Statistical Methods for Studying the West Antarctic Ice Sheet

Murali Haran

Department of Statistics, Pennsylvania State University

ISBA 2016
Cagliari, Sardinia. June 2016

Collaborators:
Won Chang (Dept of Statistics, University of Chicago)
Yawen Guan (Dept of Statistics, Penn State University)
Patrick Applegate (Penn State Institutes for Energy and Environment)
David Pollard (Penn State Institutes for Energy and Environment)

Statistics and Ice Sheets

Formulating reasonable hypotheses regarding climatic change requires physical insight and ingenuity, but subsequently testing these hypotheses demands quantitative computation. – Edward N. Lorenz (1970)

- My translation: to study future climate we need physical models, data, and statistics. Innovation on all three fronts.
- Focus here: The West Antarctic Ice Sheet (WAIS).
- 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, 2012).

The New York Times | http://nyti.ms/1qjj7Xl

SCIENCE

Climate Model Predicts West Antarctic Ice Sheet Could Melt Rapidly

By JUSTIN GILLIS MARCH 30, 2016

For half a century, climate scientists have seen the West Antarctic ice sheet, a remnant of the last ice age, as a sword of Damocles hanging over human civilization.

The great ice sheet, larger than Mexico, is thought to be potentially vulnerable to disintegration from a relatively small amount of global warming, and capable of raising the sea level by 12 feet or more should it break up. But researchers long assumed the worst effects would take hundreds — if not thousands — of years to occur.

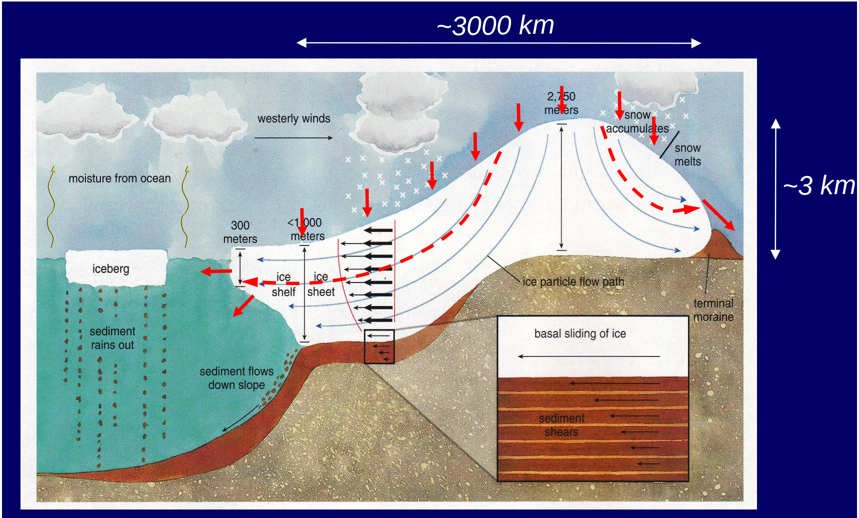
Now, new research suggests the disaster scenario could play out much sooner.

Continued high emissions of heat-trapping gases could launch a disintegration of the ice sheet within decades, according to a study published Wednesday, heaving enough water into the ocean to raise the sea level as much as three feet by the end of this century.

Talk Summary

- Our research: methods to use observations of the ice sheet to infer important parameters of the ice sheet model.
- Observations:
  1. Satellite data on the modern ice sheet.
  2. Paleo reconstructions of ice sheet from 25,000 years ago to present time.
- Challenges
  1. Two sets of data: spatial and temporal binary/non-Gaussian data. “Data” = Observations and computer model output.
  2. High-dimensional spatial data
- I will outline methods to address these issues.

Ice Sheet Physics

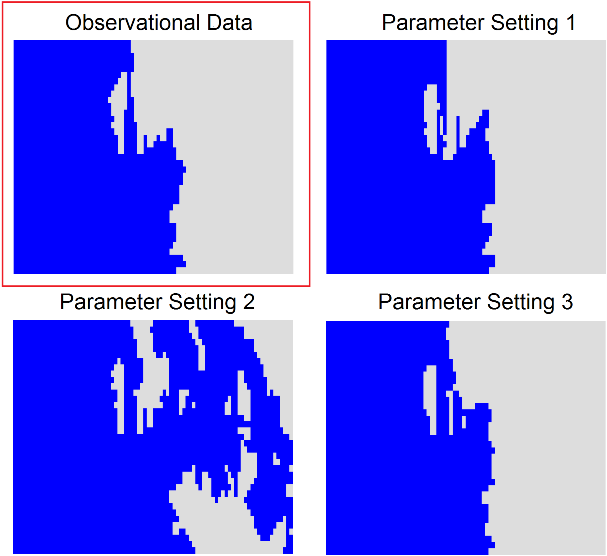


Siegert, 2002, Amer. Sci.

