SAND2018-4233C

Methodologies for Enabling Bayesian Calibration in Land-ice Modeling Towards Probabilistic Projections of Sea-level Change

J. Jakeman[1], M. Perego[1], I. K. Tezaur[1]
S. Price[2] , G. Stadler[3]

[1]Sandia National Laboratories, NM, USA
[2]Los Alamos National Laboratory, NM, USA
[3]Courant Institute, NY, USA

SIAM Conference on Uncertainty Quantification

Garden Grove, April 18, 2018

Center for Computing Research

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Funded by (ProSPect)

SciDAC Scientific Discovery through Advanced Computing

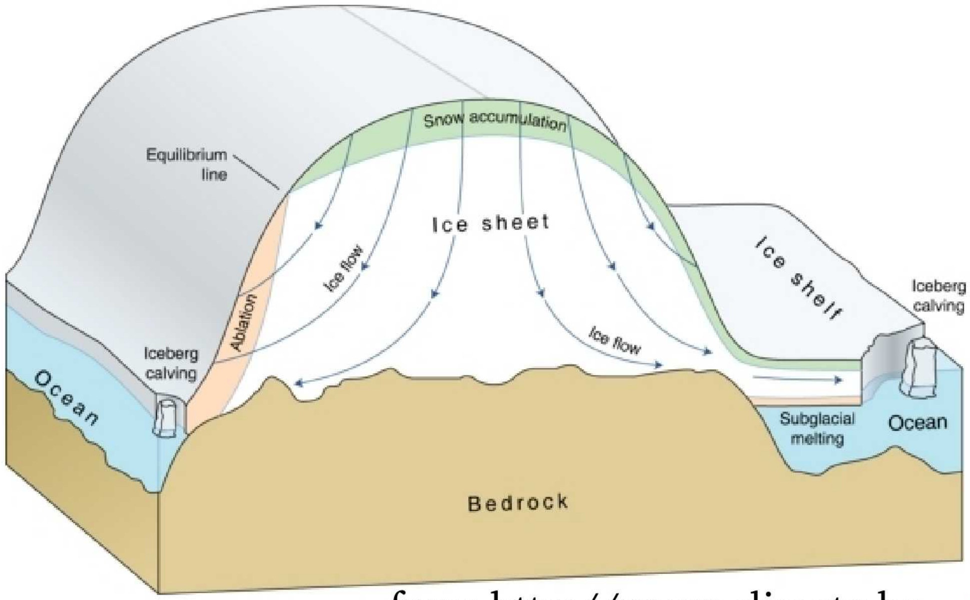
Sandia National Laboratories

Brief introduction and motivation

- Modeling ice sheets (Greenland and Antarctica) dynamics is essential to provide estimates for sea level rise* in next decades to centuries.

- Ice behaves like a very viscous shear-thinning fluid (similar to lava flow) and can be modeled with nonlinear Stokes equation.

- Greenland and Antarctica ice sheets store most of the fresh water on hearth.
  They have a shallow geometry (thickness up to 4km, horizontal extensions of thousands of km).



from http://www.climate.be

*DOE SciDAC project **ProSPect** (**Pro**babilistic **S**ea Level **P**roj**ect**ion from Ice Sheet and Earth System Models), Institutes: LANL, LBNL, SNL, ONL, NYU, Univ. of Michigan

Problem definition

Our Quantity of Interest (QoI) in ice sheet modeling:

total ice mass loss/gain by, e.g., 2100 → **sea level rise prediction**

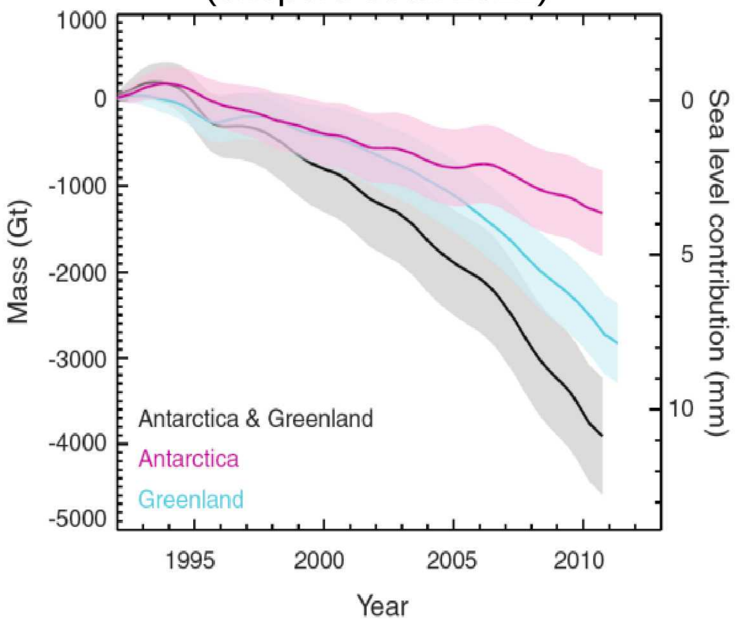
Main sources of uncertainty:

- climate forcing (e.g. Surface Mass Balance)
- **basal friction**
- bedrock topography (noisy and sparse data)
- geothermal heat flux
- modeling errors
- model parameters (e.g. Glen's Flow Law exponent)

Problem definition

Ultimate goal:
quantify the QoI and related uncertainties

(sheperd et al. 2012)



Work flow:

- Perform *adjoint-based deterministic inversion* to estimate initial ice sheet state (i.e. characterize the present state of ice sheet to be used for performing prediction runs).
- *Bayesian inference*: Gaussian posterior low-rank approximation; use deterministic inversion to characterize the parameter distribution (i.e, use the inverted field as mean field of the parameter distribution and approximate its covariance using sensitivities/Hessian).
- *Forward Propagation*

Ice Sheet Modeling

Ice momentum equations

- Ice flow equations (momentum and mass balance)

$$\begin{cases} -\nabla \cdot \sigma = \rho \mathbf{g} \\ \nabla \cdot \mathbf{u} = 0 \end{cases}$$

with:

$$\sigma = 2\mu \mathbf{D} - pI, \qquad \mathbf{D}_{ij}(\mathbf{u}) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

Nonlinear viscosity:

$$\mu = \frac{1}{2}\alpha(T)\,|\mathbf{D}(\mathbf{u})|^{\frac{1}{n}-1}, \quad n \geq 1, \quad (\text{tipically } n \simeq 3)$$

Viscosity is singular when ice is not deforming

Stiffening/Damage factor

$$\mu^*(x,y,z) = \phi(x,y)\,\mu(x,y,z)$$

ϕ : stiffening factor that accounts for modeling errors in rheology

Stokes approximations in different regimes

$\text{Stokes}(\mathbf{u}, p)$

$$\begin{cases} -\nabla \cdot (2\mu \mathbf{D}(\mathbf{u}) - p\mathbf{I}) = \rho \mathbf{g} \\ \nabla \cdot \mathbf{u} = 0 \end{cases}$$

$\text{FO}(u, v)$

First Order* or
Blatter-Pattyn model

$$-\nabla \cdot \left(2\mu \tilde{\mathbf{D}} - \rho g (s - z)\mathbf{I}\right) = \mathbf{0}$$

Dukowicz, Price and Lipscomb, 2010. J. Glaciol.

Stokes approximations in different regimes

$$\text{Stokes}(\mathbf{u}, p)$$

$$\begin{cases} -\nabla \cdot (2\mu \mathbf{D}(\mathbf{u}) - p\mathbf{I}) = \rho \mathbf{g} \\ \nabla \cdot \mathbf{u} = 0 \end{cases}$$

Drop terms using **scaling argument** based on the fact that ice sheets are shallow

$$\mathbf{D}(\mathbf{u}) = \begin{bmatrix} u_x & \frac{1}{2}(u_y + v_x) & \frac{1}{2}(u_z + \cancel{w_x}) \\ \frac{1}{2}(u_y + v_x) & v_y & \frac{1}{2}(v_z + \cancel{w_y}) \\ \frac{1}{2}(u_z + \cancel{w_x}) & \frac{1}{2}(v_z + \cancel{w_y}) & w_z \end{bmatrix} \qquad \mathrm{u} := \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$\mu = \mu(|\mathbf{D}(\mathbf{u})|)$$

$$\text{FO}(u, v)$$

First Order* or Blatter-Pattyn model

*Dukowicz, Price and Lipscomb, 2010. J. Glaciol

Stokes approximations in different regimes

$$\mathrm{Stokes}(\mathbf{u}, p)$$

$$\begin{cases} -\nabla \cdot (2\mu \mathbf{D}(\mathbf{u}) - p\mathbf{I}) = \rho \mathbf{g} \\ \nabla \cdot \mathbf{u} = 0 \end{cases}$$

