SAND2019-9964PE

Bayesian Calibration of the Thermal Battery

1544 Department Meeting

Summer Seminar

07/25/2019

PRESENTED BY

Kyle Neal

Mentors: Josh Mullins & Ben Schroeder

Manager: Angel Urbina, Sandia National Labs

Advisor: Sankaran Mahadevan, Vanderbilt University

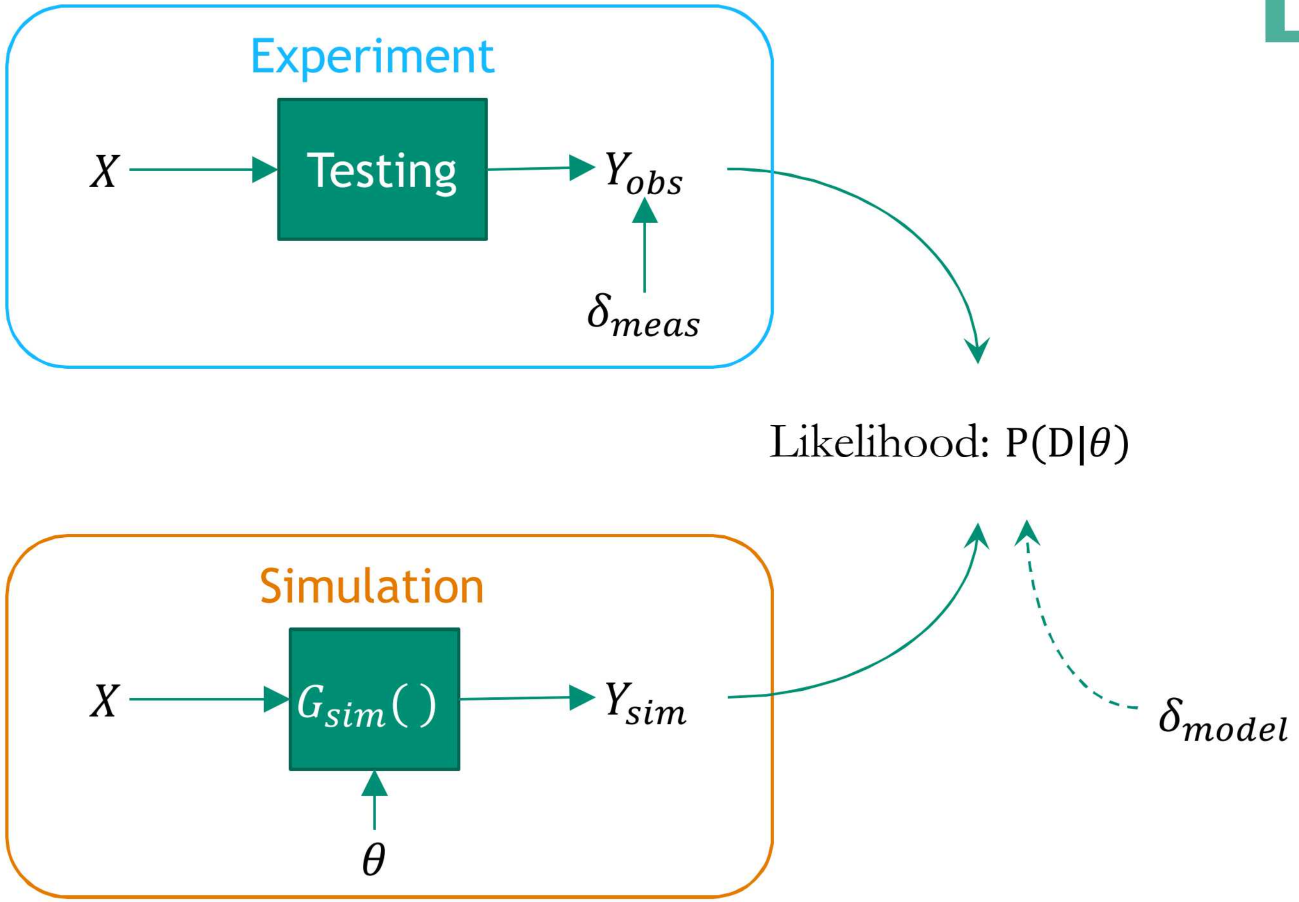
Colleague: Abhinav Subramanian, Vanderbilt University

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

- Motivation
  - Thermal battery
    - Spatio-temporal data
    - Challenges
- Principal component analysis (PCA)
  - Calibration in latent response (eigen) space
- Novel Sampling Algorithm
  - IISGA
    - Iterative importance sampling
    - Genetic algorithm
- Future Work
  - Model discrepancy formulations

Robust Importance Sampling for Bayesian Model Calibration with Spatio-Temporal Data
(2019, in preparation)



X – test settings

θ – model parameters (will be calibrated)

θ^* – sample/realization

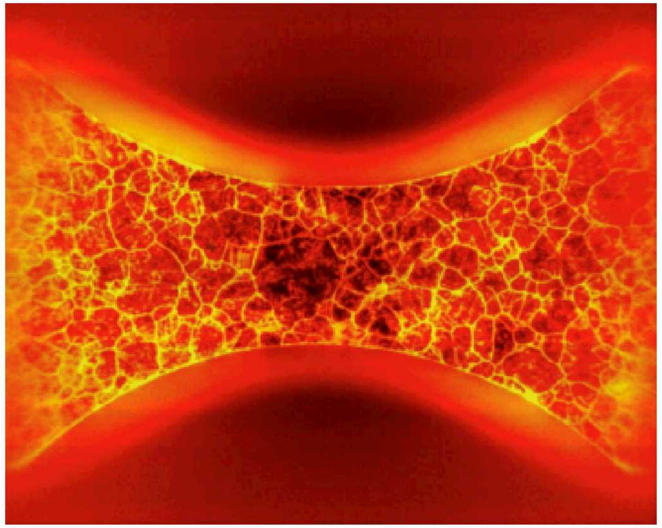
δ_{model} – model discrepancy

Motivation

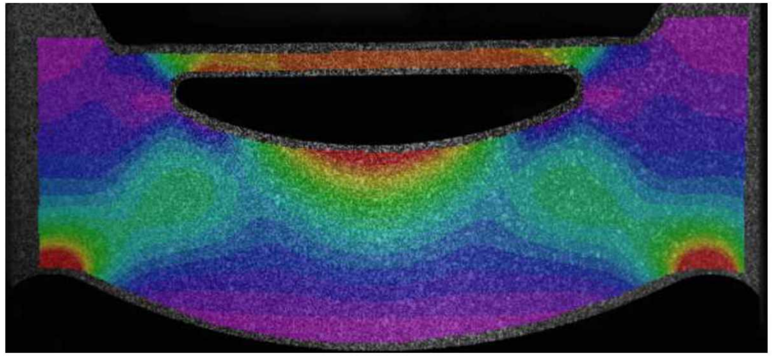
Spatio-temporal outputs "Field Data" are common

Examples

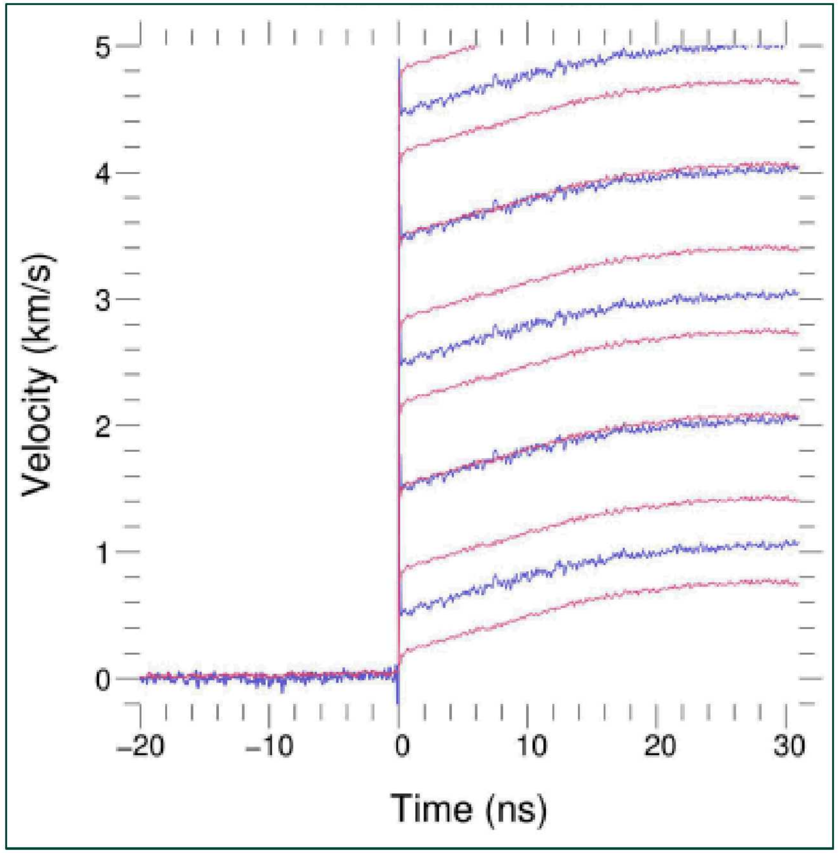
- Digital image correlation
- Thermal heatmap
- Time series data



Thermography for coupon test[1]



DIC of D-Specimen[2]



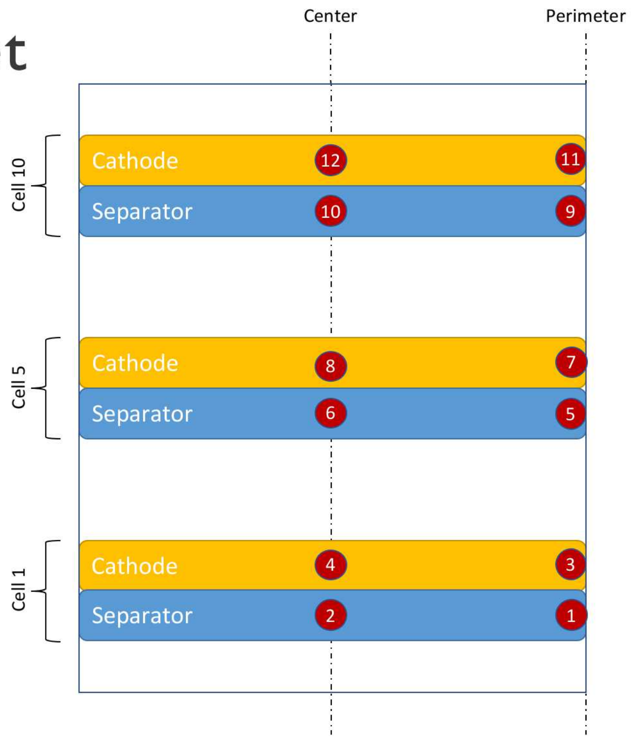
Time series realizations[3]

