Bayesian Calibration with Spatio-Temporal Data

Kyle Neal[1]

Sankaran Mahadevan[1], Abhinav Subramanian[1], Josh Mullins[2], Ben Schroeder[2]

[1] Vanderbilt University

[2] Sandia National Labs

IMAC XXXVIII

February 12, 2020

Outline

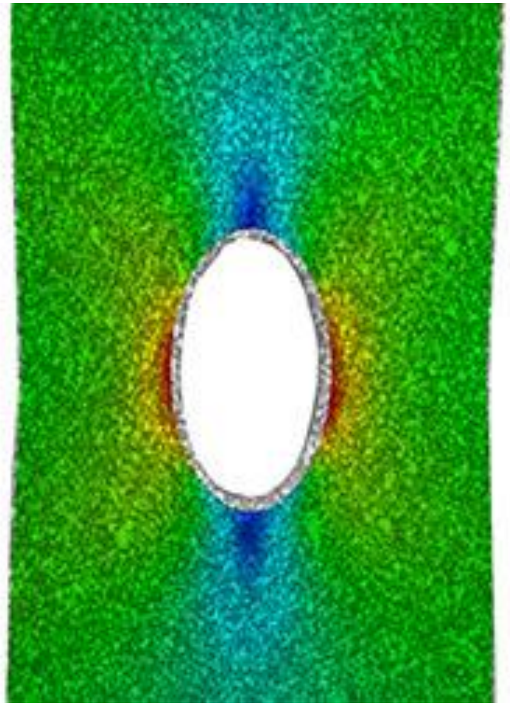
- Motivation
  - Spatio-temporal data
  - Challenges
- Mathematics
  - Principal component analysis (PCA)
- Illustrative problem
  - “Peaks” in MATLAB
  - Model discrepancy estimation of erroneous models

- Emphasis placed here on Bayesian calibration
- But ideas not exclusively Bayesian

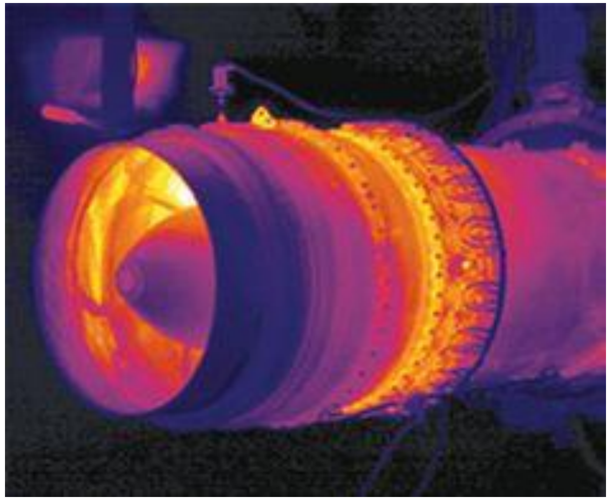
Spatio-temporal outputs "Field Data" are common

<u>Examples</u>

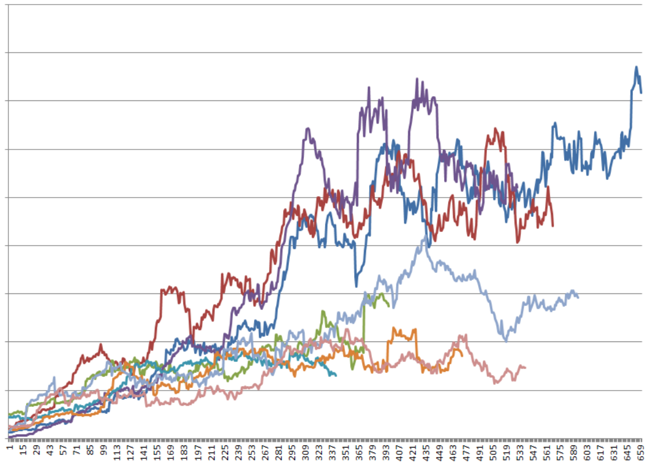
- Digital image correlation
- Thermal heatmap
- Time series data



DIC of coupon test[1]



NDE thermography on jet engine[2]



Stock market data[3]

1 www.gom.com
2 www.caparotesting.com
3 www.stackexchange.com

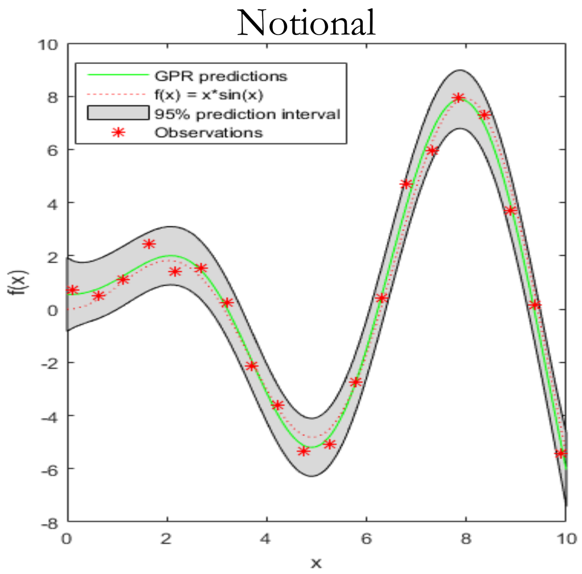
Challenges with Field Data

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

Building surrogate models

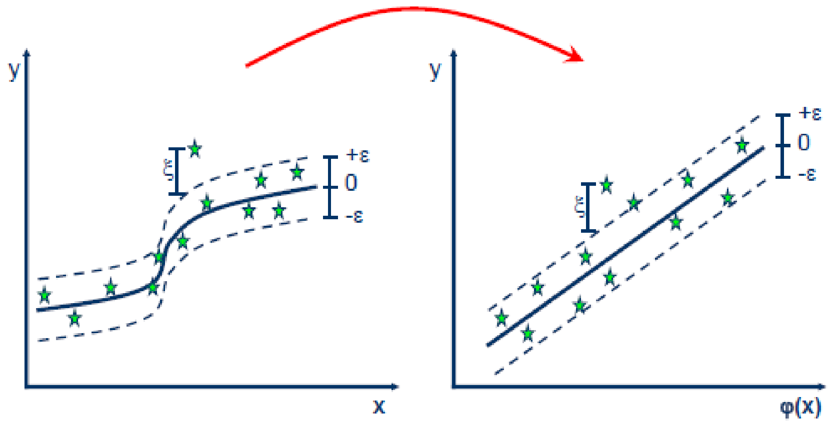
- Gaussian process, SVM, polynomial chaos, ANN, etc.
- One-to-many mapping
  - Inputs are RVs and output is random process/field
- Approaches:
  - Build separate surrogate for each location
  - Include time/space as an input
  - Feature selection
  - Decomposition / dimension reduction technique

Gaussian Process [1]



Support Vector Machine [2]

