Statistical Methods for Ice Sheet Models

Murali Haran

Department of Statistics, Pennsylvania State University

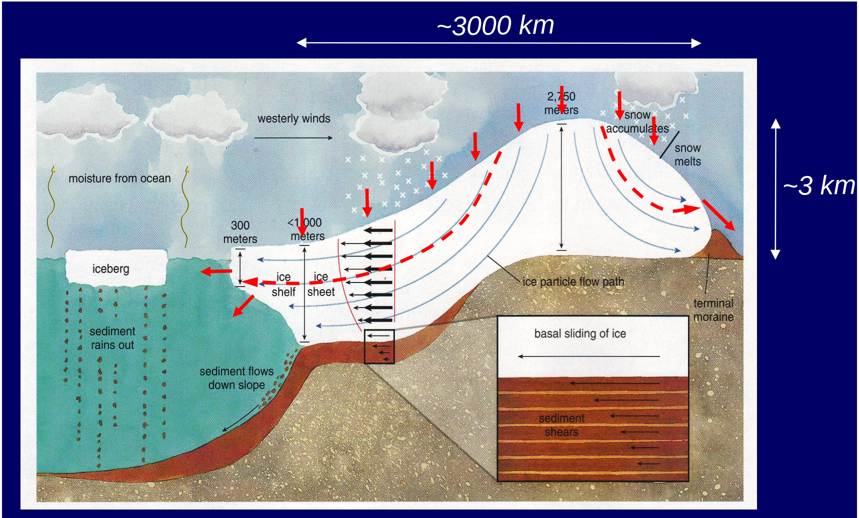
Brigham Young University
Department of Statistics

Collaborators:
Won Chang (University of Chicago Statistics)
Patrick Applegate and David Pollard (Penn State Institutes for Energy and Environment)
Klaus Keller (Penn State Geosciences)
Roman Olson (U of New South Wales Climate Change Research Centre)

Talk Summary

- The West Antarctic Ice Sheet (WAIS) has the potential to be a significant contributor to future sea level change.
- How can we project the future behavior of WAIS? One approach: PSU-ICE model (Pollard and DeConto, 2009).
- Uncertain about key model input parameters.
- Our research: methods to use observations of the ice sheet to infer parameters.
- Challenges
  1. spatial binary/non-Gaussian data
  2. high-dimensional
  3. data-model discrepancies
- I will describe methods to address these issues.

Ice Sheet Physics

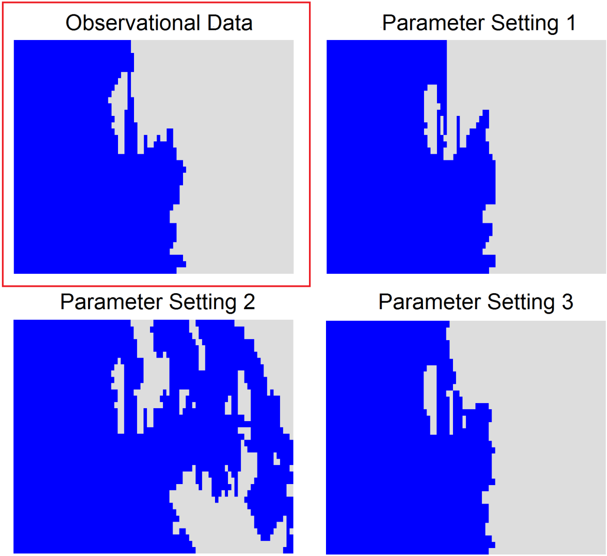


Siegert, 2002, Amer. Sci.