Drop terms using **scaling argument** based on the fact that ice sheets are shallow

$$\mathbf{D}(\mathbf{u}) = \begin{bmatrix} u_x & \frac{1}{2}(u_y + v_x) & \frac{1}{2}(u_z + \cancel{w_x}) \\ \frac{1}{2}(u_y + v_x) & v_y & \frac{1}{2}(v_z + \cancel{w_y}) \\ \frac{1}{2}(u_z + \cancel{w_x}) & \frac{1}{2}(v_z + \cancel{w_y}) & w_z \end{bmatrix} \qquad \mathrm{u} := \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$\mu = \mu(|\mathbf{D}(\mathbf{u})|)$$

Quasi-hydrostatic approximation

3rd momentum equation

$$-\cancel{\partial_x(\mu u_z)} - \cancel{\partial_y(\mu v_z)} - \partial_z(2\mu w_z - p) = -\rho g,$$

continuity equation

$$w_z = -(u_x + v_y)$$

$$\Longrightarrow p = \rho g(s - z) - 2\mu(u_x + v_y)$$

$$\mathrm{FO}(u, v)$$

First Order* or Blatter-Pattyn model

*Dukowicz, Price and Lipscomb, 2010. J. Glaciol

Stokes approximations in different regimes

$\text{Stokes}(\mathbf{u}, p)$

$$\begin{cases} -\nabla \cdot (2\mu \mathbf{D}(\mathbf{u}) - p\mathbf{I}) = \rho \mathbf{g} \\ \nabla \cdot \mathbf{u} = 0 \end{cases}$$

Drop terms using **scaling argument** based on the fact that ice sheets are shallow

$$\mathbf{D}(u,v) = \begin{bmatrix} u_x & \frac{1}{2}(u_y + v_x) & \frac{1}{2}(u_z + \cancel{w_x}) \\ \frac{1}{2}(u_y + v_x) & v_y & \frac{1}{2}(v_z + \cancel{w_y}) \\ \frac{1}{2}(u_z + \cancel{w_x}) & \frac{1}{2}(v_z + \cancel{w_y}) & -(u_x + v_y) \end{bmatrix} \qquad \mathrm{u} := \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

$$\mu = \mu(|\mathbf{D}(u,v)|)$$

Quasi-hydrostatic approximation

3rd momentum equation

$$-\cancel{\partial_x(\mu u_z)} - \cancel{\partial_y(\mu v_z)} - \partial_z(2\mu w_z - p) = -\rho g,$$

continuity equation

$$w_z = -(u_x + v_y)$$

$$\Longrightarrow p = \rho g(s - z) - 2\mu(u_x + v_y)$$

$\text{FO}(u, v)$

First Order* or Blatter-Pattyn model

$$-\nabla \cdot \left(2\mu \tilde{\mathbf{D}} - \rho g(s - z)\mathbf{I}\right) = \mathbf{0}$$

$$\text{with } \tilde{\mathbf{D}}(u,v) = \begin{bmatrix} 2u_x + v_y & \frac{1}{2}(u_y + v_x) & \frac{1}{2}u_z \\ \frac{1}{2}(u_y + v_x) & u_x + 2v_y & \frac{1}{2}v_z \end{bmatrix}$$

Dukowicz, Price and Lipscomb, 2010. J. Glaciol

Deterministic Inversion

GOAL

Find ice sheet initial state that

- matches observations (e.g. surface velocity, temperature, etc.)
- is in compliance with flow model and climate forcing

Bibliography

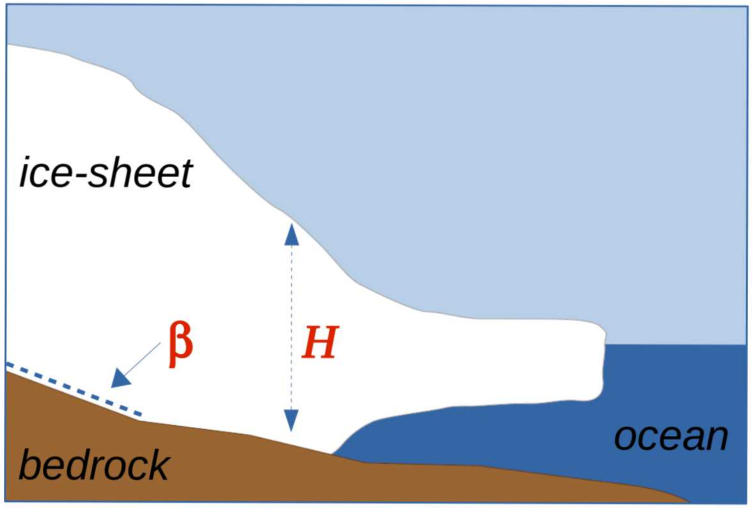
- *Arthern, Gudmundsson*, J. Glaciology, 2010
- *Price, Payne, Howat and Smith*, PNAS, 2011
- *Petra, Zhu, Stadler, Hughes, Ghattas*, J. Glaciology, 2012
- *Pollard DeConto*, TCD, 2012
- W. J. J.Van Pelt et al., The Cryosphere, 2013
- *Morlighem et al.* Geophysical Research Letters, 2013
- *Goldberg and Heimbach*, The Cryosphere, 2013
- *Michel et al.*, Computers & Geosciences, 2014
- *Perego, Price, Stadler,* Journal of Geophysical Research, 2014
- *Goldberg et al.,* The Cryosphere Discussions, 2015

Deterministic Inversion

Problem details

Available data/measurements

- *ice extension and surface topography*
- ***surface velocity***
- *surface air temperature*
- ...



Fields to be estimated

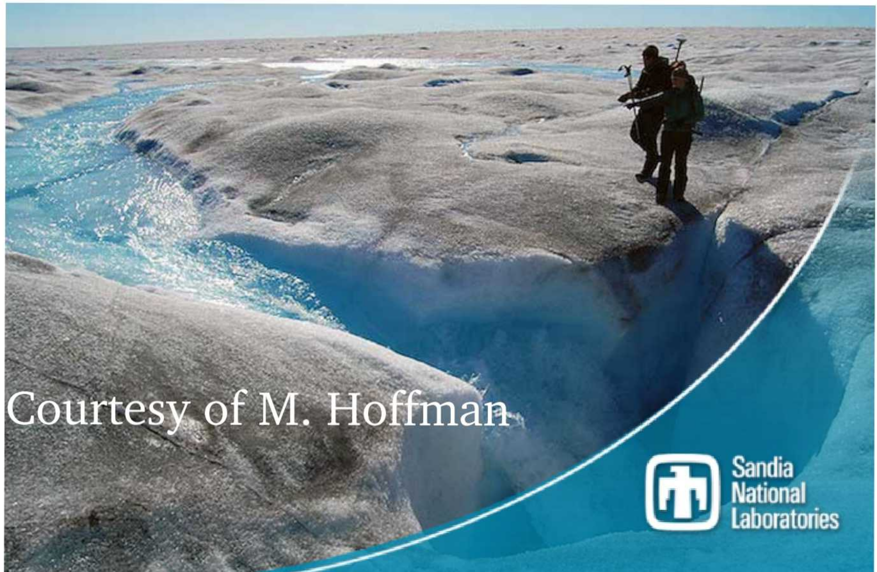
- ***basal friction β (spatially variable proxy for all basal processes)***
- ***ice stiffening (accounts for modeling error in rheology or imperfections in ice)***

Modeling Assumptions

- *ice flow described by* ***nonlinear Stokes equation***
- *ice close to* ***mechanical equilibrium***

Additional Assumption (for now)

- *given* ***temperature field***



Courtesy of M. Hoffman

Deterministic Inversion

PDE-constrained optimization problem: cost functional

Problem: find initial conditions such that the ice matches available observations.

Optimization problem:

find β and H that minimize the functional* $\mathcal{J}$

$$\begin{aligned} \mathcal{J}(\beta,\phi) &= \int_\Omega \frac{1}{\sigma_u^2}|\mathbf{u}-\mathbf{u}^{obs}|^2\,ds && \text{surface velocity mismatch} \\ &+ \int_\Omega \frac{1}{\sigma_\phi^2}|\phi-1|^2\,ds && \text{stiffening factor mismatch} \\ &+\mathcal{R}(\beta,\phi) && \text{regularization terms.} \end{aligned}$$

subject to ice sheet model equations (FO or Stokes)

$\boldsymbol{U}$: computed depth averaged velocity
ϕ: stiffening factor
β: basal sliding friction coefficient
$\mathcal{R}(\beta,\phi)$ regularization term

Sandia National Laboratories

**Perego, Price, Stadler,* Journal of Geophysical Research, 2014

Inverse Problem

Estimation of ice-sheet initial state

PDE-constraint optimization problem: gradient computation

$$\text{Find } (\beta, H) \text{ that minimize } \quad \mathcal{J}(\beta, H, \mathbf{u})$$
$$\text{subject to } \quad \mathcal{F}(\mathbf{u}, \beta, H) = 0 \quad \leftarrow \quad \text{flow model}$$

How to compute **total derivatives** of the functional w.r.t. the parameters?