Nonlinear Regression



GP

- Nonparametric kernel based model
- Probabilistic

SVM

- 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

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

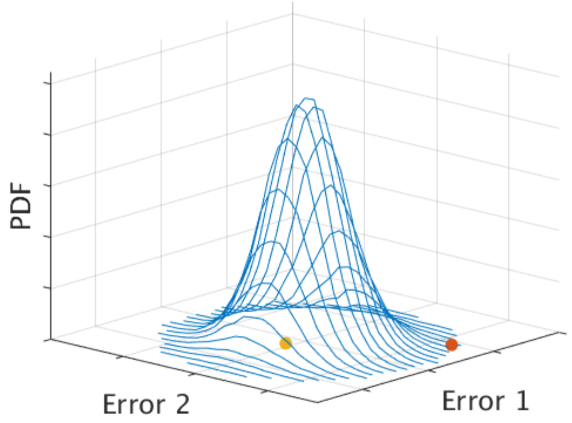
Field data has thousands of unique outputs

- 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

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



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

$$A = USV^T$$

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

- 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:

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

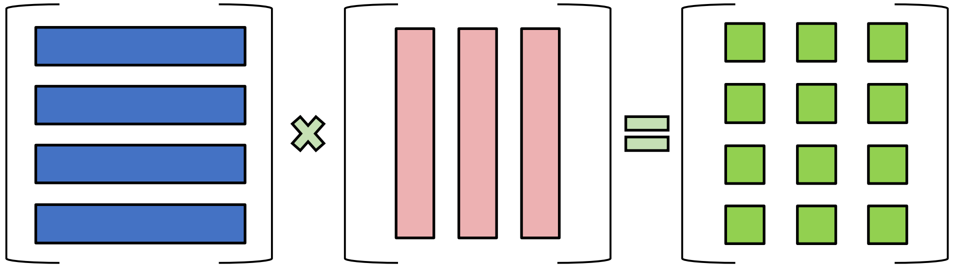
[1] https://stats.stackexchange.com/questions/134282/relationship-between-svd-and-pca-how-to-use-svd-to-perform-pca

Converting to latent response space

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

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

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



- 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

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

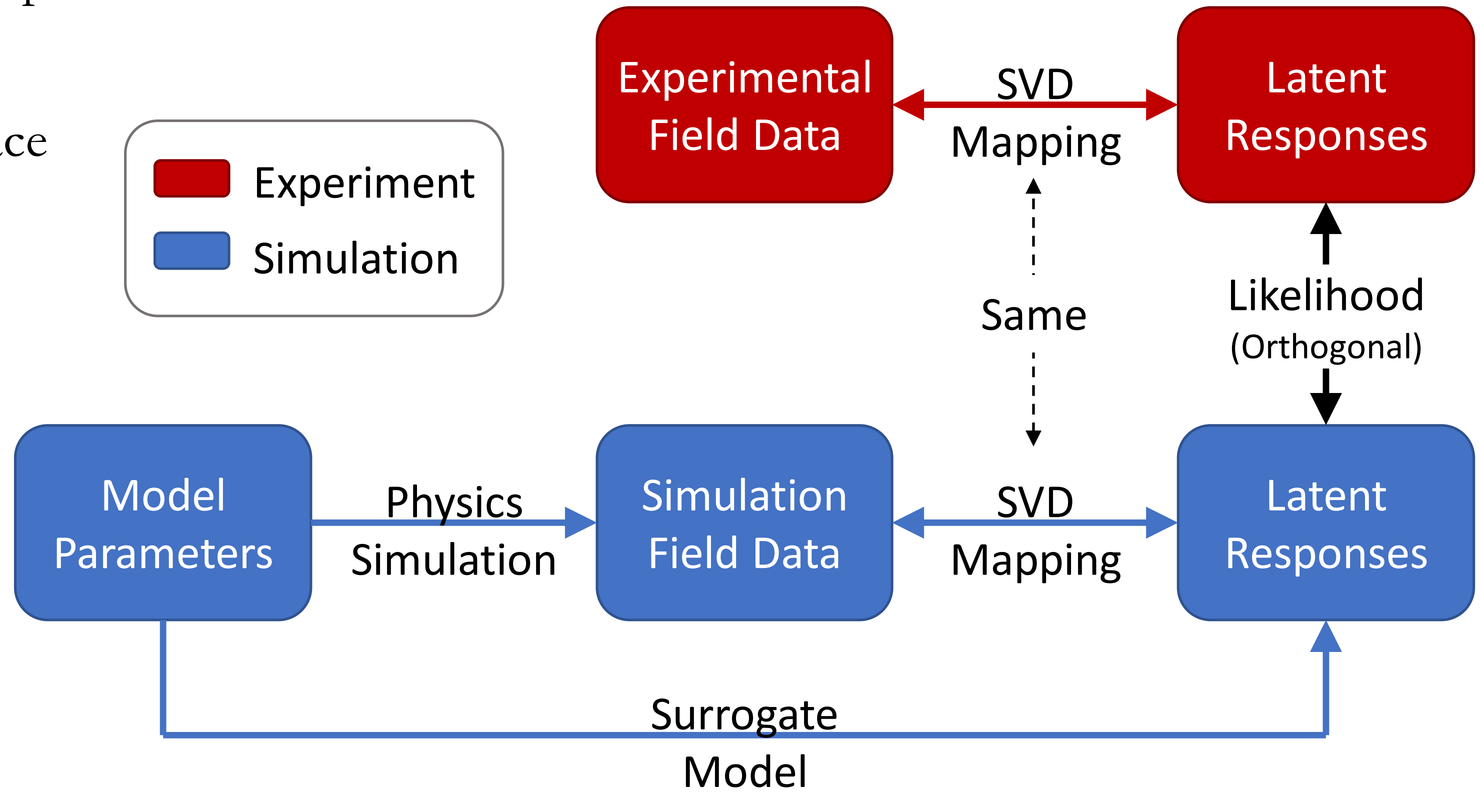
Recap: Bayesian Calibration in Latent Response Space

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

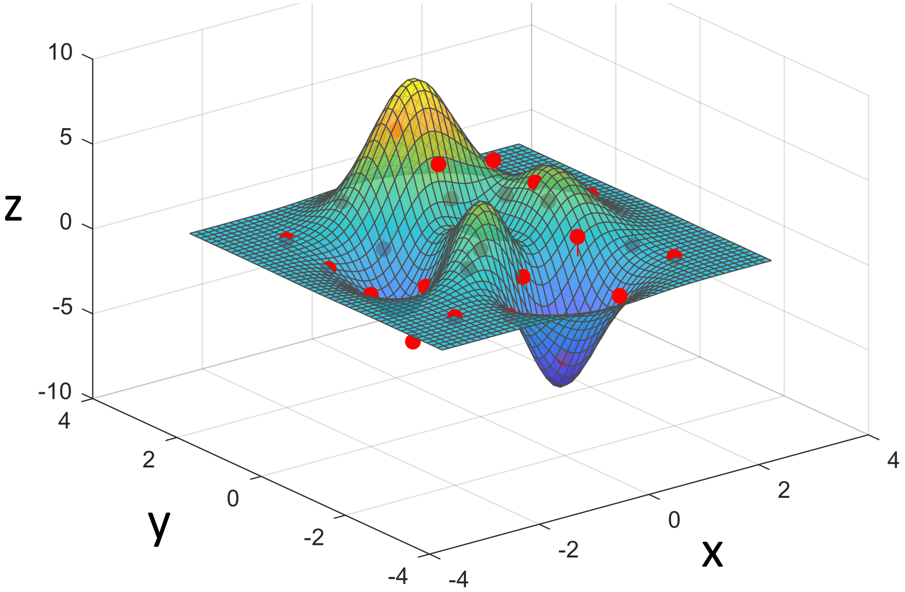


Illustrative Problem

Illustrative Problem (“Peaks” in MATLAB)