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

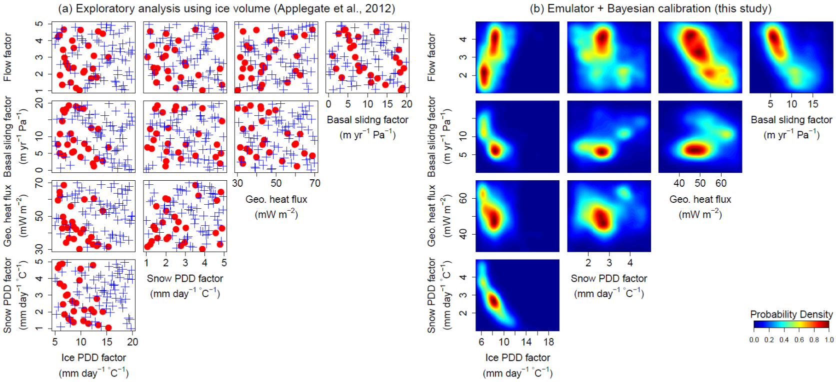
West Antarctic Ice Sheet Example



How Does Statistical Rigor Help Scientists?

1. We account for (epistemic) uncertainties in emulation
2. We provide *real* probability distributions, very important for impacts/risk quantification.
3. We use all available information (no aggregation): often reduces uncertainties.
4. We provide sharper/more useful results.

Example of Sharper Results



Left: previous ad-hoc methods. Right: statistical calibration

Two-stage Approach to Emulation-Calibration

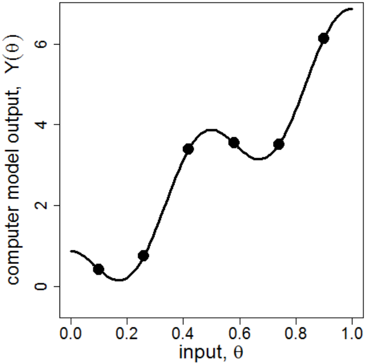
Outline of our approach

1. Emulation step: Find fast approximation for computer model using a Gaussian process (GP).
2. Calibration step: Infer climate parameter using emulator and observations, while accounting for data-model discrepancy

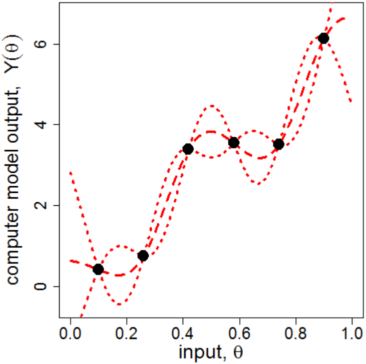
- Two stage: Liu, Bayarri and Berger (2009), Bhat, Haran, Olson, Keller (2012)
- Joint model approach: Sanso et al. (2008); Higdon et al. (2008).

Emulation Step

Toy example: model output is a scalar, and continuous.



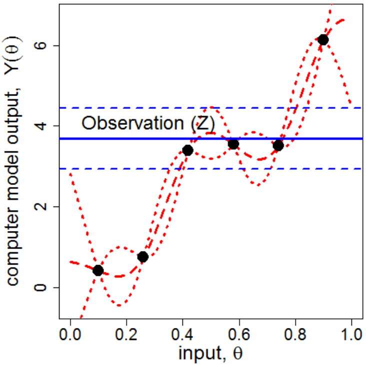
Computer model output (y-axis) vs. input (x-axis)



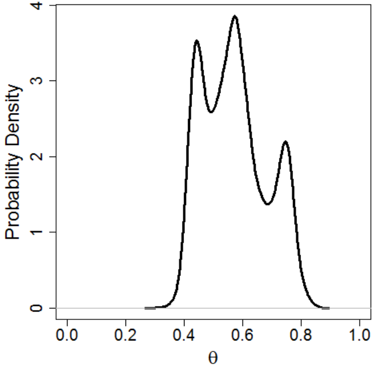
Emulation (approximation) of computer model using GP

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** $\mathbf{Y}(\theta_1), \ldots, \mathbf{Y}(\theta_p)$, where each $\mathbf{Y}(\theta_i) = (Y(\theta_i, \mathbf{s}_1), \ldots, Y(\theta_i, \mathbf{s}_n))^T$ is a vector of spatial data

Step 1: Computer Model Emulation Basics

- Fit Gaussian process model for computer model output **Y** to interpolate the values at the parameter settings $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$ *and* the spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$

$$\text{vec}\,(\mathbf{Y}) \sim N(\mathbf{X}\beta, \boldsymbol{\Sigma}(\xi_y)),$$

  vec($\cdot$) concatenates columns into one vector
- β and ξ_y estimated by maximum likelihood, $\hat{\beta}, \hat{\xi}_y$.
- Covariance interpolates across spatial surface *and* input space.