Solve State System $\quad \mathcal{F}(\mathbf{u}, \beta, H) = 0$

Solve Adjoint System $\quad \langle \mathcal{F}^*_{\mathbf{u}}(\boldsymbol{\lambda}),\, \boldsymbol{\delta}_{\mathbf{u}} \rangle = \mathcal{J}_{\mathbf{u}}(\boldsymbol{\delta}), \quad \forall \boldsymbol{\delta}_{\mathbf{u}}$

Total derivatives $\quad \mathcal{G}(\delta_\beta, \delta_H) = \mathcal{J}_{(\beta,H)}(\delta_\beta, \delta_H) - \langle \boldsymbol{\lambda}, \mathcal{F}_{(\beta,H)}(\delta_\beta, \delta_H) \rangle$

Derivative w.r.t. β $\quad \mathcal{G}_1(\delta_\beta) = \alpha_\beta \int_\Omega \nabla\beta \cdot \nabla\delta_\beta \, ds - \int_\Omega \delta_\beta \, \mathbf{u} \cdot \boldsymbol{\lambda} \, ds$

Estimation of ice sheet initial state

Algorithm and Software from **MPAS-Albany Landice model (MALI)**

ALGORITHM	SOFTWARE TOOLS
Linear Finite Elements on test/hexas	Albany
Quasi-Newton optimization (L-BFGS)	ROL
Nonlinear solver (Newton method)	NOX
Krylov linear solvers/Prec	AztecOO/ML
Uncertainty Quantification	Dakota

Albany: C++ finite element library built on Trilinos to enable multiple capabilities:

- Jacobian/adjoints assembled using automatic differentiation (SACADO).
- nonlinear and parameter continuation solvers (NOX/LOCA)
- large scale PDE constrained optimization (Piro/ROL)
- Uncertainty Quantification (using Dakota)
- linear solver and preconditioners (Belos/AztecOO, ML/MeuLu/Ifpack)

Optimization algorithm:

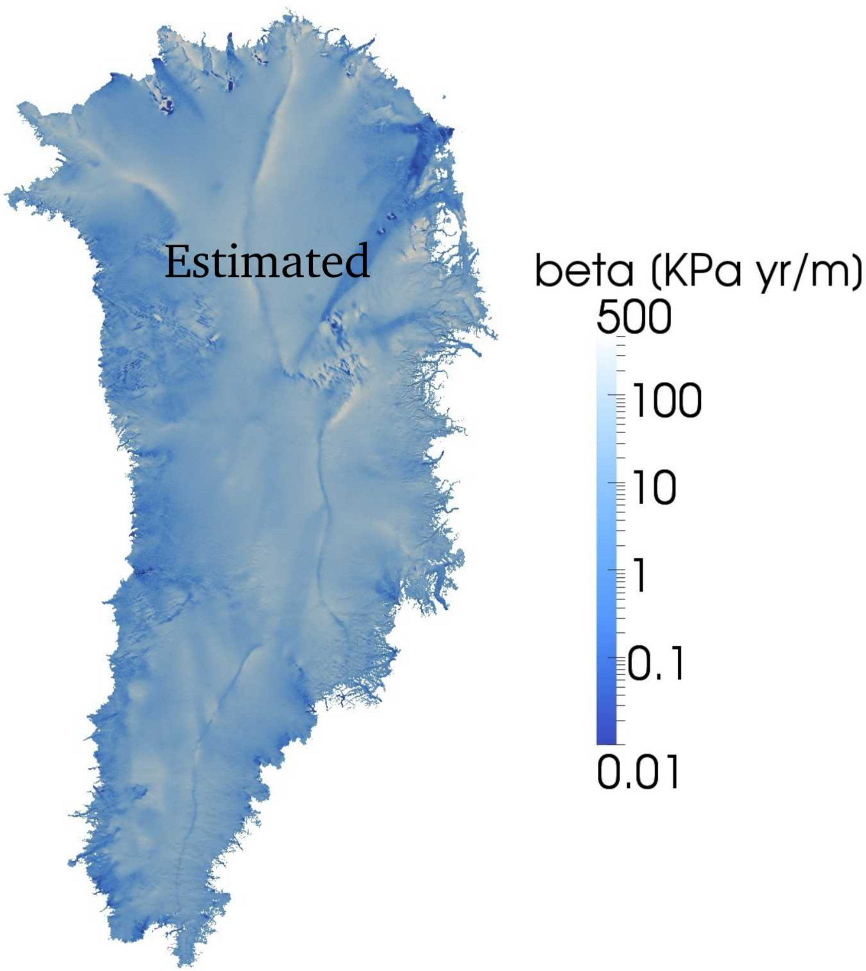
Reduce Gradient optimization, using L-BFGS.
Storage: 200, Linesearch: backtrack

Tuminaro, Perego, Tezaur, Salinger, Price, SISC, 2016.
Tezaur, Perego, Salinger, Tuminaro, Price, Hoffman, GMD, 2015
Perego, Price, Stadler, JGR, 2014

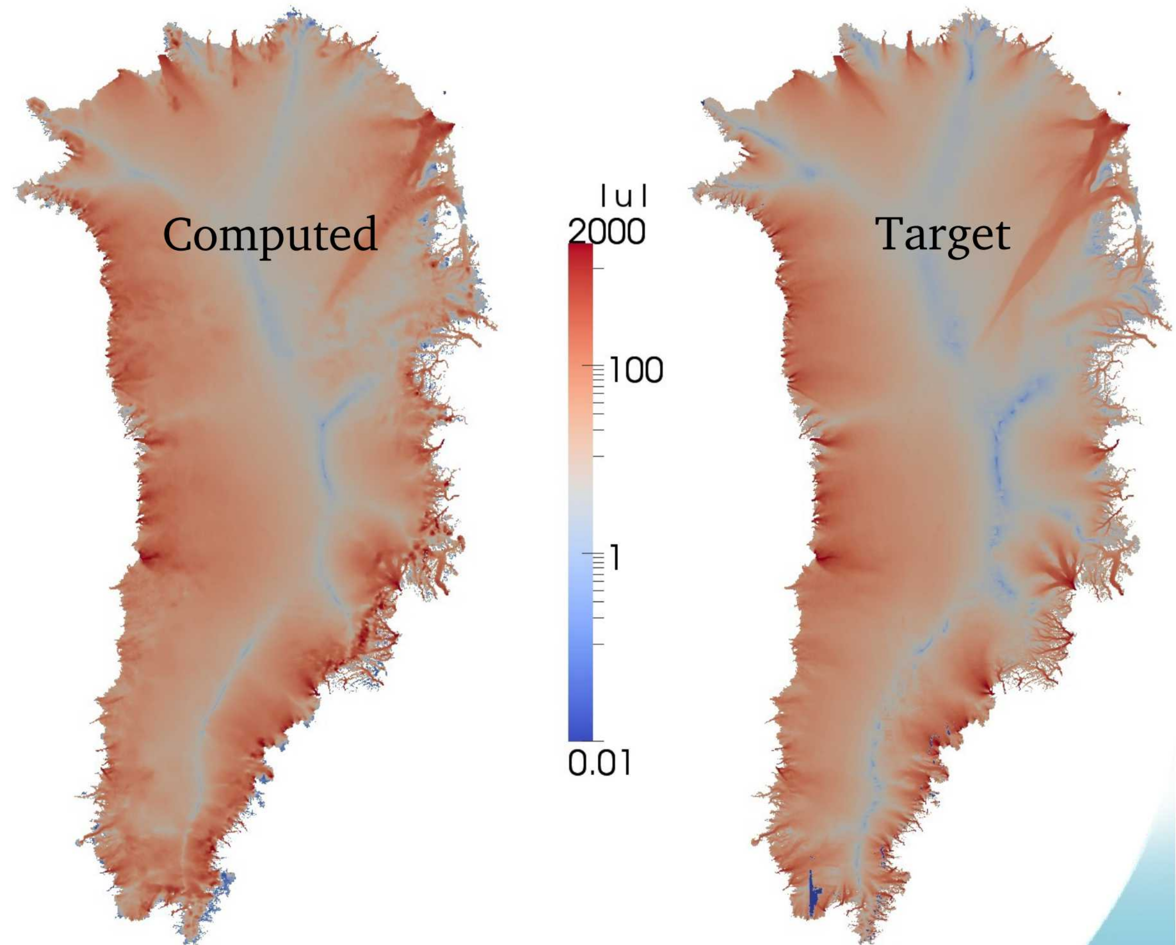
Greenland Inversion

velocity mismatch only, tuning basal friction

Inversion with 1.6M parameters



Basal friction coefficient (kPa yr/m)

