Toward Ice Sheet Model Calibration with Spatial Data

Murali Haran

Department of Statistics, Pennsylvania State University

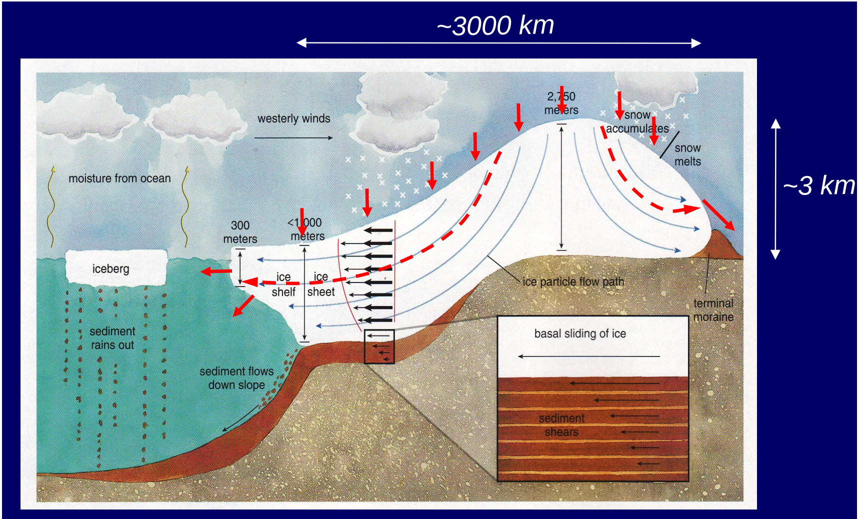
Department of Statistics, Ohio State University
Columbus, November 2014.

Collaborators:
Won Chang (University of Chicago Statistics)
Patrick Applegate, Klaus Keller, Roman Olson (Penn State Geosciences)

This Talk

- The West Antarctic Ice Sheet (WAIS) has the potential to be a significant contributor to future sea level change.
- The PSU-Ice model (Pollard and DeConto, 2009, 2012) may be used to make projections about future WAIS behavior.
- However, there is considerable uncertainty in the projections due to uncertaintly about key model input parameters.
- We would like to use observations of the ice sheet to learn about these parameters
- Several challenges posed by data: (1) high-dimensional, (2) spatial, (3) complicated errors/discrepancies, (4) non-Gaussian
- I will describe new methods to address these issues

Ice Sheet



Siegert, 2002, Amer. Sci.

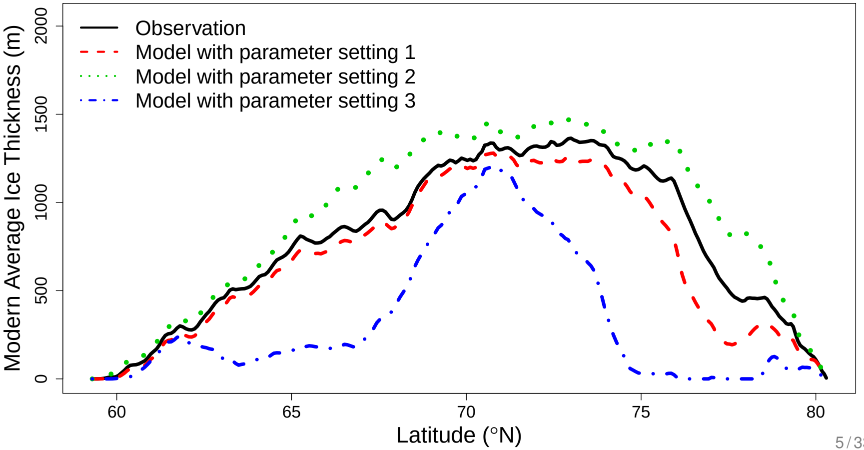
Ice Sheet Model Parameters

- The ice sheet's behavior is complex.
- Model equations predict ice flow, thickness, temperatures, and bedrock elevation, through thousands to millions of years.
- Examples of key model parameters:
  - Ocean melt coefficient: sensitivity of ice sheet to temperature change in the surrounding ocean
  - Strength of the "calving" process. Calving = where ice breaks off and transitions from attached to floating
  - "Slipperiness" of the ocean floor