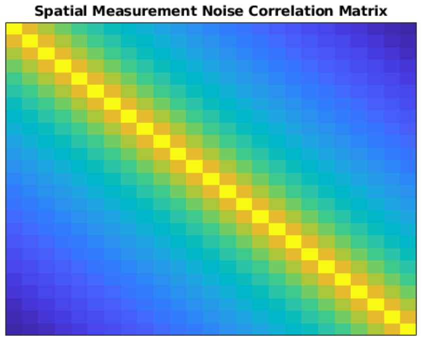
Peaks – a challenging optimization problem used by MATLAB for benchmarking
- I parameterized the function to make it a calibration problem

Observations & True Function



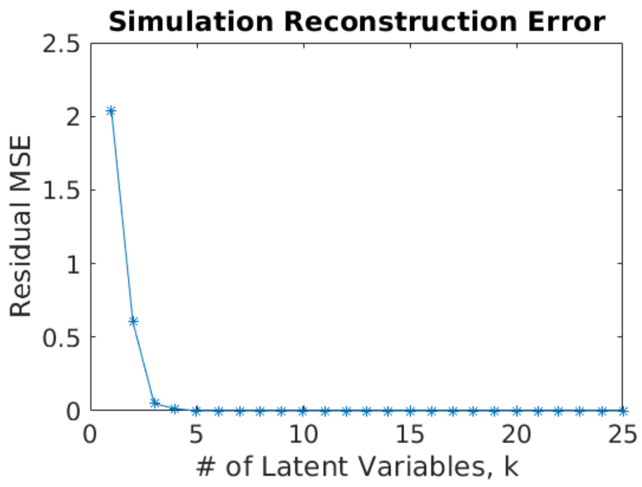
Model: Peaks function in Matlab
- $z = a_1(c_1 - x)^2 \exp(-x^2 - (y+1)^2) + a_2\left(\frac{x}{5} - x^3 - y^5\right)\exp(-x^2 - y^2) + a_3\exp(-(x+1)^2 - y^2)$
- True values
  - $a_1 = 3, a_2 = -10, a_3 = -\frac{1}{3}, c_1 = 1$

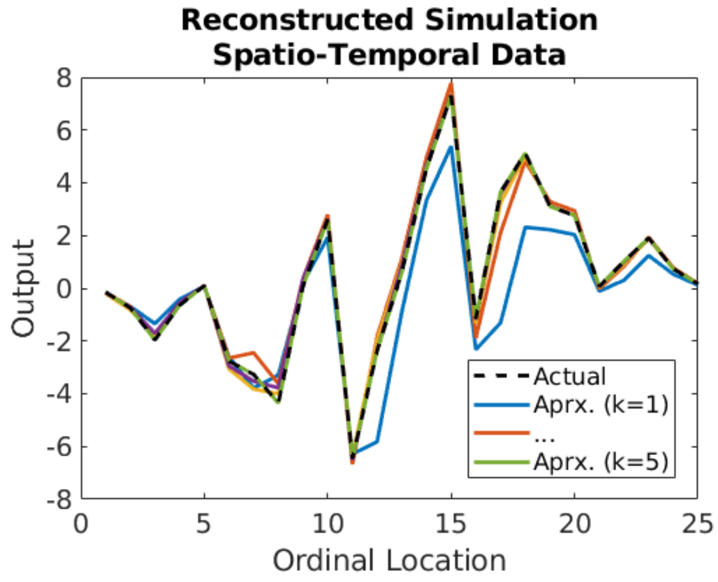
Observations
- $z_{obs} = z + N(0, \sigma_{meas})$
  - $\sigma_{meas} = 0.5$ with plotted correlation structure
- $N_{obs} = 5x5\ grid$
- 10 repeat measurements

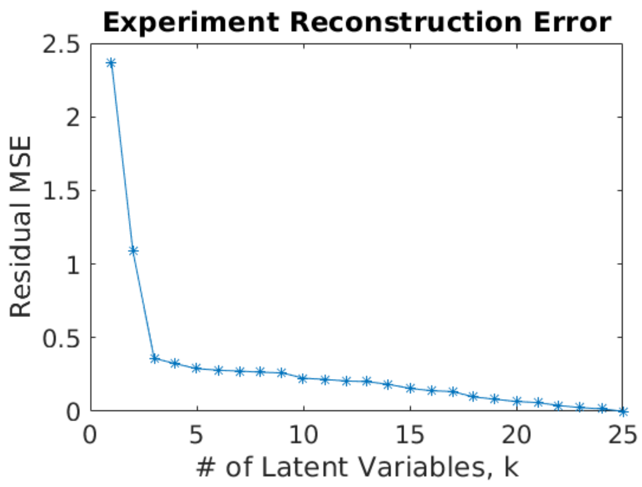


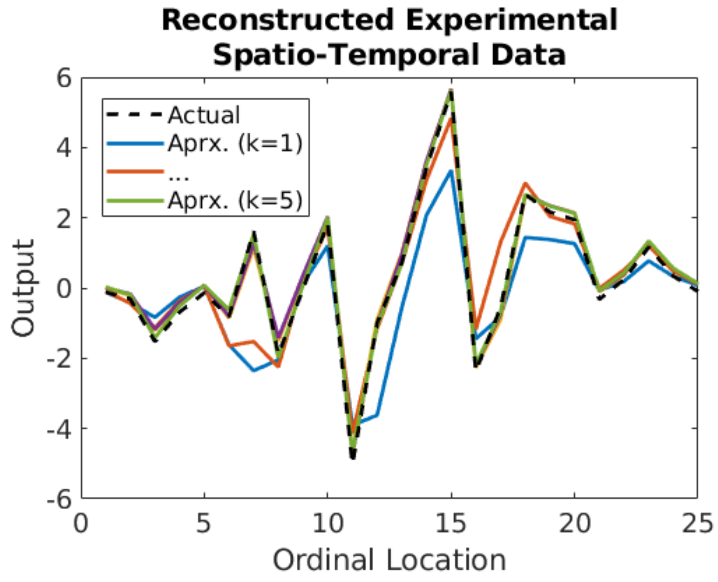
https://www.mathworks.com/examples/matlab/community/35112-optimization-tips-tricks

PCA/SVD


Simulation Reconstruction Error
Residual MSE
# of Latent Variables, k
0
5
10
15
20
25
0.5
1
1.5
2
2.5