Result: Obtain a probability model (= predictive distribution) for model output at any input parameter θ, $\eta(\boldsymbol{\theta}, \mathbf{Y})$.

Step 2: Calibration Basics

- Discrepancy $\approx$ mismatch between computer model output and data when parameters are perfectly calibrated and there is no observational error.
- Probability model for observations $\mathbf{Z}$ is then

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

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

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

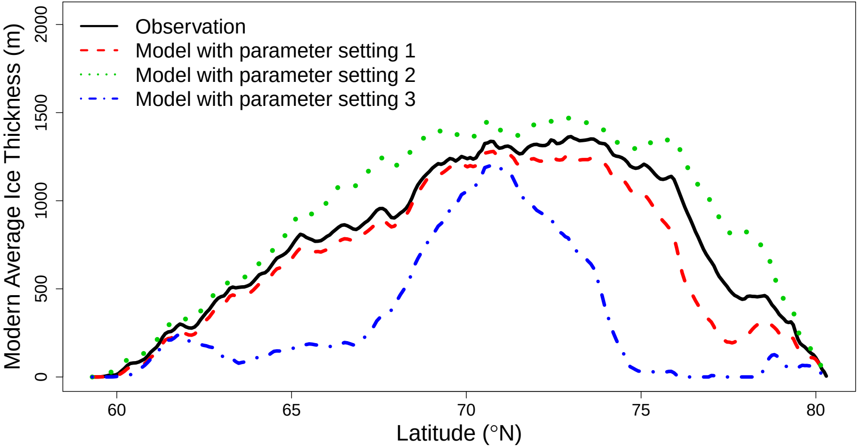
with emulator parameter $\hat{\boldsymbol{\xi}}_y$ fixed at value estimated in emulation step.

Why Data Aggregation is Popular

- The approach outlined works except for the computational issues posed by the size of the data.
- Can side-step computational problems by aggregatoin:
  - Computational problem solved.
  - Often more amenable to Gaussian process models.
  - Sometimes justified by the models: may only be appropriate (have “good skill”) at an aggregate level.

Example: Aggregated Greenland Ice Sheet Data

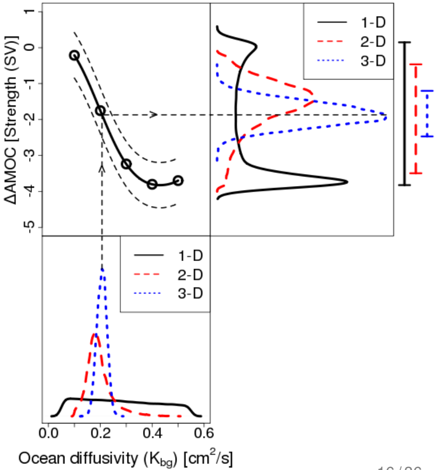
- To avoid binary spatial data: aggregate across longitude.



Avoid Aggregation to Reduce Uncertainties

(Chang, Haran, Olson, Keller, 2014)
In some cases, obtain sharper results if we can avoid aggregation.

- 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.



Emulation-Calibration with Binary Spatial Output

- Now $Y(\boldsymbol{\theta}, \mathbf{s})$ is binary (0-1) model ouput, $Z(\mathbf{s})$ is data.
- Let $\mathbf{\Gamma}_{p \times n}$ be matrix of natural parameters for model output: $\gamma_{ij}^{Y} = \log\left(\frac{p_{ij}}{1-p_{ij}}\right)$ is logit for ith parameter setting at jth spatial location and $p_{ij} = P(Y(\boldsymbol{\theta}_i, \mathbf{s}_j) = 1)$.
- Given $\mathbf{\Gamma}$, $Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$'s are conditionally independent Bernoulli.
- Approach (sketch):
  1. Assume it is possible to estimate $\mathbf{\Gamma}$ from the $n \times p$ matrix of computer model output.
  2. Emulate computer model by *interpolating natural parameters* using a Gaussian process across input parameter space and spatial locations.
  3. Calibration by using fitted Gaussian process $\eta(\boldsymbol{\theta}, \mathbf{Y})$ + discrepancy δ to obtain a likelihood function for the *natural parameter vector for observations*.