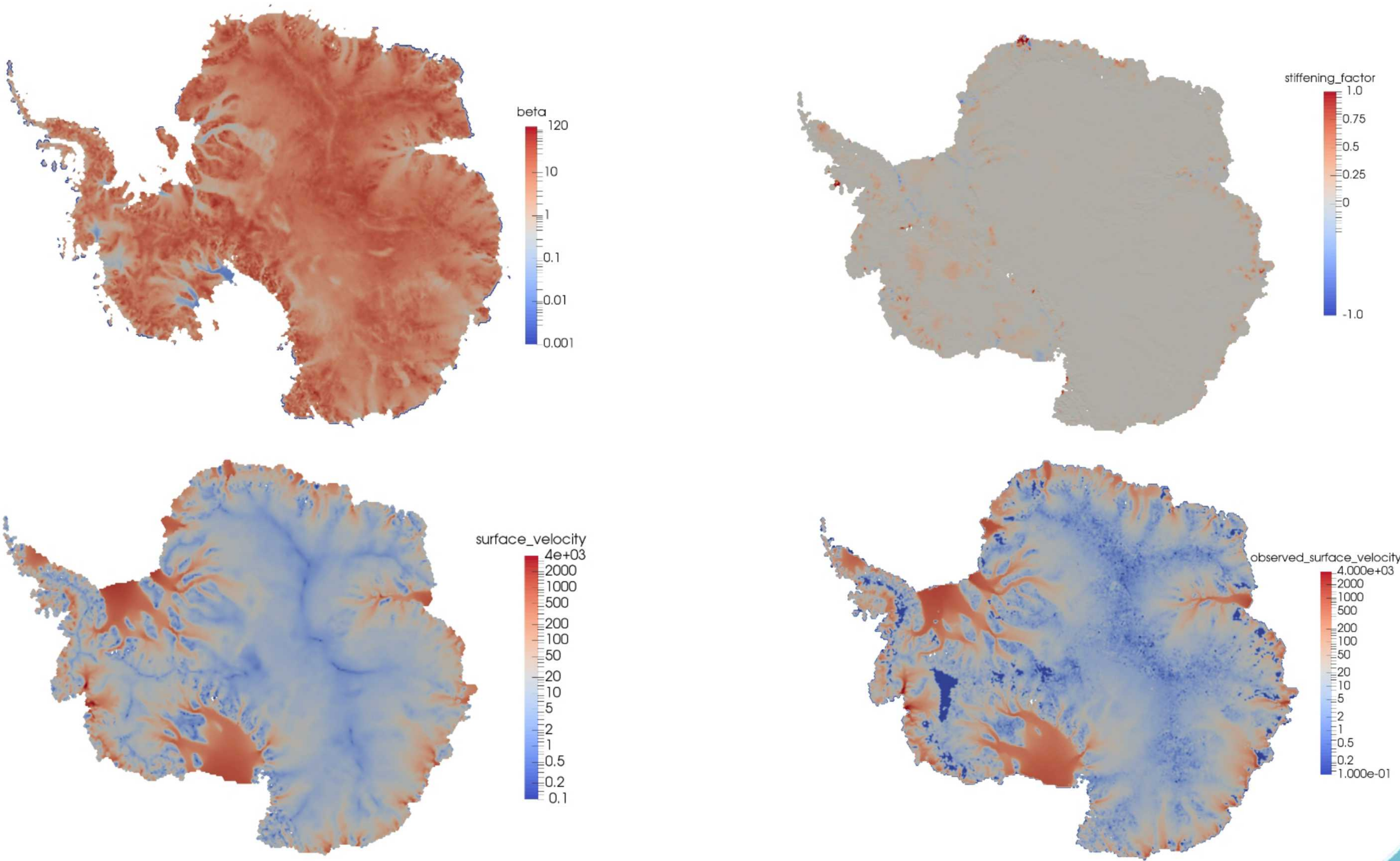


surface velocity magnitude (m/yr)

Geometry (Morlighem et al., Nature Geo., 2014)

Antarctica Inversion

velocity and stiffening mismatches, tuning basal friction and stiffening



Top left: estimated basal friction coefficient [kPa yr/m],
Top right: estimated softness parameter [adim],
Bottom left: computed surface velocity magnitude [m/yr],
Bottom right: observed surface velocity magnitude [m/yr]

Bayesian Inference

Bayes:

$$\boxed{\pi_{\text{post}}(x|d) = \frac{\pi_{\text{like}}(d|x)\,\pi_{\text{prior}}(x)}{\pi(d)} \propto \pi_{\text{like}}(d|x)\,\pi_{\text{prior}}(x)}$$

Relation model/data:

$$d = \mathcal{M}(x) + \eta, \qquad \eta \sim \mathcal{N}(0, \Sigma_{\text{noise}})$$

parameter-to-observation map

Likelihood p.d.f

$$\pi_{\text{like}}(d|x) = \frac{1}{(2\pi)^k|\Sigma_{\text{noise}}|}\exp\left(-\frac{1}{2}(\mathcal{M}(x)-d)^T\,\Sigma_{\text{noise}}^{-1}\,(\mathcal{M}(x)-d)\right)$$

Prior p.d.f

$$\pi_{\text{prior}}(x) \propto \exp\left(-\frac{1}{2}(x-x_{\text{prior}})^T\,\Sigma_{\text{prior}}^{-1}\,(x-x_{\text{prior}})\right)$$

Posterior p.d.f:

$$\pi_{\text{post}}(x|d) \propto \exp\left(-\frac{1}{2}(\mathcal{M}(x)-d)^T\,\Sigma_{\text{noise}}^{-1}\,(\mathcal{M}(x)-d) - \frac{1}{2}(x-x_{\text{prior}})^T\,\Sigma_{\text{prior}}^{-1}\,(x-x_{\text{prior}})\right)$$

Bayesian Inference

Posterior p.d.f:

$$\pi_{\text{post}}(x|d) \propto \exp\left(-\frac{1}{2}(\mathcal{M}(x)-d)^T\,\Sigma_{\text{noise}}^{-1}\,(\mathcal{M}(x)-d)) - \frac{1}{2}(x-x_{\text{prior}})^T\,\Sigma_{\text{prior}}^{-1}\,(x-x_{\text{prior}})\right)$$

$\mathcal{M}(x)$ is nonlinear, so posterior is not Gaussian

Computing the posterior is *unfeasible* for the high dimensionality of the parameter space and because the forward model is very expensive

Gaussian approximation*: we perform a quadratic approximation, at the mean of the parameter x, of the negative log of the posterior p.d.f

The mean of the distribution, or maximum a posterior point (MAP) is given by

$$x_{\text{MAP}} = \arg\min_x \left(\frac{1}{2}(\mathcal{M}(x)-d)^T\,\Sigma_{\text{noise}}^{-1}\,(\mathcal{M}(x)-d)) + \frac{1}{2}(x-x_{\text{prior}})^T\,\Sigma_{\text{prior}}^{-1}\,(x-x_{\text{prior})}\right)$$

misfit regularization

*T. Isaac, N. Petra, G. Stadler, O. Ghattas, JCP, 2015

Bayesian Inference

regularization* (prior)

$$\boxed{\mathcal{R}(\beta) = \frac{\alpha}{2}\int_\Omega (-\gamma\Delta\beta + \delta\beta)^2, \qquad \mu\beta + \frac{\partial\beta}{\partial\mathbf{n}} = 0}$$

Laplacian squared regularization with Robin boundary conditions

$$\mathcal{R}(\beta) = \frac{\alpha}{2}\int_\Omega s^2 \quad \longleftarrow \quad \int_\Omega s\varphi = \int_\Omega (\gamma\nabla\beta\cdot\nabla\varphi + \delta\beta\varphi) + \int_{\partial\Omega} \alpha\,\beta\varphi$$