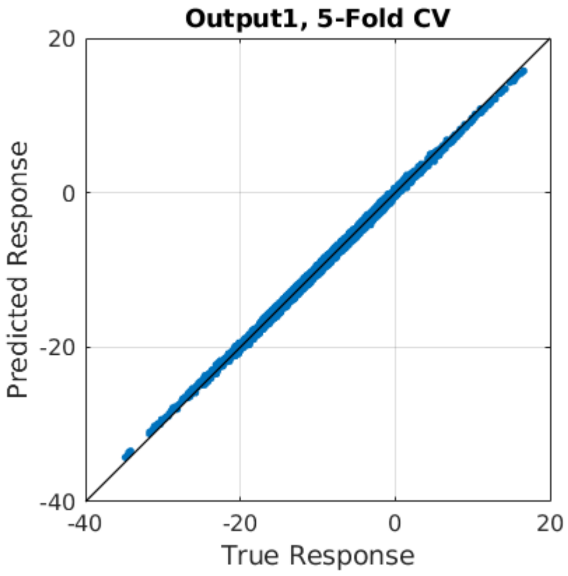

Reconstructed Simulation
Spatio-Temporal Data
Output
Ordinal Location
Actual
Aprx. (k=1)
...
Aprx. (k=5)

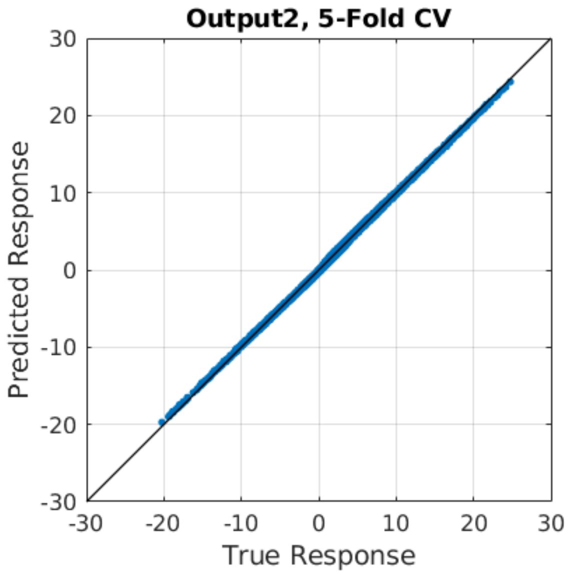

Experiment Reconstruction Error
Residual MSE
# of Latent Variables, k

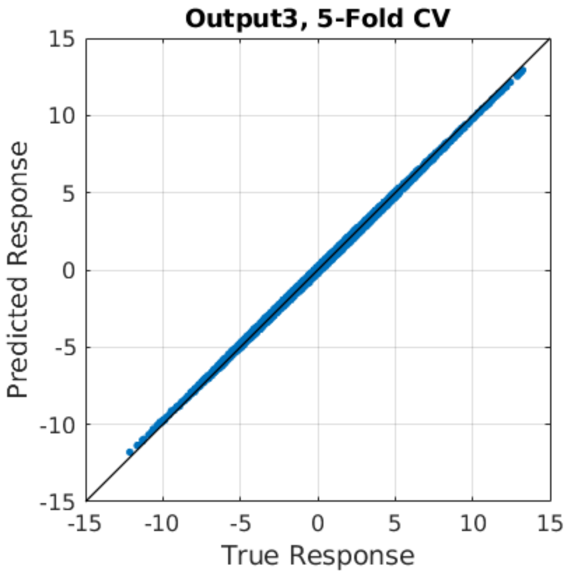

Reconstructed Experimental
Spatio-Temporal Data
Output
Ordinal Location
Actual
Aprx. (k=1)
...
Aprx. (k=5)