Ice Sheet Model Calibration

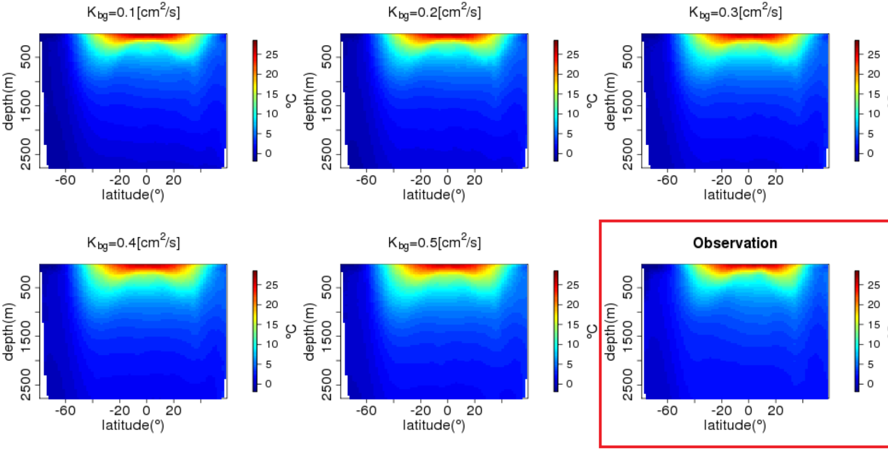
- We can "tune" the model parameters so that the model reproduces the observed behavior of the ice sheet.
- Utilize observations of the modern ice sheet: satellite data.

Which parameter settings best match observations?



Another Calibration Problem

Which parameters best match observations? This is for projection of the Atlantic Meridional Overturning Circulation (AMOC). 2D slices of high-dimensional 3D data



Two-stage Approach to Emulation-Calibration

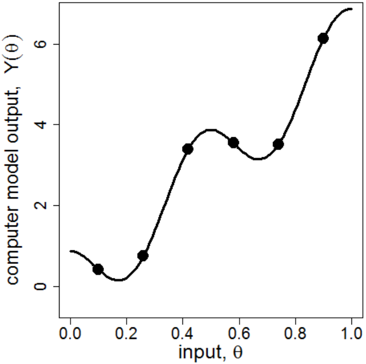
1. Emulation step: Find fast approximation for climate model using Gaussian process (GP)

2. Calibration step: Infer climate parameter using emulator and observations, while accounting for data-model discrepancy

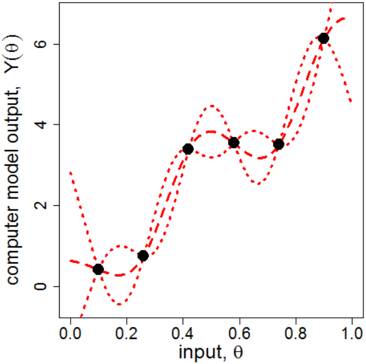
   (Bhat, Haran, Olson, Keller, 2012; Liu, Bayarri and Berger, 2009)

Emulation Step

Toy example: model output is a scalar



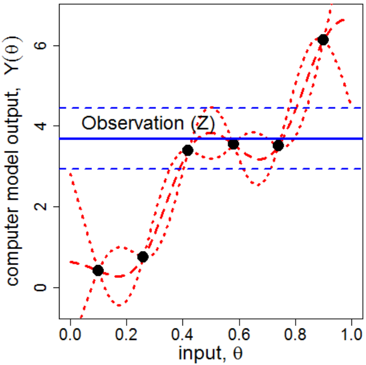
Computer model output (y-axis) vs. input (x-axis)



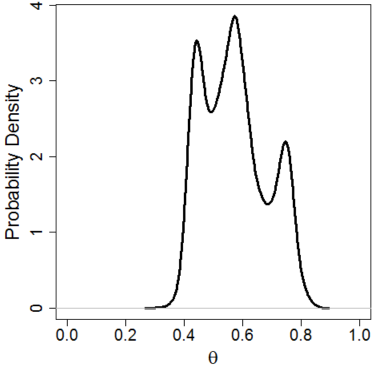
Emulation (approximation) of computer model using GP