Challenges

- Step 1 (obtaining $\boldsymbol{\Gamma}$) is ill-posed: *np* parameters for *np* data points.
- Step 2 (emulation) is 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
- Step 3 (calibration): involves having to perform a high-dimensional integration + expensive matrix operations.

We propose dimension-reduction to address both ill-posedness and computational issues.

Principal Components for Emulation: Basic 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 $\theta_1, \ldots, \theta_p$.
  Surprisingly flexible approach, including non-separable covariance.

Emulation Using PCs

Simple case: everything is well modeled by Gaussian processes.

- 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 $\mathcal{O}(J_yp^3)$ (1.2×10^{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 eigenvectors
- 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

Efficient Emulation: Outline

More complicated case: binary spatial data.

- Think in terms of generalized linear mixed models.
- Rewrite $\boldsymbol{\Gamma}$ in terms of logistic principal components (Lee et al., 2010).
- Use maximum likelihood to perform logistic principal components. Non-trivial, requires majorization-minimization (MM) algorithm (Lange et al. 2000; Hunter and Lange, 2004).
- Emulate each column of (reduced-dimensional) principal component matrix with an independent Gaussian process. Very fast and easy to do.
- We can obtain an emulator for $\boldsymbol{\Gamma}$ by emulating these principal components.

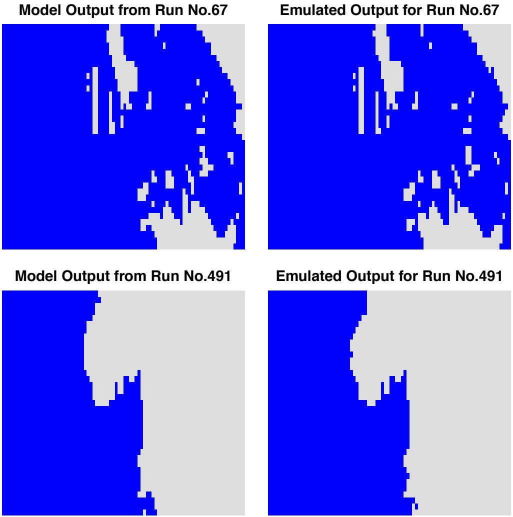
Dimension-reduction

- 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.

Emulation Examples



Efficient Calibration Outline

- Model for n-dimensional vector of natural parameters $\boldsymbol{\lambda}$ for observational data using emulator $\mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{W})$:

$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{W}) + \boldsymbol{\delta},$$

  $\boldsymbol{\delta}$ is discrepancy term that represents structural error between the model output and observational data.
- To get around the challenges in integrating out $\boldsymbol{\delta}$, use a basis representation.
- Discrepancy basis obtained from common discrepancy pattern across parameter settings.

Efficient Calibration: Details

- Calibration model for n-dimensional vector of natural parameters $\boldsymbol{\lambda}$ for observational data using the emulator $\mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{W})$:

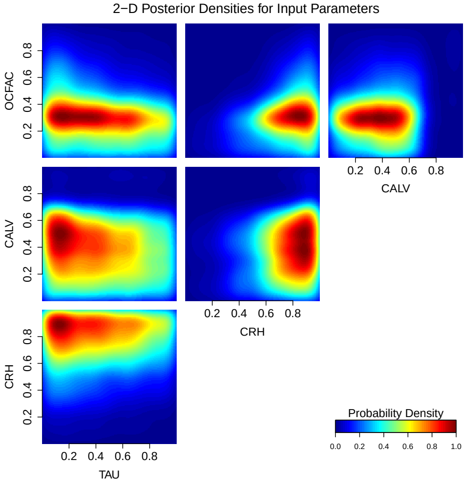
$$\boldsymbol{\lambda} = \boldsymbol{\mu} + \mathbf{K}_y\boldsymbol{\eta}(\boldsymbol{\theta}^*, \mathbf{W}) + \boldsymbol{\delta},$$