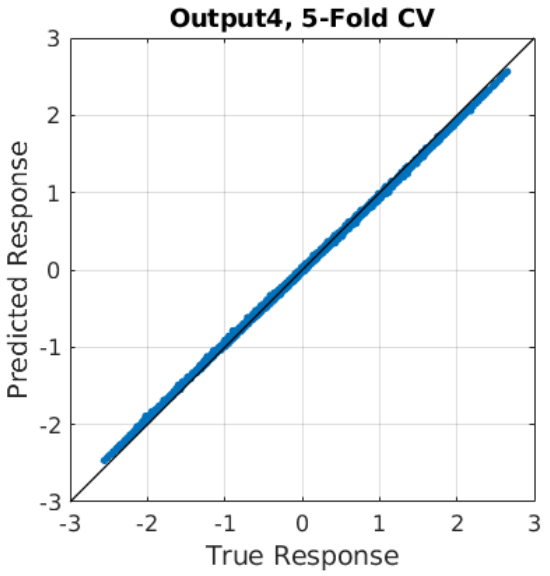
Surrogates

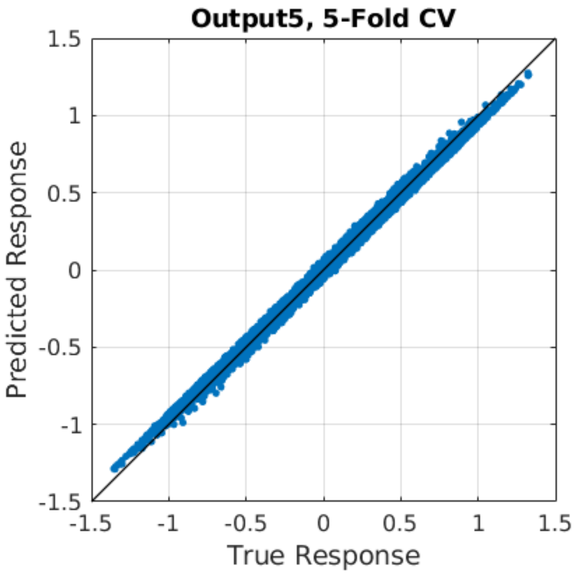
- SVM
  - Training
  - Cross-validation





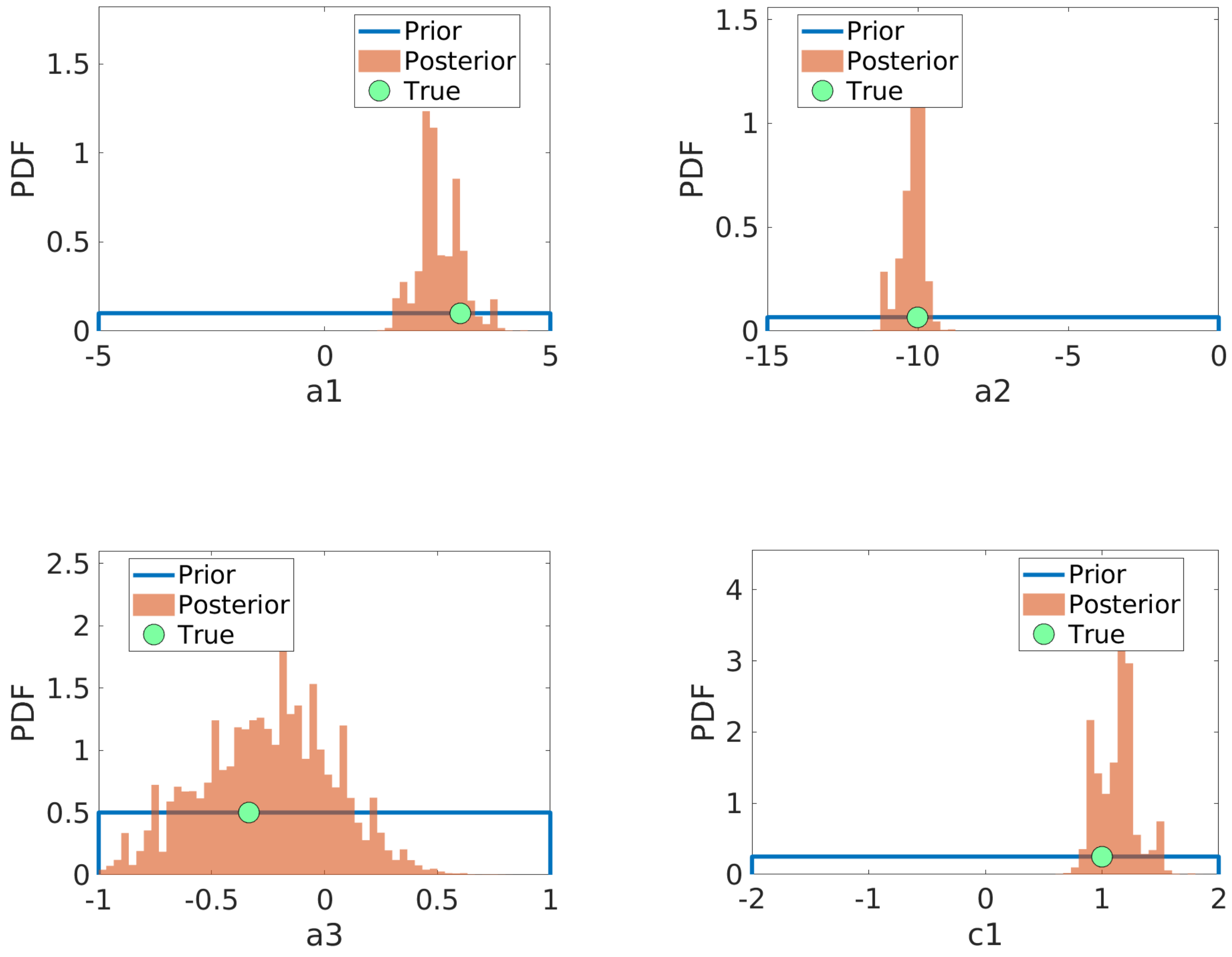






Calibration - Posteriors

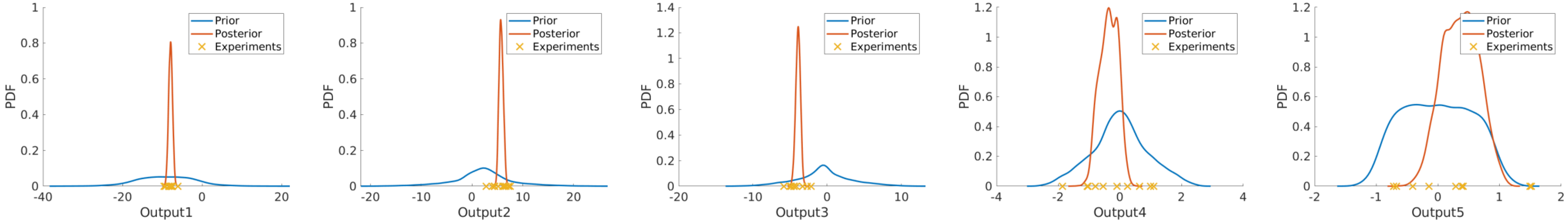
- IISGA [1] – iterative importance sampling with genetic algorithm
  - Mhsample is too slow
  - 100E3 samples ~ 7sec

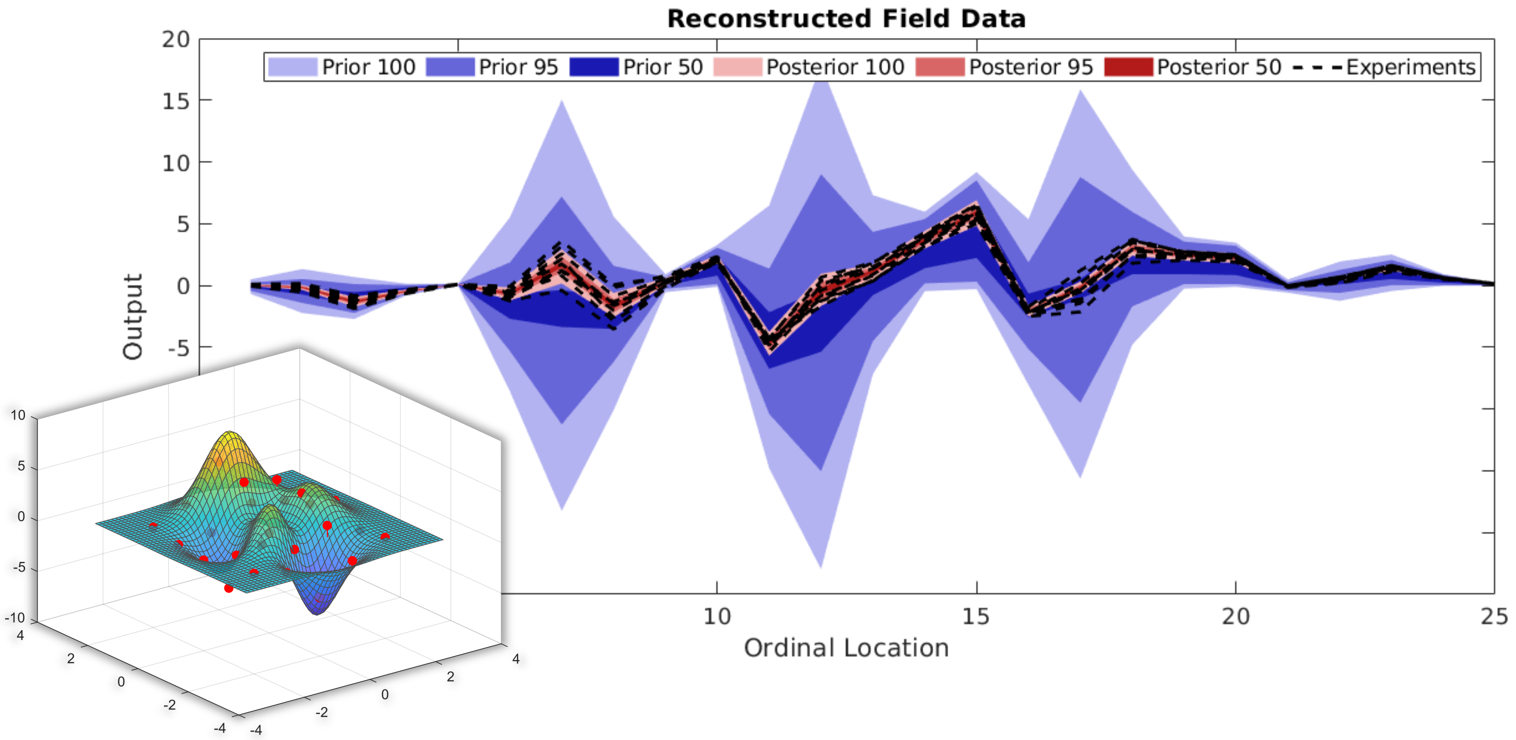


Posterior Sample Correlation			
1	0.1198	-0.5064	-0.8968
0.1198	1	-0.0771	-0.227
-0.5064	-0.0771	1	0.1312
-0.8968	-0.227	0.1312	1