Calibration Step

Toy example: model output, observations are scalars



Combining observation
and emulator



Posterior PDF of θ
given model output and observation

Summary of Statistical Problem

- **Goal**: Learn about θ based on two sources of information:
  - **Observations**: $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$, where $\mathbf{s}_1, \ldots, \mathbf{s}_n$ locations (1D, 2D or 3D)
  - **Model output** $\mathbf{Y}(\boldsymbol{\theta}_1), \ldots, \mathbf{Y}(\boldsymbol{\theta}_p)$, where each $\mathbf{Y}(\boldsymbol{\theta}_i) = (Y(\mathbf{s}_1, \boldsymbol{\theta}_i), \ldots, Y(\mathbf{s}_n, \boldsymbol{\theta}_i))^T$ is spatial field

  $\mathbf{Z}$ and $\mathbf{Y}(\boldsymbol{\theta}_i)$'s are n-dimensional vectors

- Important: output at each $\boldsymbol{\theta}_i$ is often a high-dimensional spatial field and number of runs (p) may be large
  - Example 1 (Chang, Haran, Olson, Keller, 2014): $n = 61,051$ locations, $p = 250$ runs.
  - Example 2 (Chang, Haran, Applegate, Pollard, 2014a, b): $n = 264$ (aggregated) locations, or 3,182 (unaggregated) for $p = 250$ to 500 model runs.

GP for Computer Model Emulation

- Fit GP to *np*-dimensional data $\mathbf{Y} = \left(\mathbf{Y}(\boldsymbol{\theta}_1)^T, \ldots, \mathbf{Y}(\boldsymbol{\theta}_p)^T\right)^T$ for interpolation.

- Covariance used for:
  1. non-linear relationship between parameter and model output (model output as a function of parameter)
  2. non-linear spatial surface (model output as a function of location)

- Covariance function example:

$$\begin{aligned}\text{Cov}\left(Y(\mathbf{s},\boldsymbol{\theta}), Y(\mathbf{s}',\boldsymbol{\theta}'); \boldsymbol{\xi}\right) =& \kappa \exp\left(-\frac{g\left(\mathbf{s},\mathbf{s}'\right)}{\phi_{\mathbf{s}}}\right) \exp\left(-\frac{\|\boldsymbol{\theta}-\boldsymbol{\theta}'\|}{\phi_\theta}\right) \\ &+ \zeta I(\boldsymbol{\theta}=\boldsymbol{\theta}') I(\mathbf{s}=\mathbf{s}')\end{aligned}$$

where g is geodesic distance, and $\boldsymbol{\xi} = (\kappa, \phi_s, \phi_\theta, \zeta)$ is a vector of covariance parameters.

Step 1: Emulation (Approximating Computer Model)

- Find MLE for covariance parameter $\boldsymbol{\xi}$, denoted by $\hat{\boldsymbol{\xi}}$
- Get $\boldsymbol{\eta}(\boldsymbol{\theta}_{NEW}, \mathbf{Y})$ for prediction at any $\boldsymbol{\theta}_{NEW} \in \Theta$:
  - GP gives

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}(\boldsymbol{\theta}_{NEW}) \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}_{n(p+1)\times 1}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{n(p+1)\times n(p+1)} \right)$$

  - Emulator:

$$\boldsymbol{\eta}\left(\boldsymbol{\theta}_{NEW}, \mathbf{Y}\right) = \mathbf{Y}(\boldsymbol{\theta}_{NEW})|\mathbf{Y} \sim N\left(\Sigma_{21}\Sigma_{11}^{-1}\mathbf{Y}, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

Step 2: Calibration (Inferring Input Parameter)

- Probability model for **Z** based on

$$\mathbf{Z} = \eta(\boldsymbol{\theta}, \mathbf{Y}) + \boldsymbol{\delta},$$

where n-dimensional spatial field $\boldsymbol{\delta}$ is model-observation discrepancy with covariance parameter $\boldsymbol{\xi}_\delta$.

- Inference for $\boldsymbol{\theta}$ based on posterior distribution

$$\pi(\boldsymbol{\theta}, \boldsymbol{\xi}_\delta | \mathbf{Z}, \mathbf{Y}, \hat{\boldsymbol{\xi}}) \propto \underbrace{L(\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}, \boldsymbol{\xi}_\delta, \hat{\boldsymbol{\xi}})}_{\text{likelihood given by above}} \times \underbrace{p(\boldsymbol{\theta}) \times p(\boldsymbol{\xi}_\delta)}_{\text{priors for } \boldsymbol{\theta} \text{ and } \boldsymbol{\xi}_\delta}$$

with emulator parameter $\hat{\boldsymbol{\xi}}$ fixed at value estimated in emulation step.