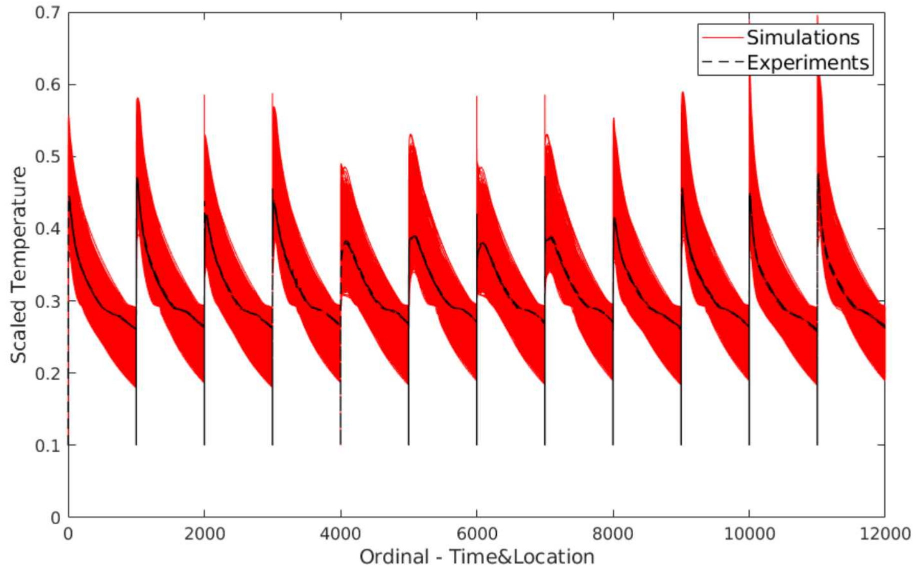
1 SAND2014-2227P
2 SAND2017-10365PE
3 SAND2019-7427C

Thermal battery dataset

Temperature is spatially and temporally varying

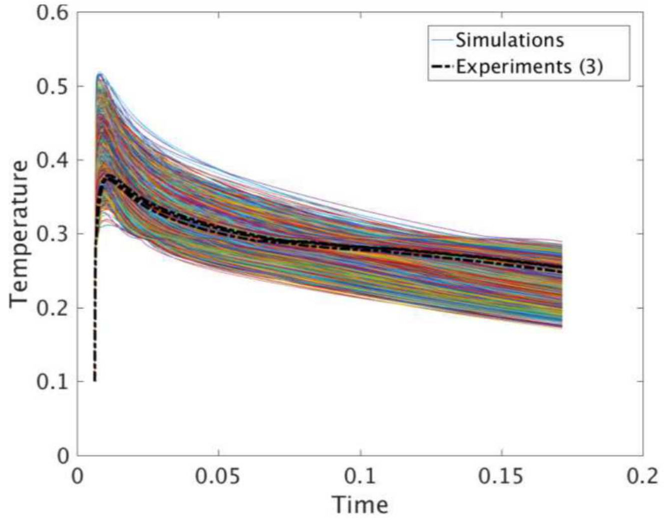
- High dimensional output space





Data preprocessing

- Sim & exp same length
  - Truncate traces
  - Zero padding is also an option
- Time steps the same
  - Align traces
  - Interpolation



How to organize data in a matrix?

- Separate matrices for exps and sims

$$location1, location2, \ldots, location12$$

$$[t_1, \ldots, t_N] \quad [t_1, \ldots, t_N] \quad [t_1, \ldots, t_N]$$

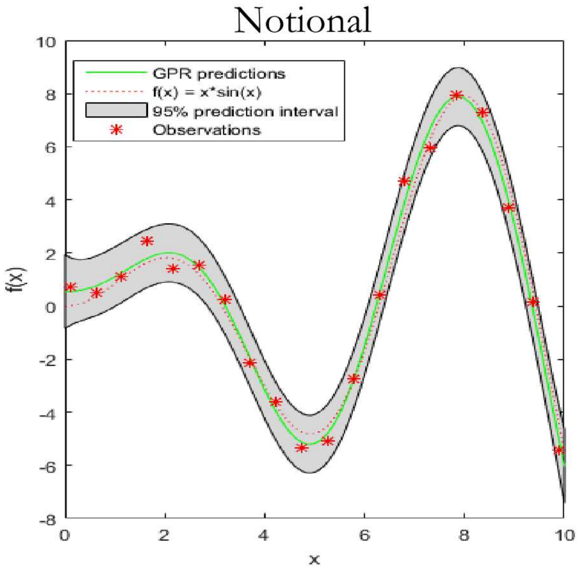
samples ↓ $\left[\quad \ldots \quad \right]$

Why spatio-temporal data must be processed before Bayesian Calibration a) surrogate model construction

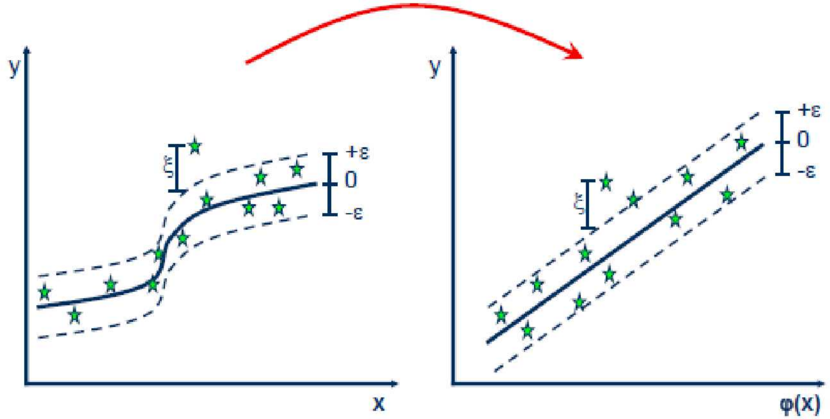
Building surrogate models

- Gaussian process, SVM, polynomial chaos, ANN, etc. [3]
- One-to-many mapping
  - Inputs are RVs and output is random process/field
- Approaches:
  - Build separate surrogate for each location
  - Include time/space as an input
  - Feature selection
  - Decomposition / dimension reduction technique

Gaussian Process [1]



Support Vector Machine [2]
Nonlinear Regression



GP

- Nonparametric kernel based model
- Probabilistic

SVM Regression

- Hyperplane that maximizes the margin
- Nonlinear – map data (via kernel transform) to higher dimension where linearly separable
- Parametric/nonparametric depending on linear/kernel based
- Typically deterministic

[1] MathWorks
[2] saedsayad.com by Dr. Saed Sayad

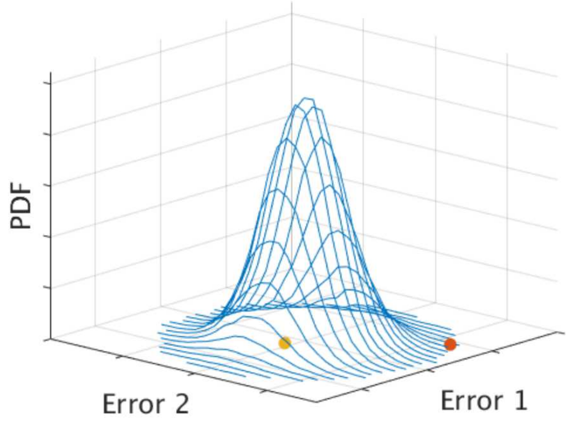
[3] Neal, K., Hu, Z., Mahadevan, S., & Zumberge, J. (2019). Discrepancy prediction in dynamical system models under untested input histories. Journal of Computational and Nonlinear Dynamics, 14(2), 021009.

Why spatio-temporal data must be processed before Bayesian Calibration b) likelihood covariance

$$L(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\sigma}) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} exp\left(-\frac{1}{2}\left(\left(\boldsymbol{y}_{obs,i} - \boldsymbol{y}_{sim}(\boldsymbol{\theta})\right) - \boldsymbol{\mu}\right)^T \Sigma_y^{-1}\left(\left(\boldsymbol{y}_{obs,i} - \boldsymbol{y}_{sim}(\boldsymbol{\theta})\right) - \boldsymbol{\mu}\right)\right)$$

Field data has thousands of unique outputs

- High-dimensional joint PDF for likelihood function
- In the case of a Gaussian likelihood (Gaussian error sources),
  - Determinant and inverse of covariance matrix required
  - "As the number of data points [output quantities] increases, this covariance matrix may become ill-conditioned and lead to significant numerical errors in the computation of the likelihood function" [1]



Correlation of error terms

- Correlation in covariance matrix required
  - If a model makes a poor prediction at one output location (spatial or temporal), it is probable that it will also fail at a nearby output location, which suggests statistical correlation between model discrepancies at these two output locations
- Correlation can be calibrated, but increases number of calibration terms
  - Pulls from same experimental data used to update model parameters – which is what we actually want
  - Makes sampling (e.g., MCMC) more difficult due to curse of dimensionality

How to reduce dimension of output?

- Feature selection
  - Doesn't address correlations
- Mathematical decomposition
  - If eigen based, removes correlation

[1] Ling, Y., Mullins, J., and Mahadevan, S., "Selection of model discrepancy priors in Bayesian calibration," *J. Comput. Phys.*, vol. 276, pp. 665–680, Nov. 2014.

PCA

Principal Component Analysis (PCA)

Difference between PCA and singular value decomposition (SVD)?

- SVD is a matrix decomposition (mathematical)
  - Generalized Eigen decomposition for rectangular matrices
- PCA is a strategy to remove correlation by mapping data onto principal directions (data science)
  - Eigen decomposition of covariance matrix
  - SVD on centered data matrix

→ Equivalent [1]

Mathematics of SVD [2]

$$A = USV^T$$

$$[n \times p] = [n \times r][r \times r][r \times p]$$

- n – number of samples
- p – number of dimensions (timesteps)
- r – rank of A, number of linearly independent rows or columns, or the dimension of the space that is spanned by the vectors it contains
  - maximum value for r is min(n,p)
- U and V are both column-orthonormal
- S – diagonal with singular values

Dimension Reduction

Latent response:

$$\gamma_{[n\times k]} = U^*_{[n\times k]} S^*_{[k\times k]}$$

where $k \ll r$

Convert back to trace:

$$A^*_{[n\times p]} = \gamma_{[n\times k]} {V^*}^T_{[k\times p]}$$

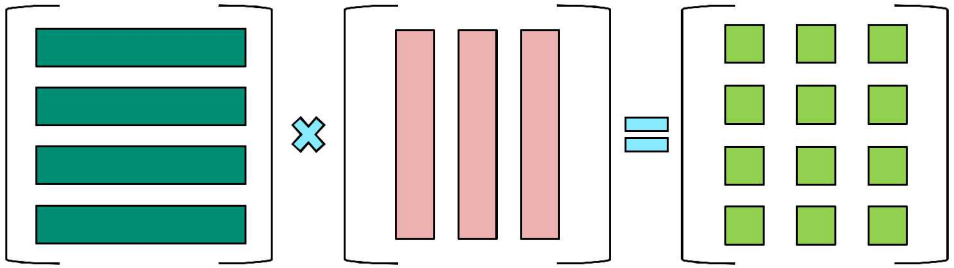
[1] https://stats.stackexchange.com/questions/134282/relationship-between-svd-and-pca-how-to-use-svd-to-perform-pca
[2] Van Buren, K., Reilly, J., Neal, K., Edwards, H., & Hemez, F. (2017). Guaranteeing robustness of structural condition monitoring to environmental variability. Journal of Sound and Vibration, 386, 134-148.