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

Calibration Fit





- Posterior predictions have significantly reduced variance and bias
- Note, measurement error isn't included in simulation predictions

Illustrative Problem ("Peaks" in MATLAB) with model error

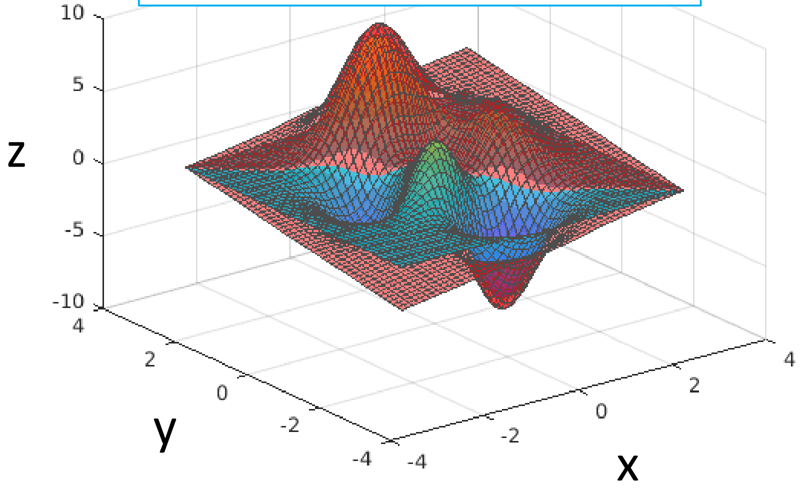
Model: Peaks function in Matlab

- $z = a_1(c_1 - x)^2 \exp(-x^2 - (y+1)^2) + a_2\left(\frac{x}{5} - x^3 - y^5\right)\exp(-x^2 - y^2) + a_3 \exp(-(x+1)^2 - y^2)$
- $z_{sim} = \left(z + \frac{1}{2}\mathrm{x} + \frac{1}{2}y\right)$
- $a_1 = 3, a_2 = -10, a_3 = -\frac{1}{3}, c_1 = 1$

Observations

- $z_{obs} = z + N(0, \sigma_{meas})$
  - $\sigma_{meas} = 0.5$ with previous correlation structure
- $N_{obs} = 5x5\ grid$
- 10 repeat measurements

Erroneous Model Overlaid on True Function

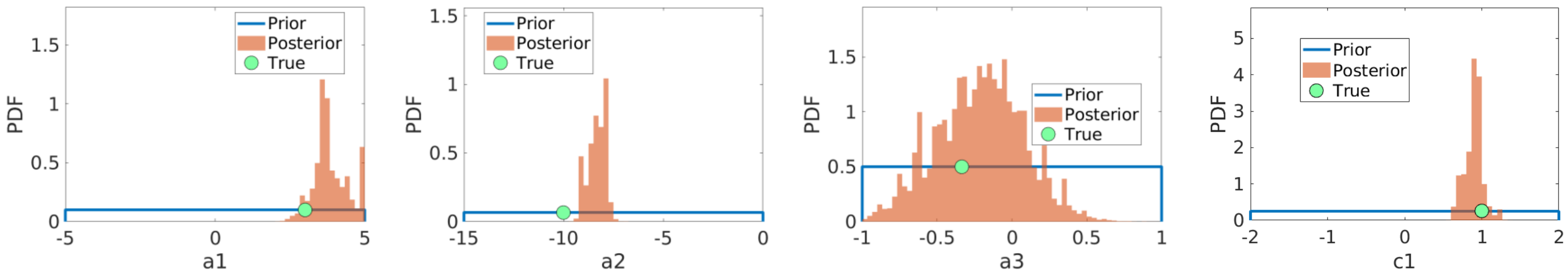


https://www.mathworks.com/examples/matlab/community/35112-optimization-tips-tricks

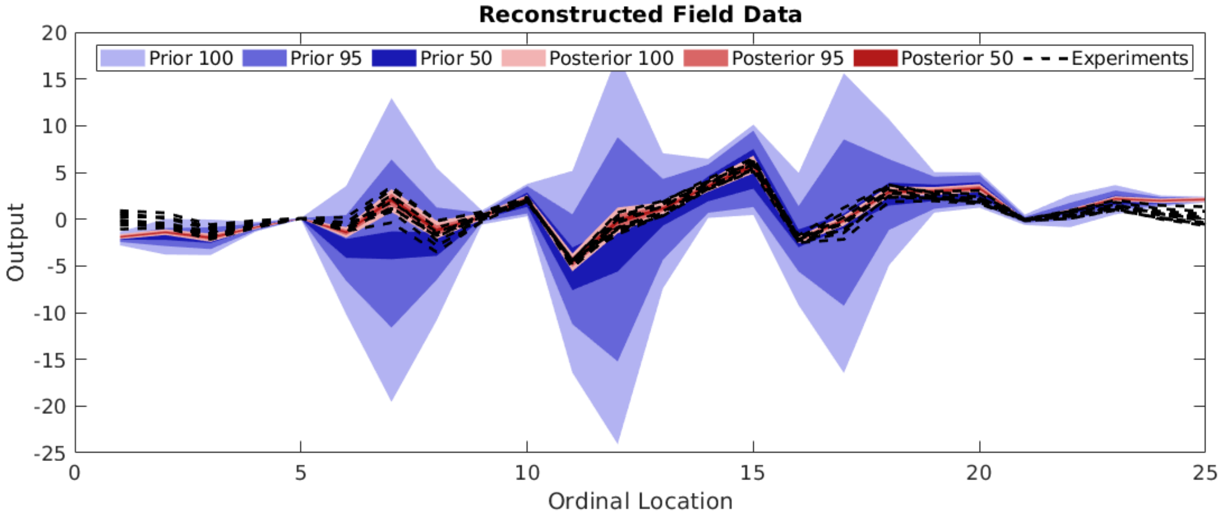
Calibration

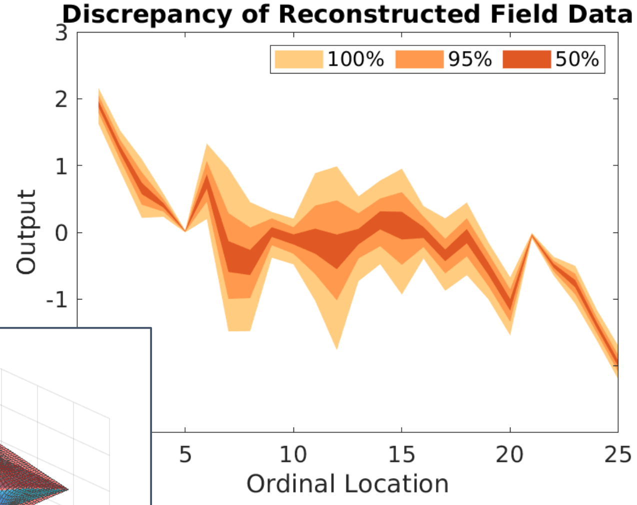
Erroneous Simulation Model

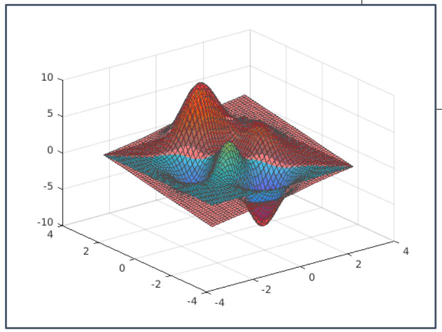
Posteriors



Prediction Fit







Include a model discrepancy term in calibration?