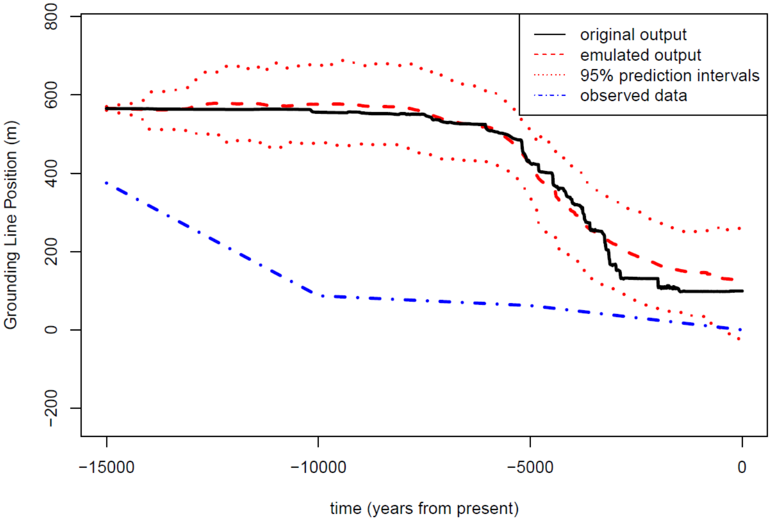
Ice Sheet Model Parameters

- 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

Modern Data


Observational Data
Parameter Setting 1
Parameter Setting 2
Parameter Setting 3

Paleo Data



original output
emulated output
95% prediction intervals
observed data
800
600
400
200
0
−200
Grounding Line Position (m)
−15000
−10000
−5000
0
time (years from present)

Two-stage Approach to Emulation-Calibration

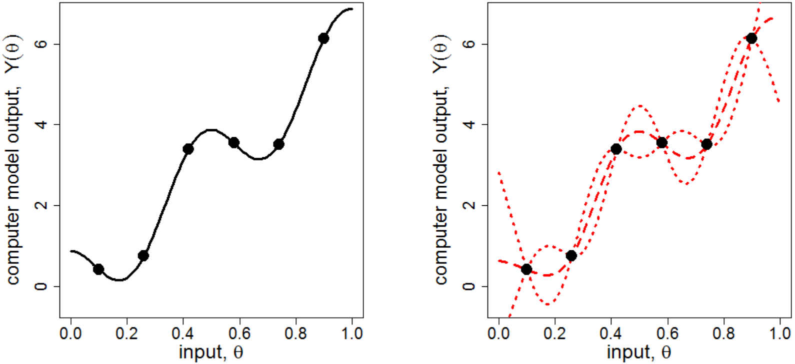
Outline of our approach

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.

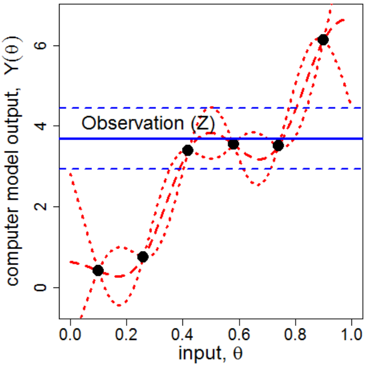


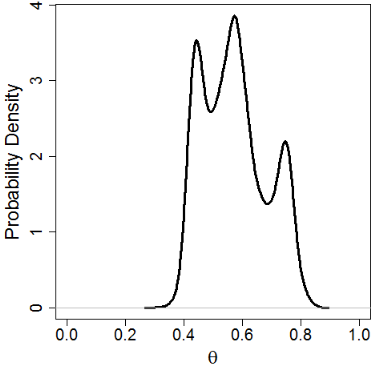
Computer model output (y-axis) vs. input (x-axis)

