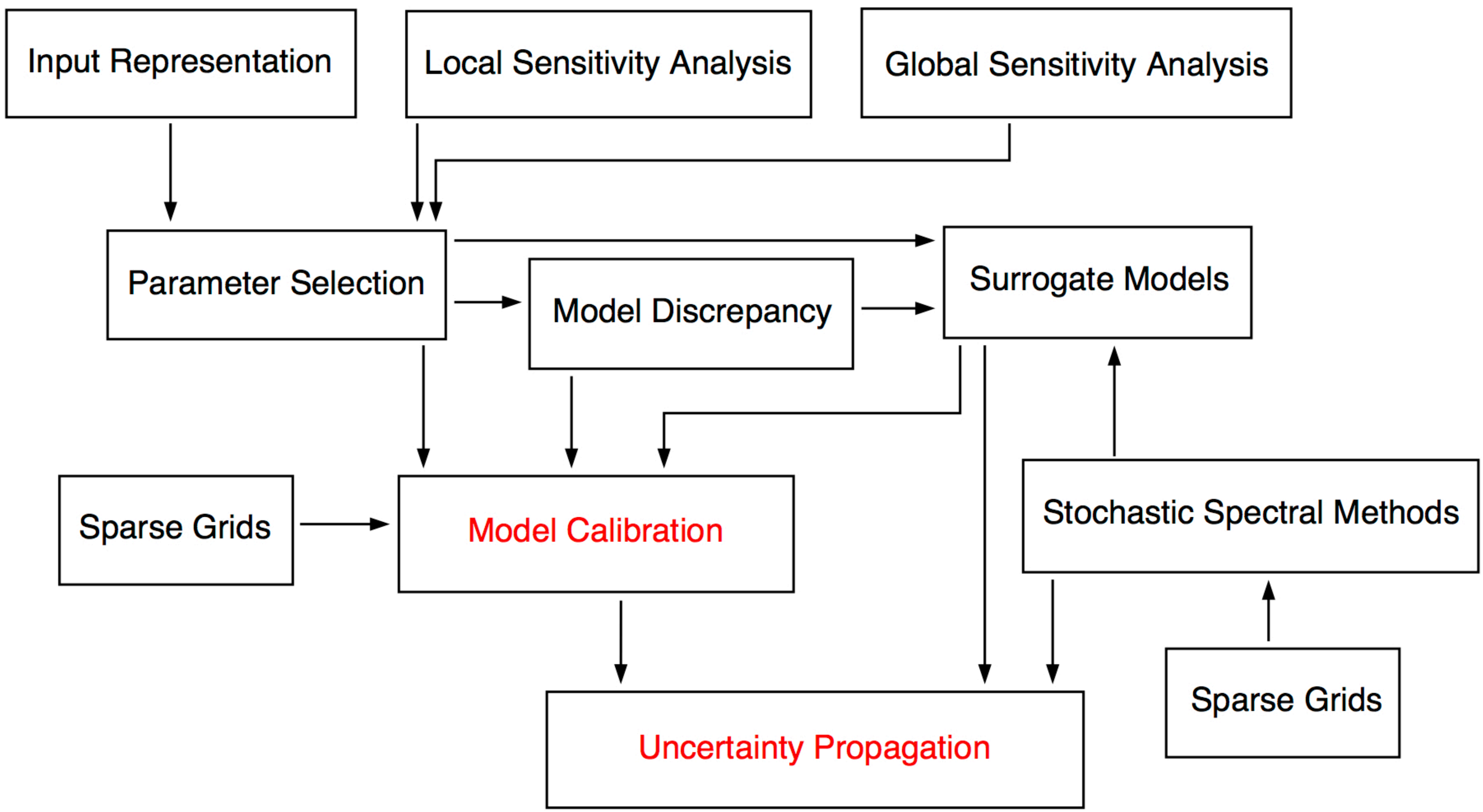
Example: Shearon Harris outside Raleigh

UQ Questions:

- What is peak operating temperature?
- What is expected level of CRUD buildup?
- What is risk associated with operating regime?
- What is expected profit for new design?

Steps in Uncertainty Quantification

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



Model Calibration and Uncertainty Propagation

Sources of Uncertainty:

- Model
- Parameters
- Sensor measurements
- Initial conditions

Strategy:

- Quantify uncertainty in parameters
- Propagate uncertainty through model

Example: HIV model

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

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

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

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

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

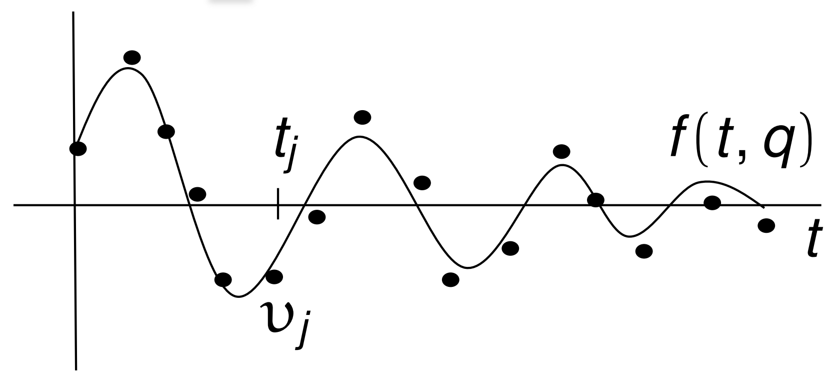
$$\dot{E} = \lambda_E + \frac{\underline{b_E}(T_1^* + T_2^*)}{T_1^* + T_2^* + \underline{K_b}} E - \frac{d_E(T_1^* + T_2^*)}{T_1^* + T_2^* + K_d} E - \delta_E E$$

Parameters: Reduced set

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

Point Estimates: Ordinary least squares

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



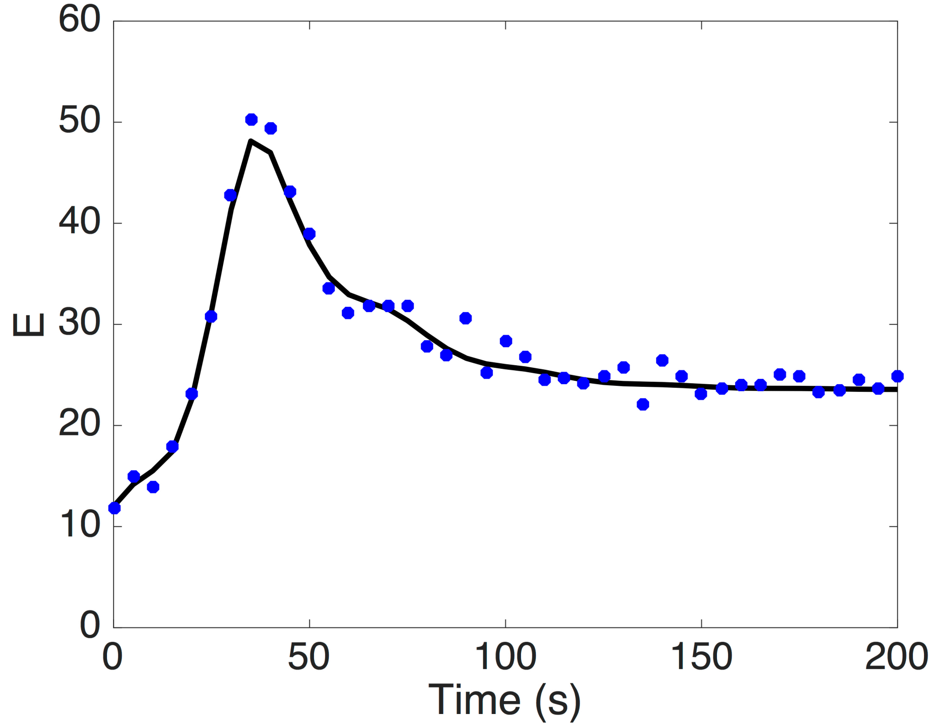
Note: Scaling critical since parameter values vary by 8 orders of magnitude.

Model Calibration and Predictions

Optimization Results:

b_E	δ	d_1	k_2	λ_1	K_b
0.30	0.68	9.1×10^{-3}	1.22×10^{-4}	9.95×10^{3}	88.5

Data and Prediction of Immune Effector Response E:



Note: Point estimates but no quantification of uncertainty in:

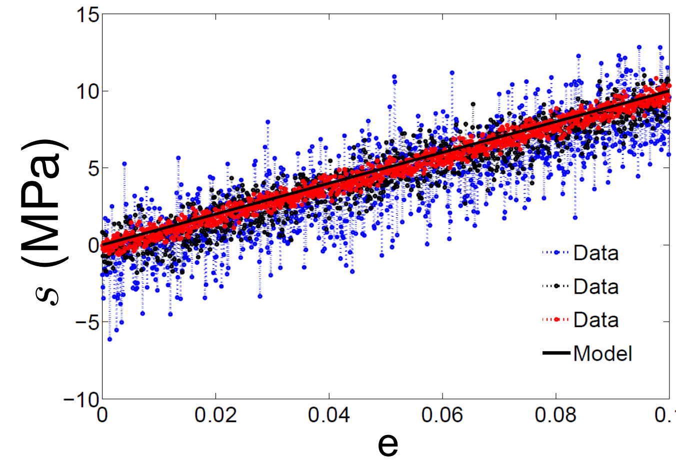
- Model
- Parameters
- Data

Goals:

- Replace point estimates with distributions.
- Construct credible and prediction intervals.
- Natural in a Bayesian framework

Bayesian Inference: More General Model

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

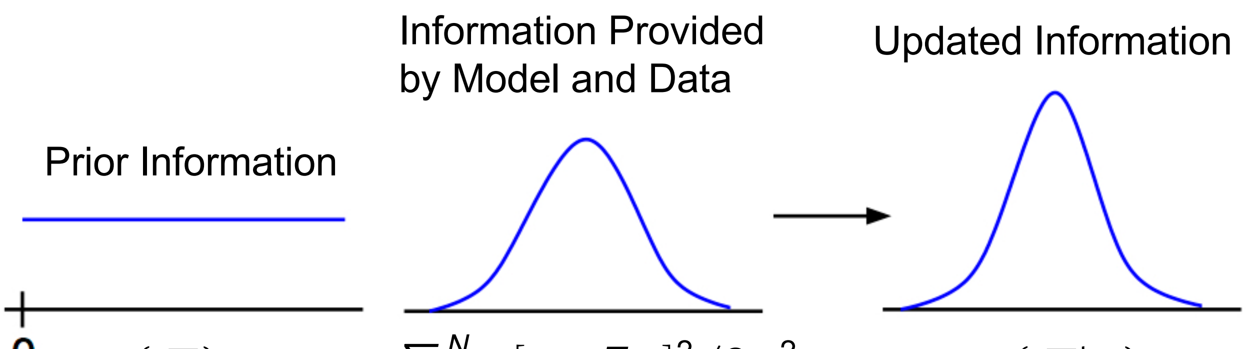


$$s_i = Ee_i + \varepsilon_i \;,\; i = 1, \ldots, N$$

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

Parameter: Stiffness E

Strategy: Use model fit to data to update prior information



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

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

- Prior Distribution: Quantifies prior knowledge of parameter values
- Likelihood: Probability of observing a data given set of parameter values.
- Posterior Distribution: Conditional distribution of parameters given observed data.

Problem: Can require high-dimensional integration

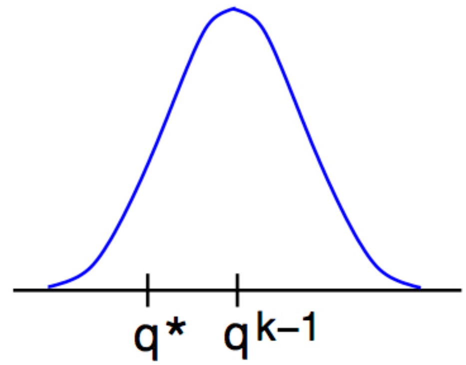
• e.g., HIV Model: p = 6 - 23!

• 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.

Delayed Rejection Adaptive Metropolis (DRAM)

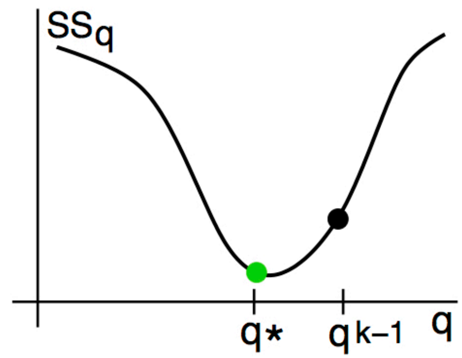
Algorithm: [Haario et al., 2006] – MATLAB, Python

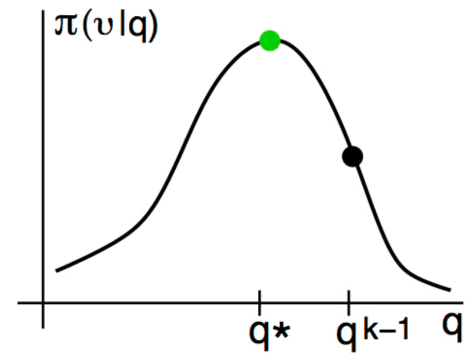


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$
2. For $k = 1, \cdots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

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

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

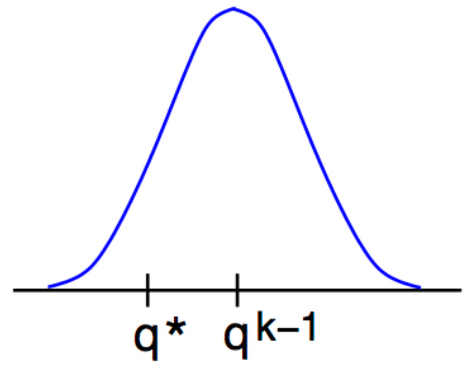




   (c) Accept q^* with probability dictated by likelihood

Delayed Rejection Adaptive Metropolis (DRAM)