Converting to latent response space

- Each row of A is projected onto the k orthonormal column vectors in V

$$A_{[n\times p]}V^{*}_{[p\times k]} = \gamma_{[n\times k]} \ ,$$

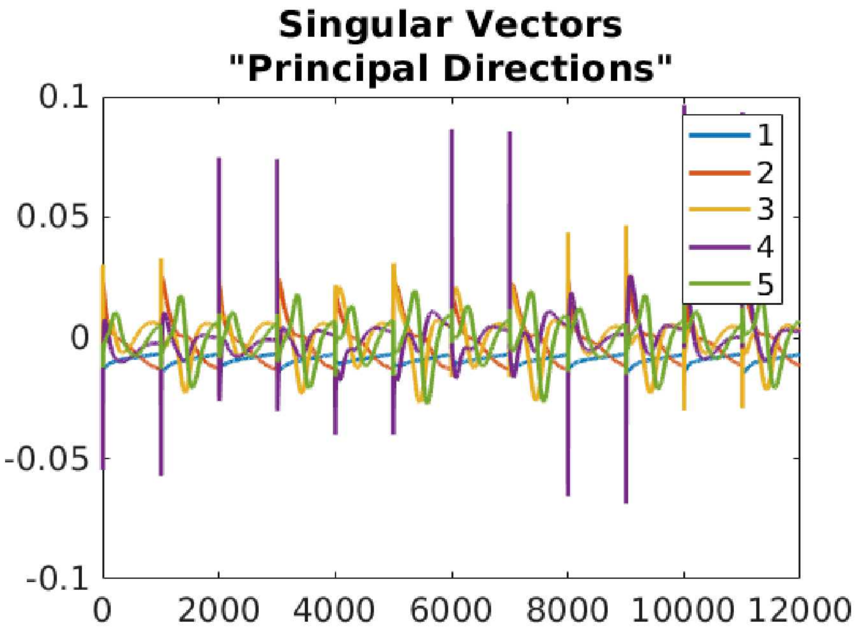
$$\left(V^{*T}_{[k\times p]}\right)^{-1} = V^{*}_{[p\times k]}$$

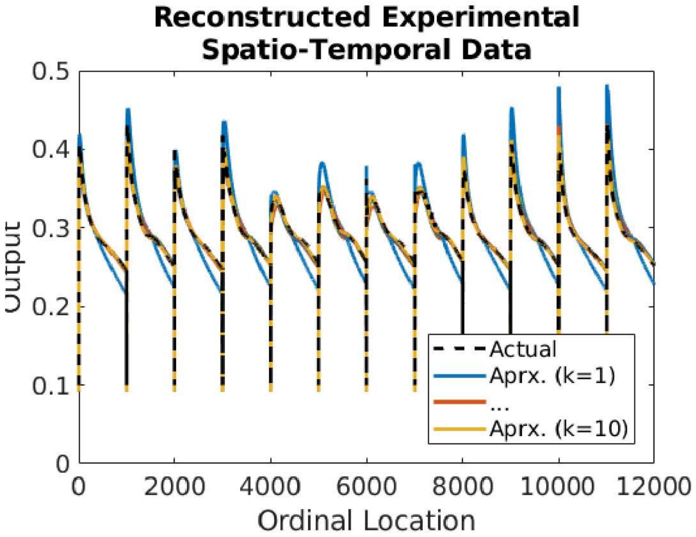


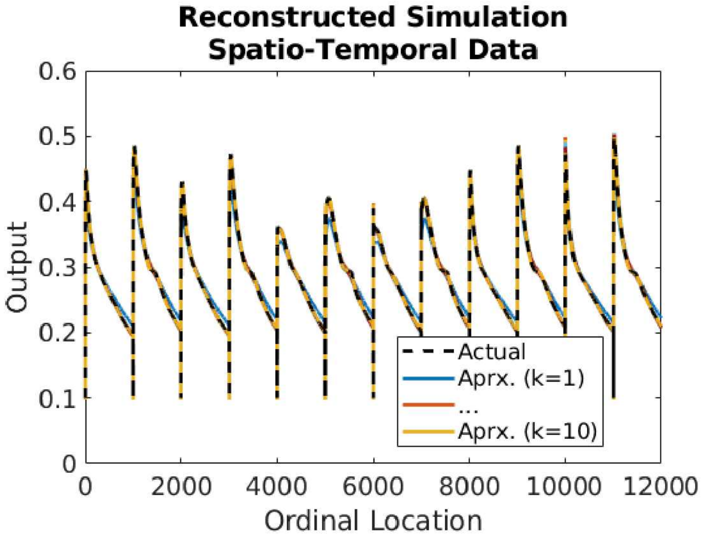
- This new space is termed "latent response", γ, and has k dimensions instead of p
- We want both simulation and experimental data to be in same latent response space for calibration
- Use latent response as outputs for calibration

$$A^{sim}_{[N\times p]}V^{*}_{[p\times k]} = \gamma^{sim}_{[N\times k]} \qquad A^{exp}_{[m\times p]}V^{*}_{[p\times k]} = \gamma^{exp}_{[m\times k]}$$

PCA on Thermal Battery





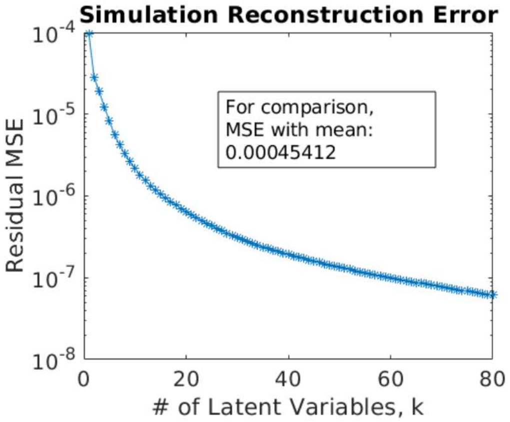


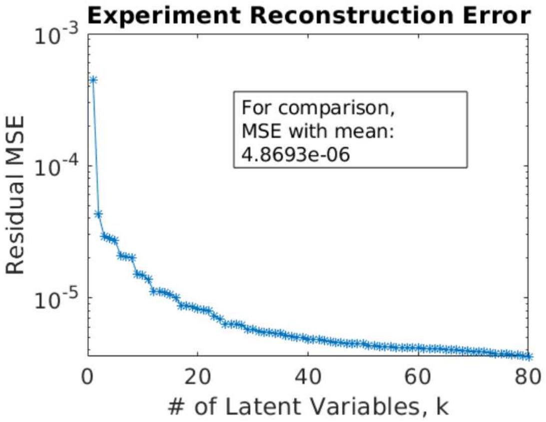
Selecting number of latent variables

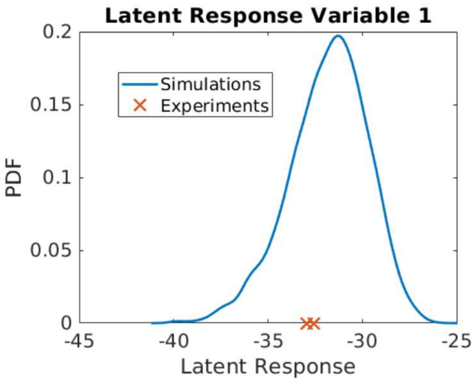
- Computed MSE of residual

Higher latent response variables may mostly be capturing noise

- Difficult to build a surrogate for these







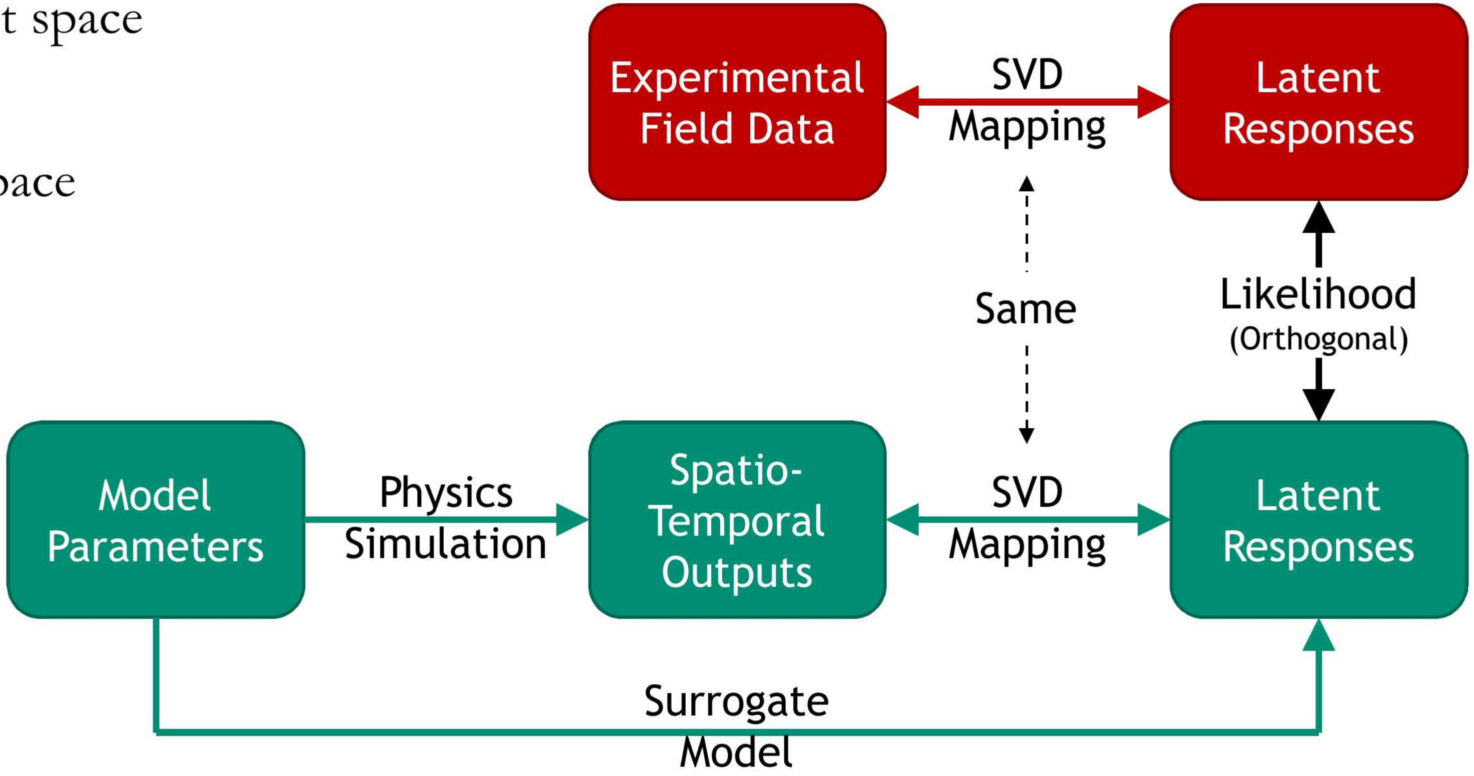
Recap: Bayesian Calibration in Latent Response Space