L^2 projection

$$M_{ij} = \int_\Omega \varphi_i\varphi_j \qquad K_{ij} = \int_\Omega (\gamma\nabla\varphi_i\cdot\nabla\varphi_j + \delta\varphi_i\varphi_j) + \int_{\partial\Omega}\mu\varphi_i\varphi_j$$

$$\Sigma_{\text{prior}} := K^{-1}MK^{-1} = LL^T, \quad L := K^{-1}M^{\frac{1}{2}}$$

$$\pi_{\text{prior}}(x) \propto \exp\left(-\frac{1}{2}(x - x_{\text{prior}})^T\,\Sigma_{\text{prior}}^{-1}\,(x - x_{\text{prior}})\right)$$

*T. Isaac, N. Petra, G. Stadler, O. Ghattas, JCP, 2015

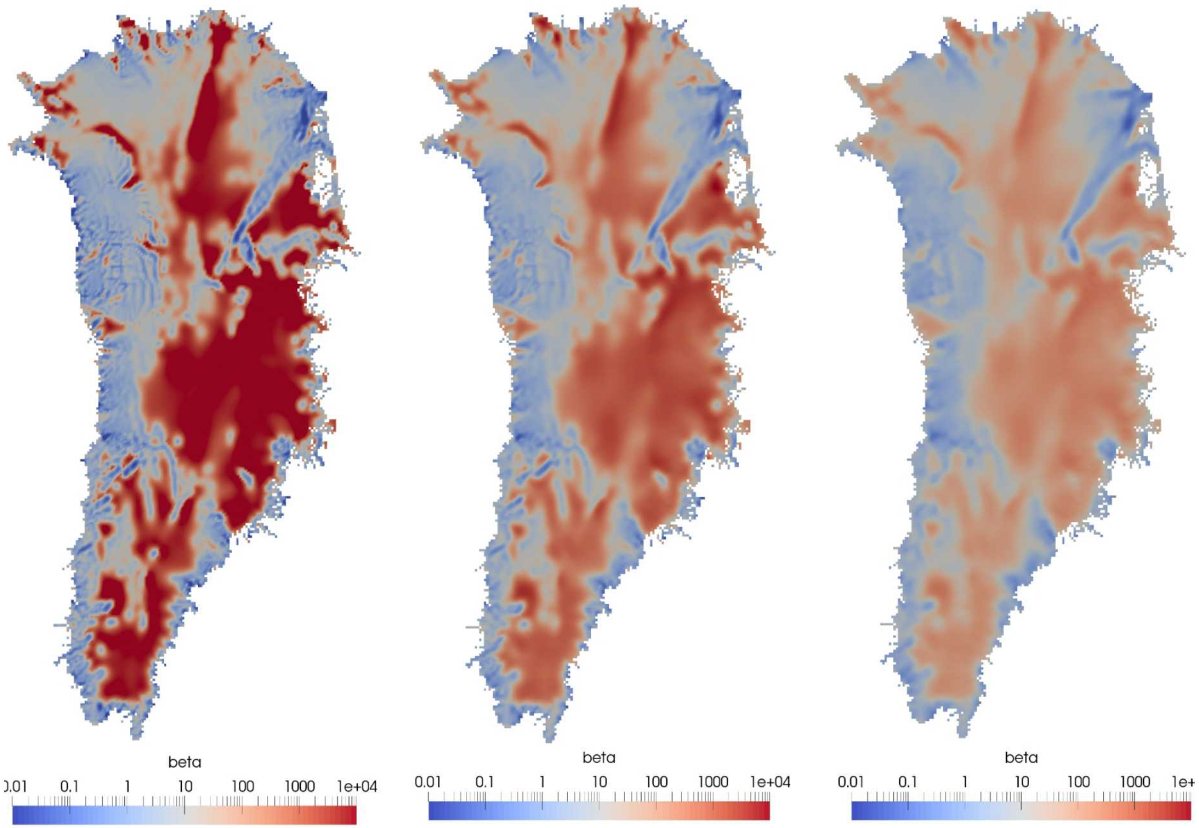
Sandia National Laboratories

Bayesian Inference

regularization, deterministic inversion

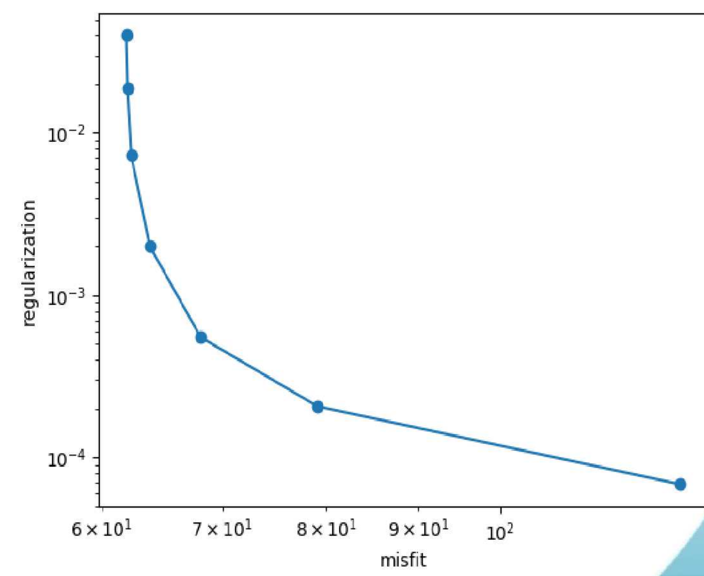
$$\mathcal{R}(\beta) = \frac{\alpha}{2}\int_\Omega (-\gamma\Delta\beta + \delta\beta)^2, \qquad \mu\beta + \frac{\partial\beta}{\partial\mathbf{n}} = 0$$

Laplacian squared regularization with Robin boundary conditions



Basal friction coefficient (MAP) [kPa yr /m]
for increasing (from left to right) regularization

L-curve



Bayesian Inference

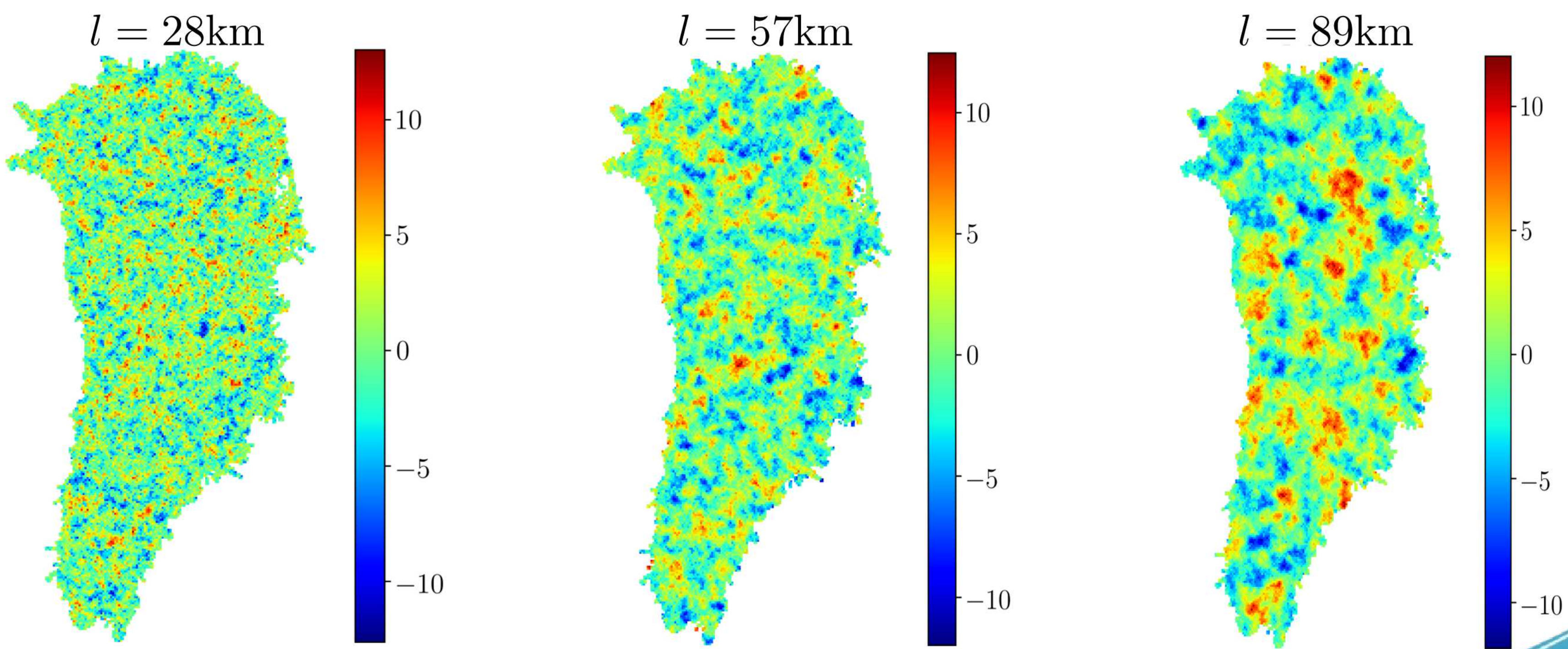
regularization, correlation lenght

$$\boxed{\mathcal{R}(\beta) = \frac{\alpha}{2}\int_\Omega (-\gamma\Delta\beta + \delta\beta)^2, \qquad \mu\beta + \frac{\partial\beta}{\partial\mathbf{n}} = 0}$$