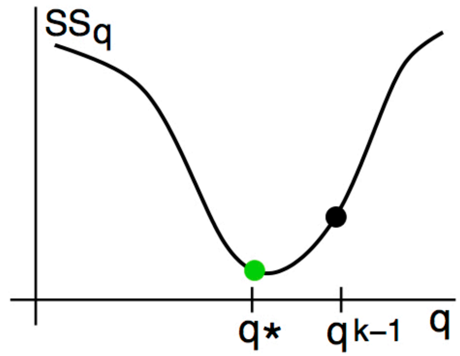
Algorithm: [Haario et al., 2006] – MATLAB, Python

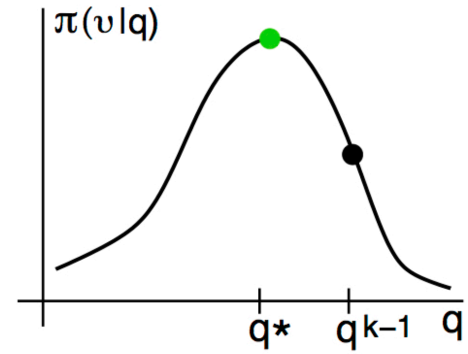


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$
2. For $k = 1, \cdots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

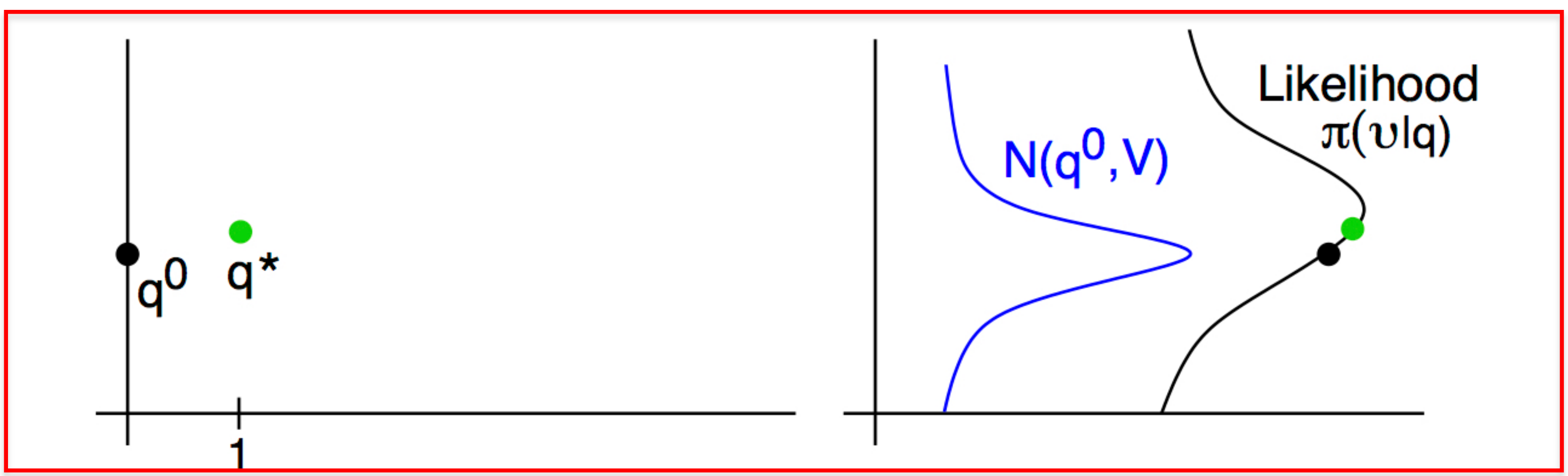
$$SS_{q^*} = \sum_{i=1}^{N} [\upsilon_i - f(t_i, q^*)]^2$$

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



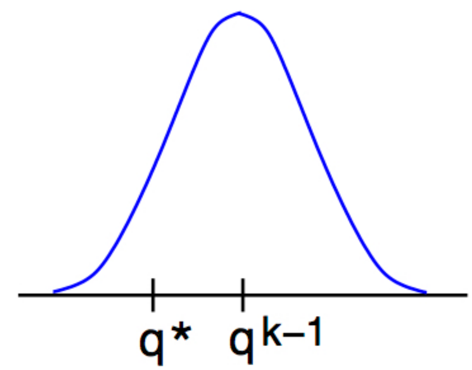


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

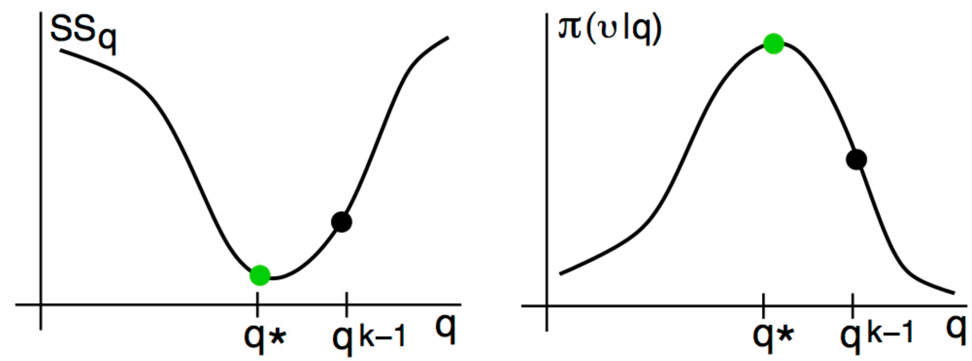
Algorithm: [Haario et al., 2006] – MATLAB, Python



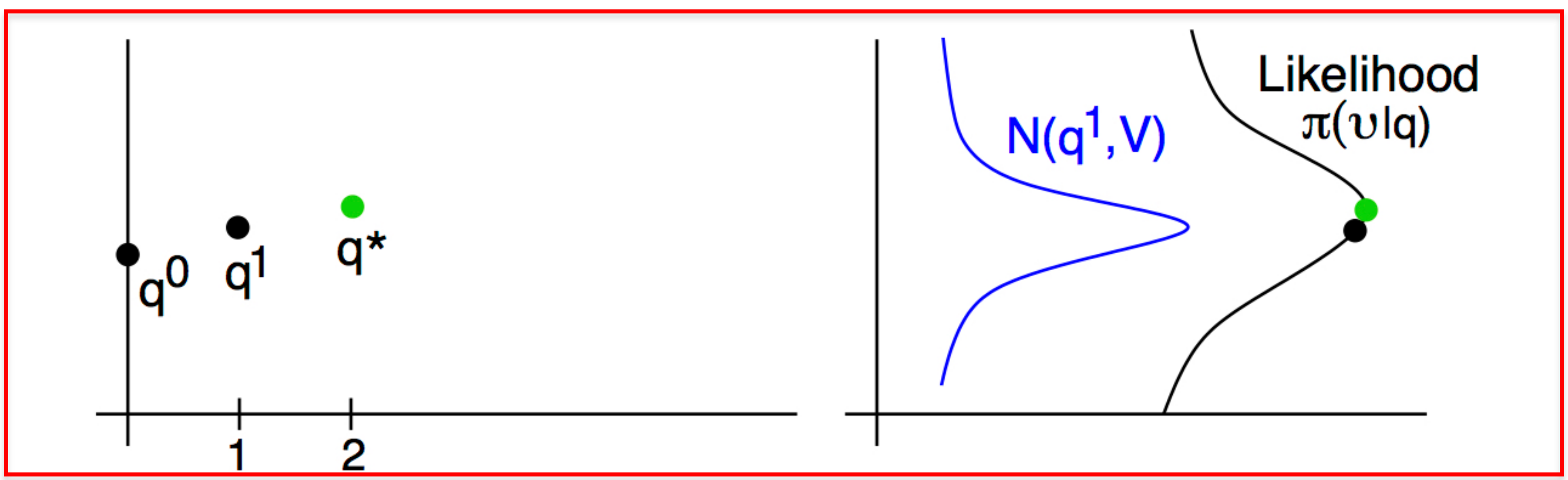
1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$
2. For $k = 1, \cdots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

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

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

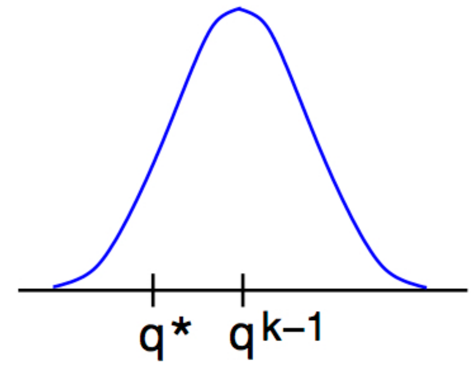


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

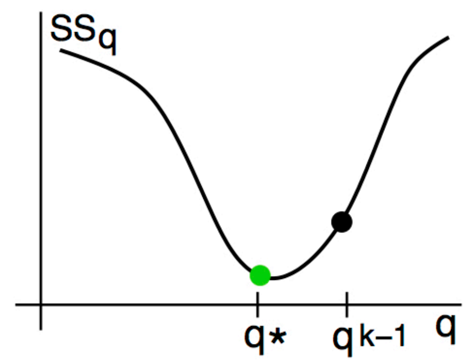
Algorithm: [Haario et al., 2006] – MATLAB, Python

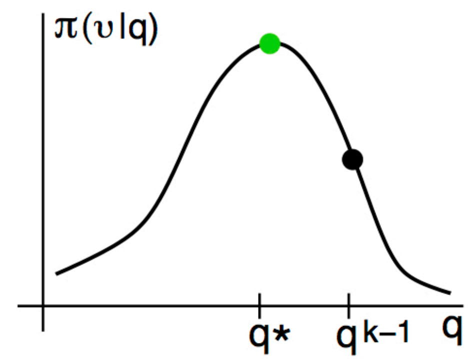


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$
2. For $k = 1, \cdots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

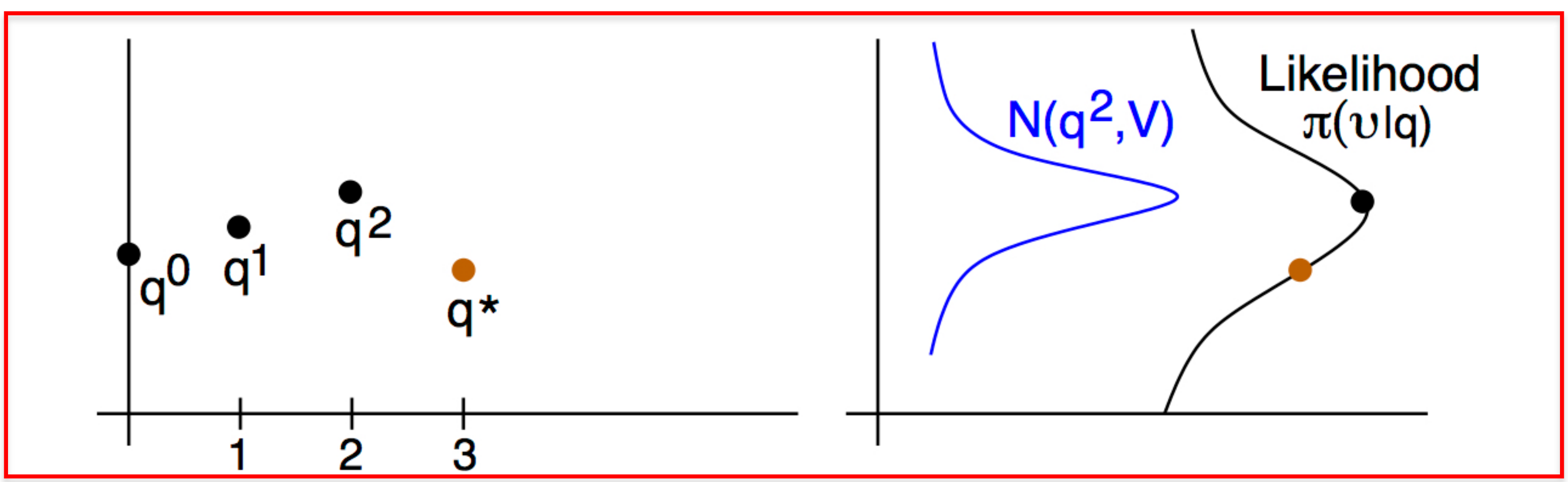
$$SS_{q^*} = \sum_{i=1}^{N} [\upsilon_i - f(t_i, q^*)]^2$$

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



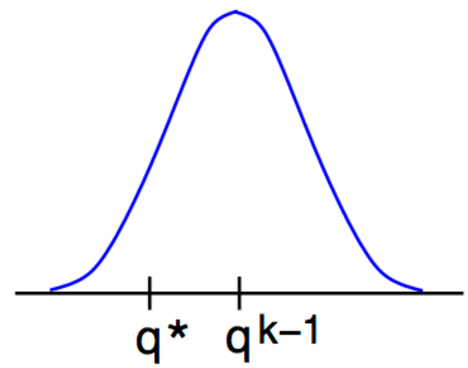


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

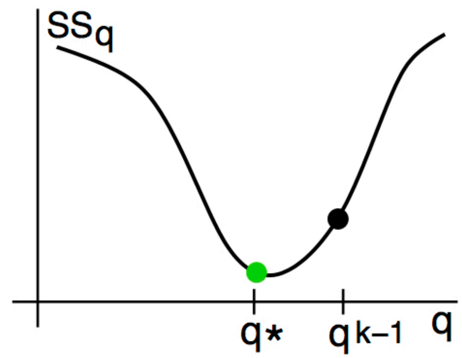
Algorithm: [Haario et al., 2006] – MATLAB, Python

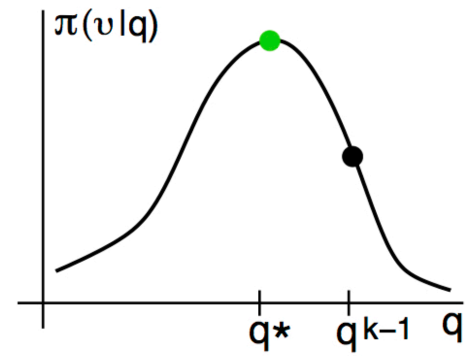


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$
2. For $k = 1, \cdots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