Convert simulations to latent response space
- Use SVD

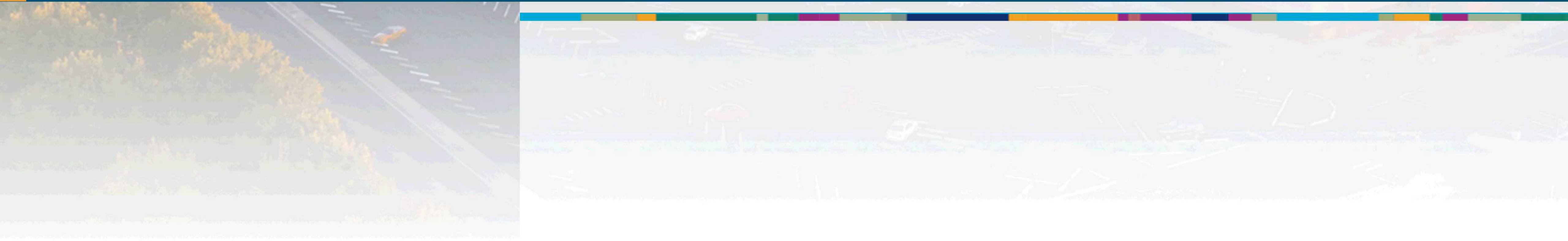
Build surrogates mapping model parameters to latent response

Convert experimental data to same latent space
- This is a change of basis

Perform calibration in latent response space
- Likelihood
  - Covariance will be diagonal
    → latent response space is orthogonal
  - Error terms must be calibrated

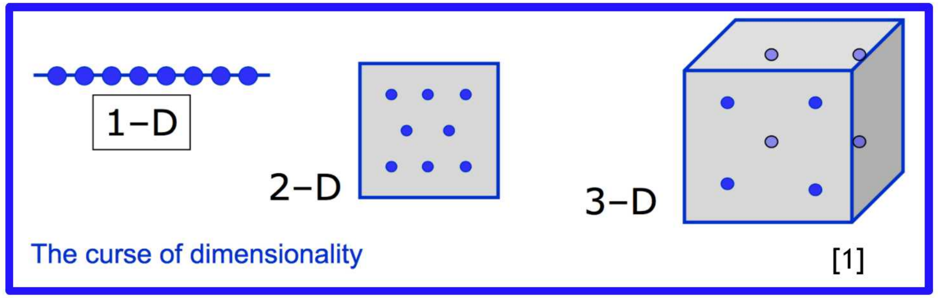


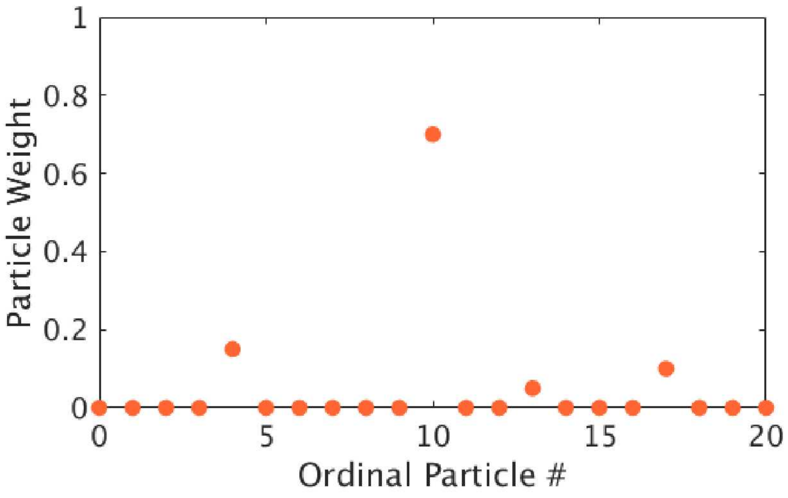
IISGA



Why IISGA is Needed

- Curse of dimensionality
  - Large number of parameters to calibrate
- Sample degeneration
  - Afflicts sequential Monte Carlo methods
- Leverage high performance computing
  - Parallelization





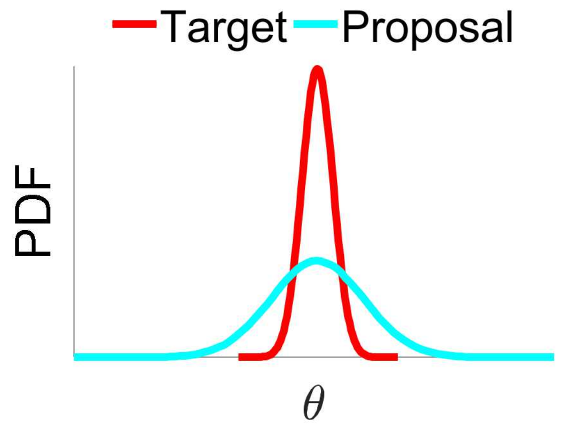
[2]

[1] "Bi-Clustering" by Jinze Liu,
Liu, J., & Wang, W. (2003, November). Op-cluster: Clustering by tendency in high dimensional space. In Third IEEE international conference on data mining (pp. 187-194). IEEE.
[2] Sandia.gov

IISGA – Iterative Importance Sampling with Genetic Algorithm

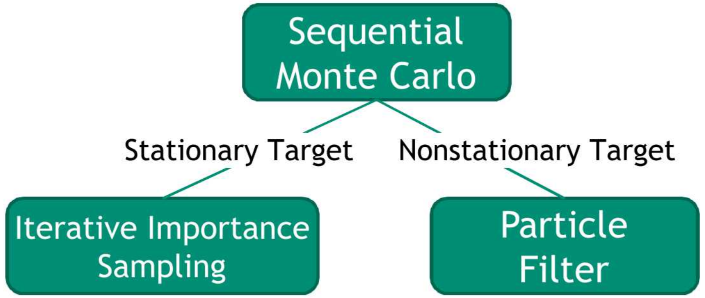
Iterative Importance Sampling

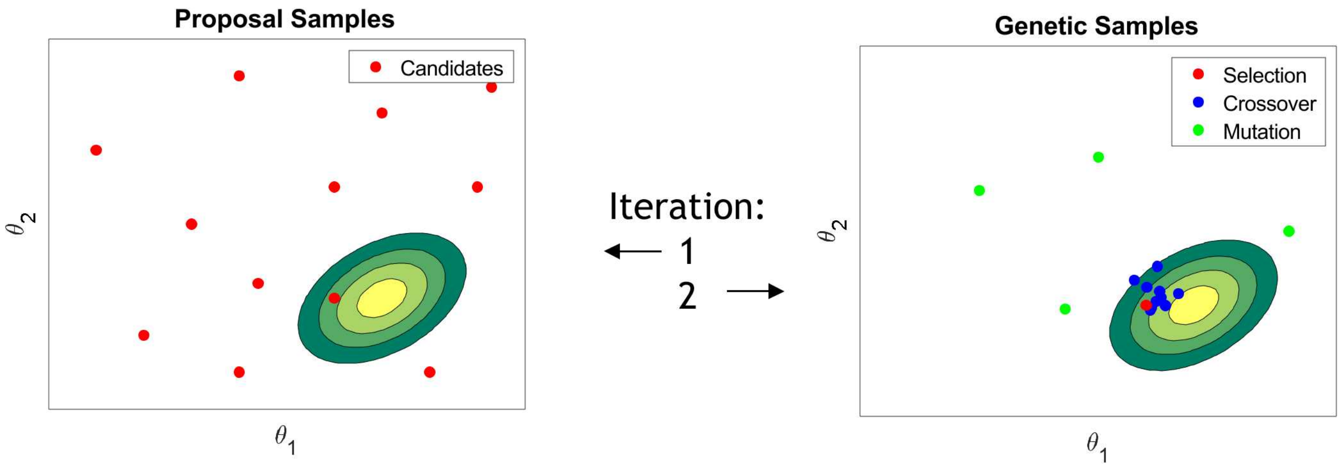
- Draw samples from known "Proposal" distribution $q(\theta)$
- Assign a weight to each sample
  - $w(\theta) \propto \frac{p(\theta|D)}{q(\theta)} = \frac{p(D|\theta)p(\theta)}{q(\theta)}$
- Iterative? → gradually moves the proposal towards the target



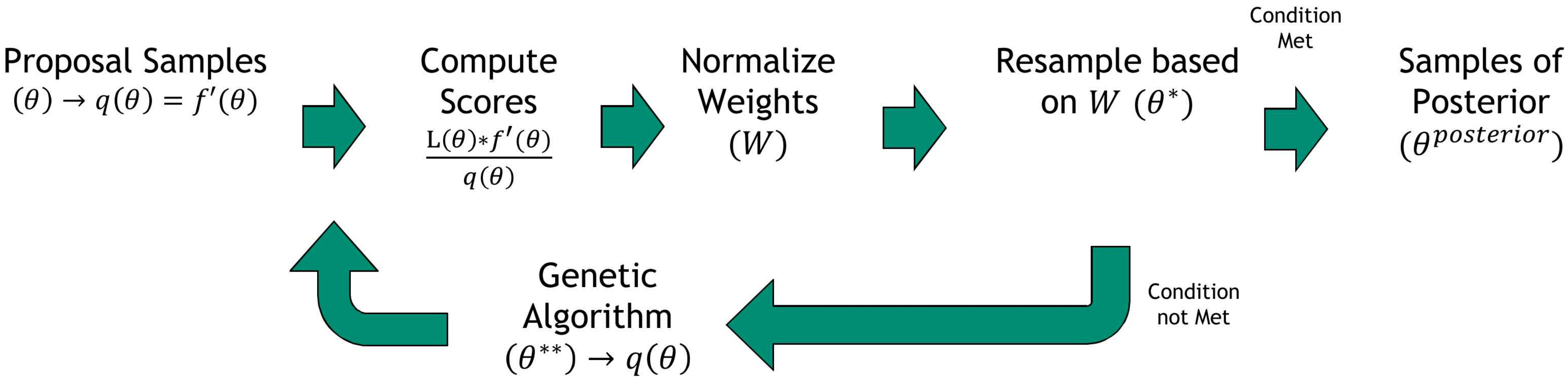
Genetic Algorithm

- Selection – retain best performing samples
- Crossover – add noise to existing samples
- Mutation – randomly sample from prior





IIGSA – Procedure described schematically



3 million posterior samples	
IISGA (40 cores)	Slice sampling (1 core)
10 hrs	495 days (predicted)

Significant speedup for previous calibration study

- Vectorized likelihood – most surrogate models are vectorized by default
  - MCMC can only evaluate samples serially
- Multi-core
  - MCMC can run separate chains on multiple cores

Not so fast.. Computational Challenges with IISGA

How to estimate proposal density?

