Combining High-dimensional Data from Climate Models and Observations to Sharpen Climate Projections

Murali Haran

Department of Statistics, Penn State University

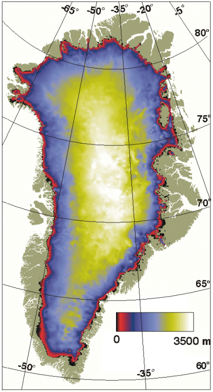
Collaborators:
Won Chang (Penn State Statistics)
Patrick Applegate, Klaus Keller (Penn State Geosciences)
Roman Olson (Climate Change Research Centre, University of New South Wales)

What This Talk Is About

- Physical models may be used to make climate projections. Examples:
  - Ice sheet models for projecting future ice sheet behavior
  - Models for Atlantic Meridional Overturning Circulation (AMOC)
- A major source of uncertainty about these projections is due to uncertainty about climate model input parameters.
- We propose a method for inferring model parameters from spatial model outputs and observations.
- Challenges: Data in the form of high-dimensional spatial processes. Complicated error structures.
- I will describe novel computationally efficient approaches based on principal components (PC) and kernel convolution.
- Fast automated approach for calibration, uncertainty quantification.

Greenland Ice Sheet

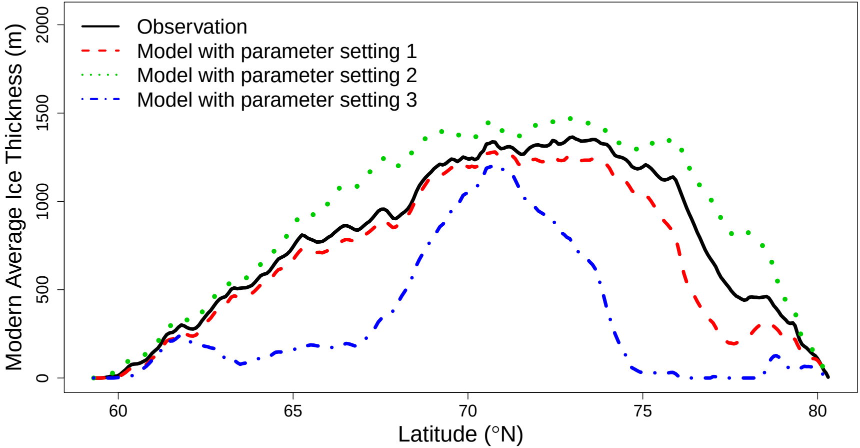
Important contributor to sea level rise. Total melting would result in sea level rise of 7m.



Bamber et al. (2001)

Calibration Problem

Which model parameter settings best match observations?



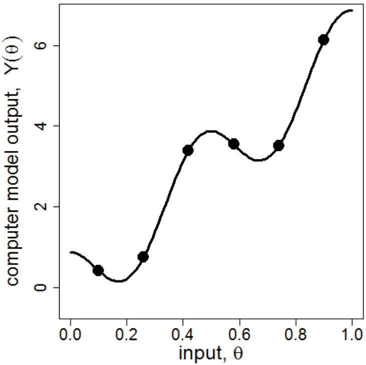
Two-stage Approach to Emulation-Calibration

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

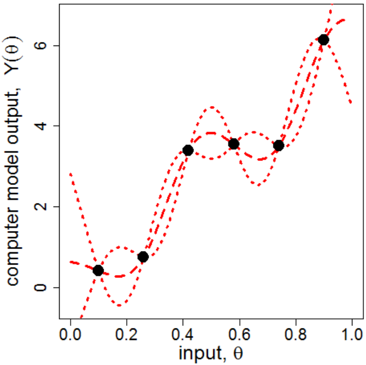
2. Calibration step: Infer climate parameter using emulator and observations, while accounting for data-model discrepancy

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

Emulation Step: Toy Example

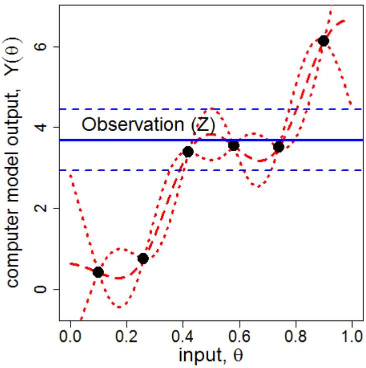


Computer model output (y-axis) vs. input (x-axis)