Calibrate a model discrepancy term

True function

- $z = a_1(c_1 - x)^2 \exp(-x^2 - (y+1)^2) + a_2\left(\frac{x}{5} - x^3 - y^5\right)\exp(-x^2 - y^2) + a_3 \exp(-(x+1)^2 - y^2)$

Erroneous simulation model

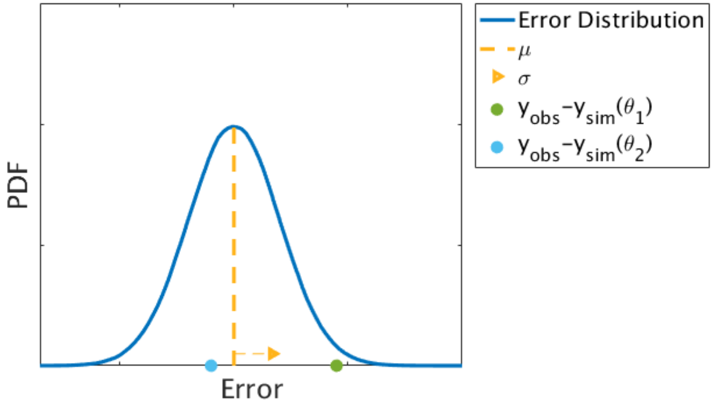
- $z_{sim} = \left(z + \frac{1}{2}\mathrm{x} + \frac{1}{2}y\right)$

Include model discrepancy

- $\lambda_{sim} = \left(z + \frac{1}{2}\mathrm{x} + \frac{1}{2}y\right)V^* + \delta_{sim}$

Discrepancy Types, δ_{sim}

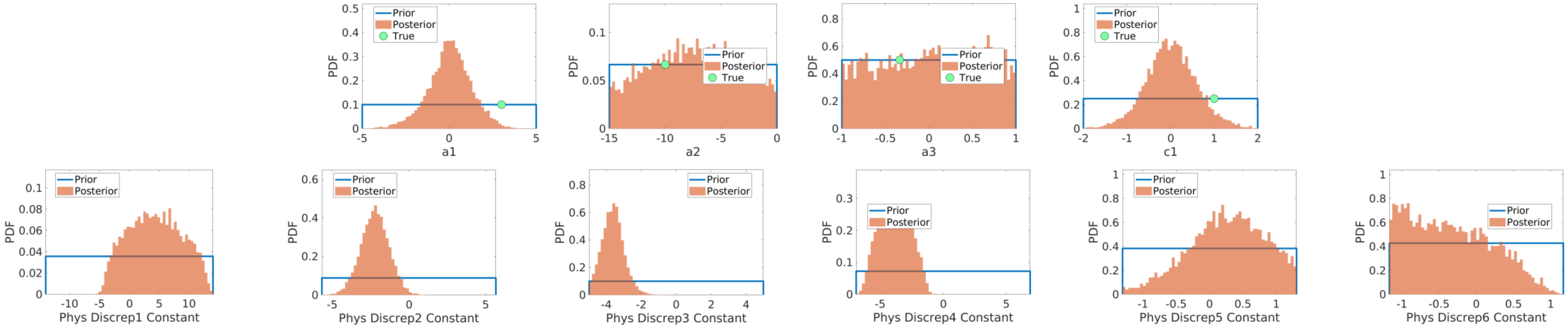
1. Constant
2. 0-mean Gaussian (variance only)
3. Function of inputs (x & y)



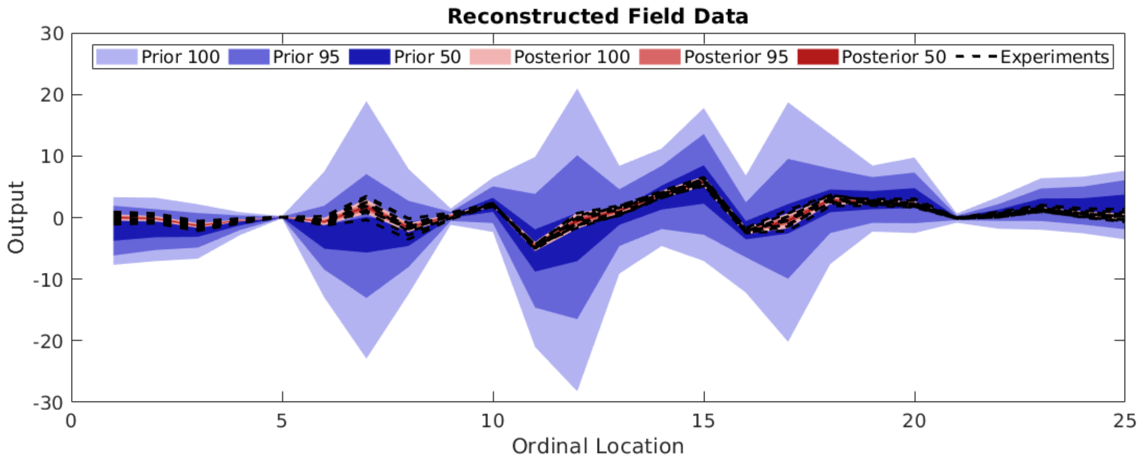
Calibration

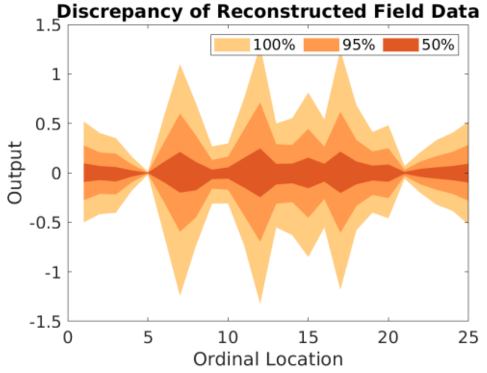
Erroneous Simulation Model | Constant Bias Model Discrepancy