- Through the GA we have proposal samples but no distribution

Curse of dimensionality persists

- Few samples from proposal land in target
- Posterior sample set is sparse

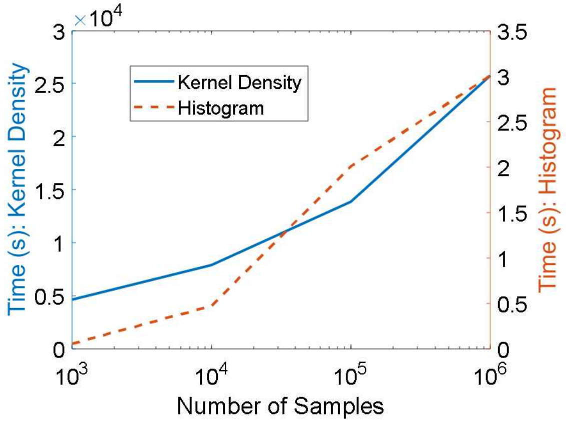
Frequency Histogram – Adaptive Binning

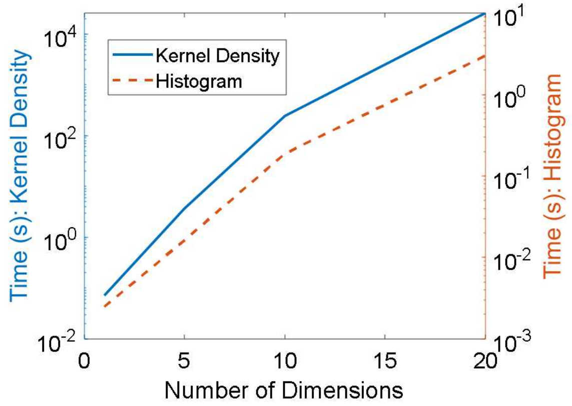
Assign a weight to each sample

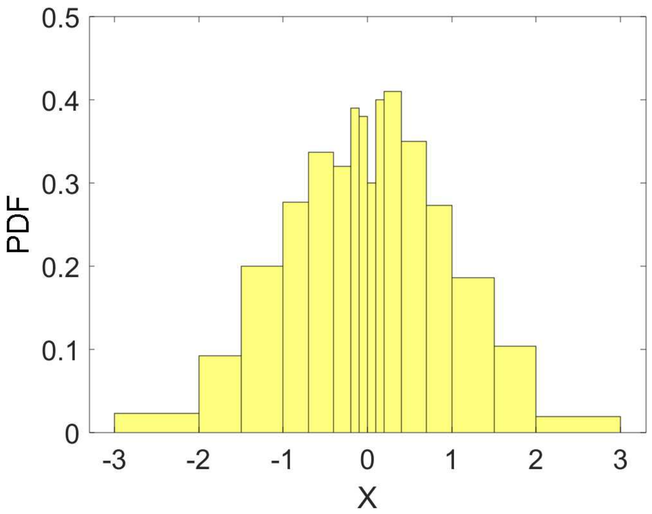
$$w(\theta) \propto \frac{p(\theta|D)}{q(\theta)} = \frac{p(D|\theta)p(\theta)}{q(\theta)}$$

Accurately estimate density of proposal distribution

- Kernel estimators are slow
- No built-in N-dimensional histogram function
  - Modified open source code to prevent overflow error
    - Grid of $(N_{bins})^{n_{dim}}$ was sparse, disposed of empty bins
- How to decide on the number of bins?







Adaptive binning

- Initially using coarse bins
- For areas of high density, do bin size refinement

Challenges with Curse of Dimensionality (1) Posterior becomes hard to find

Most samples have likelihood weights of ~0, and these samples are lost in resampling

When prior covers a much larger region than posterior, particularly in high dimensions

- $\{\theta^{prior}\} >> \{\theta^{post}\} \therefore \{Y_{sim}(\theta^{prior}, x)\} >> \{Y_{sim}(\theta^{post}, x)\}$
- See plot at right
- Recall that the posterior is proportional to the likelihood

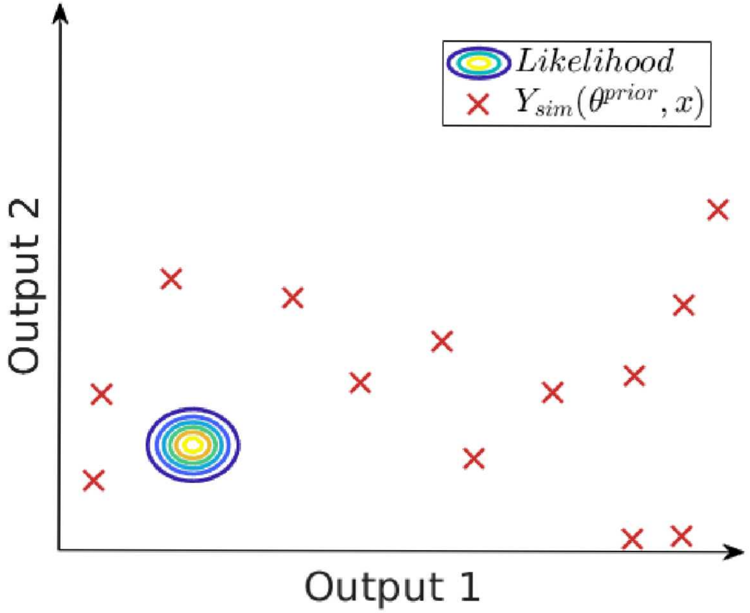
This effect becomes more pronounced as θ becomes higher dimensional

- If for example,
  - $range(\theta_i^{prior}) = 2 * range(\theta_i^{post}) \; \forall \, i \in \{1:16\}$
  - $\frac{Volume(\theta^{post})}{Volume(\theta^{prior})} = \frac{.5^{16} * Volume(\theta^{prior})}{Volume(\theta^{prior})} = 0.000015$

Essentially, it is difficult for samples from the prior to find (land in) the posterior

Solution: artificially inflate the likelihood

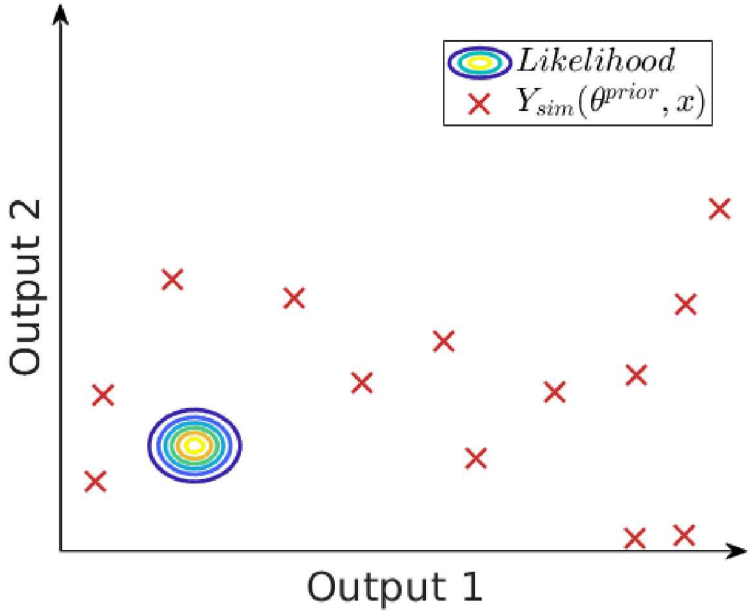
Note, inflating the likelihood works for SMC samplers, not for MCMC



Inflated Likelihood

Idea: scale the likelihood covariance matrix
- This will flatten the likelihood function
- More samples will have non-zero weights
- Reduce scale factor over iterations to avoid a bias posterior

Similar to simulated annealing
- "Temperature" acts as a SF for the distance (in terms of objective function) a proposed x' can be from current best solution x
  - Decreases over iterations



Optimization Terminology

$$\max f(x)$$
$$\text{s.t. } x \in \Omega$$

f – objective function
x – decision variable
Ω – feasible set

Calibration Parallels

f → L – likelihood function
x → θ – model parameters
Ω – → π' – prior distribution

Simulated Annealing (maximize)

Iterate over values of x

$\Delta = f(x') - f(x)$
if $\Delta \geq 0$
    $x = x'$
else
    if $\exp\left(\frac{\Delta}{Temp}\right) \geq \text{Uniform}(0,1)$
        $x = x'$
    end
end

Reduce $Temp$

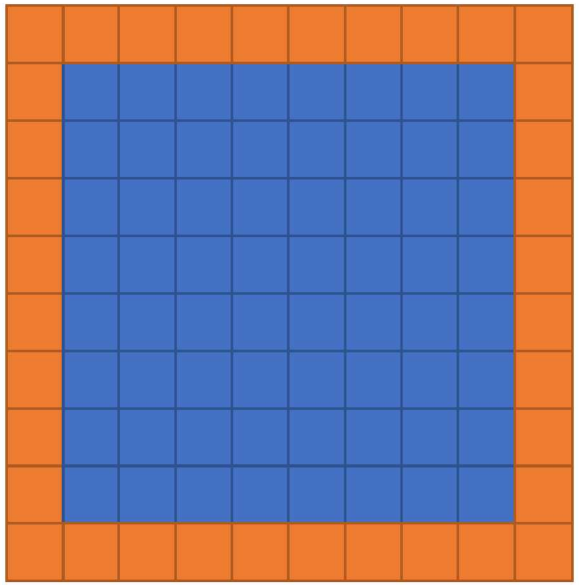
Challenges with Curse of Dimensionality (2) More samples become close to perimeter