Computational Challenges and Our Approach

- Emulation requires dealing with $np \times np$ covariance matrix of $\mathbf{Y}$. E.g. if $n = 61,051$ $p = 250$
  - Cholesky decomposition costs $\frac{1}{3}n^3p^3 = 1.185 \times 10^{21}$ flops.
  - Covariance matrix is of size $8 \times \frac{250^2 \times 61051^2}{1024^3} = 1,735,624$
- Calibration faces similar challenges for dealing with $n \times n$ covariance matrix.

Our fast reduced-dimension approach: Fast computation using PC and Kernel Convolution

Main Idea

- Consider model outputs at $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$ as if they were replicates of a multivariate process, thereby obtaining their PCs

$$\begin{pmatrix} Y(\mathbf{s}_1, \boldsymbol{\theta}_1) & \ldots & Y(\mathbf{s}_n, \boldsymbol{\theta}_1) \\ \vdots & \ddots & \vdots \\ Y(\mathbf{s}_1, \boldsymbol{\theta}_p) & \ldots & Y(\mathbf{s}_n, \boldsymbol{\theta}_p) \end{pmatrix}_{p \times n} \Rightarrow \begin{pmatrix} Y_1^R(\boldsymbol{\theta}_1) & \ldots & Y_{J_y}^R(\boldsymbol{\theta}_1) \\ \vdots & \ddots & \vdots \\ Y_1^R(\boldsymbol{\theta}_p) & \ldots & Y_{J_y}^R(\boldsymbol{\theta}_p) \end{pmatrix}_{p \times J_y}$$

- PCs pick up characteristics of model output that vary most across input parameters $\theta_1, \ldots, \theta_p$.

Emulation Using PCs

- Fit 1-dimensional GP for each series $Y_j^R(\boldsymbol{\theta}_1), \ldots, Y_j^R(\boldsymbol{\theta}_p)$
- $\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)$: J_y-dimensional emulation process for PCs, $\mathbf{Y}^R$ is collection of PCs
- Computation reduces from $\mathcal{O}(n^3p^3)$ to $\mathcal{O}(J_y p^3)$ $(1.2 \times 10^{21}$ to 1.0×10^8 flops).
- Emulation for original output: compute $\mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)$ where $\mathbf{K}_y$ is matrix of scaled eignvectors
- Flexible emulator

Dimension Reduction for Discrepancy Process

- Kernel convolution: Specifying n-dimensional discrepancy process δ using J_d-dimensional knot process $\boldsymbol{\nu}$ ($J_d < n$) and kernel functions
- Kernel basis matrix $\mathbf{K}_d$ links grid locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ to knot locations $\mathbf{a}_1, \ldots, \mathbf{a}_{J_d}$;

  $$\{\mathbf{K}_d\}_{ij} = \exp\left(-\frac{g(\mathbf{s}_i, \mathbf{a}_j)}{\phi_d}\right)$$

  with $\phi_d > 0$. Fix ϕ_d at large value determined by expert judgment
- Results in better identifiability: Overly flexible discrepancy process may be confounded with emulator

Calibration in Reduced Dimensions

- Probability model for dimension-reduced observation $\mathbf{Z}^R$:

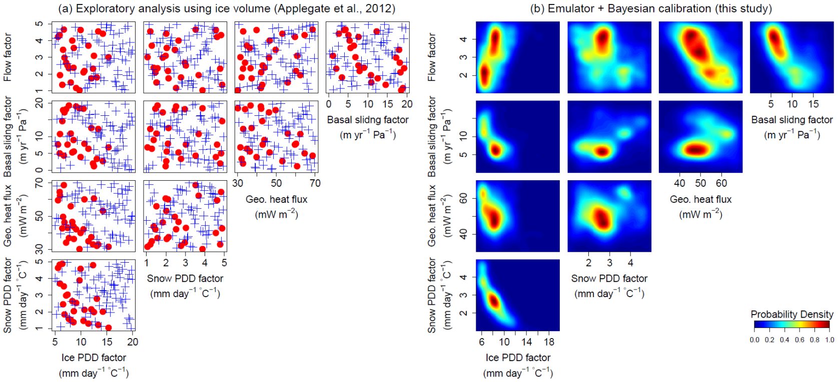
$$\begin{aligned}\mathbf{Z} &= \underbrace{\mathbf{K}_y \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)}_{\text{emulator}} + \underbrace{\mathbf{K}_d \boldsymbol{\nu}}_{\text{discrepancy}} + \underbrace{\boldsymbol{\epsilon}}_{\text{observation error}} ,\\ \Rightarrow \mathbf{Z}^R &= (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{Z} = \begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R) \\ \boldsymbol{\nu} \end{pmatrix} + (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\boldsymbol{\epsilon},\end{aligned}$$