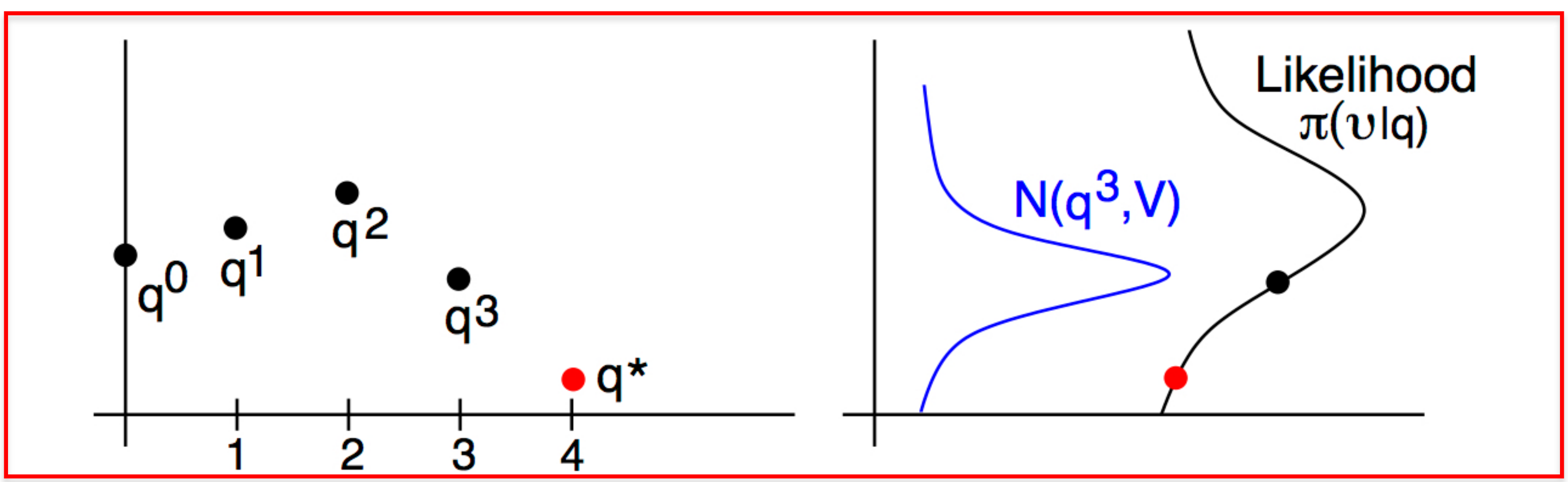
$$SS_{q^*} = \sum_{i=1}^{N} [\upsilon_i - f(t_i, q^*)]^2$$

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



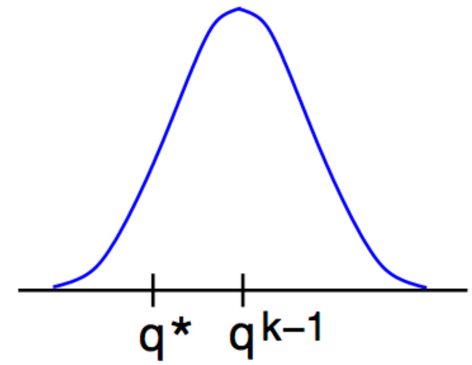


   (c) Accept q^* with probability dictated by likelihood



Delayed Rejection Adaptive Metropolis (DRAM)

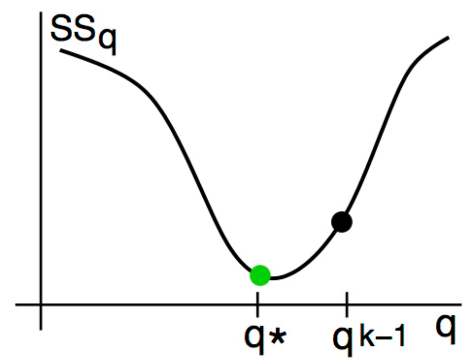
Algorithm: [Haario et al., 2006] – MATLAB, Python

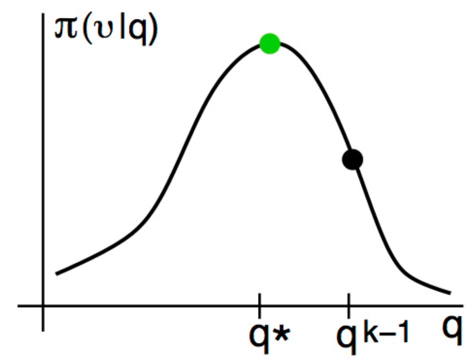


1. Determine $q^0 = \arg\min_q \sum_{i=1}^{N} [\upsilon_i - f(t_i, q)]^2$
2. For $k = 1, \cdots, M$
   (a) Construct candidate $q^* \sim N(q^{k-1}, V)$
   (b) Compute likelihood

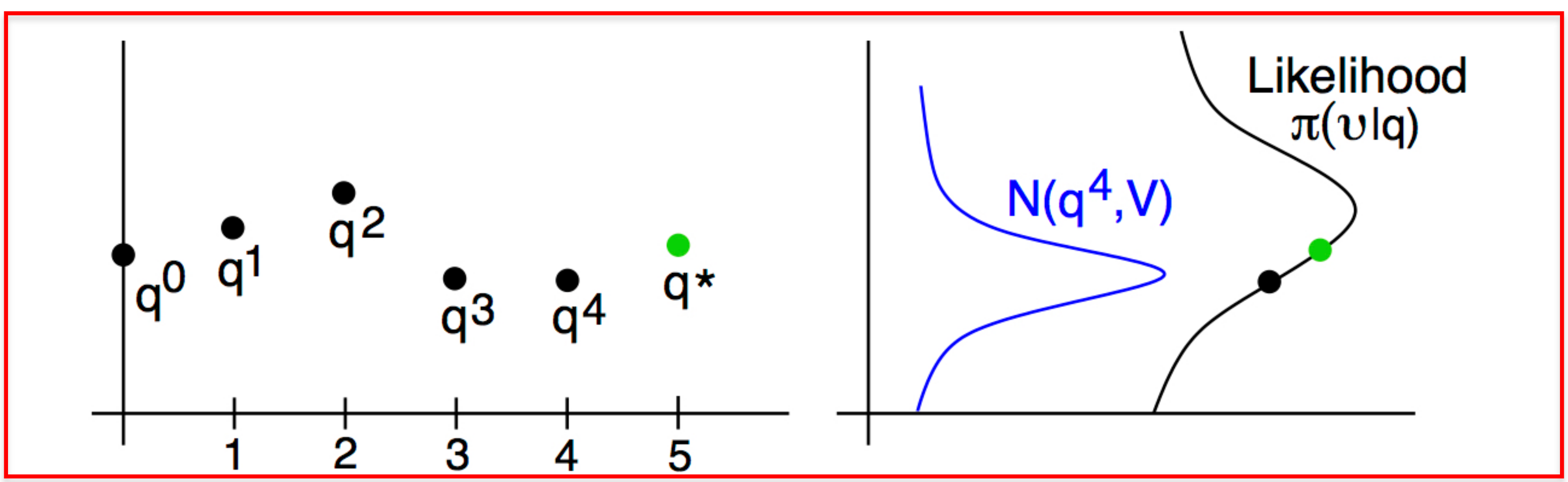
$$SS_{q^*} = \sum_{i=1}^{N} [\upsilon_i - f(t_i, q^*)]^2$$

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





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



Bayesian Model Calibration – HIV Example

Model:

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

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

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

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

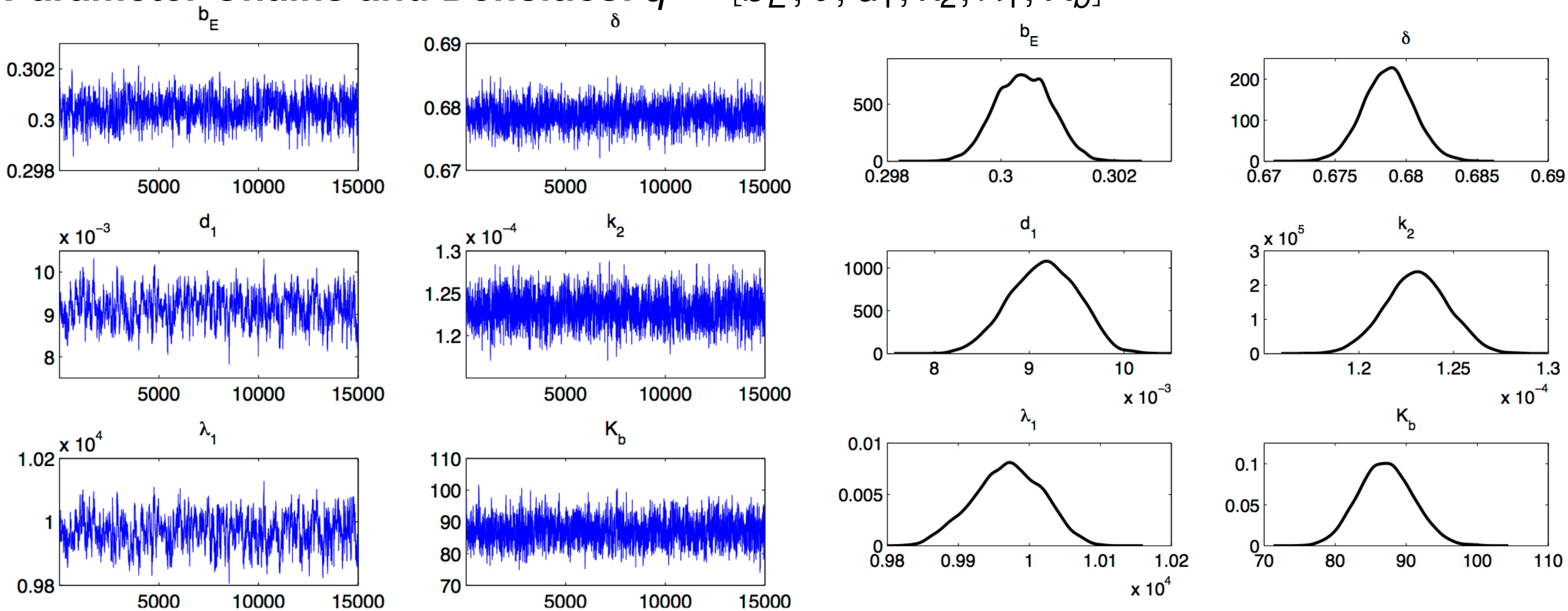
$$\dot{V} = \underline{N_T} \delta (T_1^* + T_2^*) - cV - [(1-\varepsilon)\rho_1 k_1 T_1 + (1 - f\varepsilon)\rho_2 k_2 T_2] V$$

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

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, *Geosci. Model Dev. 11, 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.

Wetland Methane Model

Subset of Governing Relations:

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

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

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

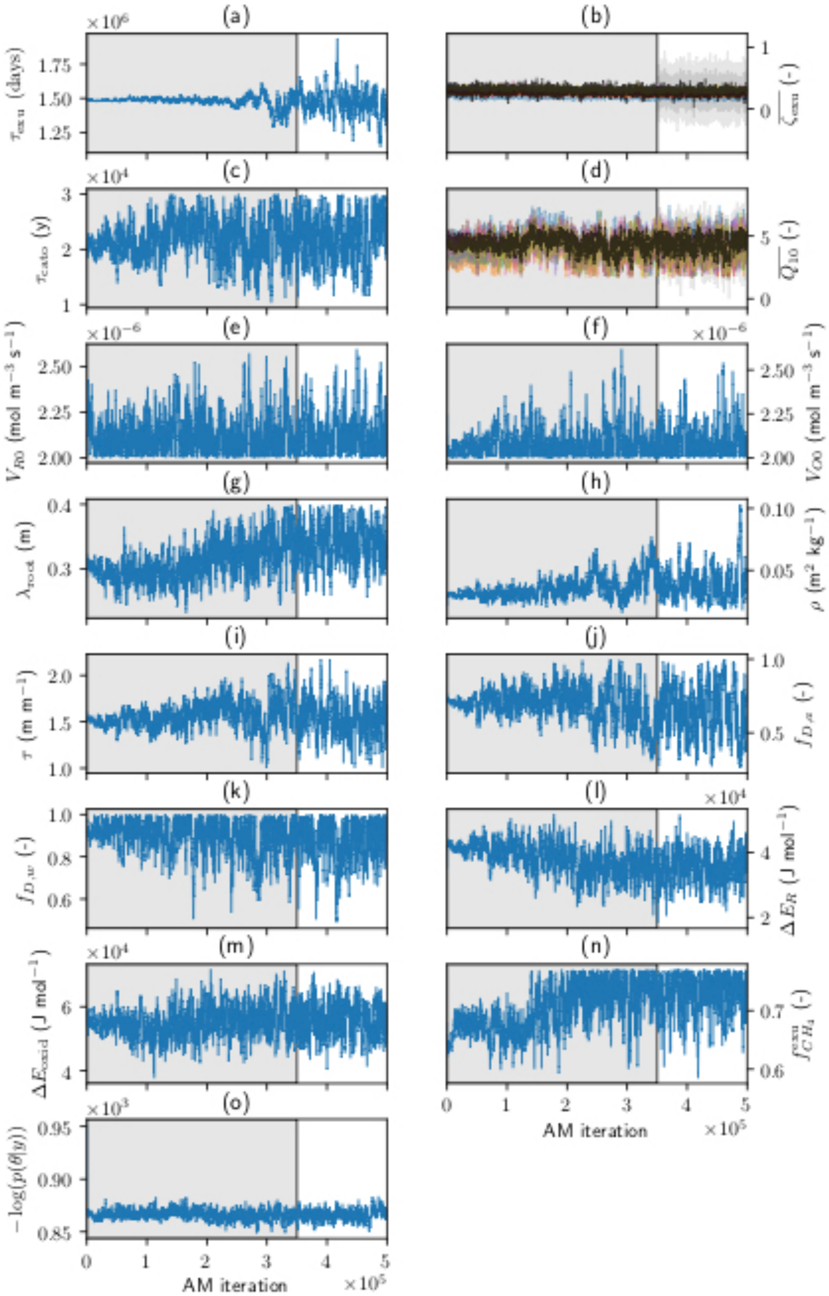
$$\frac{\partial[CO_2]}{\partial t}(t,z) = -T_{CO_2} + R_{CO_2}^{exu} + R_{CO_2}^{peat} + R_{CH_4}^{oxid} + R_{aerob}^{peat}$$

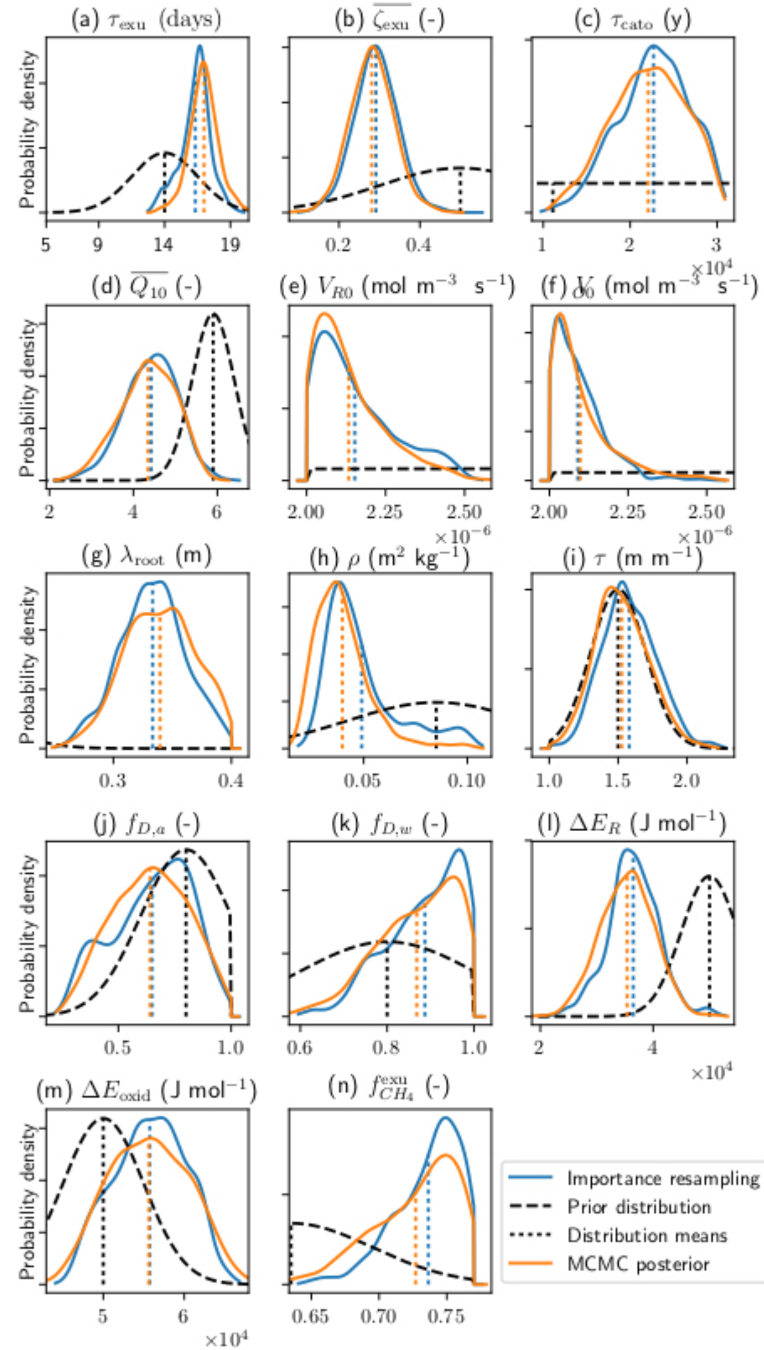
Initial Calibration Parameters: 14

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

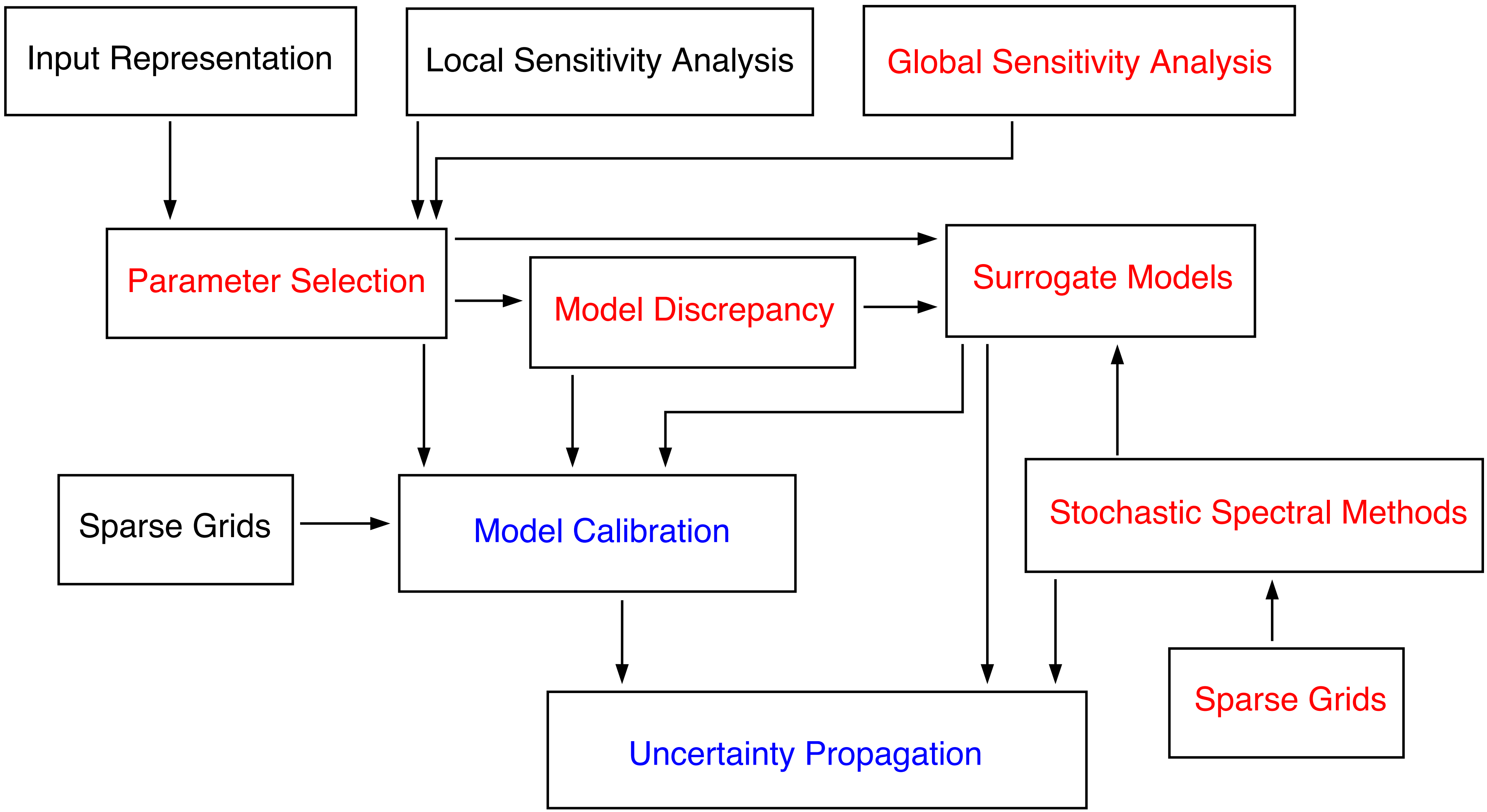
Wetland Methane Model: Chains and Marginal Distributions





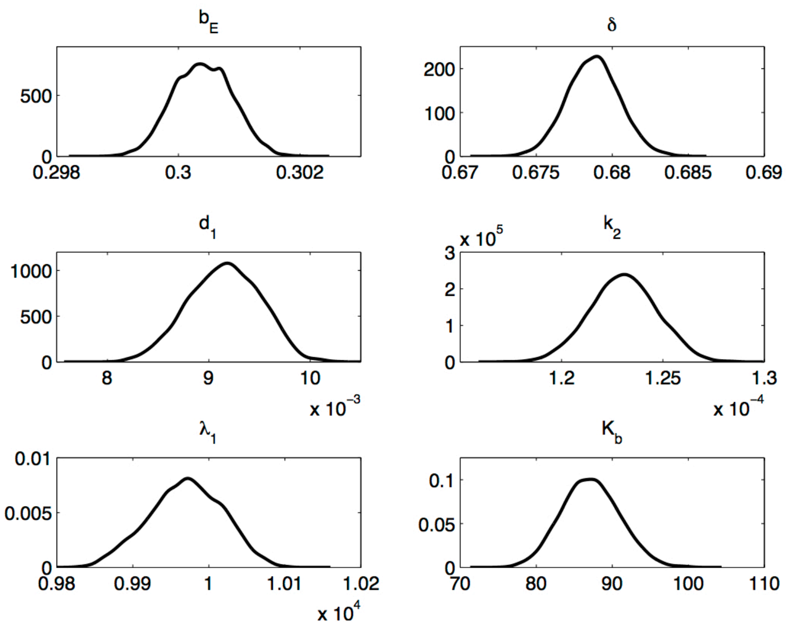
Steps in Uncertainty Quantification

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



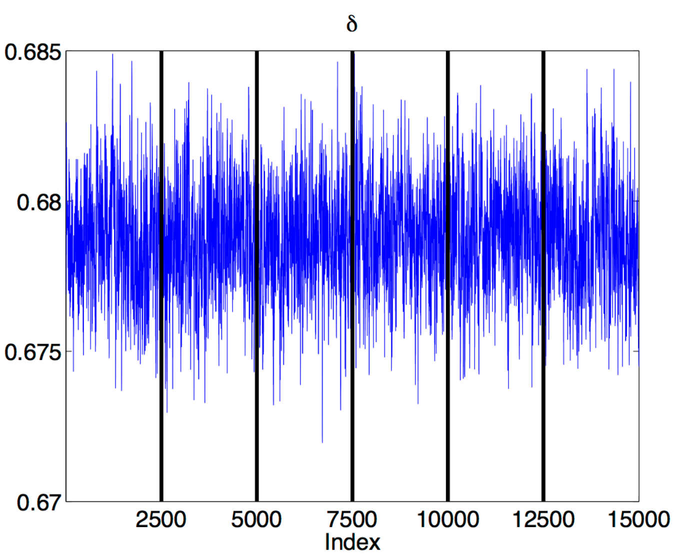
Propagation of Uncertainty in Models – HIV Example

Parameter Densities:

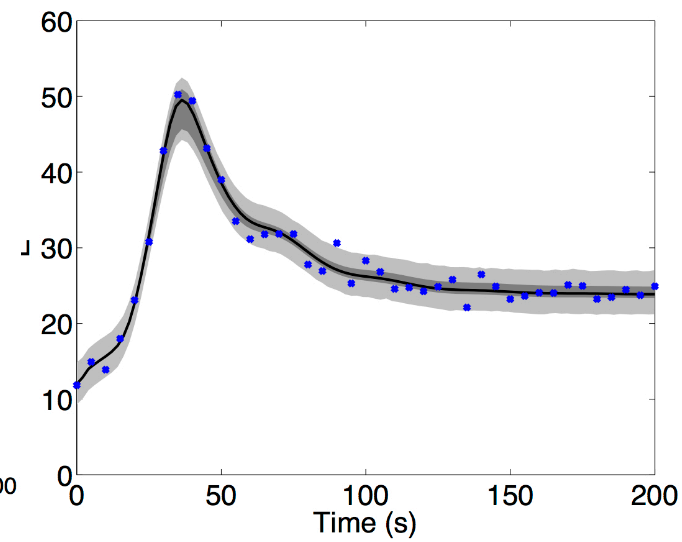


Techniques:

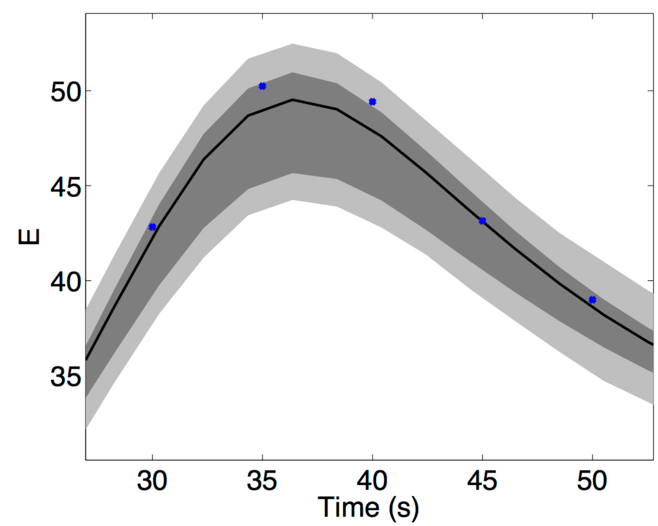
- Sample from parameter densities to construct prediction intervals for QoI.
- Slow convergence rate $\mathcal{O}(1/\sqrt{M})$
- 100-fold more evaluations required to gain additional place of accuracy.
- Significant numerical analysis used to efficiently propagate densities.