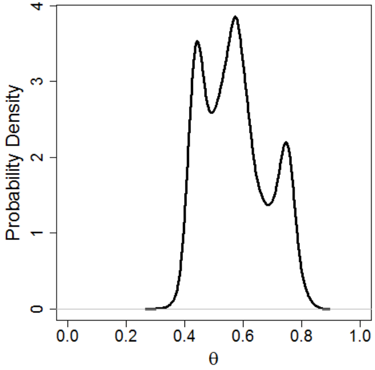


Emulation (approximation) of computer model using GP

- Accounts for interpolation uncertainty
- Provides a probability model that connects θ to $Y(\theta)$



Combining observation
and emulator



Posterior PDF of θ
given model output and observations

Probabilistic calibration provides full (possibly multi-modal) distribution, accounting for uncertainties

Summary of Statistical Problem

- **Goal**: Learning about $\boldsymbol{\theta}$ based on two sources of information:
  - **Observations**: Mean ice thickness profile† $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^T$, where $\mathbf{s}_1, \ldots, \mathbf{s}_n$ are latitude points.
  - **Ice model output*** for mean ice thickness $\mathbf{Y}(\boldsymbol{\theta}_1), \ldots, \mathbf{Y}(\boldsymbol{\theta}_p)$, 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 process (Applegate et al 2012).

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

- Important: output at each $\boldsymbol{\theta}_i$ is a spatial process. $n = 264$ locations, $p = 100$ runs.

†Averaged over longitude
*SICOPOLIS (Greve, 1997; Greve et al., 2011)

Step 1: Dimension Reduction

- Consider model outputs at $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p$ as replicates and obtain 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 $\theta_1, \ldots, \theta_p$.
- Note: from cross validation it is apparent that a separability assumption (which greatly speeds up computing time) is untenable.

Step 2: 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 $\mathcal{O}(n^3 p^3)$ to $\mathcal{O}(J_y p^3)$ (6.13×10^{12} to 3.33×10^6 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

Dimension Reduction for Discrepancy Process

- Kernel convolution: Specifying n-dimensional discrepancy process $\boldsymbol{\delta}$ 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 $\mathbf{a}_1, \ldots, \mathbf{a}_{J_d}$;

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

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

Calibration in Reduced Dimensions

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

$$\mathbf{Z} = \underbrace{\mathbf{K}_y \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R)}_{\text{emulator}} + \underbrace{\mathbf{K}_d \boldsymbol{\nu}}_{\text{discrepancy}} + \underbrace{\boldsymbol{\epsilon}}_{\text{observation error}},$$

$$\Rightarrow \mathbf{Z}^R = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{Z} = \begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{Y}^R) \\ \boldsymbol{\nu} \end{pmatrix} + (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\boldsymbol{\epsilon},$$

  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

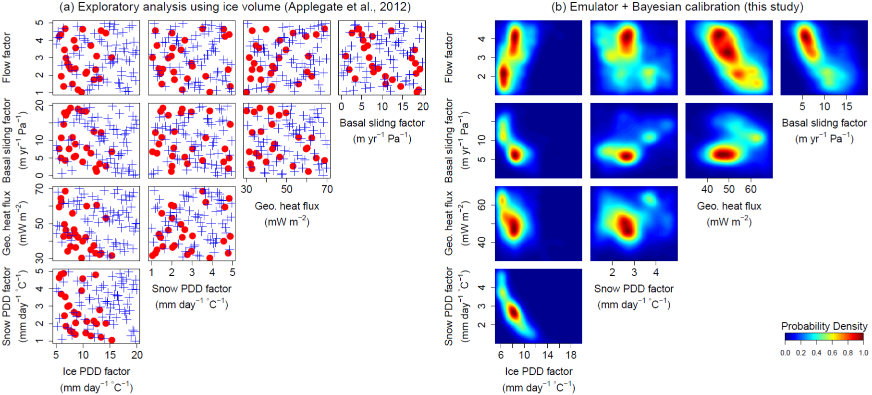
$$\pi(\boldsymbol{\theta}, \kappa_d, \sigma^2 | \mathbf{Z}^R, \mathbf{Y}^R) \propto \underbrace{L(\mathbf{Z}^R | \mathbf{Y}^R, \boldsymbol{\theta}, \kappa_d, \sigma^2)}_{\text{likelihood given by above}} \underbrace{p(\boldsymbol{\theta}) p(\kappa_d) p(\sigma^2)}_{\text{priors}}$$

Perfect Model Experiment

Test if our calibration method can recover the "truth".