  with combined basis $[\mathbf{K}_y\ \mathbf{K}_d]$, knot process $\boldsymbol{\nu} \sim N(\mathbf{0}, \kappa_d \mathbf{I})$, and observational error $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.
- Infer $\boldsymbol{\theta}$ through posterior distribution

$$\pi(\boldsymbol{\theta}, \kappa_d, \sigma^2 | \mathbf{Z}^R, \mathbf{Y}^R) \propto \underbrace{L(\mathbf{Z}^R | \mathbf{Y}^R, \boldsymbol{\theta}, \kappa_d, \sigma^2)}_{\text{likelihood given by above}} \underbrace{p(\boldsymbol{\theta}) p(\kappa_d) p(\sigma^2)}_{\text{priors}}$$

How Does Statistical Rigor Help?

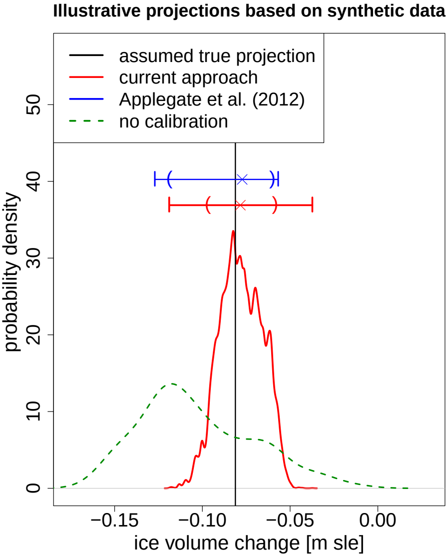
Left: non-rigorous vs right: statistical calibration



"Underneath the hood": (i) we account for (epistemic) uncertainties in emulation, (ii) real probability distributions.

Ice Volume Change Projection

Real probability distribution ⇒ genuine projection uncertainty quantification



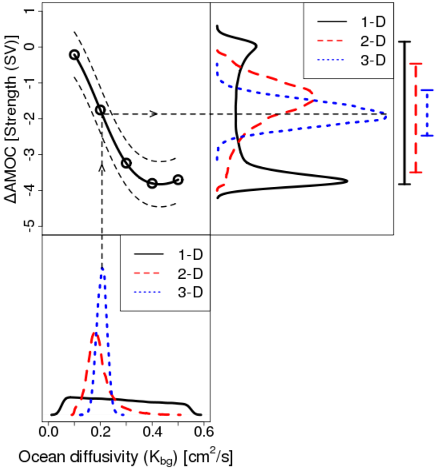
What is the Effect of Data Aggregation?

- Common practice: Calibration using aggregated data (e.g. zonal average)
  - Avoiding computational issues
  - Limited skill of climate model in reproducing spatial patterns
- Using unaggregated data may result in
  - perhaps less uncertainty due to using more data?
  - perhaps more uncertainty due to poor model skill?
- Largely unanswered due to inability to handle unaggregated data

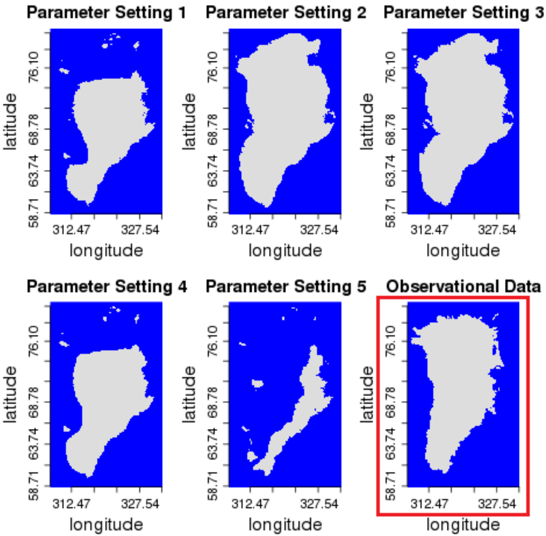
Results

Computational efficiency allows us to calibrate using unaggregated data.