Posteriors



Prediction Fit





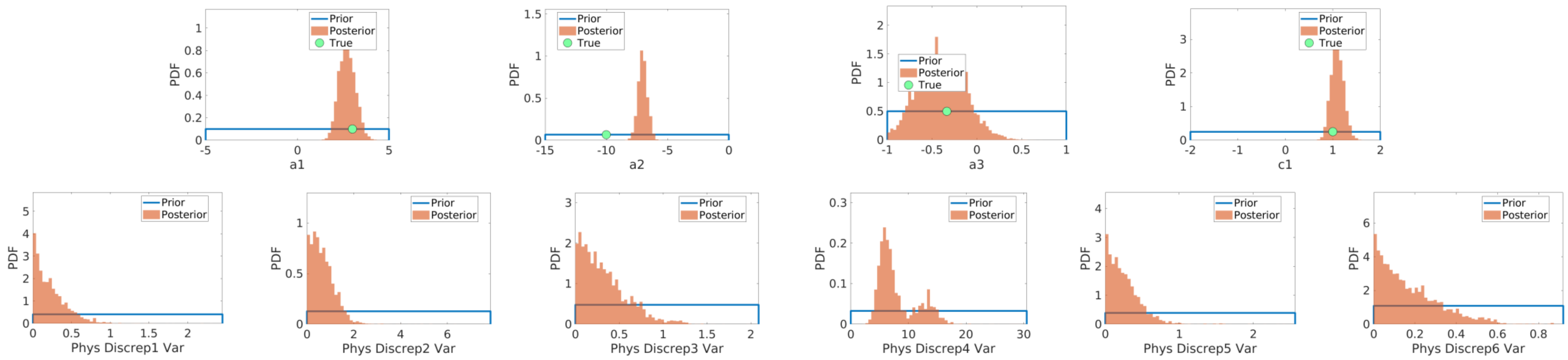
<u>Identifiability</u>

- Given C, find A and B?
  - $A + B = C$
  - $Y_{sim}(\theta) + \delta_{model} = Y_{obs} + \epsilon_{meas}$

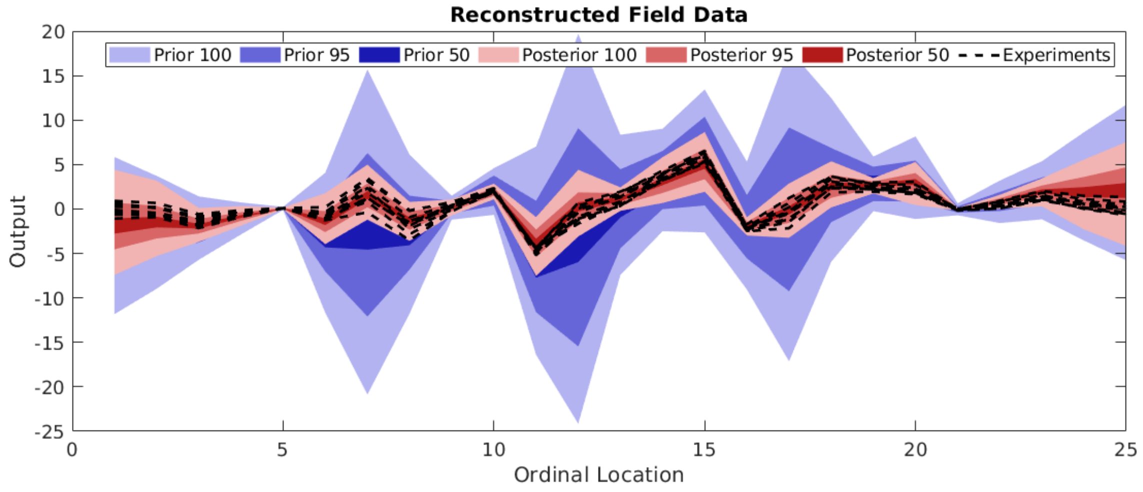
Calibration

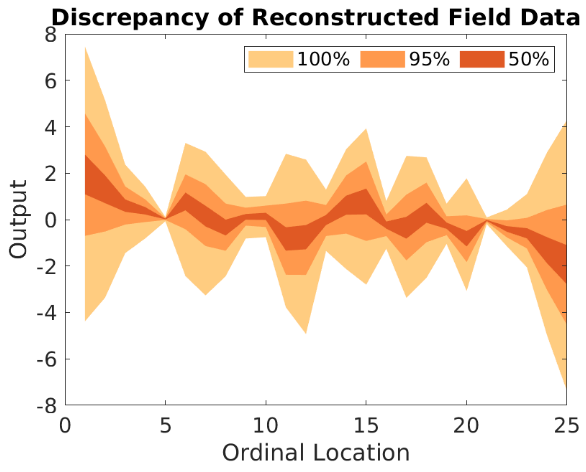
Erroneous Simulation Model | Variance Only Model Discrepancy

Posteriors



Prediction Fit





Conclusions
"Bayesian Calibration with Spatio-Temporal Data"

- Used PCA (SVD) to convert to lower dimensional, orthogonal space
- Methodology demonstrated on "peaks" problem
- Challenges calibrating in presence of model form error discussed

Thank you

Questions?

kyle.d.neal@vanderbilt.edu

Backup Slides

Bayesian Calibration

Posterior Likelihood Prior

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

Data

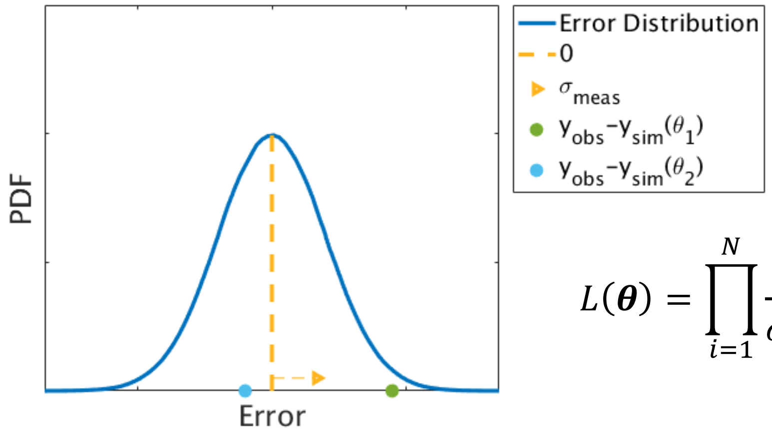
- **Bayes' Theorem**
  - Use observations to update beliefs
- **Posterior**
  - Find values of theta that are most probable given the data
- **Uncertainty sources**
  - Aleatory – irreducible, naturally varying
    - Measurement noise
  - Epistemic – reducible, lack of knowledge
    - Model parameter uncertainty (calibration)
    - Model bias
- **Likelihood (case 1):**
  - $Y_{obs} = Y_{sim} + \epsilon_{meas}$
  - Measurement noise: i.i.d. with $N(0, \sigma^2)$

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

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

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

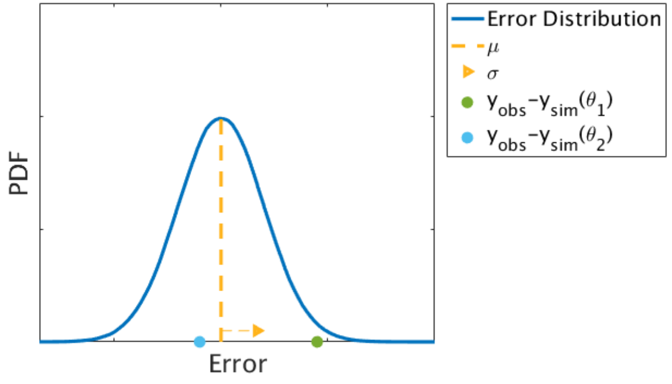
Notional 1-D Likelihood



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

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

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

- **Solving for posterior**
  - Markov chain Monte Carlo (MCMC)
    - Samples from chain approach posterior distribution
  - Iterative Importance Sampling with a Genetic Algorithm
    - 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