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**:
    1. Observed time series of past grounding line positions reconstructed from paleo records: $\mathbf{Z}_1 = (Z_1(t_1), \ldots, Z_1(t_n))^T$, $t_1, \ldots, t_n$ are time points locations.
    2. Observed modern ice-no ice from satellite data: $\mathbf{Z}_2 = [Z_2(\mathbf{s}_1), \ldots, Z_2(\mathbf{s}_m)]$, locations $\mathbf{s}_1, \ldots, \mathbf{s}_m$.
  - **Model output**
    1. $\mathbf{Y}_1(\boldsymbol{\theta}_1), \ldots, \mathbf{Y}_1(\boldsymbol{\theta}_p)$, where each $\mathbf{Y}_1(\boldsymbol{\theta}_i) = (Y_1(\boldsymbol{\theta}_i, t_1), \ldots, Y_1(\boldsymbol{\theta}_i, t_n))^T$ is a time series of grounding line positions at parameter setting $\boldsymbol{\theta}_i$.
    2. $\mathbf{Y}_2(\boldsymbol{\theta}_1), \ldots, \mathbf{Y}_2(\boldsymbol{\theta}_p)$, where each $\mathbf{Y}_2(\boldsymbol{\theta}_i) = (Y_2(\boldsymbol{\theta}_i, \mathbf{s}_1), \ldots, Y_2(\boldsymbol{\theta}_i, \mathbf{s}_m))^T$ is a vector of spatial data at parameter setting $\boldsymbol{\theta}_i$.

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 $\mathbf{s}_1, \ldots, \mathbf{s}_n$

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

  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 MLE.

Challenge: Two sets of non-Gaussian, dependent data.

Principal Components for Emulation

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

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

  A version of this in Higdon et al. (2008).
- Reduced rank approach that *obtains basis from data*.
- PCs pick up characteristics of model output that vary most across input parameters $\theta_1, \ldots, \theta_p$.
  Surprisingly flexible, e.g. non-separable covariance.
- Directly applied to paleo data.
- Modern data: more complicated because spatial and non-Gaussian.

Emulation-Calibration with Binary Spatial Output

- Now $Y(\boldsymbol{\theta}, \mathbf{s})$ is binary (0-1) model ouput, $Z(\mathbf{s})$ is data.
- Let $\boldsymbol{\Gamma}_{p \times n}$ 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 $\boldsymbol{\Gamma}$, $Y(\boldsymbol{\theta}_i, \mathbf{s}_j)$'s are conditionally independent Bernoulli.
- Approach (sketch):
  1. Assume it is possible to estimate $\boldsymbol{\Gamma}$ 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*.

Efficient Emulation: Outline

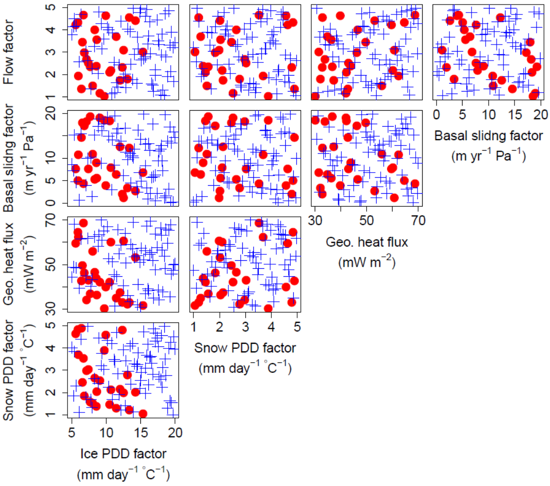
- Think in terms of generalized linear mixed models.
- Rewrite $\boldsymbol{\Gamma}$ in terms of logistic principal components (Lee et al., 2010).
- Use maximum likelihood to perform logistic principal components. Non-trivial, requires majorization-minimization (MM) algorithm (Lange et al. 2000; Hunter 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.
- Calibration: a few more details with discrepancies. Then standard MCMC...

Challenges

- Step 1 (obtaining $\mathbf{\Gamma}$) 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.

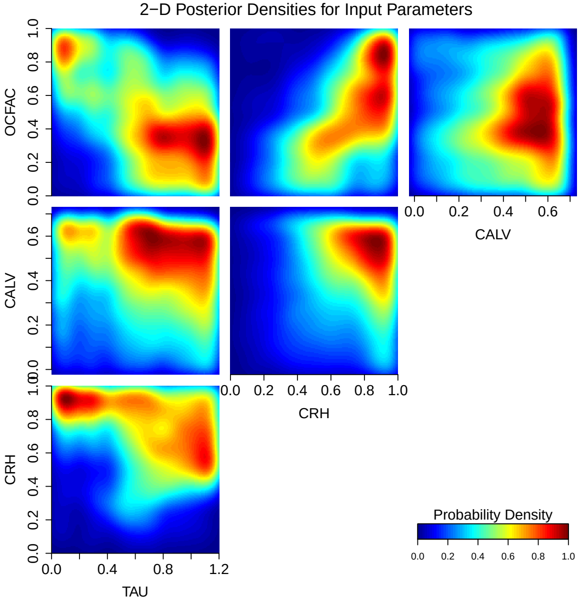
We propose dimension-reduction to address both ill-posedness and computational issues.