- We compare 1D (depth profile) and 2D (zonal average) with 3D (unaggregated) data.
- Inference with 3D data leads to sharper inference for $\boldsymbol{\theta}$.
- Inference using 3D data is more robust to changes in prior specifications for discrepancy parameters.



Spatial Binary Output



Calibration with Binary Output

- Standard Gaussian process approach does not apply.
- Let $\mathbf{\Gamma}$ be $p \times n$ matrix of natural parameters for model output, its (i, j)th element $\gamma_{ij}^{Y} = \log \frac{p_{ij}}{1-p_{ij}}$ is logit for the ith parameter setting at the jth spatial location and $p_{ij} = P(Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1)$.
- Emulate computer model by interpolating natural parameters at different input parameter settings using a zero-mean Gaussian process with covariance function:

$$\text{Cov}\left(\gamma_{ij}, \gamma_{kl}\right) = C\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_k, \mathbf{s}_j, \mathbf{s}_l; \boldsymbol{\xi}_y\right),$$

where $\boldsymbol{\xi}_y$ is a vector of the covariance parameters.

Challenges

- Ill-posed problem: Estimation of $\{\gamma_{ij}, i = 1 \ldots, p, j = 1, \ldots, n\}$ is not possible by simple maximization of log-likelihood. *np* parameters for *np* data points.
- Computationally infeasible: Cholesky factorization has computational cost of $\frac{1}{3} \times p^3 \times n^3 = \frac{1}{3} \times 499^3 \times 3,182^3 = 1.33 \times 10^{18}$ flops
- Calibration also involves having to perform a very high-dimensional integration + expensive matrix operations
- We propose dimension-reduction to address both ill-posedness and computational issues.

Dimension-reduction

- Consider $\mathbf{\Gamma}$ the $p \times n$ matrix of natural parameters for model output. Using logistic principal components (Lee et al., 2010), rewrite as:

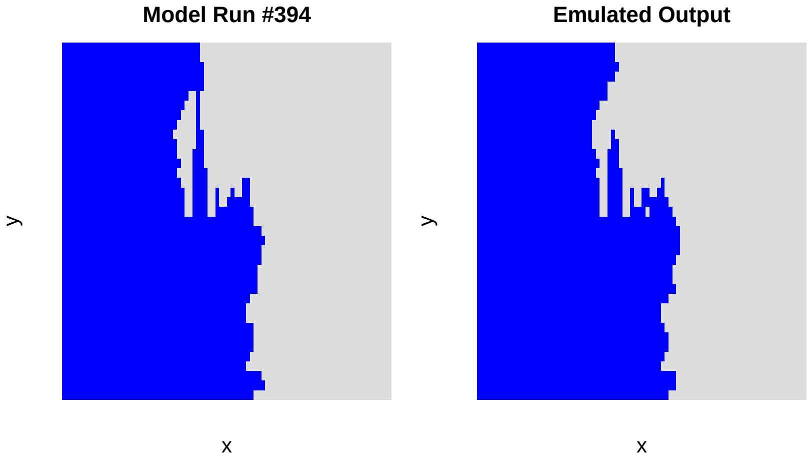
$$\mathbf{\Gamma} = \mathbf{1}_p \otimes \boldsymbol{\mu}^T + \mathbf{W}\mathbf{K}_y^T, \tag{1}$$

  where $\mathbf{K}_y$ is an $n \times J_y$ orthogonal basis matrix, $\mathbf{W}$ is the $p \times J_y$ principal component matrix with (i, j)th element $w_j(\boldsymbol{\theta}_i)$, and $\boldsymbol{\mu}$ is the $n \times 1$ mean vector.
- Non-trivial and computationally challenging optimization to obtain matrices $\mathbf{W}$, $\mathbf{K}_y$ by maximizing log-likelihood. Use majorization-minimization (MM) algorithm (Lange et al. 2000; Hunter and Lange, 2004).
- Emulate each column of $\mathbf{W}$ using a separate Gaussian process.
- (Analogous to Gaussian emulation) By emulating these principal components we can emulate the original process.