Samples from Chain



Data, Credible Intervals and Prediction Intervals

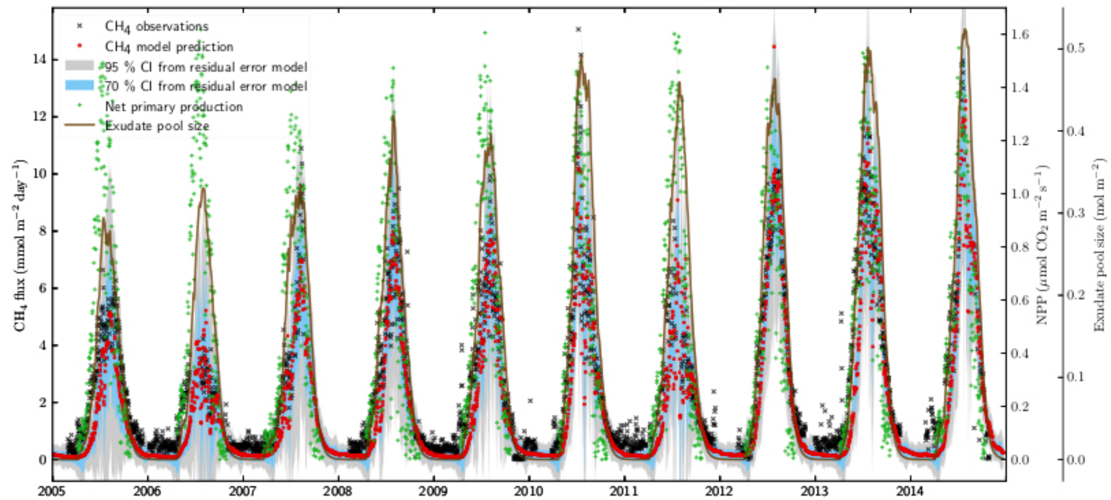


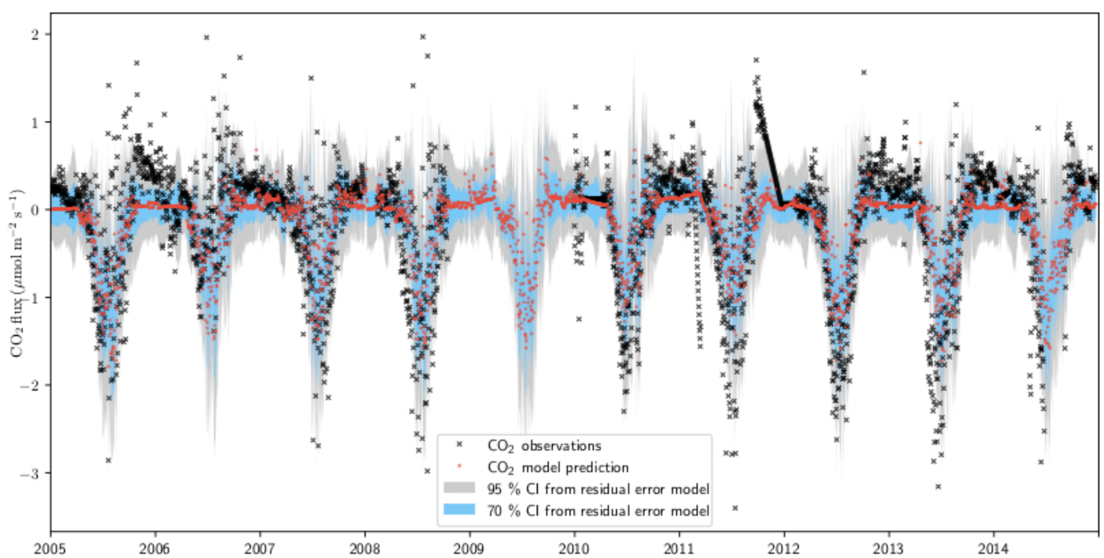
Non-Gaussian Credible and Prediction Intervals

Wetland Methane Model

Authors: Susiluoto, Raivonen, Backman, Laine, Makela, Peltola, Vesala, Aalto, *Geosci. Model Dev. 11, 2018*

Observation: Prediction intervals consistent for methane and carbon dioxide





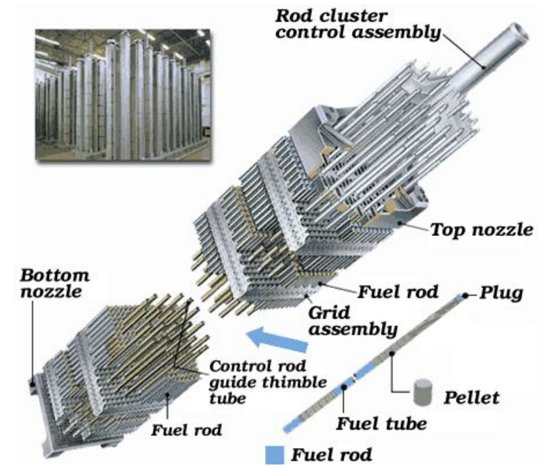
Use of Prediction Intervals: Nuclear Power Plant Design

Subchannel Code (COBRA-TF): numerous closure relations, ~70 parameters

e.g., Dittus—Boelter Relation

$$Nu = \color{red}{0.023} Re^{0.8} Pr^{0.4}$$

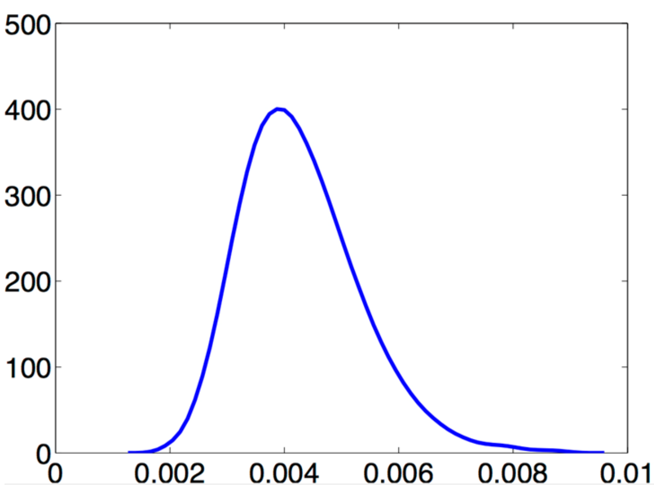
Nu: Nusselt number
Re: Reynolds number
Pr: Prandtl number



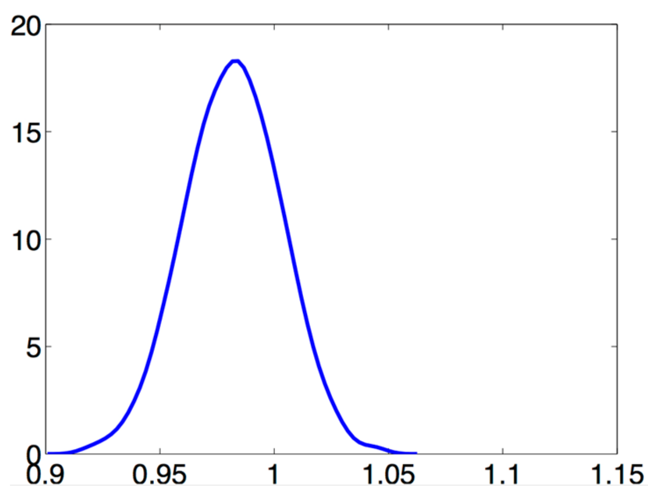
Industry Standard: Employ conservative, uniform, bounds

i.e., [0, 0.046], [0, 1.6], [0,0.8]

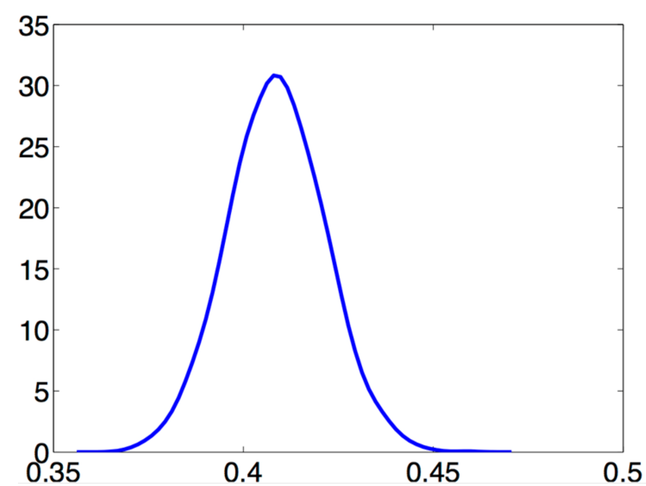
Bayesian Analysis: Employ conservative bounds as priors



$2\sigma \approx 0.0035$



$2\sigma \approx 0.06$

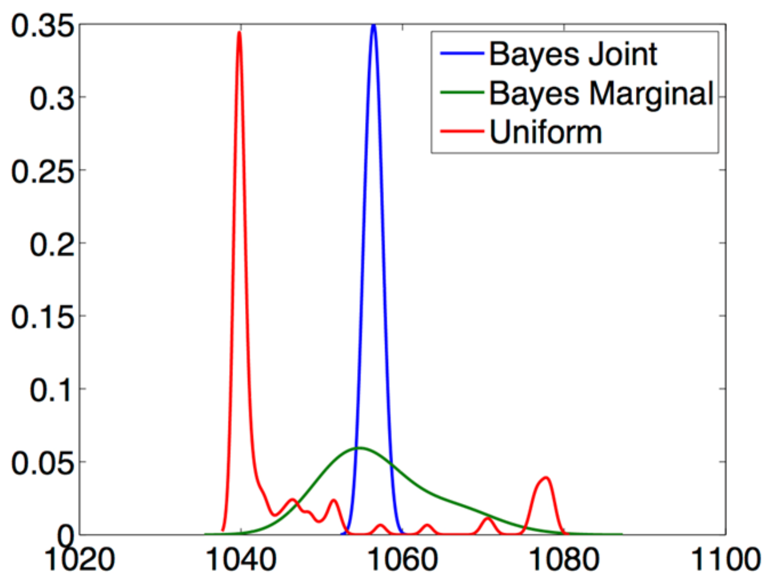


$2\sigma \approx 0.03$

Note: Substantial reduction in parameter uncertainty

Use of Prediction Intervals: Nuclear Power Plant Design

Strategy: Propagate parameter uncertainties through COBRA-TF to determine uncertainty in maximum fuel temperature



Notes:

• Temperature uncertainty reduced from 40 degrees to 5 degrees

• Can run plant 20 degrees hotter, which significantly improves efficiency

Ramification: Savings of 10 billion dollars per year for US power plants

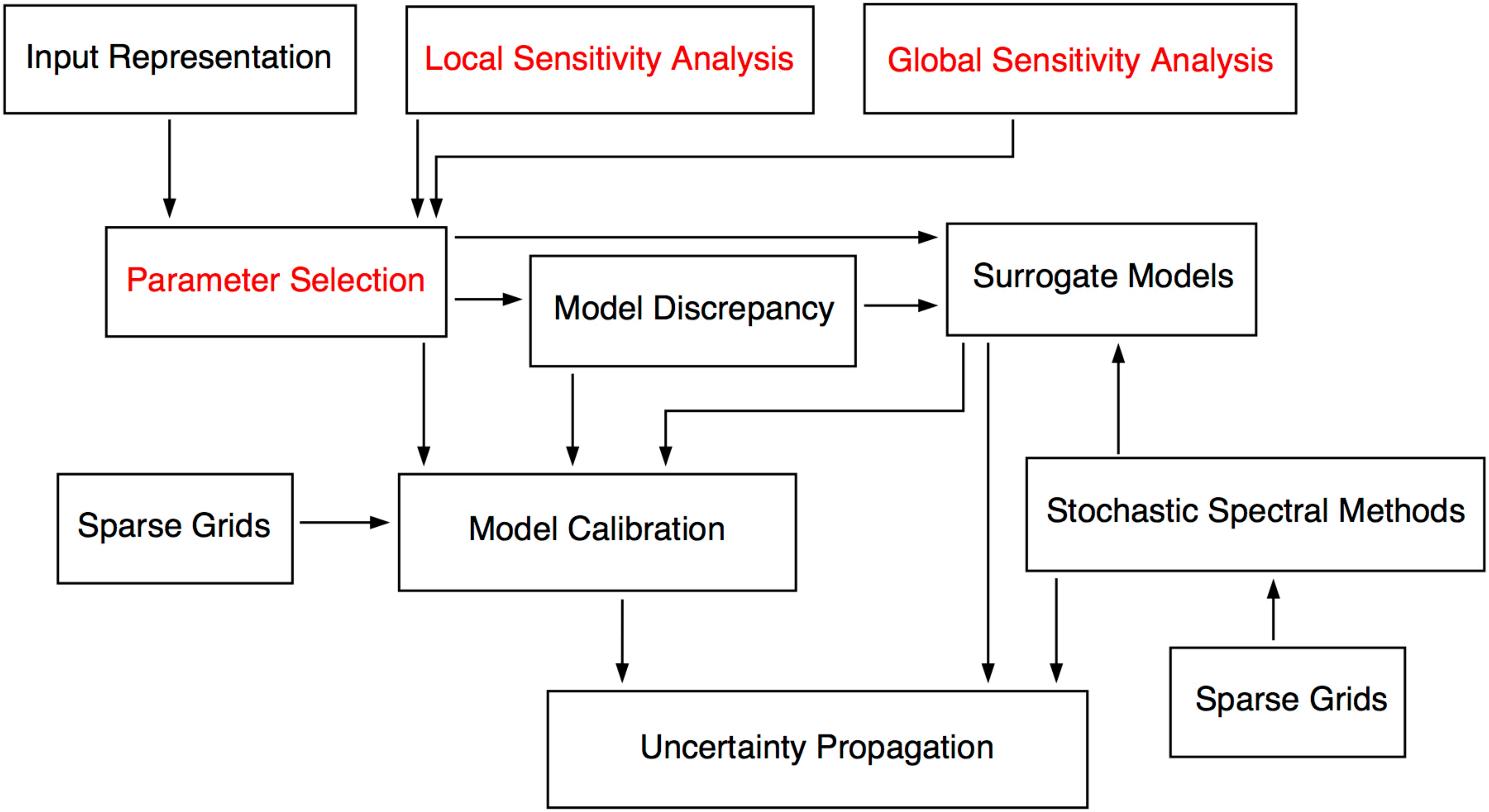
Issues:

• We considered only one of many physical relations

• Nuclear regulatory commission takes years to change requirements and codes

Good News: We were able to work with Westinghouse to reduce uncertainties.