Laplacian squared regularization with Robin boundary conditions

Correlation length $l \propto \sqrt{\dfrac{\gamma}{\delta}}$

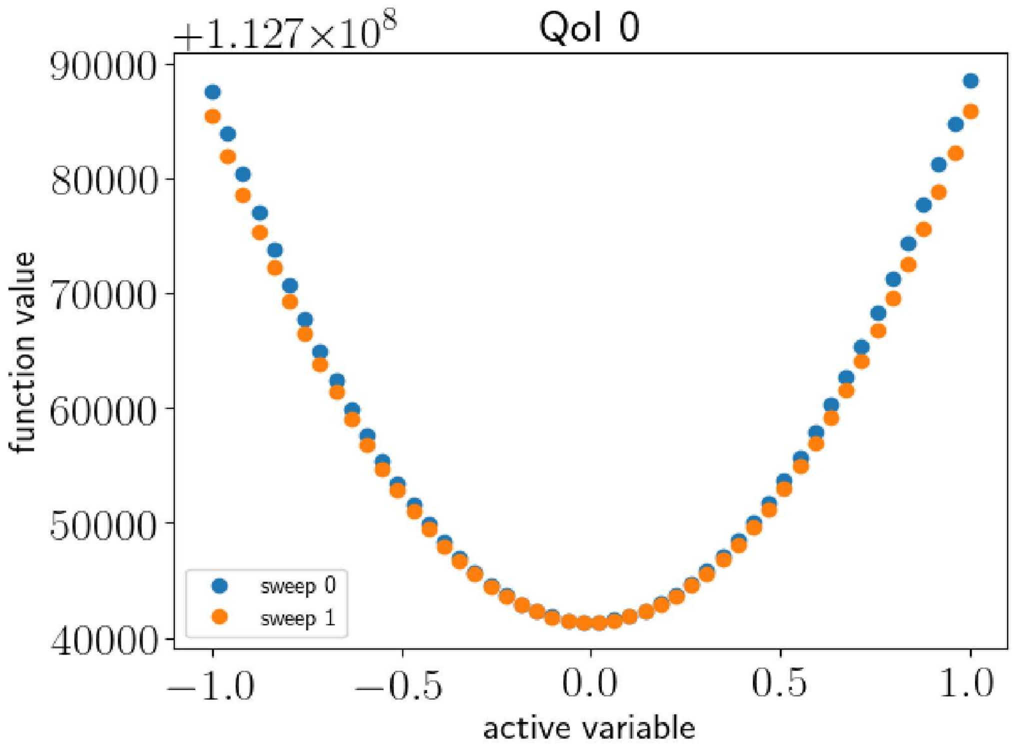
$$\Sigma_{\text{prior}} = LL^T$$
$$\beta = \beta_0 + Ln, \qquad n \sim \mathcal{N}(0,1)$$

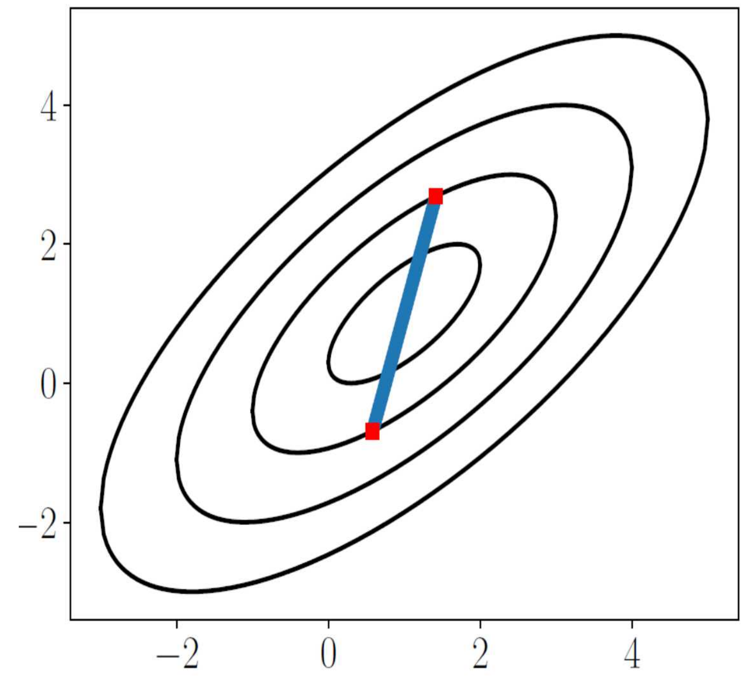


Bayesian Inference

misfit functional

Checking smoothness of misfit functional.
Evaluations along two orthogonal random directions





Sandia National Laboratories

Bayesian Inference

model reduction*

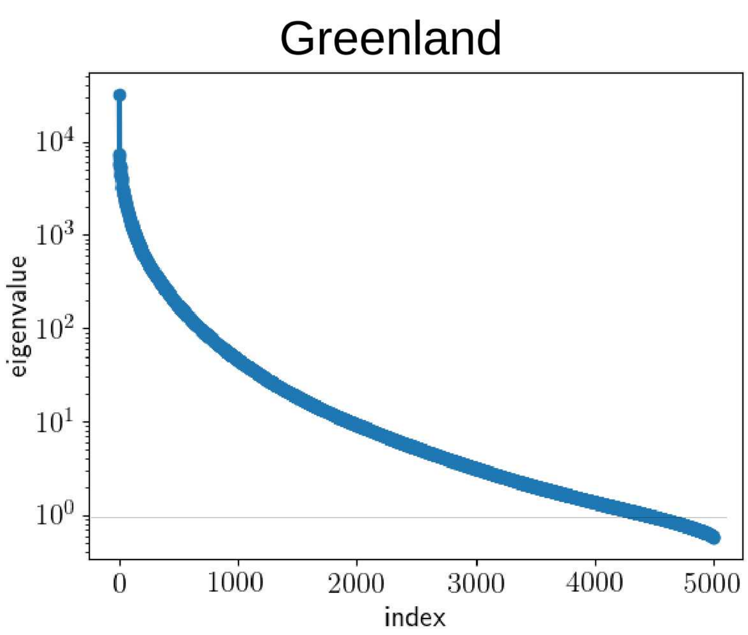
$$A := \nabla \mathcal{M}|_{x=x_{\mathrm{MAP}}}\,, \qquad d \approx \mathcal{M}(x_{\mathrm{MAP}}) + A(x - x_{\mathrm{MAP}}) + \eta$$

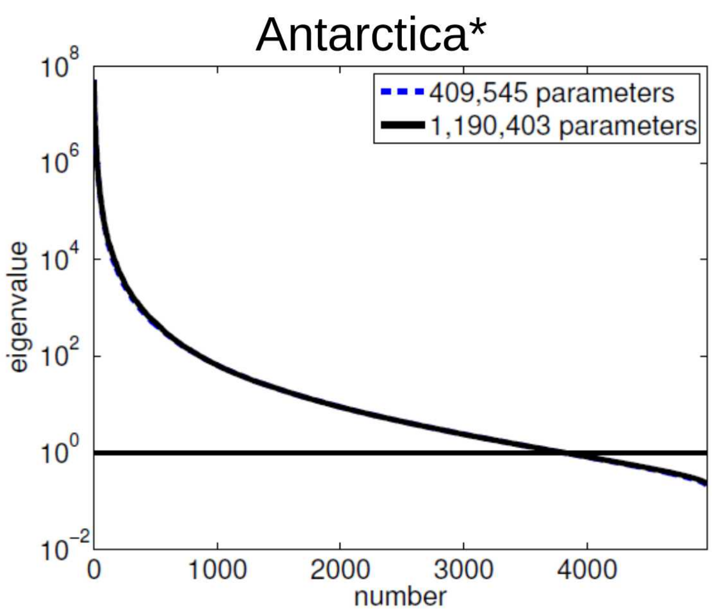
$$\Sigma_{\text{post}} = \left(H + \Sigma_{\text{prior}}^{-1}\right)^{-1} = \left(\Sigma_{\text{prior}} H + I\right)^{-1} \Sigma_{\text{prior}} \qquad\qquad H = A^T \Sigma_{\text{noise}}^{-1} A$$

$$\Sigma_{\text{prior}} H \approx W_r \Lambda_r V_r^T \quad \text{(low rank approximation - randomized SVD)}$$

$$\Sigma_{\text{post}} = \Sigma_{\text{prior}} - W_r \Lambda_r W_r^T + \mathcal{O}\left(\sum_{i=r+1}^{N} \frac{\lambda_i}{1+\lambda_i}\right)$$

(Sherman-Morrison-Woodbury formula)





*T. Isaac, N. Petra, G. Stadler, O. Ghattas, JCP, 2015

Bayesian Inference

model reduction

$$A := \nabla \mathcal{M}|_{x=x_{\text{MAP}}}, \qquad d \approx \mathcal{M}(x_{\text{MAP}}) + A(x - x_{\text{MAP}}) + \eta$$

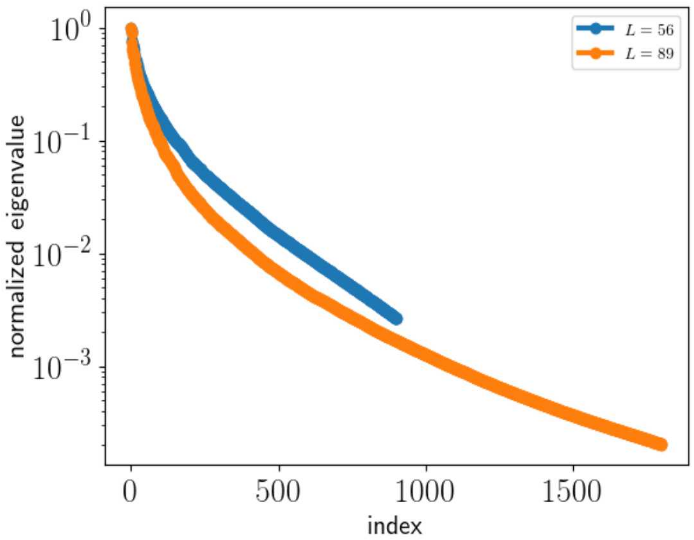
$$\Sigma_{\text{post}} = \left(H + \Sigma_{\text{prior}}^{-1}\right)^{-1} = \left(\Sigma_{\text{prior}} H + I\right)^{-1} \Sigma_{\text{prior}} \qquad H = A^T \Sigma_{\text{noise}}^{-1} A$$

$$\Sigma_{\text{prior}} H \approx W_r \Lambda_r V_r^T \quad \text{(low rank approximation - randomized SVD)}$$

$$\Sigma_{\text{post}} = \Sigma_{\text{prior}} - W_r \Lambda_r W_r^T + \mathcal{O}\left(\sum_{i=r+1}^{N} \frac{\lambda_i}{1+\lambda_i}\right)$$

(Sherman-Morrison-Woodbury formula)

Greenland



Eigenvalue decay,
dependence on correlation length

Bayesian Inference

model reduction

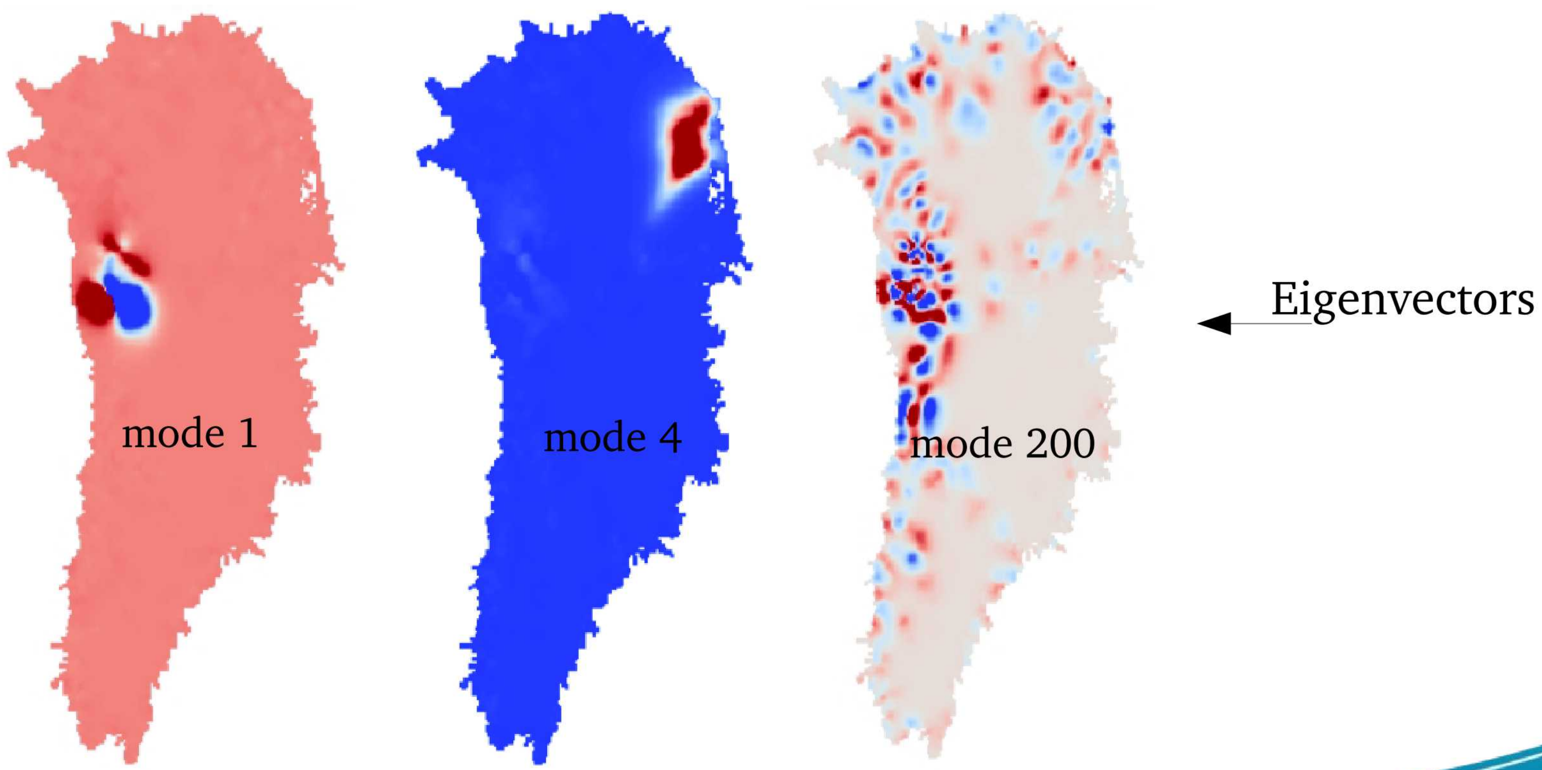
$$A := \nabla \mathcal{M}|_{x=x_{\text{MAP}}}, \qquad d \approx \mathcal{M}(x_{\text{MAP}}) + A(x - x_{\text{MAP}}) + \eta$$

$$\Sigma_{\text{post}} = \left(H + \Sigma_{\text{prior}}^{-1}\right)^{-1} = (\Sigma_{\text{prior}} H + I)^{-1}\,\Sigma_{\text{prior}} \qquad H = A^T \Sigma_{\text{noise}}^{-1} A$$

$\Sigma_{\text{prior}} H \approx W_r \Lambda_r V_r^T$ (low rank approximation - randomized SVD)

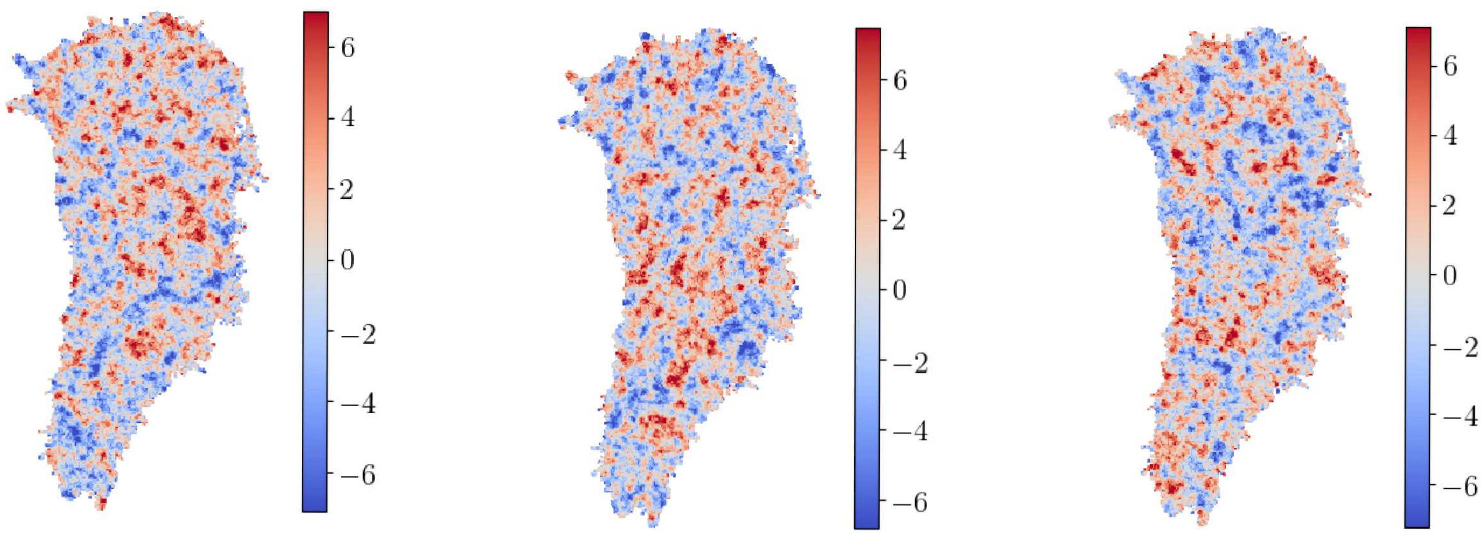
$$\Sigma_{\text{post}} = \Sigma_{\text{prior}} - W_r \Lambda_r W_r^T + \mathcal{O}\left(\sum_{i=r+1}^{N} \frac{\lambda_i}{1+\lambda_i}\right)$$

(Sherman-Morrison-Woodbury formula)

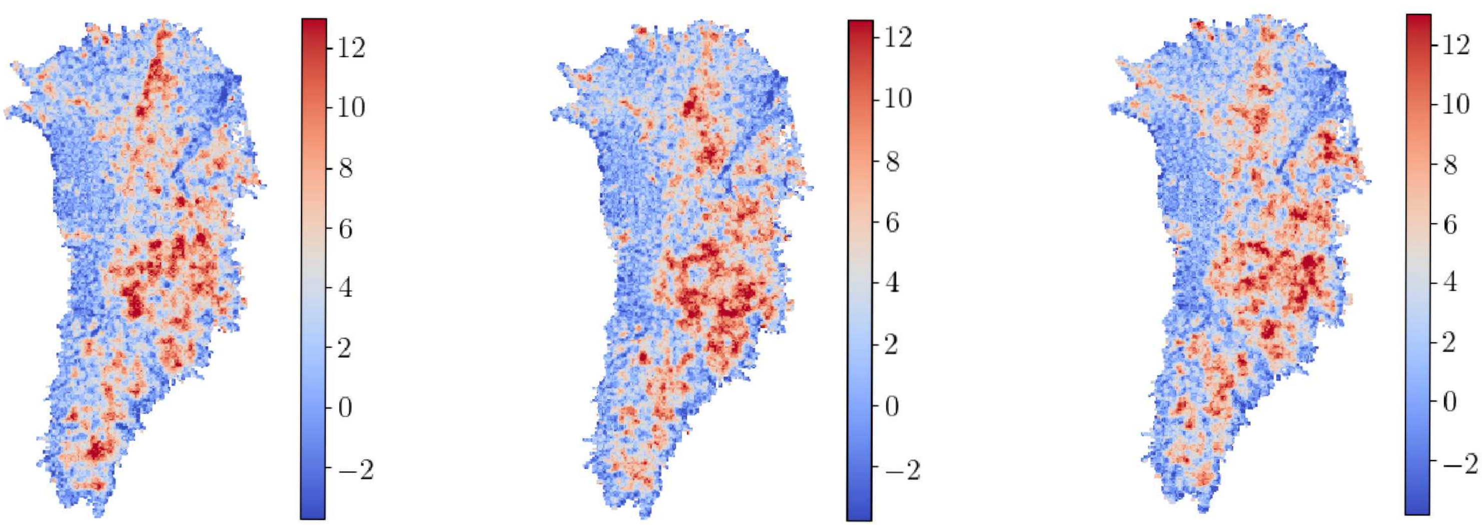


Bayesian Inference

prior/posterior sampling



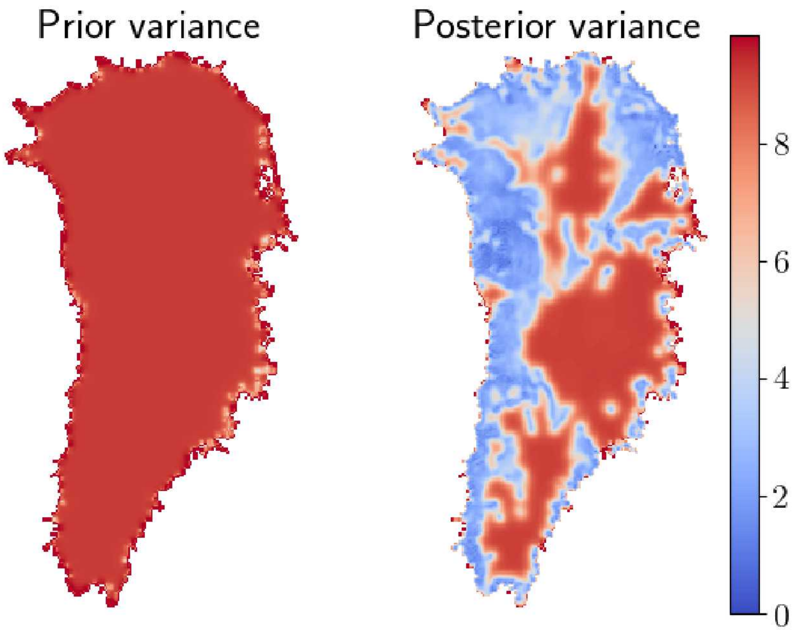
Samples from prior distribution



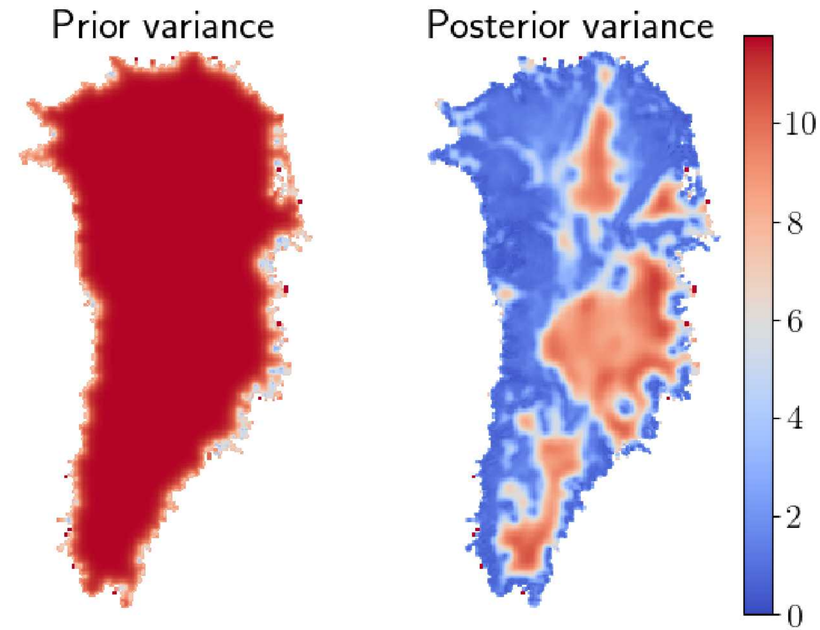
samples from posterior distribution

Bayesian Inference

prior/posterior point-wise variance



$l = 57\text{km}$



$l = 89\text{km}$

The use of data has drastically reduced the variance.
Higher correlation length reduces variance

Conclusion

Presented a framework for Bayesian inference based on a Gaussian approximation of the posterior, applied to Greenland ice sheet problem

Posterior distribution can be efficiently sampled using a low-rank approximation of the covariance matrix, although the effective rank of the covariance matrix is high O(1000)

TODO: Check whether the Gaussian approximation of the posterior is a good one

TODO: Consider high resolution grids and investigate how the effective rank depends on the mesh resolution

Next Step: Forward Propagation

Thanks!