  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

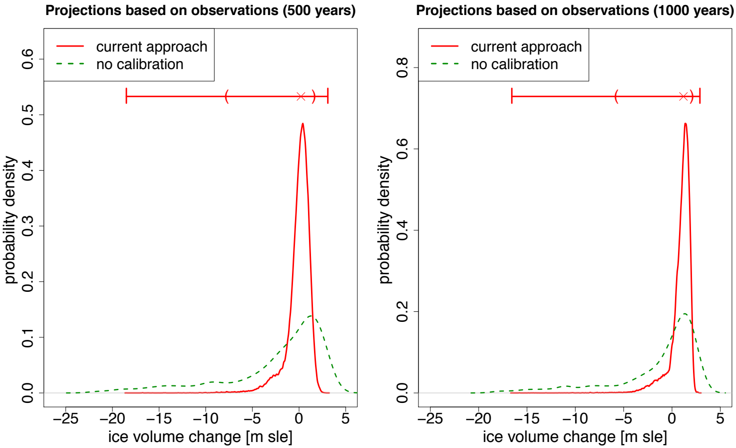
$$\boldsymbol{\delta} = \mathbf{K}_d\mathbf{v},$$

  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.

Calibration Results



Example Projections



Concluding Remarks

- I have described a computationally expedient approach for computer model emulation and calibration for high-dimensional binary spatial output.
- This tool is valuable for ice sheet model projections.
- Promising results when we incorporate paleo data.
- Caveats: we are using simple projection scenarios, we are ignoring floating ice information...
- Method tested extensively using simulated data examples, multiple data sets.
- Potentially useful for other (non-climate) complex computer model applications.

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:

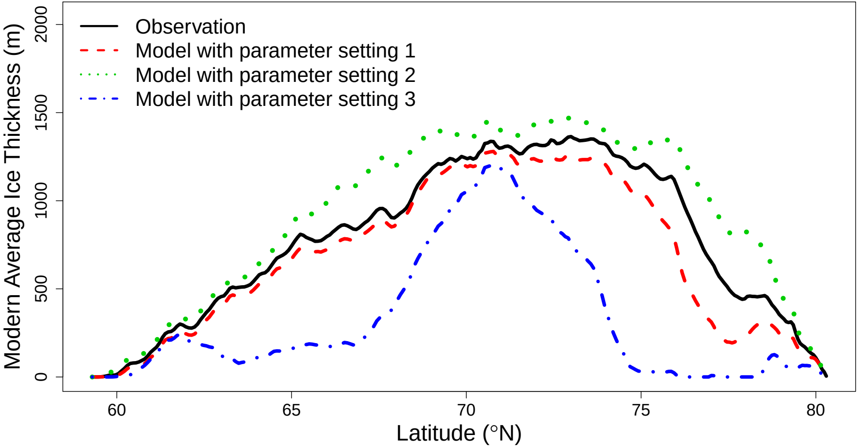
- The Network for Sustainable Climate Risk Management (SCRiM), **NSF GEO-1240507.**
- **NSF CDSE/DMS-1418090** Statistical Methods for Ice Sheet Projections

Relevant Manuscripts

- Chang, W., Haran, M, Applegate, P., Pollard, D. (2016): Calibrating an Ice sheet model using high-dimensional non-Gaussian spatial data, *Journal of the American Statistical Association*
- Chang, W., Haran, M, Applegate, P., Pollard, D. (2016): Improving Ice Sheet Model Calibration Using Paleoclimate and Modern Data, *on arxiv.org*.
- Chang, W., M. Haran, R. Olson, and K. Keller (2015): 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*

Aggregated Ice Sheet Data: Example

- To avoid binary spatial data: aggregate across longitude.



Appendix 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:

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

Appendix: Computational Challenges for Gaussian Emulation-Calibration

- 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

Appendix: 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

Appendix: Cross-Validation for Emulator

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