Steps in Uncertainty Quantification



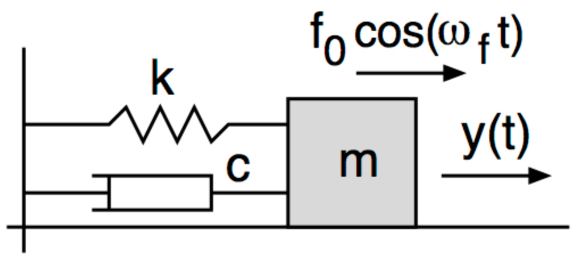
Parameter Selection: Required for models with unidentifiable or noninfluential inputs

- e.g., HIV and SIR model

Parameter Selection Techniques

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

Example: Spring model



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

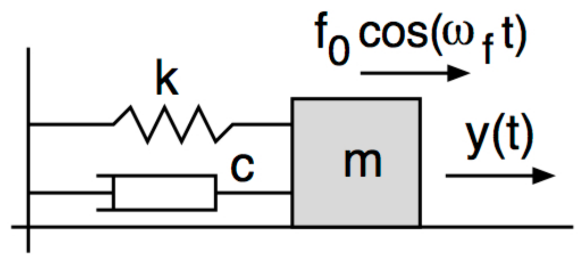
$$z(0) = z_0 \; , \; \frac{dz}{dt}(0) = z_1$$

Problem: Parameters $q = [m, c, k, f_0]$ and $q = \left[1, \frac{c}{m}, \frac{k}{m}, \frac{f_0}{m}\right]$ yield same displacements

Parameter Selection Techniques

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

Example: Spring model



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

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

Problem: Parameters $q = [m, c, k, f_0]$ and $q = \left[1, \frac{c}{m}, \frac{k}{m}, \frac{f_0}{m}\right]$ yield same displacements

Solution: Reformulate problem as

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

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

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

Techniques for General Models:

- Sensitivity analysis: See tutorial by Pierre Gremaud!!
- Active Subspaces

Parameter Selection: Nuclear Power Plant Design

Subchannel Code (COBRA-TF): numerous closure relation and parameters

parameter	partial correlation	simple correlation	morris main	morris interaction	CPS variation
k_eta	0.07	0.03			
k_gama	-0.03	0.04			
k_sent	-0.03	-0.02			
k_sdent	-0.07	-0.01			
k_tmasv	-0.03	0.00			
k_tmasl	0.11	0.00	6.48E-05	2.28E-05	medium
k_tmasg	-0.19	-0.01			
k_tmomv	-0.12	-0.01			
k_tmome	0.02	0.00			
k_tmoml	0.02	-0.02	2.23E-04	1.30E-04	medium
k_xk	0.08	-0.02			
k_xkes	-0.05	0.00			
k_xkge	-0.07	0.01			
k_xkl	0.04	-0.01			
k_xkle	-0.03	0.00			
k_xkvls	0.11	-0.01			
k_xkwvw	-0.10	0.01			
k_xkwlw	0.14	0.01			
k_xkwew	-0.01	0.03			
k_qvapl	-0.09	-0.01			
k_tnrgv	-0.03	0.00			
k_tnrgl	-0.01	0.03	9.00E-06	9.49E-06	low
k_rodqq	0.02	-0.01			
k_qradd	-0.02	0.00			
k_qradv	-0.01	0.00			
k_qliht	-0.01	0.00			
k_sphts	-0.05	0.03			
k_cond	-0.04	0.00			
k_xkwvx	0.03	-0.02			
k_xkwlx	1.00	0.88	1.80E-01	7.07E-03	high
k_cd	1.00	0.46	9.59E-02	7.88E-03	high
k_cdfb	-0.02	-0.01			
k_wkr	0.02	0.02			

5 Identified Active Inputs:

k_cd: Pressure loss coefficient of space in sub-channel

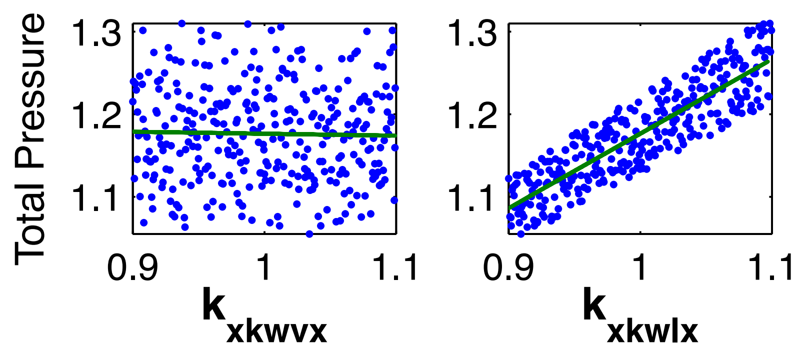
k_xkwlx: Vertical liquid wall drag coefficient

k_tmasl: Loss of liquid mass due to mixing and void drift

k_tmoml: Loss of liquid momentum due to mixing and void drift

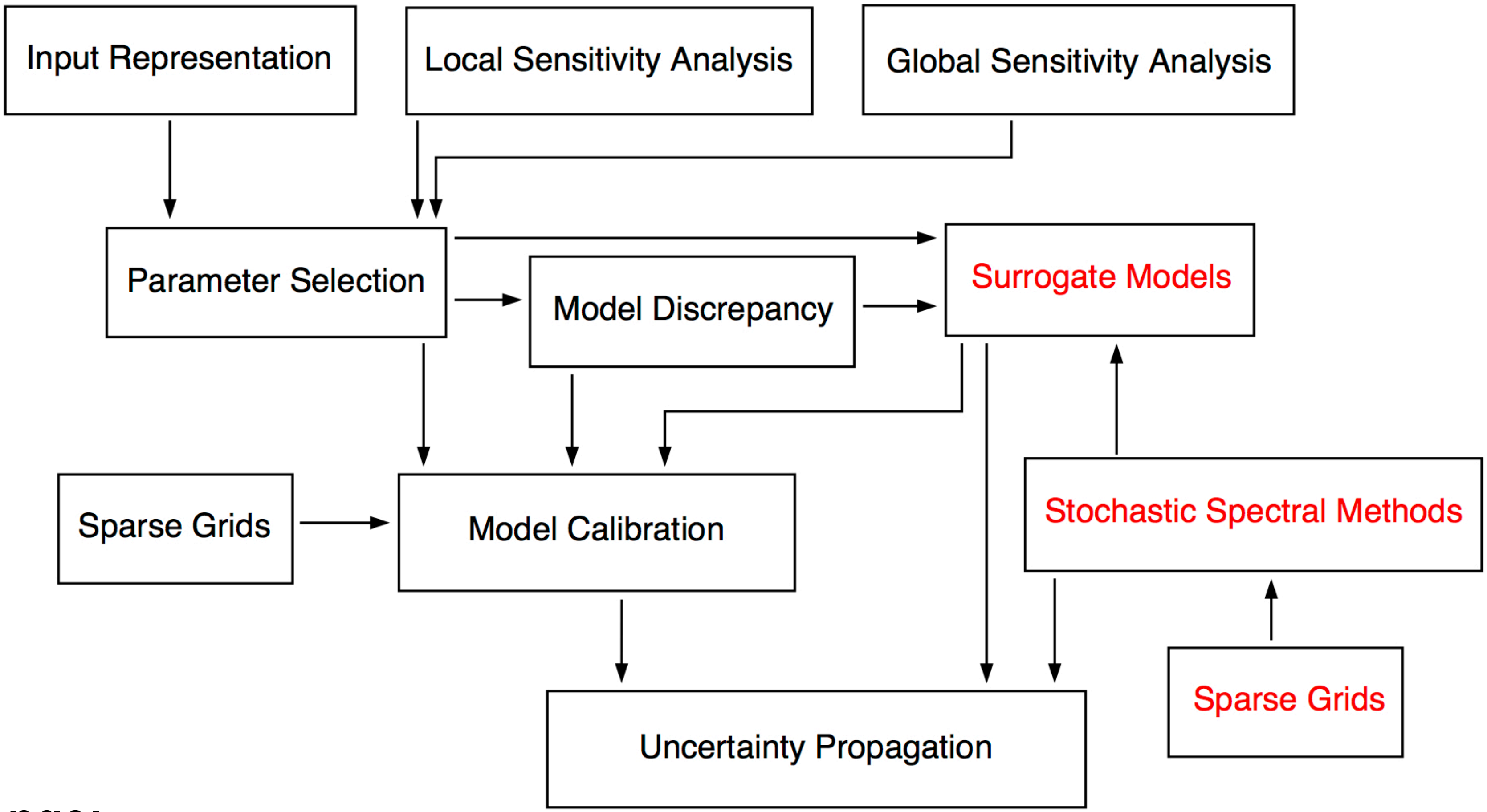
k_tnrgl: Loss of liquid enthalpy due to mixing and void drift

Partial Correlation:



Note: 33 initial VUQ parameters reduced to 5 via sensitivity analysis

Steps in Uncertainty Quantification

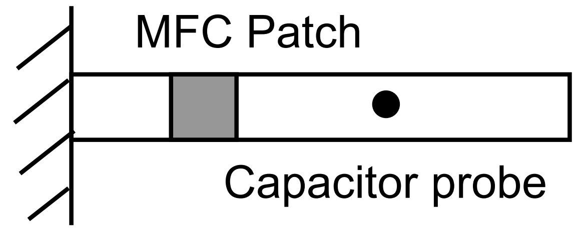


Challenge:

- How do we do uncertainty quantification for computationally expensive models?
- Example:
  - We have a computational budget of 5000 model evaluations.
  - Bayesian inference and uncertainty propagation require 120,000 evaluations.

Uncertainty Quantification Challenges

Example: MFC model – Fourth-order PDE

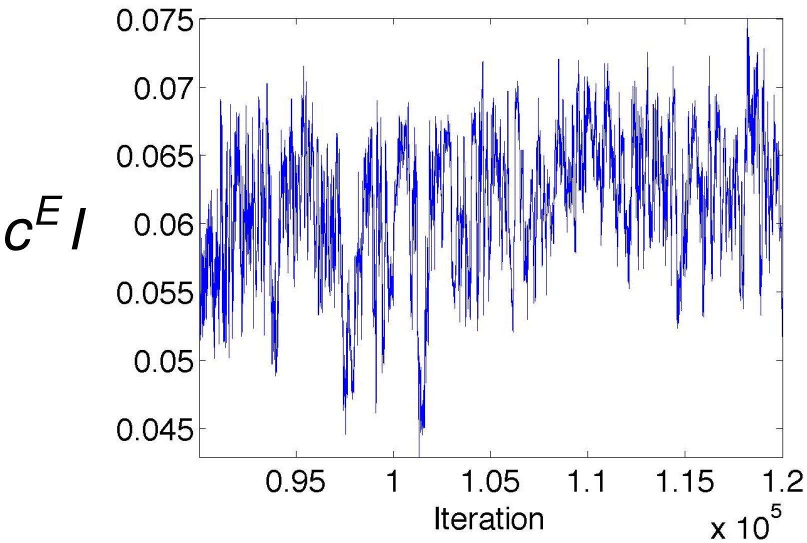


$$\rho \frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 M}{\partial x^2} = f$$

$$M = \quad \underline{-c^E I} \frac{\partial^2 w}{\partial x^2} - c_D I \frac{\partial^3 w}{\partial x^2 \partial t}$$
$$- \left[k_1 e(E, \sigma_0) E + k_2 \varepsilon_{irr}(E, \sigma_0)\right] \chi_{MFC}(x)$$

Macro-Fiber Composite

Bayesian Inference: Took 6 days!



Problem:

1.2×10^5 PDE solutions

Solution: Highly efficient surrogate models

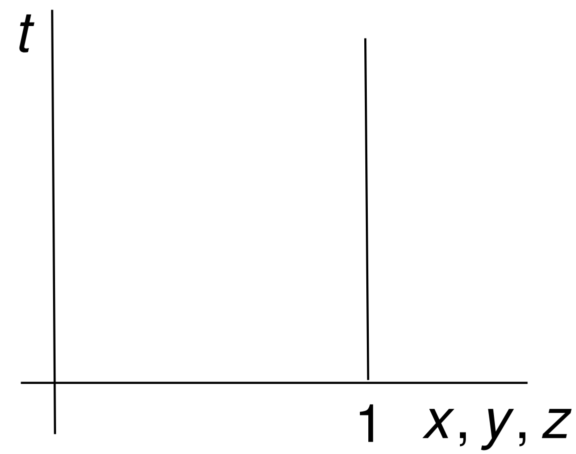
Surrogate Models: Motivation

Example: Consider the heat equation

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

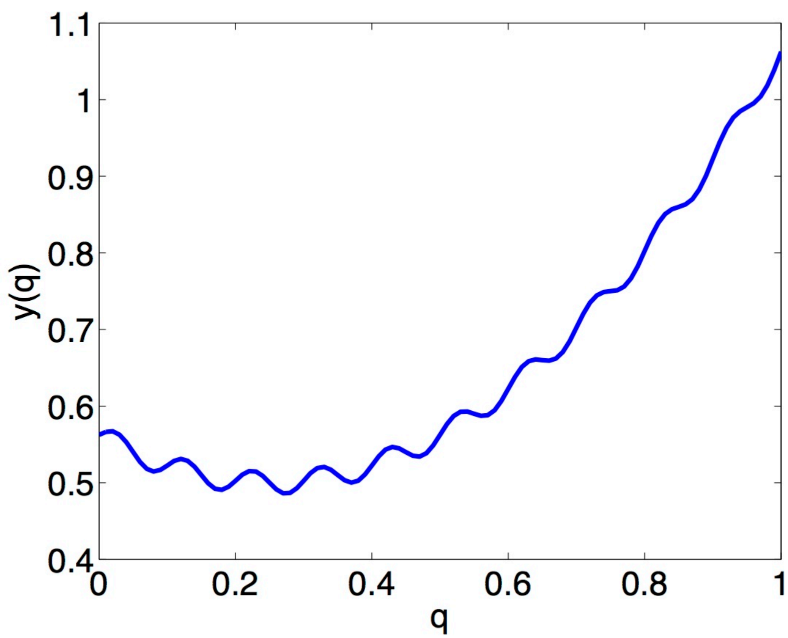
Boundary Conditions

Initial Conditions



with the response

$$y(q) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 u(t,x,y,z)\,dxdydzdt$$



Notes:

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

Surrogate Models: Motivation

Example: Consider the heat equation

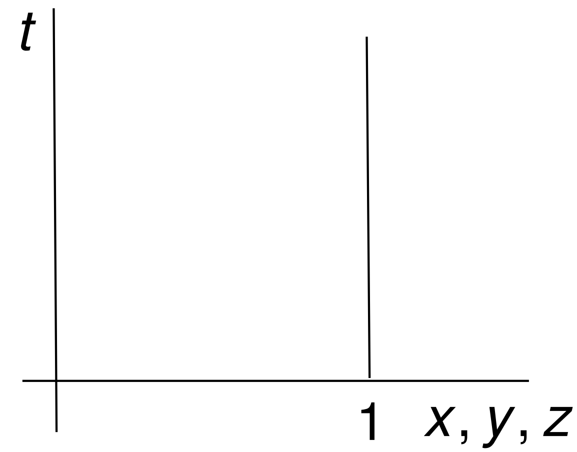
$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + f(q)$$

Boundary Conditions

Initial Conditions

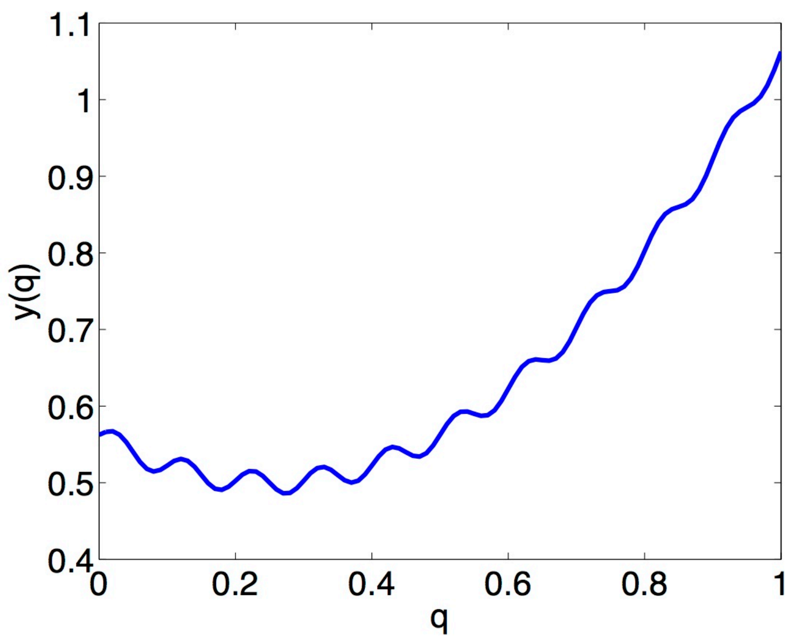
with the response

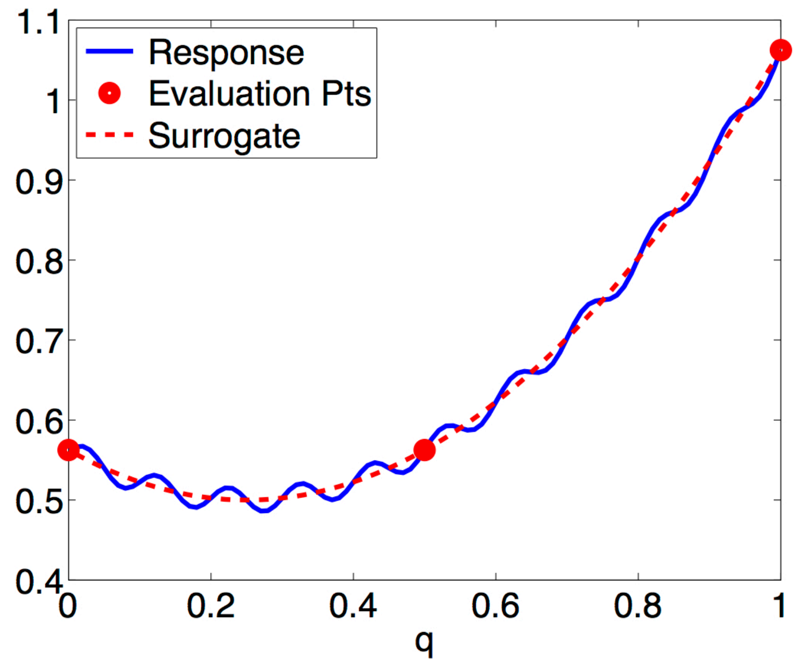
$$y(q) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 u(t,x,y,z)\,dxdydzdt$$



Surrogate: Quadratic

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





Data-Fit Models: Polynomial Surrogate

Quadratic Surrogate: Regression

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

Deterministic System:

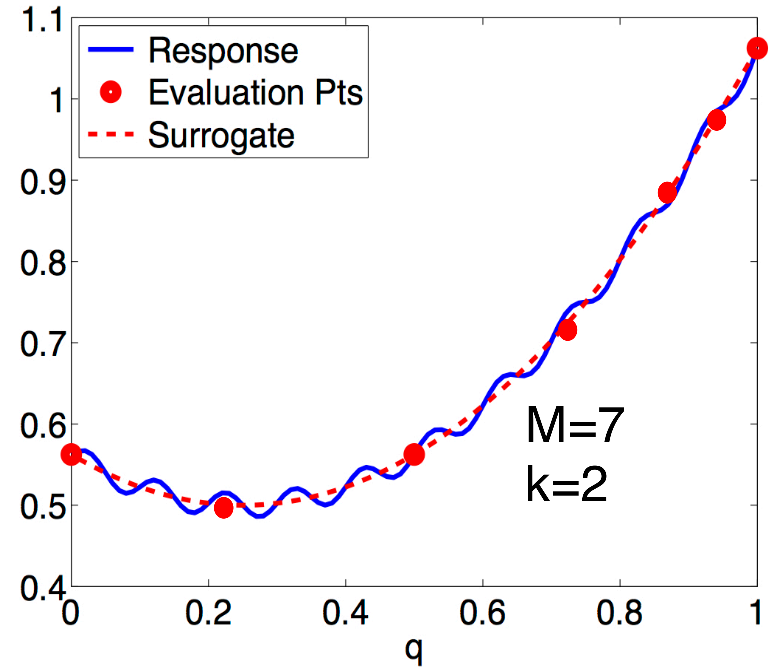
$$\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} 1 & q^1 & (q^1)^2 \\ \vdots & & \vdots \\ 1 & q^M & (q^M)^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Least Squares Estimate:

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

MATLAB:

$$>> beta = X \backslash y_{obs}$$



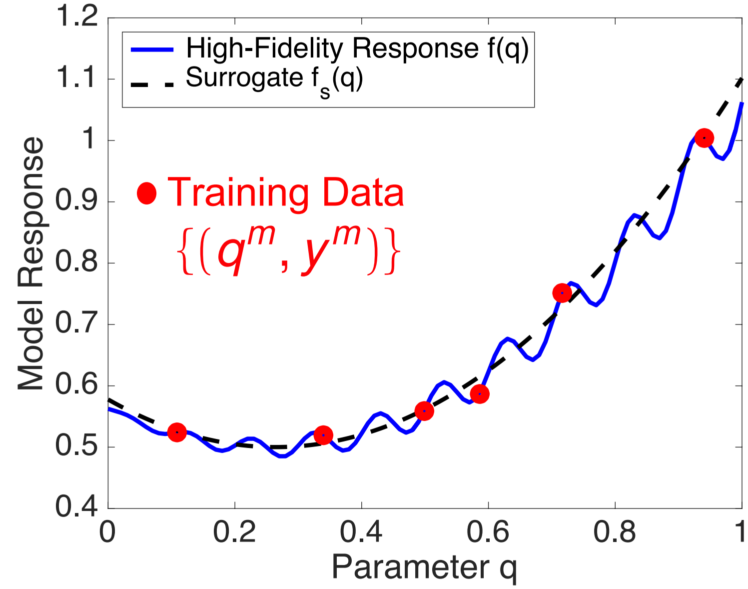
Notes:

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

Gaussian Process (GP) Emulators

Strategy: Consider deterministic model evaluations

$$y^m = f(q^m) + \varepsilon^m$$



Note: A Gaussian process is essentially a Gaussian distribution for functions

$$f(q) \sim GP(m(q), c(q, q'))$$

Example: Take

$$m(q) = 0$$

$$c(q, q') = e^{-(q-q')^2/L^2}$$

Question: How do we build model based on training data $\{(q^m, y^m)\}$?

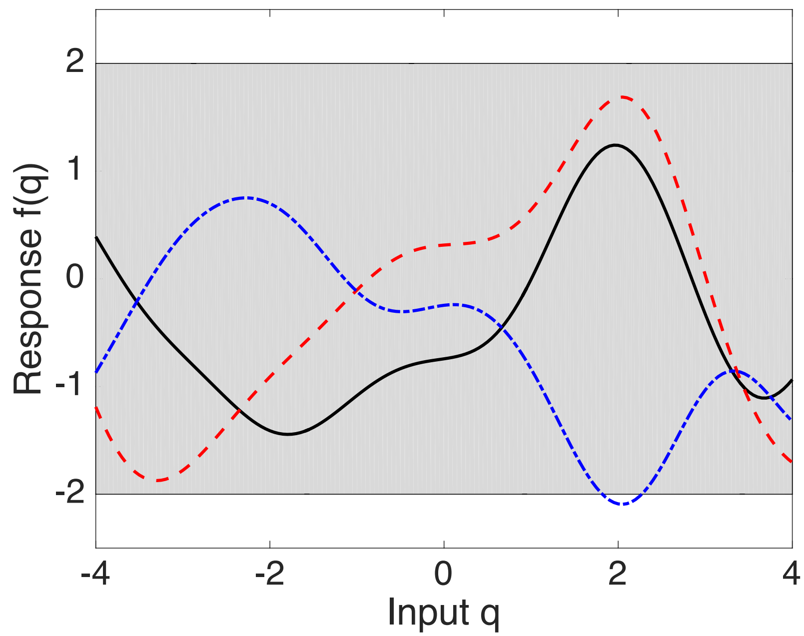
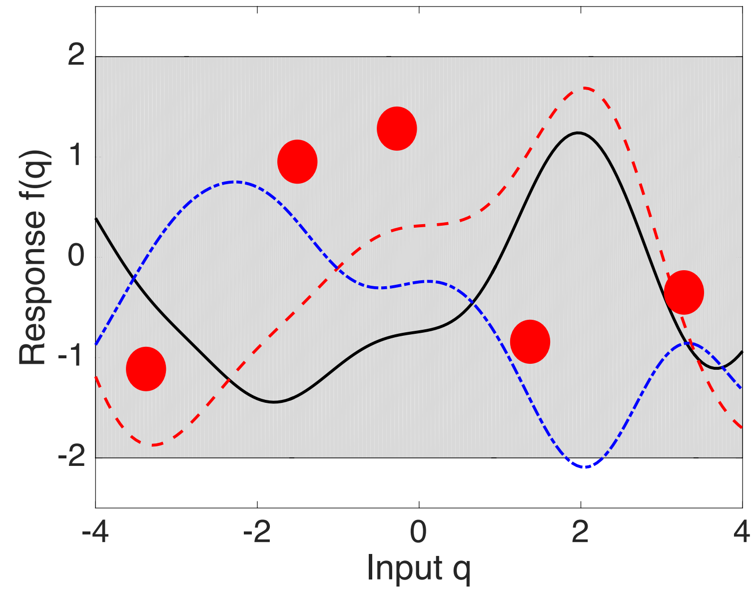


Figure: 3 realizations with L = 1

Gaussian Process (GP) Emulators

Solution: Restrict functions to those that interpolate or regress by conditioning on training data $\{(q^m, y^m)\}$ ●



Example: Five data points ●

Note: This can be done analytically for Gaussian functions!

Joint Prior Distribution:

Covariance Matrices

Training

$$\begin{bmatrix} y \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_0 1 \\ \beta_0 1^* \end{bmatrix}, \begin{bmatrix} C & C_* \\ C_*^T & C_{**} \end{bmatrix} \right)$$

Predictions

Conditional Predictive Distribution: Conditioned to fit training data

$$f^* | Q, Q^*, y \sim \mathcal{N}\left[\beta_0 1^* + C_*^T C^{-1}(y - \beta_0 1), C_{**} - C_*^T C^{-1} C_*\right]$$

Training

Mean

Covariance

Gaussian Process (GP) Emulators

Solution: Restrict functions to those that interpolate or regress by conditioning on training data $\{(q^m, y^m)\}$

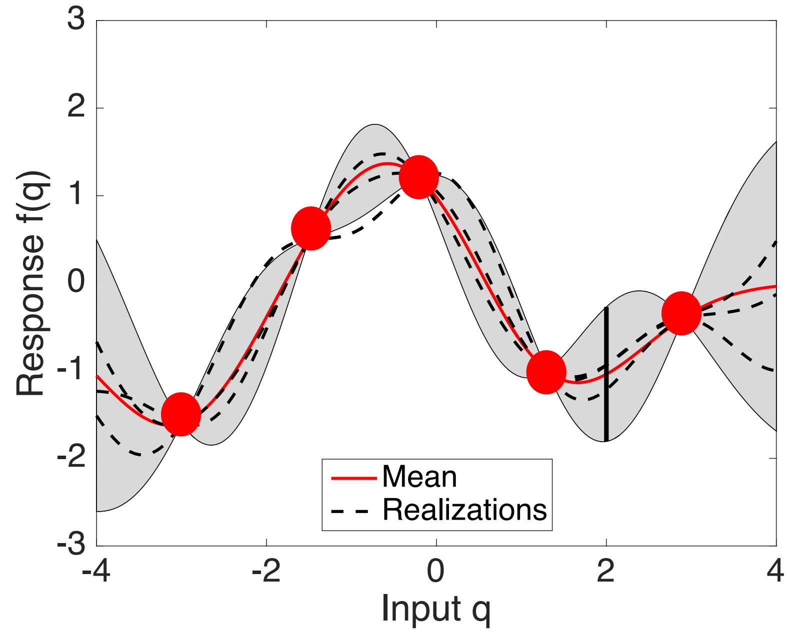
Note: This can be done analytically for Gaussian functions!

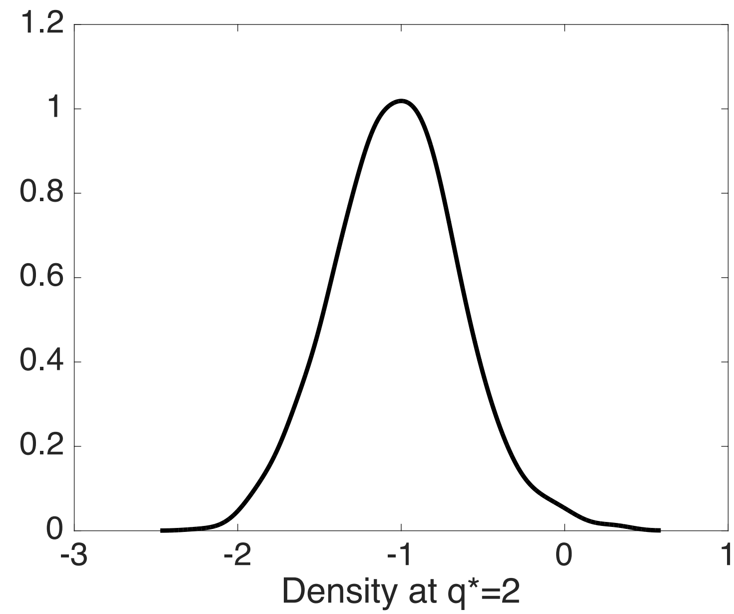
Conditional Predictive Distribution:

$$E[f^*] = \beta_0 1 + C_*^T C^{-1}(y - \beta_0 1)$$

$$\text{cov}[f^*] = C_{**} - C_*^T C^{-1} C_*.$$

Result: This quantifies uncertainty in future predictions!

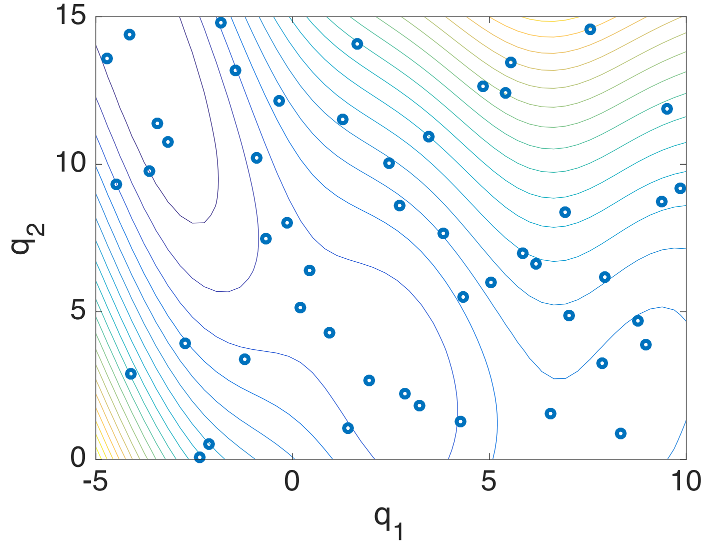




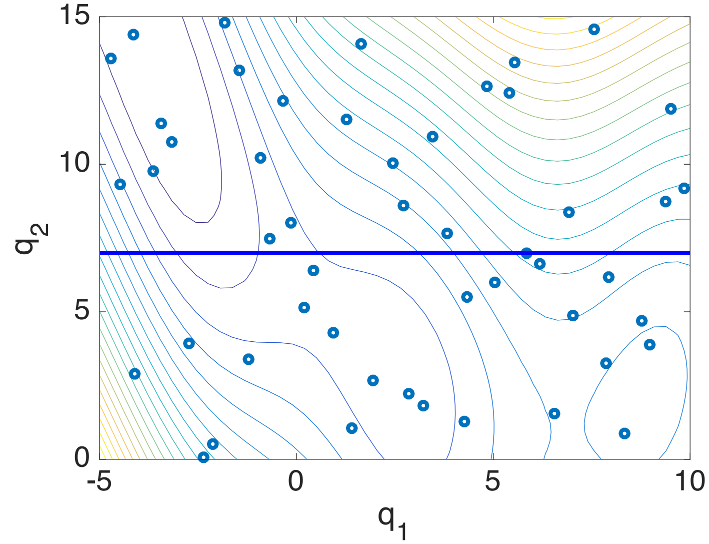
Gaussian Process (GP) Emulators

Example: Consider the modified Branin function

$$f(q) = \left[q_2 - \frac{5.1}{4\pi^2} q_1^2 + \frac{5}{\pi} q_1 - 6 \right]^2 + 10 \left[\left(1 - \frac{1}{8\pi} \right) \cos q_1 + 1 \right] + 5 q_1$$



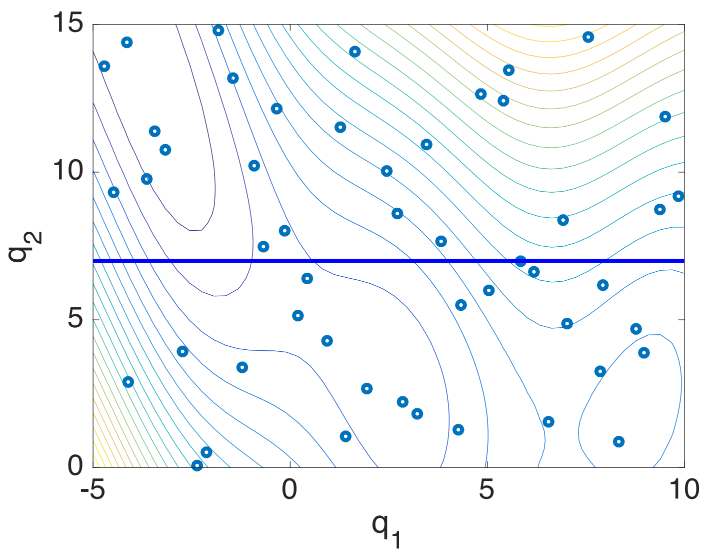
Function and Data

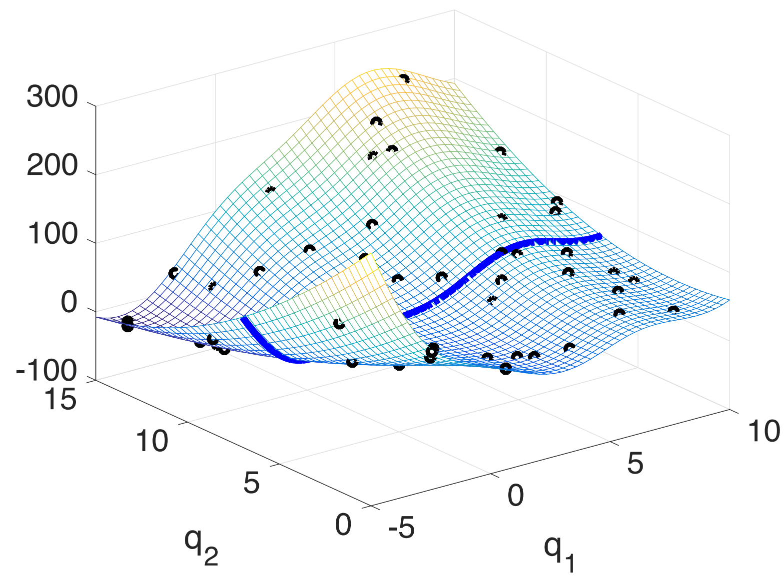


Data and GP Emulator

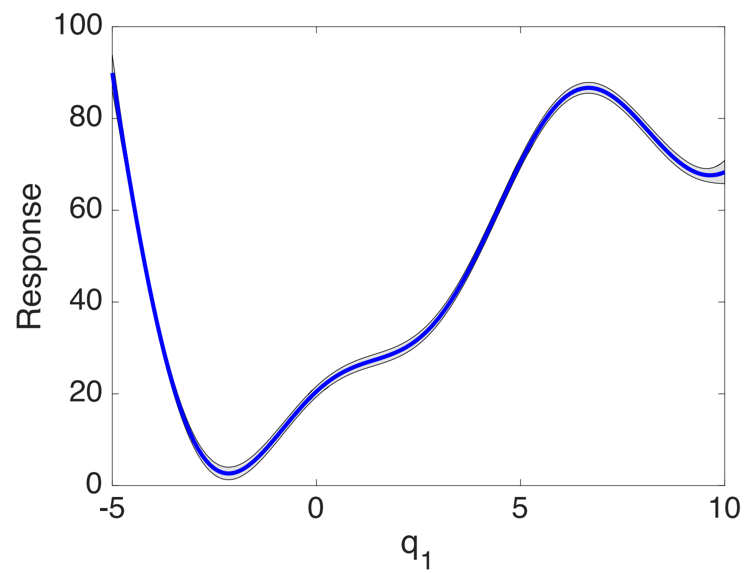
Gaussian Process (GP) Emulators

Example: Consider the modified Branin function





Note: GP emulator construction in multiple dimensions is quite straight-forward



Example: Modeling of Volcanic Pyroclastic Flows

Authors: Bayarri, Berger, Calder Dalbey, Lunagomez, Patra, Pitman, Spiller, Wolpert; *Technometrics,* 51(4), 2009; Gu and Berger, *The Annals of Applied Statistics*, 2016.

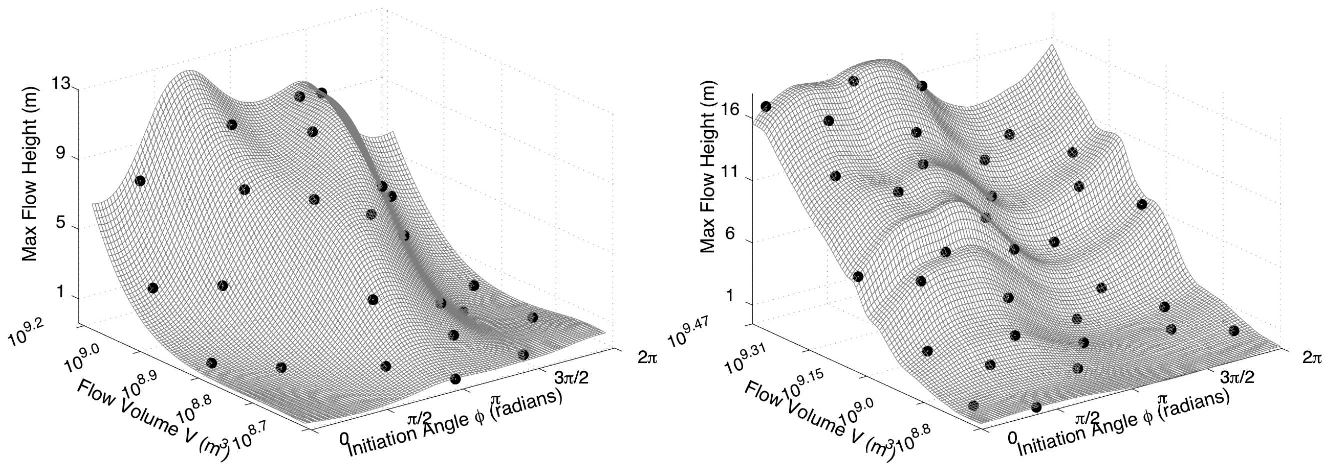
Objectives:

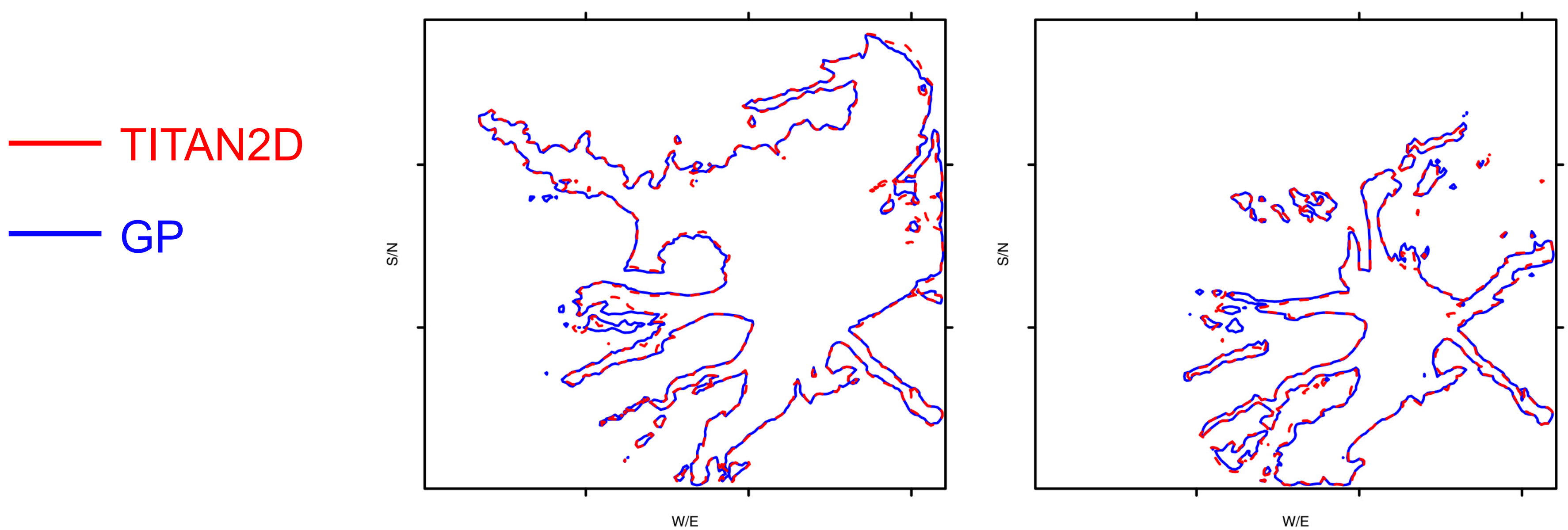
- Employ simulation models and surrogates to assess risk of *rare* catastrophic events; e.g., volcanic eruption.
- Employed TITAN2D to simulate flows.
- Test Case: Soufrière Hills Volcano on Island of Montserrat.
- Use emulator to identify threshold inputs – e.g., critical flow depth – that define catastrophic event.
- Compared GP and mathematical surrogates; GP advantageous for this application.

Example: Modeling of Volcanic Pyroclastic Flows

Objectives:

- Use emulator to identify threshold inputs – e.g., critical flow depth – that define catastrophic event. Employed TITAN2D and GP surrogates.





Concluding Remarks

Notes:

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

• 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 is critical to isolate identifiable and influential parameters.

• Surrogate models critical for computationally intensive simulation codes.

• Codes and packages: Sandia Dakota, R, MATLAB, Python, nanoHUB.

• *Prediction is very difficult, especially if it's about the future*, Niels Bohr.