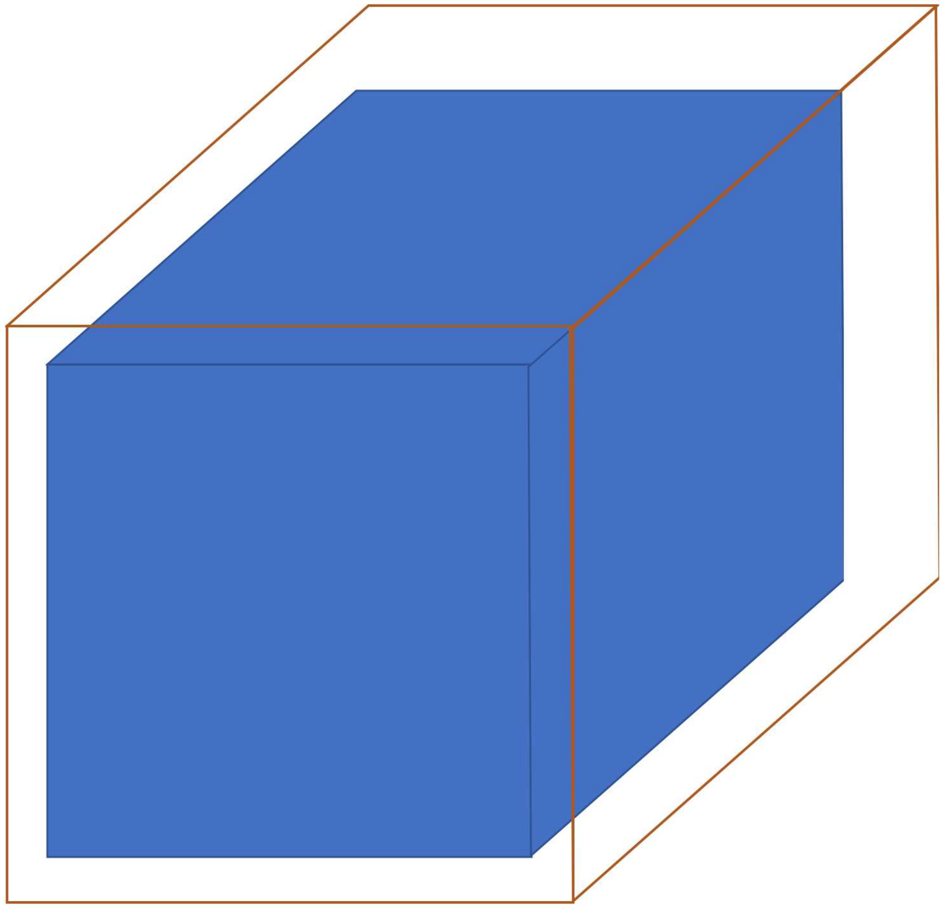
Solution:

- Gaussian noise in crossover
- add check that sample is inside bounds

Computed below: samples in a perimeter cell

Causes problems with noise added in crossover (GA)





$$1-\left(\frac{8}{10}\right)^{1}=20\%$$

$$1-\left(\frac{8}{10}\right)^{2}=36\%$$

$$1-\left(\frac{8}{10}\right)^{3}=48.8\%$$

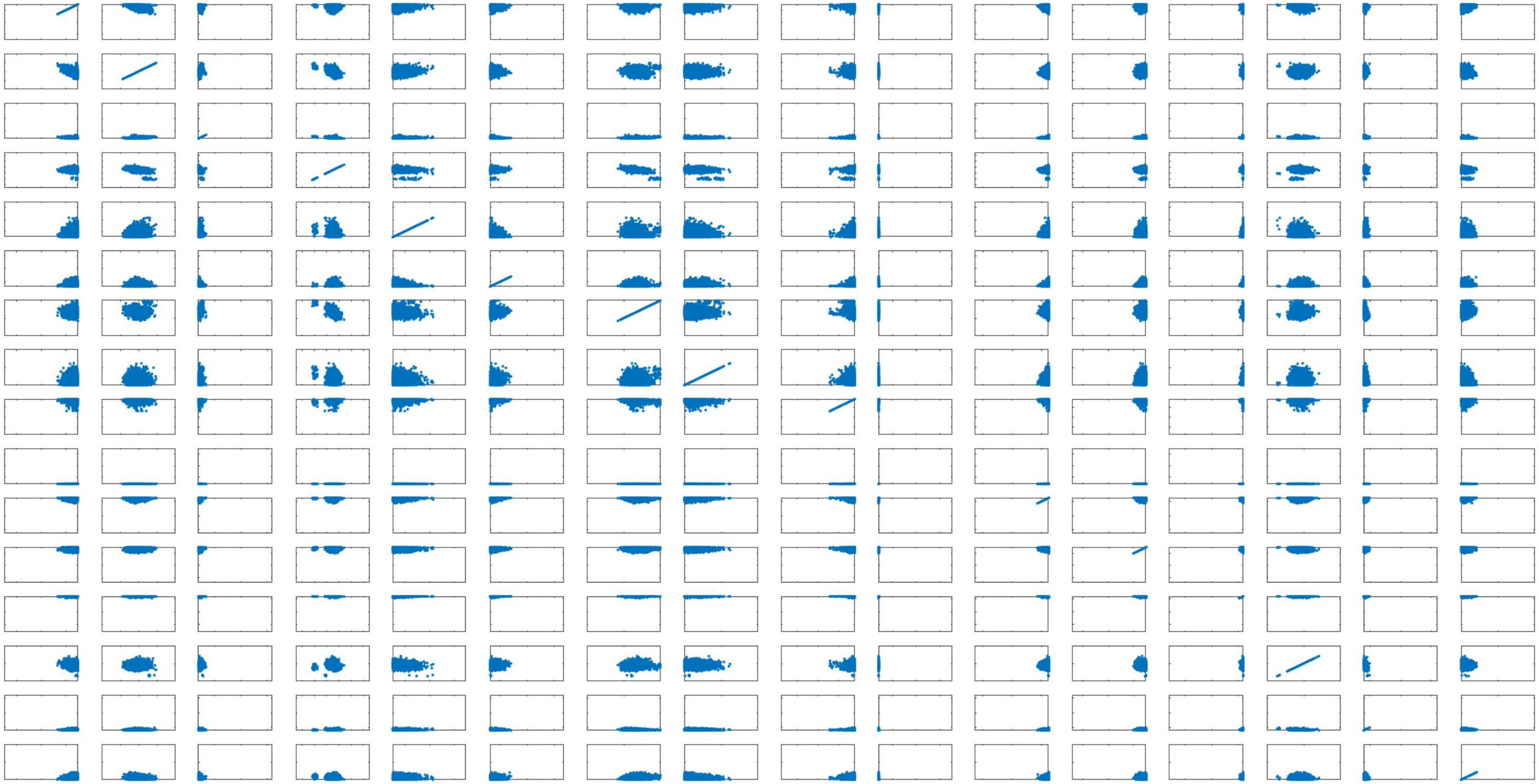
$$1-\left(\frac{8}{10}\right)^{26}=99.7\%$$

Results from Thermal Battery

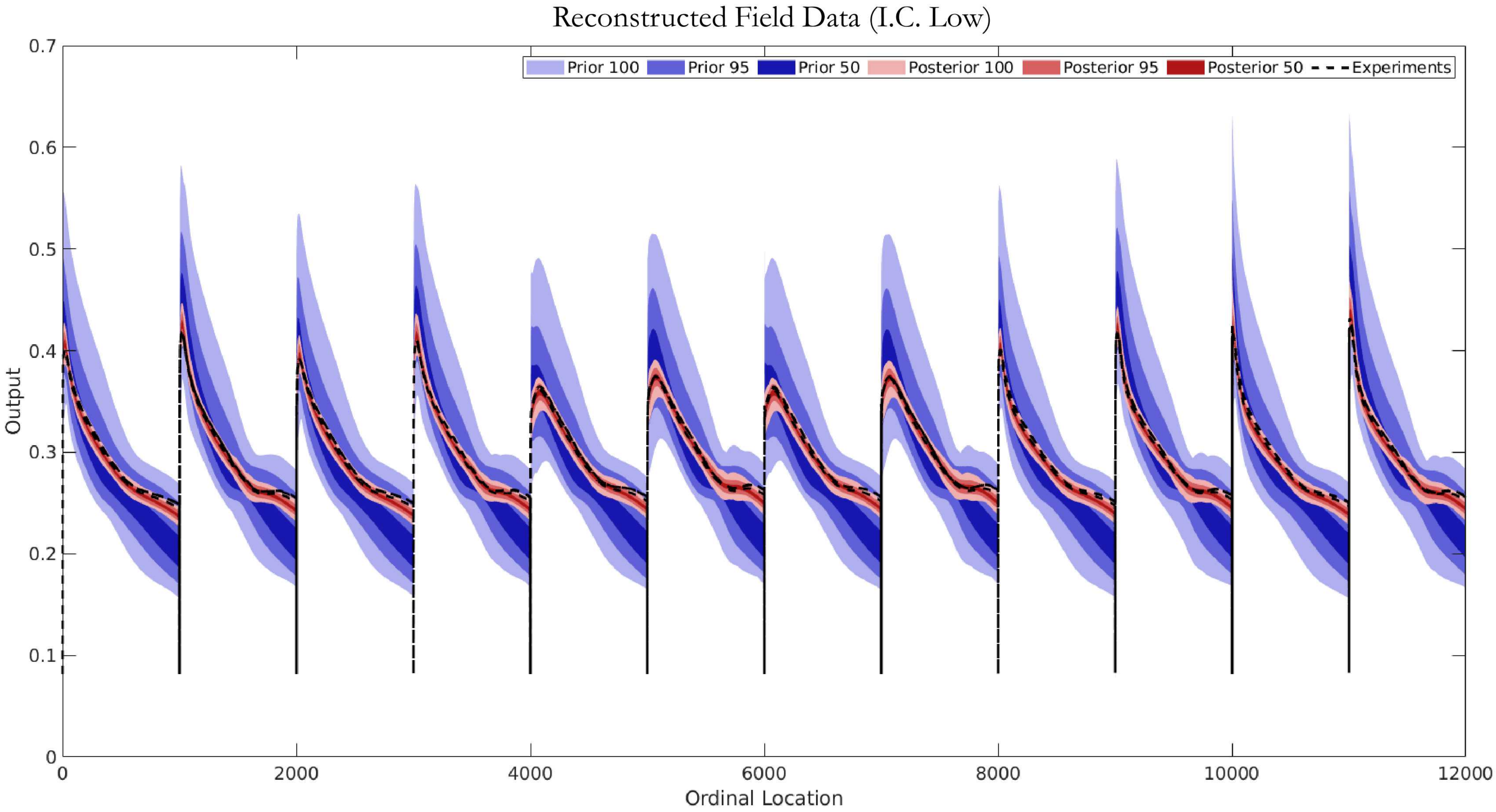
Posterior Samples

Axis limits are range of uniform priors

- 16 model parameters, plotted to show correlations



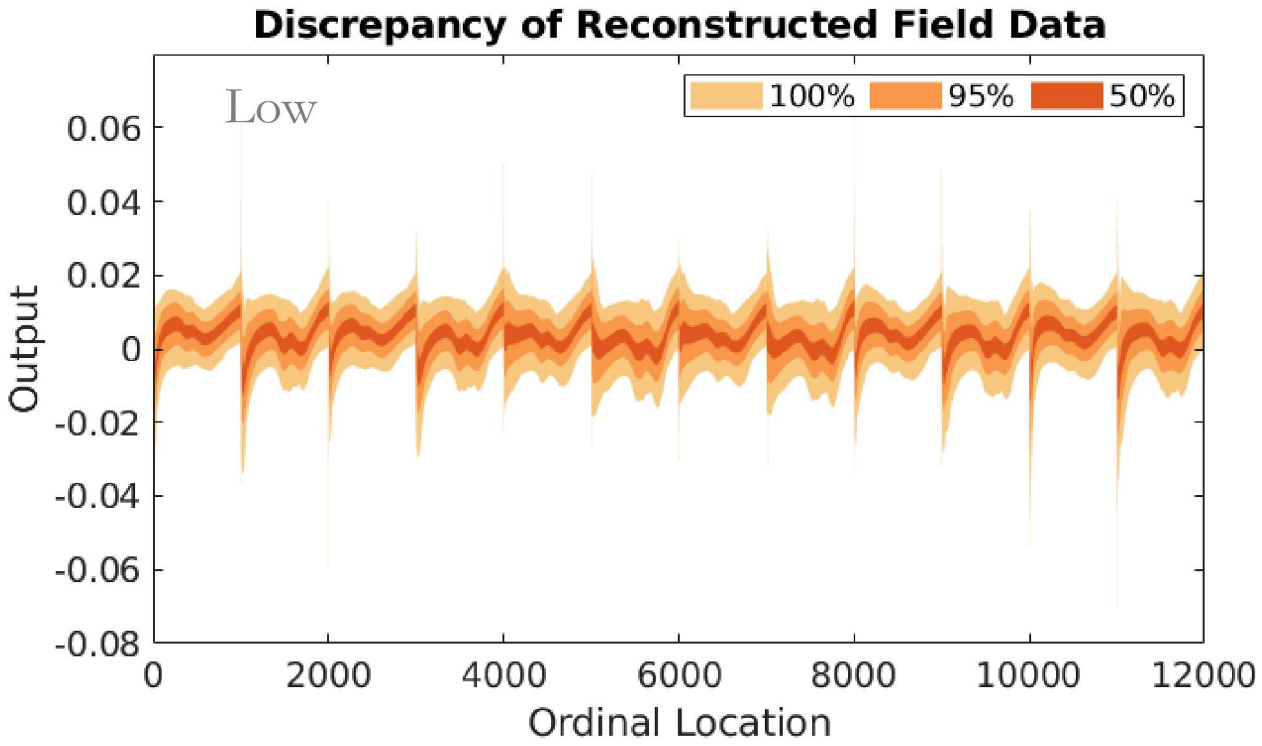
Calibration Fit

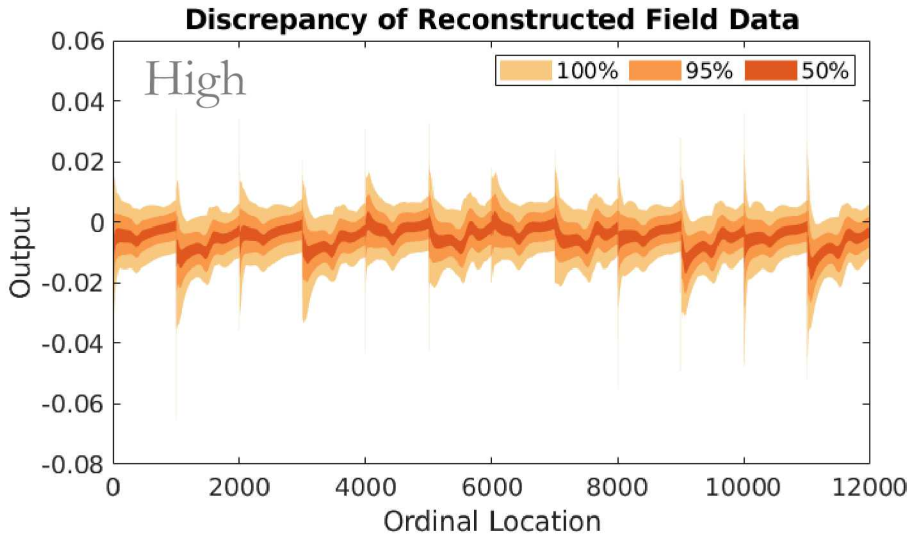


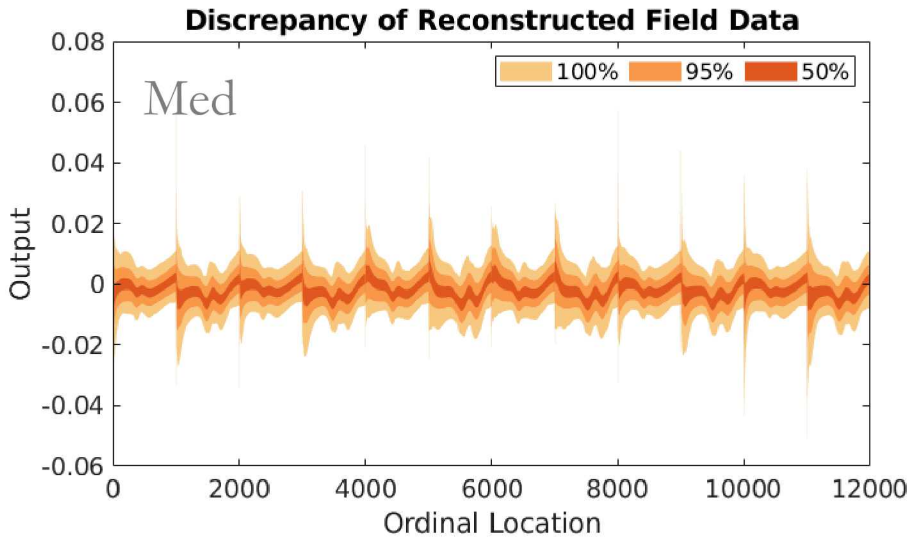
Future Work

Model Discrepancy (after calibration)

- $\delta_{model}(x) = mean\big(Y_{obs}(x)\big) - Y_{sim}(\theta^{post}, x)$
- δ_{model} varies over time, location, and initial temperature







Simultaneous Parameter Calibration and Model Discrepancy Estimation

Multiple discrepancy formulations

Extend these to latent response space

- Can map back to spatio-temporal space

KOH framework

- A Gaussian process is used to represent the dependence of model discrepancy on input conditions (initial temperature)

Seminar (Early Fall): Non-intrusive state estimation of model error

- Model error – differential equations
  - Can be extrapolated
- Model discrepancy – output quantities
  - May only be applicable to calibration test settings
- Work with Abhinav to apply his methods to thermal battery

Thank you

Kyle Neal

kneal@sandia.gov

References

Neal, K., et al. (2019) Robust Importance Sampling for Bayesian Model Calibration with Spatio-Temporal Data (in preparation)

Neal, K., Hu, Z., Mahadevan, S., & Zumberge, J. (2019). Discrepancy prediction in dynamical system models under untested input histories. Journal of Computational and Nonlinear Dynamics, 14(2), 021009.

Ling, Y., Mullins, J., and Mahadevan, S., "Selection of model discrepancy priors in Bayesian calibration," *J. Comput. Phys.*, vol. 276, pp. 665–680, Nov. 2014.

Van Buren, K., Reilly, J., Neal, K., Edwards, H., & Hemez, F. (2017). Guaranteeing robustness of structural condition monitoring to environmental variability. Journal of Sound and Vibration, 386, 134-148.

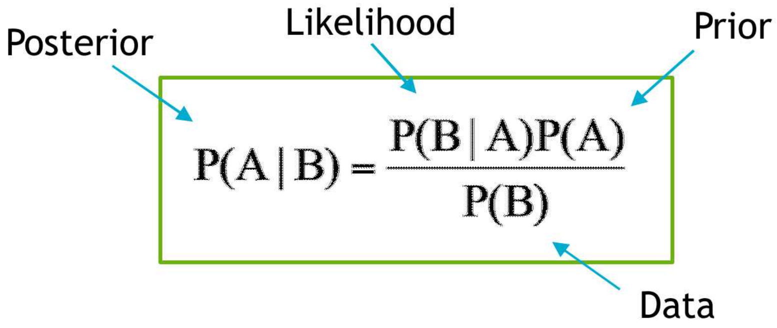
Liu, J., & Wang, W. (2003, November). Op-cluster: Clustering by tendency in high dimensional space. In Third IEEE international conference on data mining (pp. 187-194). IEEE.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464.

Morzfeld, M., Day, M. S., Grout, R. W., Heng Pau, G. S., Finsterle, S. A., & Bell, J. B. (2018). Iterative importance sampling algorithms for parameter estimation. *SIAM Journal on Scientific Computing*, *40*(2), B329-B352.

Backup Slides

Bayesian Calibration



Bayes' Theorem

- Use observations to update beliefs

Posterior

- Find values of theta that are most probable given the data

$$P(\boldsymbol{\theta}|D) = \frac{P(D|\boldsymbol{\theta}) * P(\boldsymbol{\theta})}{P(D)}$$

$$P(\boldsymbol{\theta}|D) \propto P(D|\boldsymbol{\theta}) * P(\boldsymbol{\theta})$$

$$\propto P(D|\boldsymbol{\theta}) = L(\boldsymbol{\theta})$$

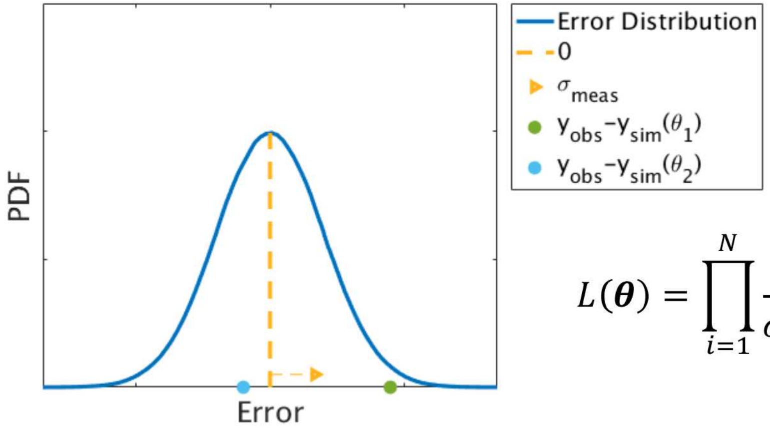
Uncertainty sources

- Aleatory – irreducible, naturally varying
  - Measurement noise
- Epistemic – reducible, lack of knowledge
  - Model parameter uncertainty (calibration)
  - Model bias

Likelihood (case 1):

- $Y_{obs} = Y_{sim} + \epsilon_{meas}$
- Measurement noise: i.i.d. with $N(0, \sigma^2)$

Notional 1-D Likelihood

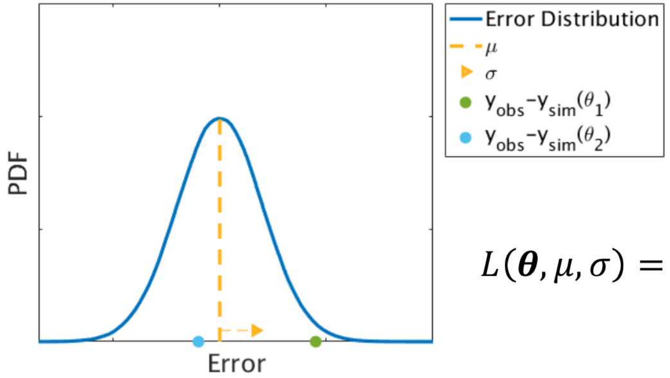


$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}\left(\left(y_{obs,i} - y_{sim}(\boldsymbol{\theta})\right) - 0\right)^2\right)$$

Bayesian Calibration Cont'd

Likelihood (case 2):

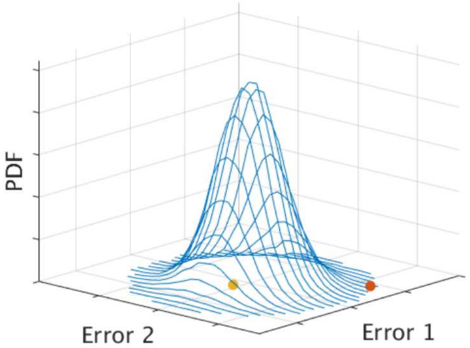
- $Y_{obs} = Y_{sim} + \epsilon_{meas} + \delta_{model}$
- Model error can take on different forms
- Kennedy O'Hagan Framework
  - Treat model error as a Gaussian random process with unknown mean and covariance



$$L(\boldsymbol{\theta}, \mu, \sigma) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}\left((y_{obs,i} - y_{sim}(\boldsymbol{\theta})) - \mu\right)^2\right)$$

Likelihood (case 3):

- Multiple output quantities
- Covariance matrix of errors needed



Solving for posterior

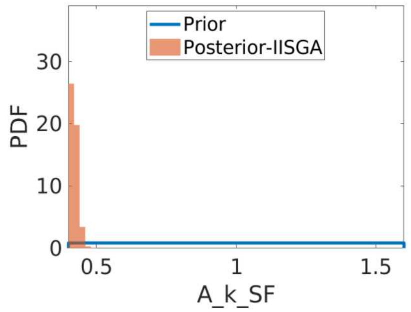
- Markov chain Monte Carlo (MCMC)
  - Samples from chain approach posterior distribution
- Particle filter (PF)
  - Particles weighted based on likelihood scores
  - Easily parallelizable
- Sample-based Bayesian methods require many model evaluations
  - Replace computationally expensive physics model with efficient surrogate model

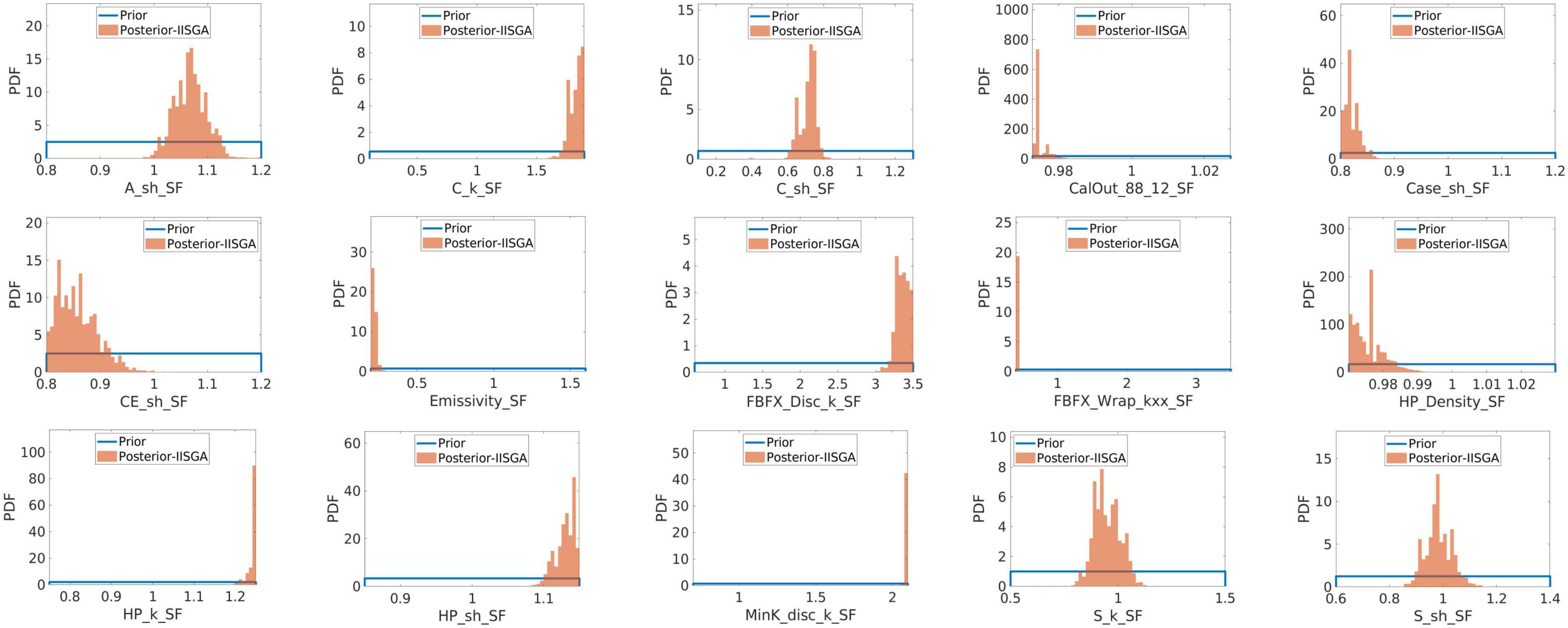
$$L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} exp\left(-\frac{1}{2}\left(\left(\boldsymbol{y}_{obs,i} - \boldsymbol{y}_{sim}(\boldsymbol{\theta})\right) - \boldsymbol{\mu}\right)^T \Sigma_y^{-1}\left(\left(\boldsymbol{y}_{obs,i} - \boldsymbol{y}_{sim}(\boldsymbol{\theta})\right) - \boldsymbol{\mu}\right)\right)$$

Current Implementation of IISGA				
17 calibration parameters	5 output variables	Samples: 10 million	Resampling algorithm 10 iterations	50 cores ~11hrs

Posteriors

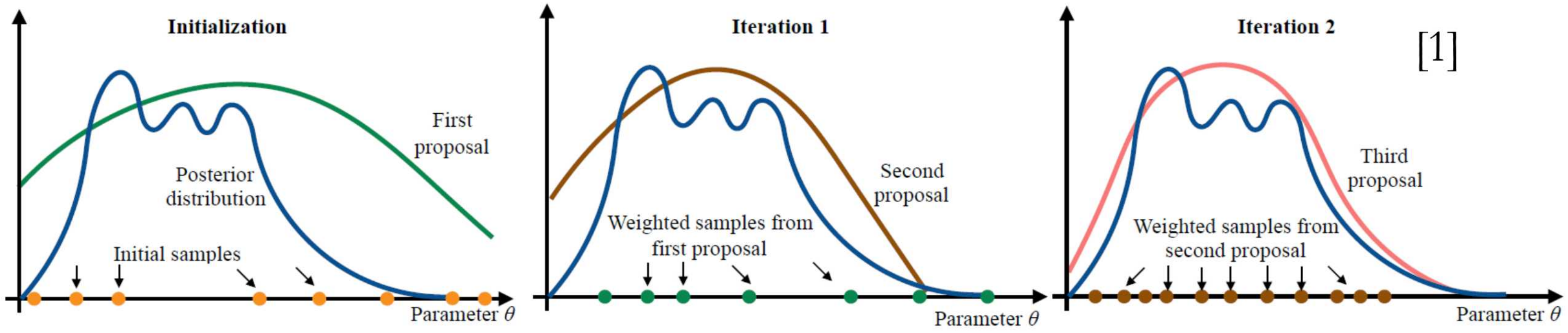
- Marginal posteriors for all 16 model parameters





Iterative Importance Sampling Algorithm (ISA) [1]

Basic Idea: iteratively improve the proposal distribution



Stopping criteria:

- $|R_{k+1} - R_k| < tol$
- where $R = \frac{var(w)}{E(w)^2} + 1$
  - w is the normalized weight,
- if $q(\theta) = p(\theta|D)$, then $var(w) = 0$ so $R = 1$

[1] Morzfeld, M., Day, M. S., Grout, R. W., Heng Pau, G. S., Finsterle, S. A., & Bell, J. B. (2018). Iterative importance sampling algorithms for parameter estimation. *SIAM Journal on Scientific Computing*, *40*(2), B329-B352.

Iterative Importance Sampling Algorithm (ISA) - Differences

Their Method

Use a Gaussian for each iteration of proposal distribution: $q_i,\ where\ i = 1{:}\,N_{iterations}$

- Update the mean and variance: $q_{i+1} = N(\mu_{i+1}, \sigma_{i+1})$, where $\mu_{i+1}\ \sigma_{i+1}$ are computed from $\{\theta_i, w_i\}$
- Generate new samples: $\theta_{i+1} \sim q_{i+1}$

General strategy seems to be start with a broad proposal distribution and hone in

Our Method

Construct empirical/kernel density distributions based on the samples

- Should better capture multimodality

Rather than draw new samples from a normal on each iteration, the same samples are carried throughout – but the samples are allowed to change *genetically*