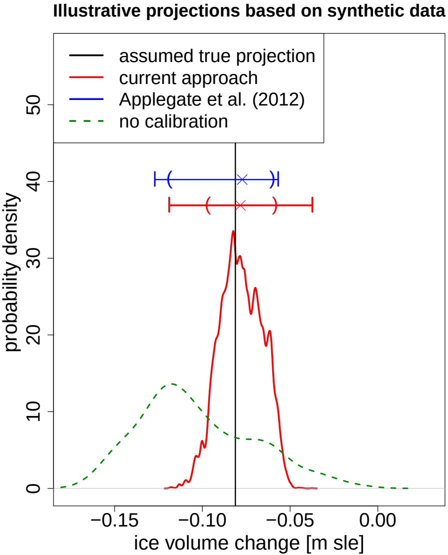
1. Pick one model output as synthetic truth.
2. Generate observational data by adding structural error.
3. Calibrate the parameters using remaining model outputs.
4. See if we get back parameter setting for synthetic truth.

Parameter Inference

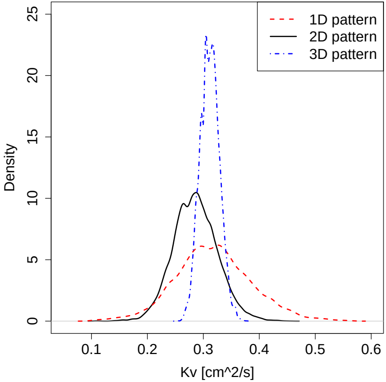


Previous approach versus our statistically rigorous approach

Ice Volume Change Projection

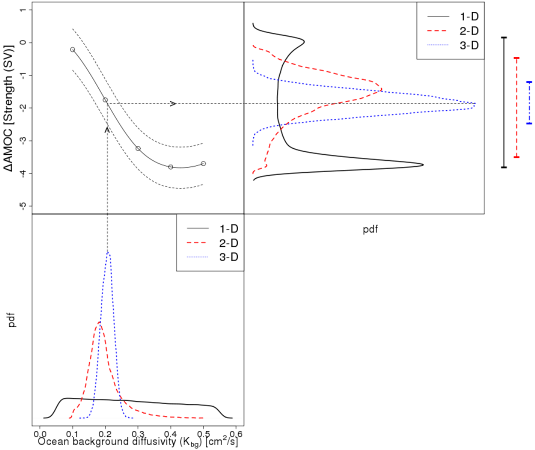


Another Example: Inferring Ocean Vertical Diffusivity K_{bg}



- K_{bg} is a key source of uncertainty in projections of the Atlantic Meridional Overturning Circulation (AMOC)
- Fast calibration with unaggregated 3D spatial data ($61,051 \times 250$ parameter settings) sharpens inference and projections

MOC Projections for 2100 Using Inferred K_{bg}



Discussion

- Dimension reduction-based approach:
  - Very fast, scales well with n, number of spatial location
  - Easy to specify model. We have applied it to various problems.
- Ice model calibration:
  - Provides sharper posterior densities for input parameters and sea level rise projections
  - Shows clear interaction between parameters
- MOC projections: much tighter with unaggregated spatial data
- Ice sheet projections: current ice sheet model is still problematic; we are investigating other models before using real data
- Open challenge (ongoing work): can we learn more from the full spatial pattern of binary or zero-inflated ice sheet data?

References

- 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, *submitted to Geophysical Model Development Discussion*
- Chang, W., Haran, M., Olson, R., and Keller, K. (2013) Fast dimension-reduced climate model calibration, *accepted for publication in the Annals of Applied Statistics, arXiv:1303.1382*.
- Applegate, P. J., Kirchner, N., Stone, E. J., Keller, K., and Greve, R., 2012, An assessment of key model parametric uncertainties in projections of Greenland Ice Sheet behavior: *The Cryosphere* 6, 589-606.

This work was supported by the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507