Non-statistical Calibration

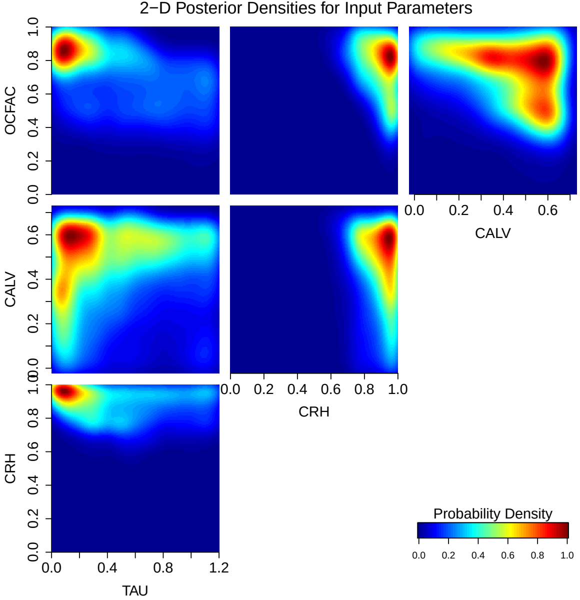


Applegate et al. (2012)

Calibration Results with Modern Data

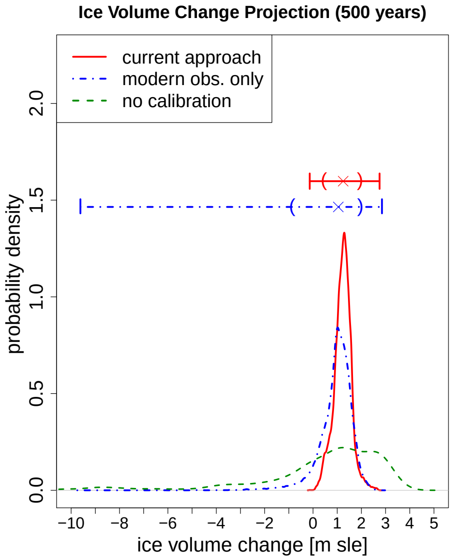


Calibration Results with Modern and Paleo Data



Paleo data eliminates unrealistic trajectories of the ice sheet.

Ice Volume Projections: Paleo Cuts Off Left Tail



Possibility of “no sea level rise” is virtually eliminated.

Summary Remarks

- Principle components-based approach to computer model calibration with two sources of data.
  - Driven by computer model output at various settings $\Rightarrow$ we choose a basis that is well suited to emulation and calibration.
  - Sidesteps issues with choice of basis, number of knots.
- Method tested extensively using simulated data examples, multiple data sets.
- Using paleo data results in an important change in sea level rise projections.
- Ongoing work: detailed understanding the current state of ice sheet thickness using simpler physical models.

Conclusion

I have a good brain and I have said a lot of things.
– Donald J. Trump

Acknowledgments

Collaborators:

- Won Chang, University of Chicago/University of Cincinnati
- Yawen Guan, Penn State Statistics
- 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): Improving Ice Sheet Model Calibration Using Paleoclimate and Modern Data, *on arxiv.org*.
- 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., M. Haran, R. Olson, and K. Keller (2014): Fast dimension-reduced climate model calibration, *Annals of Applied Statistics*

Appendix

BEGIN APPENDIX

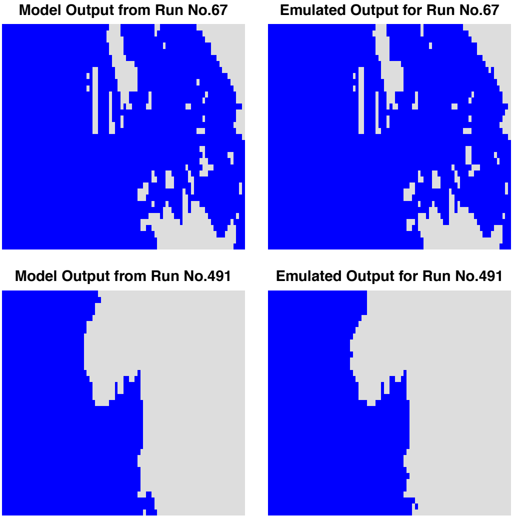
Dimension Reduction for Discrepancy Process

- 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 $\mathbf{s}_1, \dots, \mathbf{s}_n$ to knot locations $\mathbf{a}_1, \dots, \mathbf{a}_{J_d}$;

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

  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

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.

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 $\mathcal{O}(J_y p^3)$ (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

- 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.