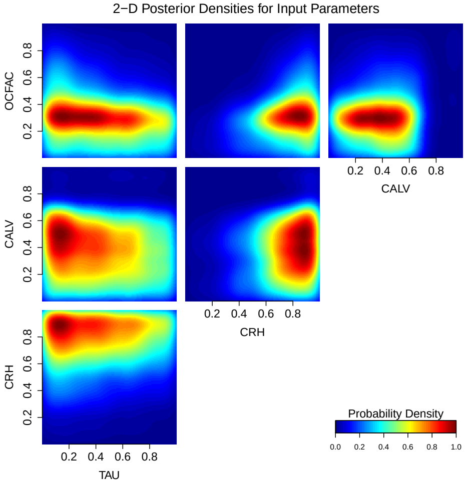
Calibration

- Calibration model for n-dimensional vector of natural parameters $\boldsymbol{\lambda}$ for observational data using the emulator $\mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{W})$:
$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{W}) + \boldsymbol{\delta},$$
where $\boldsymbol{\theta}^*$ is the "best fit" value for the observational data, $\boldsymbol{\delta}$ is discrepancy term that represents structural error between the model output and observational data.
- To get around the challenges in integrating out $\boldsymbol{\delta}$ described above we use a basis representation for the discrepancy term such that
$$\boldsymbol{\delta} = \mathbf{K}_d\mathbf{v},$$
with the $n \times J_d$ basis matrix $\mathbf{K}_d$ and the J_d-dimensional random coefficient vector $\mathbf{v} \sim N(\mathbf{0}, \sigma_d^2\mathbf{I}_{J_d})$.
- Discrepancy basis obtained from common discrepancy pattern across parameter settings, proportion mismatch at each location.

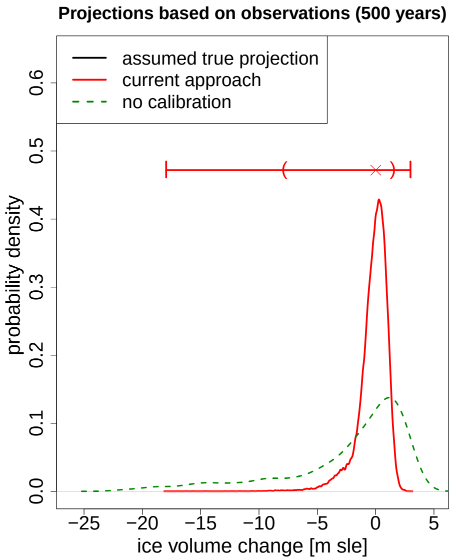
Emulation



Calibration Results



Calibration Results



Summary

- I have described computationally expedient approaches for computer model emulation and calibration for high-dimensional Gaussian and binary spatial data.
- Many potential applications beyond climate science.

Acknowledgments

Collaborators:

- Won Chang, University of Chicago
- David Pollard, Earth and Environmental Systems Institute (EESI), Penn State U.
- Patrick Applegate, EESI, Penn State U.
- Klaus Keller, Geosciences, Penn State U.
- Roman Olson, The University of New South Wales

This work was partially supported by the following grants:

- The Network for Sustainable Climate Risk Management (SCRiM), NSF GEO-1240507.
- NSF CDSE/DMS-1418090 Statistical Methods for Ice Sheet Projections

Relevant Manuscripts

- Chang, W., M. Haran, R. Olson, and K. Keller (2014): Fast dimension-reduced climate model calibration, *Annals of Applied Statistics*
- Chang, W., Applegate, P., Haran, M. and Keller, K. (2014) Probabilistic calibration of a Greenland Ice Sheet model using spatially-resolved synthetic observations: toward projections of ice mass loss with uncertainties, *Geoscientific Model Development*
- Chang, W., Haran, M, Applegate, P., Pollard, D. (2014): alibrating an Ice sheet model using high-dimensional non-Gaussian spatial data (under preparation)

Appendix: Cross-Validation for Emulator

- Example of leave-10%-out cross validation result: