SIAG CSE - Best Paper Prize 2018

Tobin Isaac, Noemi Petra, Georg Stadler, and Omar Ghattas

Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large–scale problems, with application to flow of the Antarctic ice sheet

Journal of Computational Physics, published in 2015

"… *a cornerstone paper in CS&E that demonstrates a scalable algorithmic framework for geophysical model inversion and uncertainty quantification on extreme-scale ice-sheet modeling exploiting supercomputing architectures*."

Propagating Uncertainty from Data to Prediction with a Model of the Antarctic Ice Sheet

Toby Isaac[1] · Noémi Petra[2] · Georg Stadler[3] · Omar Ghattas[4]

[1]School of Computational Science and Engineering, Georgia Tech
[2]School of Natural Sciences, UC Merced
[3]Courant Institute, NYU
[4]Oden Institute, UT Austin

February 27, 2019 · SIAM CSE · Spokane

Great Coauthors and Mentors!

Motivation

Solving the Forward Problem

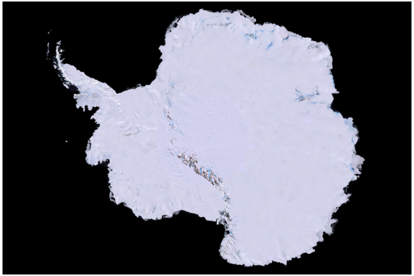
Inversion for Parameter Fields

Our paper does not make actionable predictions or projections.

It is a methodology paper.

What Quantitative Changes Should We Expect?

Antarctica



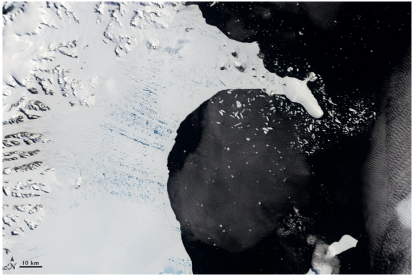
visibleearth.nasa.gov

- **Quantities of interest (QoIs): q** drive decisions, e.g.
  - Air temperature
  - Ocean salinity
  - Sea level
  - Albedo
- Measurements of **states: w**
- Unobtainable
  - Future
  - Infeasible
- Regime changes invalidate trends

$$q_{\text{obs}} \not\to q_{\text{pred}}$$

What Quantitative Changes Should We Expect?

Larsen B Ice Shelf, 2002



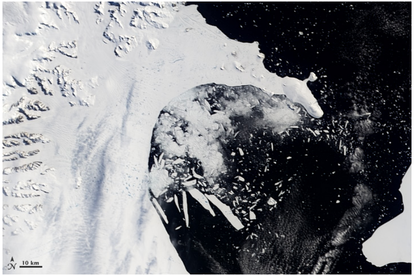
earthobservatory.nasa.gov

- **Quantities of interest (QoIs): q** drive decisions, e.g.
  - Air temperature
  - Ocean salinity
  - Sea level
  - Albedo
- Measurements of **states: w**
- Unobtainable
  - Future
  - Infeasible
- Regime changes invalidate trends

$$q_{\text{obs}} \not\to q_{\text{pred}}$$

What Quantitative Changes Should We Expect?

Larsen B Ice Shelf, 2002



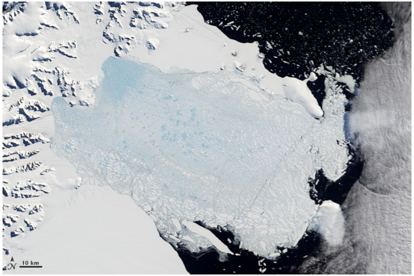
earthobservatory.nasa.gov

- **Quantities of interest (QoIs):** q drive decisions, e.g.
  - Air temperature
  - Ocean salinity
  - Sea level
  - Albedo
- Measurements of **states:** w
- Unobtainable
  - Future
  - Infeasible
- Regime changes invalidate trends

$$q_{\text{obs}} \not\to q_{\text{pred}}$$

What Quantitative Changes Should We Expect?

Larsen B Ice Shelf, 2002



earthobservatory.nasa.gov

- **Quantities of interest (QoIs): q** drive decisions, e.g.
  - Air temperature
  - Ocean salinity
  - Sea level
  - Albedo
- Measurements of **states: w**
- Unobtainable
  - Future
  - Infeasible
- Regime changes invalidate trends

$$q_{\text{obs}} \not\to q_{\text{pred}}$$

Models Inform Us When Data Can't

- PDEs are often the best **models** of states

  Stokes: Balance of Momentum and Mass

$$-\nabla \cdot \boldsymbol{\sigma} = \rho \mathbf{g}, \quad [\boldsymbol{\sigma} = \mu(T, \mathbf{u})(\nabla \mathbf{u} + \nabla \mathbf{u}^T) - \mathbf{I}p]$$
$$\nabla \cdot \mathbf{u} = 0, \qquad \text{+b.c.s}$$

- Symbolized as

$$A(\mathbf{w}) = \mathbf{0}$$

How do we choose A so $\mathbf{w}$ matches reality?

Models Inform Us When Data Can't

- PDEs are often the best **models** of states

Stokes: Balance of Momentum and Mass (Nonlinear)

$$-\nabla \cdot \boldsymbol{\sigma} = \rho \boldsymbol{g}, \quad [\boldsymbol{\sigma} = \mu(T, \boldsymbol{u})(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) - \boldsymbol{I}p]$$

$$\nabla \cdot \boldsymbol{u} = 0, \qquad \text{+b.c.s}$$

- Symbolized as

$$A(\boldsymbol{w}) = \boldsymbol{0}$$

How do we choose A so $\boldsymbol{w}$ matches reality?

Models Inform Us When Data Can't

- PDEs are often the best **models** of states

  Stokes: Balance of Momentum and Mass

$$-\nabla \cdot \boldsymbol{\sigma} = \rho \boldsymbol{g}, \quad [\boldsymbol{\sigma} = \mu(T, \boldsymbol{u})(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) - \boldsymbol{I}p]$$

$$\nabla \cdot \boldsymbol{u} = 0, \qquad \text{+b.c.s}$$

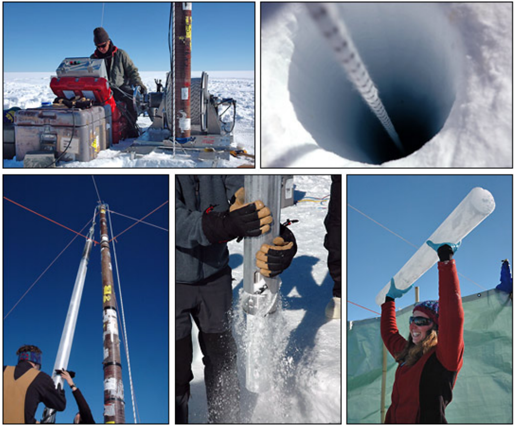
- Symbolized as

$$A(\boldsymbol{w}) = \mathbf{0}$$

How do we choose A so $\boldsymbol{w}$ matches reality?

Scientists Gather Lots of Data

Data



earthobservatory.nasa.gov

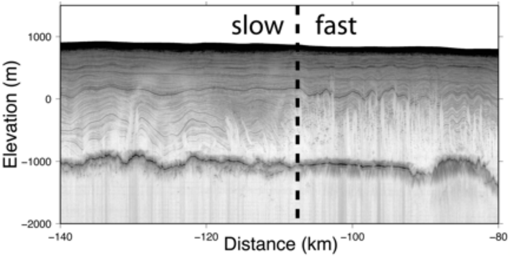
- **Observations/Data: $\boldsymbol{d}_{\text{obs}}$**, e.g.
  - Ice cores / boreholes
  - Radar / stratigraphy
  - GRACE gravity field measurements
  - Interferometric synthetic aperture radar (InSAR) / lidar
- Data is also a measurement of the state $\boldsymbol{w}$:

$$\boldsymbol{d}(\boldsymbol{w}) \approx \boldsymbol{d}_{\text{obs}}$$

suggests the model is valid.

Scientists Gather Lots of Data

Data



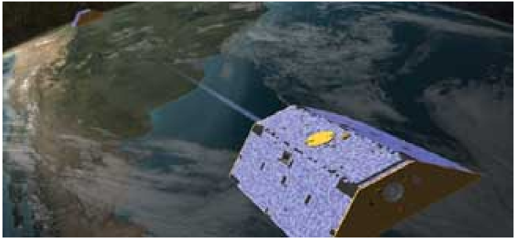
- **Observations/Data: $\boldsymbol{d}_{\text{obs}}$**, e.g.
  - Ice cores / boreholes
  - Radar / stratigraphy
  - GRACE gravity field measurements
  - Interferometric synthetic aperture radar (InSAR) / lidar
- Data is also a measurement of the state $\boldsymbol{w}$:

$$\boldsymbol{d}(\boldsymbol{w}) \approx \boldsymbol{d}_{\text{obs}}$$

suggests the model is valid.

Scientists Gather Lots of Data

Data



(UT Center for Space Research)

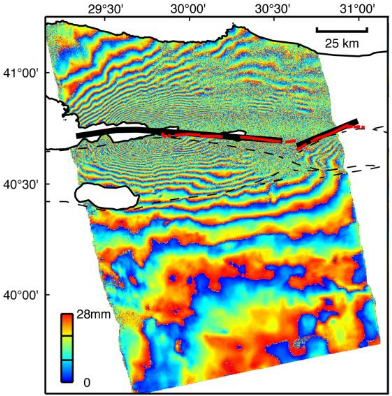
- **Observations/Data: $\boldsymbol{d}_{\text{obs}}$**, e.g.
  - Ice cores / boreholes
  - Radar / stratigraphy
  - GRACE gravity field measurements
  - Interferometric synthetic aperture radar (InSAR) / lidar
- Data is also a measurement of the state $\boldsymbol{w}$:

$$\boldsymbol{d}(\boldsymbol{w}) \approx \boldsymbol{d}_{\text{obs}}$$

suggests the model is valid.

Scientists Gather Lots of Data

Data



(NASA/JPL-Caltech)

- **Observations/Data: $\boldsymbol{d}_{\text{obs}}$**, e.g.
  - Ice cores / boreholes
  - Radar / stratigraphy
  - GRACE gravity field measurements
  - Interferometric synthetic aperture radar (InSAR) / lidar
- Data is also a measurement of the state $\boldsymbol{w}$:

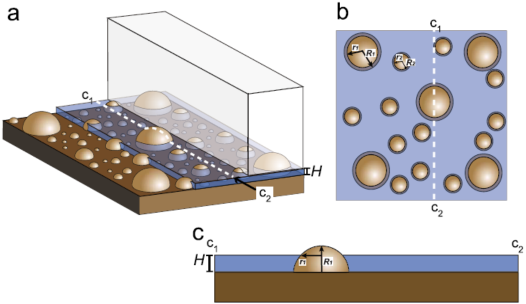
$$\boldsymbol{d}(\boldsymbol{w}) \approx \boldsymbol{d}_{\text{obs}}$$

suggests the model is valid.

- Our dataset $\boldsymbol{d}_{\mathrm{obs}} \in \mathbb{R}^{N_d}$ is the MEaSUREs surface velocity data for Antarctica (Rignot, Mouginot, and Scheuchl 2011).
- $\boldsymbol{d} : \mathcal{X} \longrightarrow \mathbb{R}^{N_d}$ traces the velocity on the surface.

Parameter Fields

Subglacial Systems



(Creyts and Schoof 2009, Figure 3)

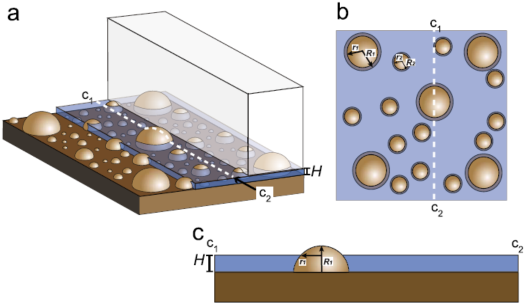
- Uncertain **parameters** $\boldsymbol{m}$

$$A(\boldsymbol{w}; \boldsymbol{m}) = \mathbf{0}$$
$$\boldsymbol{q}_{\text{pred}} = \boldsymbol{q}(\boldsymbol{w}; \boldsymbol{m})$$

- Often parameter fields, e.g.
  - Initial conditions
  - Material properties
  - Boundary conditions
- Number of parameters in computation increases with model resolution

Parameter Fields

Subglacial Systems

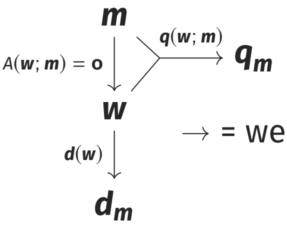


(Creyts and Schoof 2009, Figure 3)

$$T_{\parallel}(\boldsymbol{\sigma} \boldsymbol{n} + \beta(\boldsymbol{m}(x))\boldsymbol{u})|_{\Gamma_{\text{base}}} = \boldsymbol{0}$$

Inversion for Prediction

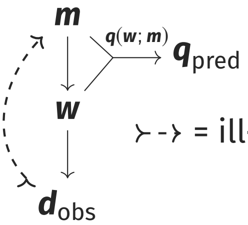
Summary



$\rightarrow$ = well-posed

- State $\boldsymbol{w}$
- QoIs $\boldsymbol{q}$
- Meas. $\boldsymbol{d}$
- Obs. $\boldsymbol{d}_{\text{obs}}$
- Param. field $\boldsymbol{m}$
- Model A ("forward")

Observations to Predictions

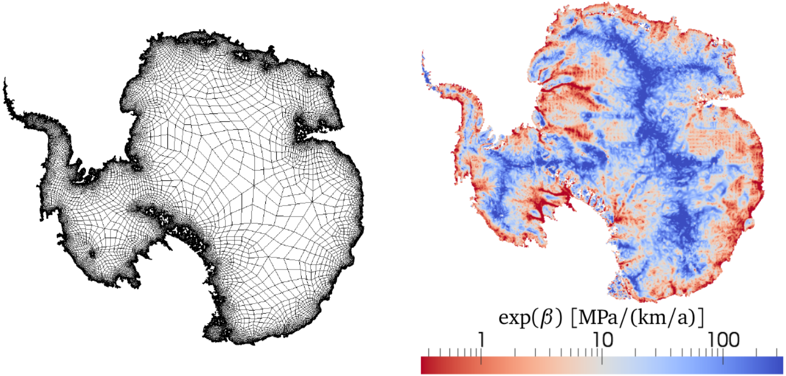


$\rightarrowtail$ - - $\rightarrow$ = ill-posed

What are the sources of error and uncertainty?

- Noisy $\boldsymbol{d}$
- Ill-posedness of inversion
- Model/discretization error

Modeling the Ice Sheet



A finite element model built on the p4est AMR library (Burstedde, Wilcox, and Ghattas 2011).

Solving the Forward Problem

Armijo-**Newton**-Eisenstat-Walker-**Krylov**-Saad-Schur Method

- **Armijo-Newton**: quadratic convergence near solution, globalized for stability
- **Eisenstat-Walker**: Adaptive tolerance for *inexact* linear solver based on nonlinear convergence history
- **Krylov-Saad**: FGMRES(k) solver allows variable preconditioning
- **Schur**: Use block upper-triangular preconditioner for Stokes operator:

$$A = \begin{pmatrix} F & B^* \\ B & 0 \end{pmatrix}, \quad P = \begin{pmatrix} \tilde{F} & B^* \\ 0 & \tilde{S} \end{pmatrix},$$

Solving the Forward Problem

- $\tilde{F}$: smoothed-aggregation algebraic multigrid (PETSc GAMG with custom plugin aggregator for high-anistropy)
- $\tilde{S} = -\mu^{-1}\hat{M}$: lumped mass matrix.

Solving the Forward Problem

All of these choices seek to achieve optimal performance: time to solution $\sim N/P$.

	#dof	#cores	#Newton	#Krylov	solve time (s) / eff (%)	setup time (s) / eff (%)	#Krylov (Poisson)
P1	38M	128	8	149	504.8 / 100	493.5 / 100	12
		256	8	153	259.6 / 97	260.4 / 95	12
		512	8	157	134.3 / 94	156.0 / 80	12
		1024	8	147	70.1 / 90	97.2 / 63	12
P2	270M	1024	9	240	796.6 / 100	735.0 / 100	12
		2048	9	245	414.3 / 96	424.6 / 87	12
		8192	9	243	130.7 / 76	229.0 / 40	13
P3	2.1B	16,384	13	314	771.5 / 100	1424.5 / $\star$	15
		65,536	13	367	504.2 / 38	1697.1 / $\star$	15
		131,072	11	340	232.9 / 42	2033.1 / $\star$	16

The Bayesian Inference Framework

1. Adopt a likelihood of $\boldsymbol{d}_{\text{obs}}$ given $\boldsymbol{m}$,

$$\pi_{\text{like}}(\boldsymbol{d}_{\text{obs}} | \boldsymbol{m})$$

- Simplest form

$$\boldsymbol{d}_{\text{obs}} \propto \mathcal{N}(\boldsymbol{d}_{\boldsymbol{m}}, C_{\text{obs}})$$

- C_{obs} characterizes noise and model error

2. Adopt a prior distribution of the parameter field

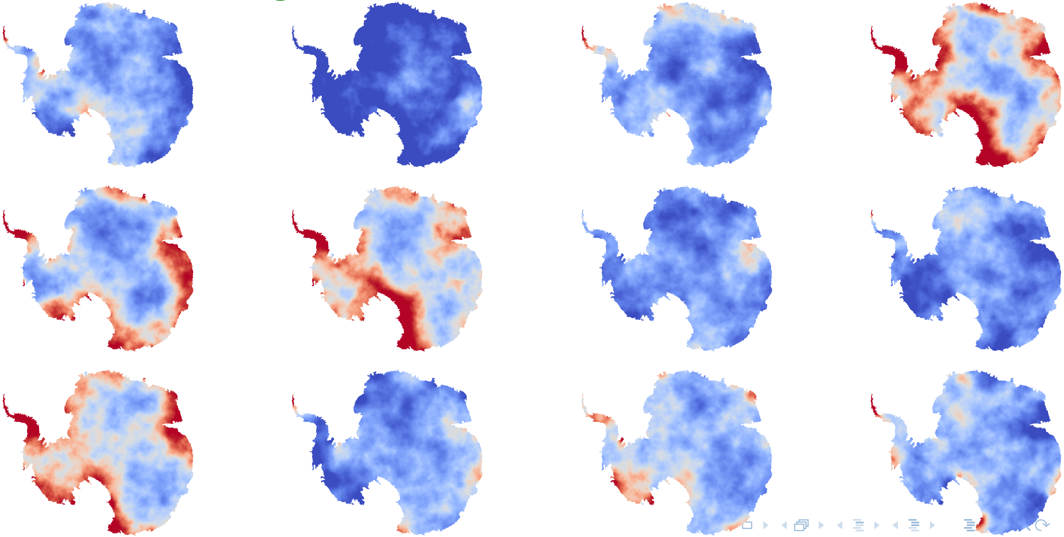
$$\pi_{\text{prior}}(\boldsymbol{m})$$

- First principles
- "Expert knowledge"
- Expedience

$$\boldsymbol{m} \propto \mathcal{N}(\boldsymbol{m}_{\text{prior}}, C_{\text{prior}})$$

C_{prior} Samples

Antarctic Sliding Parameter



The Bayesian Inference Framework

Bayes' Law

The posterior pdf of the parameter given the observations is

$$\pi_{\text{post}}(\boldsymbol{m}|\boldsymbol{d}_{\text{obs}}) = \frac{\pi_{\text{prior}}(\boldsymbol{m})\pi_{\text{like}}(\boldsymbol{d}_{\text{obs}}|\,\boldsymbol{m})}{\int_{\mathcal{M}} \pi_{\text{prior}}(\boldsymbol{m})\pi_{\text{like}}(\boldsymbol{d}_{\text{obs}}|\,\boldsymbol{m})\,d\boldsymbol{m}}.$$

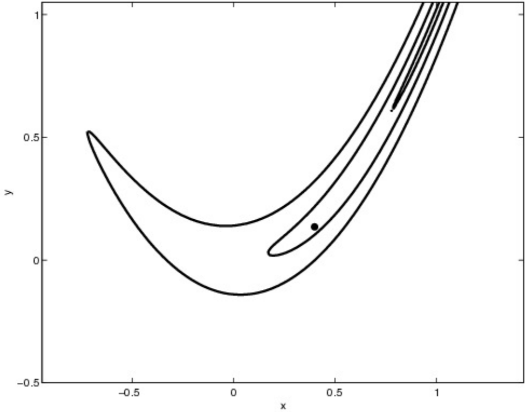
- E.g.,

$$\mathbb{E}[\boldsymbol{q}_{\text{pred}}|\boldsymbol{d}_{\text{obs}}] = \frac{\int_{\mathcal{M}} \boldsymbol{q}_{\boldsymbol{m}}\pi_{\text{prior}}(\boldsymbol{m})\pi_{\text{like}}(\boldsymbol{d}_{\text{obs}}|\,\boldsymbol{m})\,d\boldsymbol{m}}{\int_{\mathcal{M}} \pi_{\text{prior}}(\boldsymbol{m})\pi_{\text{like}}(\boldsymbol{d}_{\text{obs}}|\,\boldsymbol{m})\,d\boldsymbol{m}}$$

- Expectations require integration over $\mathcal{M}$ (high-dimensional) …
- …w.r.t. $\pi_{\text{post}}(\boldsymbol{m}|\boldsymbol{d}_{\text{obs}})$ (implicitly-defined, no direct sampling).

Markov Chain Methods

2D Rosenbrock Distribution



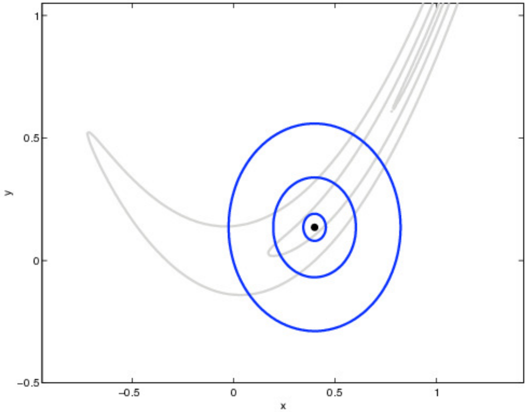
- Generate a chain of samples $\{\boldsymbol{m}_i\}$ using proposal distributions $\{P_i\}$ that can be sampled directly:

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_{\boldsymbol{m}_i} \longrightarrow \mathbb{E}[\boldsymbol{q}_{\boldsymbol{m}}] \text{ a.s.}$$

- The speed of convergence depends on the similarity of π_{post} and $\{P_i\}$.

Markov Chain Methods

2D Rosenbrock Distribution



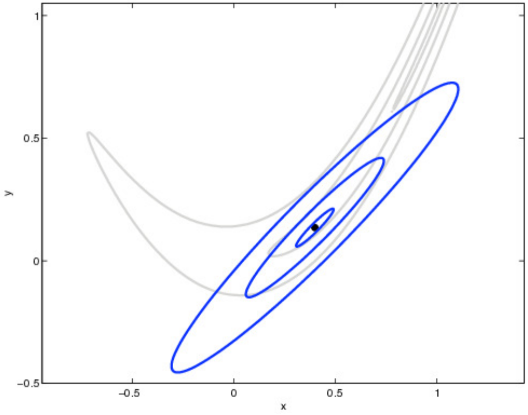
- Generate a chain of samples $\{\boldsymbol{m}_i\}$ using proposal distributions $\{P_i\}$ that can be sampled directly:

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_{\boldsymbol{m}_i} \longrightarrow \mathbb{E}[\boldsymbol{q}_{\boldsymbol{m}}] \text{ a.s.}$$

- The speed of convergence depends on the similarity of π_{post} and $\{P_i\}$.

Markov Chain Methods

2D Rosenbrock Distribution



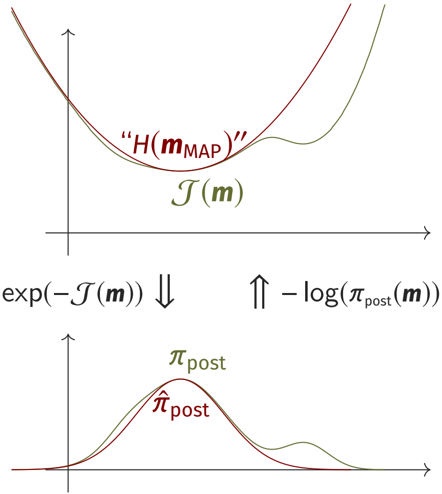
- Generate a chain of samples $\{\boldsymbol{m}_i\}$ using proposal distributions $\{P_i\}$ that can be sampled directly:

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{q}_{\boldsymbol{m}_i} \longrightarrow \mathbb{E}[\boldsymbol{q}_{\boldsymbol{m}}] \text{ a.s.}$$

- The speed of convergence depends on the similarity of π_{post} and $\{P_i\}$.

Laplace's Approximation

$$\hat{\pi}_{\text{post}}(\boldsymbol{m}) := \mathcal{N}(\boldsymbol{m}_{\text{MAP}}, H(\boldsymbol{m}_{\text{MAP}})^{-1})$$



- A quadratic fit at the maximum a posteriori likelihood (MAP) point $\boldsymbol{m}_{\text{MAP}}$
- Also provides one-shot QoI estimates

$$\mathbb{E}[\boldsymbol{q}_{\boldsymbol{m}}] \approx \boldsymbol{q}_{\boldsymbol{m}_{\text{MAP}}}$$

$$\text{Var}(\boldsymbol{q}) \approx D\boldsymbol{q}_{\boldsymbol{m}_{\text{MAP}}} H(\boldsymbol{m}_{\text{MAP}})^{-1} D\boldsymbol{q}^*_{\boldsymbol{m}_{\text{MAP}}}$$

- Requires the Hessian

$$H(\boldsymbol{m}) := D^2 \mathcal{J}(\boldsymbol{m}).$$

Laplace's Approximation to One-Shot Projection

Laplace's approximation is the basis for a one-shot estimate of QoI uncertainty:

$$\boldsymbol{q}_{\text{post}} \sim \mathcal{N}(\boldsymbol{q}(\boldsymbol{m}_{\text{MAP}}, \boldsymbol{w}_{\text{MAP}}), D_{\boldsymbol{m}}\boldsymbol{q} C_{\text{post}} D_{\boldsymbol{m}}\boldsymbol{q}^*).$$

Gaussian Assumptions and Optimization

- Given Gaussian prior and likelihood,

$$\pi_{\text{post}}(\boldsymbol{m}|\boldsymbol{d}_{\text{obs}}) \propto \exp\{-\underbrace{\tfrac{1}{2}(\|\boldsymbol{d}_{\text{obs}} - \boldsymbol{d}_{\boldsymbol{m}}\|^2_{C^{-1}_{\text{obs}}} + \|\boldsymbol{m} - \boldsymbol{m}_{\text{prior}}\|^2_{C^{-1}_{\text{prior}}})}_{\mathcal{J}(\boldsymbol{m})}\}.$$

- $\mathcal{J}(\boldsymbol{m})$ is a typical objective function in PDE-constrained optimization.
  - $\|\cdot\|_{C^{-1}_{\text{obs}}}$ misfit term (covariance smaller in slow-flow regions)
  - $\|\cdot\|_{C^{-1}_{\text{prior}}}$ regularization term (Bi-Laplacian)

Review: Algorithm and Components

Estimating Statistics of $\pi_{\text{post}}(\boldsymbol{m}|\boldsymbol{d}_{\text{obs}})$

1. Find MAP point $\boldsymbol{m}_{\text{MAP}}$ (Inexact Newton-Krylov method)
2. "Compute" $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$ (?)
3. Draw sample chain from $\mathcal{N}(\boldsymbol{m}_{\text{MAP}}, C_{\text{post}})$, compute statistics on chain

Review: Algorithm and Components

Method Requirements

1. Evaluate $\mathcal{J}(\boldsymbol{m})$
2. Compute gradient $D\mathcal{J}(\boldsymbol{m})$
3. Matrix-vector product $H(\boldsymbol{m})\hat{\boldsymbol{m}}$
4. Precondition $H(\boldsymbol{m})^{-1}$
5. Sample from $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$

Review: Algorithm and Components

Method Requirements	1–3: Use Adjoint Equations
1. Evaluate $\mathcal{J}(\boldsymbol{m})$	[Solve $A(\boldsymbol{w};\boldsymbol{m}) = \mathbf{0}$]
2. Compute gradient $D\mathcal{J}(\boldsymbol{m})$	[Solve $A^*_{\boldsymbol{w}}$ systems]
3. Matrix-vector product $H(\boldsymbol{m})\hat{\boldsymbol{m}}$	[Solve $A_{\boldsymbol{w}}, A^*_{\boldsymbol{w}}$ systems]
4. Precondition $H(\boldsymbol{m})^{-1}$	
5. Sample from $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$	

Adjoint Equations

$A_{\boldsymbol{w}}(\boldsymbol{w};\boldsymbol{m})^{*}\boldsymbol{z} = \boldsymbol{r} - D\mathcal{J}_{\text{misfit}}(\boldsymbol{m})$ in strong form

Balance of linear momentum, conservation of mass

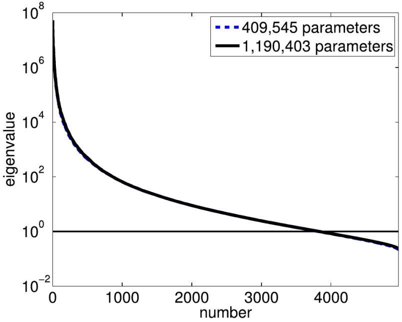
$$-\nabla \cdot \left[\boldsymbol{\mu}'(T,\boldsymbol{u})\left(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^{T}\right) - \boldsymbol{I}q\right] = \boldsymbol{r}_{\boldsymbol{u}},$$
$$\nabla \cdot \boldsymbol{v} = 0.$$

- Calculus of variations (no full-program automatic differentiation required)

Preconditioning $H(\boldsymbol{m})^{-1}$

- Each *entry* in $H(\boldsymbol{m})$ requires two PDE solves: $|\boldsymbol{m}|^2$ entries.
  - Never compute and store: only Hessian-vector products.
- $H(\boldsymbol{m})$ must be a compact perturbation of C_{prior}^{-1} (Stuart 2010)

$C_{\text{prior}}H(\boldsymbol{m}_{\text{MAP}})$ Eigenvalue Spectra, Coarse and Fine Mesh



Preconditioning $H(\boldsymbol{m})^{-1}$

C_{prior} as a preconditioner provides *mesh*-independence of number of iterations:

#sdof	#pdof	#N	#CG	avgCG	#Stokes
95,796	10,371	42	2718	65	7031
233,834	25,295	39	2342	60	6440
848,850	91,787	39	2577	66	6856
3,372,707	364,649	39	2211	57	6193
22,570,303	1,456,225	40	1923	48	5376

Review: Algorithm and Components

Method Requirements

1. Evaluate $\mathcal{J}(\boldsymbol{m})$ [Solve $A(\boldsymbol{w};\boldsymbol{m}) = \mathbf{0}$]
2. Compute gradient $D\mathcal{J}(\boldsymbol{m})$ [Solve $A_{\boldsymbol{w}}^*$ systems]
3. Matrix-vector product $H(\boldsymbol{m})\hat{\boldsymbol{m}}$ [Solve $A_{\boldsymbol{w}}, A_{\boldsymbol{w}}^*$ systems]
4. Precondition $H(\boldsymbol{m})^{-1}$
5. Sample from $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$

Review: Algorithm and Components

Method Requirements	4: Use C_{prior}
1. Evaluate $\mathcal{J}(\boldsymbol{m})$	[Solve $A(\boldsymbol{w};\boldsymbol{m}) = \mathbf{0}$]
2. Compute gradient $D\mathcal{J}(\boldsymbol{m})$	[Solve $A_{\boldsymbol{w}}^{*}$ systems]
3. Matrix-vector product $H(\boldsymbol{m})\hat{\boldsymbol{m}}$	[Solve $A_{\boldsymbol{w}}, A_{\boldsymbol{w}}^{*}$ systems]
4. Precondition $H(\boldsymbol{m})^{-1}$	[$C_{\text{prior}} \approx H(\boldsymbol{m})^{-1}$]
5. Sample from $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$	

Sampling C_{post} via Low-Rank Update

Prior Sample Modification (Accurate within ϵ)

1. Partial, randomized, generalized EVD (Halko, Martinsson, and Tropp 2011) with tolerance ϵ

$$H_{\text{misfit}} \approx V_r \Lambda_r V_r^* \quad [V_r^* C_{\text{prior}} V_r = I_r, W_r = C_{\text{prior}} V_r].$$

2. For each sample $\boldsymbol{y}$, draw sample $\boldsymbol{z}$ from C_{prior}:

$$\boldsymbol{y} = \boldsymbol{m}_{\text{MAP}} + (I - W_r(I_r - (I_r + \Lambda_r)^{-1/2})V_r^*)\boldsymbol{z}.$$

Sampling C_{post} via Low-Rank Update

Prior Sample Modification (Accurate within ϵ)

1. Partial, randomized, generalized EVD (Halko, Martinsson, and Tropp 2011) with tolerance ϵ

$$H_{\text{misfit}} \approx V_r \Lambda_r V_r^* \quad [V_r^* C_{\text{prior}} V_r = I_r, W_r = C_{\text{prior}} V_r].$$

2. For each sample $\boldsymbol{y}$, draw sample $\boldsymbol{z}$ from C_{prior}:

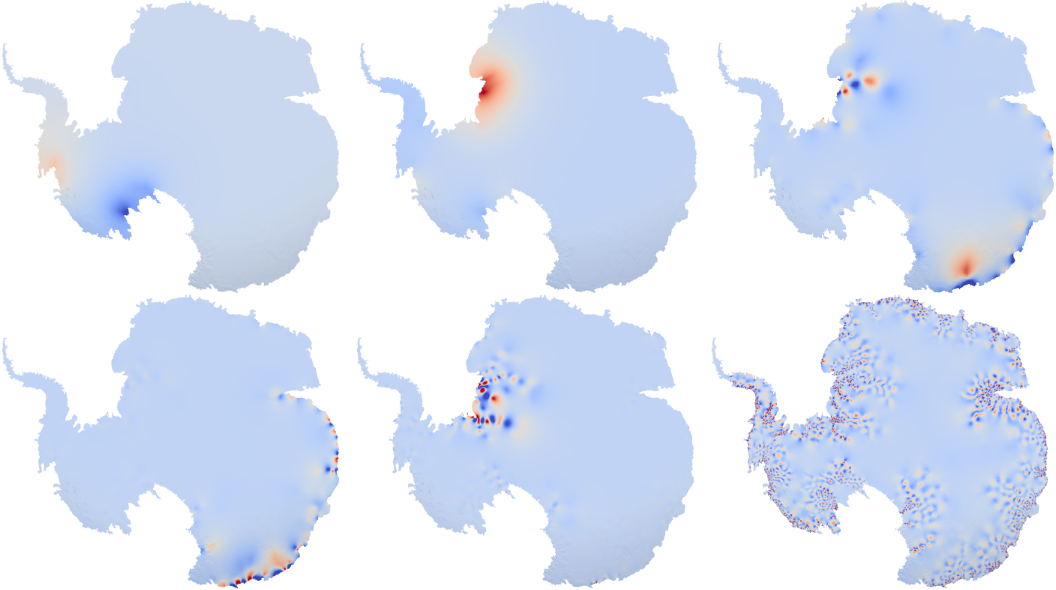
$$\boldsymbol{y} = \boldsymbol{m}_{\text{MAP}} + (I - W_r(I_r - (I_r + \Lambda_r)^{-1/2})V_r^*)\boldsymbol{z}.$$

Sampling C_{post} via Low-Rank Update

- Compactness: number of Hessian-vector products to achieve accuracy ϵ is mesh-independent
- C_{prior} treated as "black-box": no assumption about how it is sampled, or *whether we have a symmetric factor*. Compare to

$$
\begin{aligned}
C_{\text{prior}} &= LL^*, \\
L^* H_{\text{misfit}} L &\approx \tilde{V}_r \Lambda_r \tilde{V}_r^* \quad [\tilde{V}_r^* \tilde{V}_r = I_r], \\
\boldsymbol{y} &= \boldsymbol{m}_{\text{MAP}} + L\tilde{V}_r (I + \Lambda)^{-1/2} \boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(0, I).
\end{aligned}
$$

V_r vectors 1, 2, 100, 200, 500, 4000



Review: Algorithm and Components

Method Requirements

1. Evaluate $\mathcal{J}(\boldsymbol{m})$ [Solve $A(\boldsymbol{w}; \boldsymbol{m}) = \mathbf{0}$]
2. Compute gradient $D\mathcal{J}(\boldsymbol{m})$ [Solve $A^*_{\boldsymbol{w}}$ systems]
3. Matrix-vector product $H(\boldsymbol{m})\hat{\boldsymbol{m}}$ [Solve $A_{\boldsymbol{w}}, A^*_{\boldsymbol{w}}$ systems]
4. Precondition $H(\boldsymbol{m})^{-1}$ [$C_{\text{prior}} \approx H(\boldsymbol{m})^{-1}$]
5. Sample from $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$

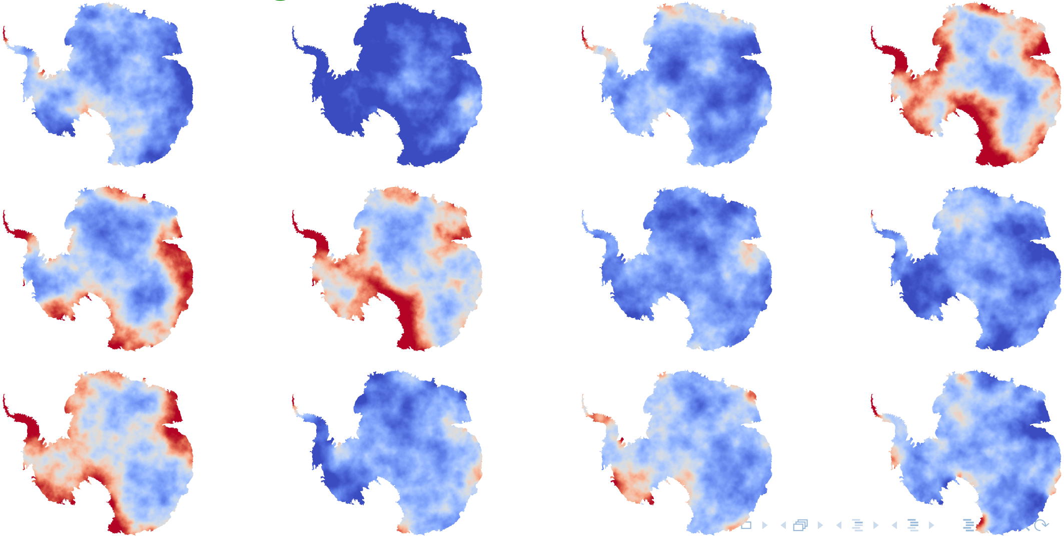
Review: Algorithm and Components

Method Requirements — **5: low-rank update pair $\{V_r, W_r\}$**

1. Evaluate $\mathcal{J}(\boldsymbol{m})$ — [Solve $A(\boldsymbol{w}; \boldsymbol{m}) = \mathbf{0}$]
2. Compute gradient $D\mathcal{J}(\boldsymbol{m})$ — [Solve $A^*_{\boldsymbol{w}}$ systems]
3. Matrix-vector product $H(\boldsymbol{m})\hat{\boldsymbol{m}}$ — [Solve $A_{\boldsymbol{w}}, A^*_{\boldsymbol{w}}$ systems]
4. Precondition $H(\boldsymbol{m})^{-1}$ — [$C_{\text{prior}} \approx H(\boldsymbol{m})^{-1}$]
5. Sample from $C_{\text{post}} = H(\boldsymbol{m}_{\text{MAP}})^{-1}$ — [Modify prior samples]

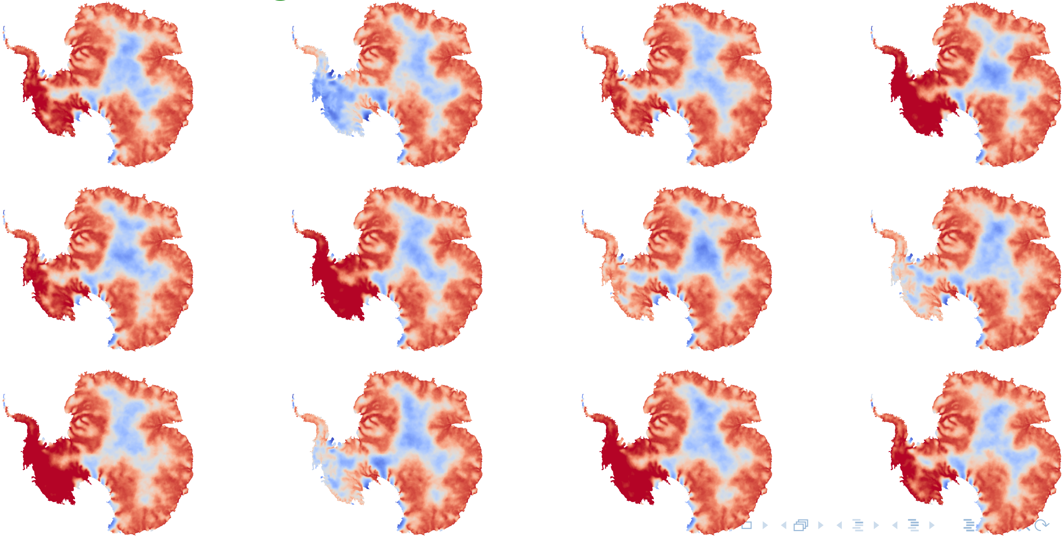
C_{prior} Samples

Antarctic Sliding Parameter

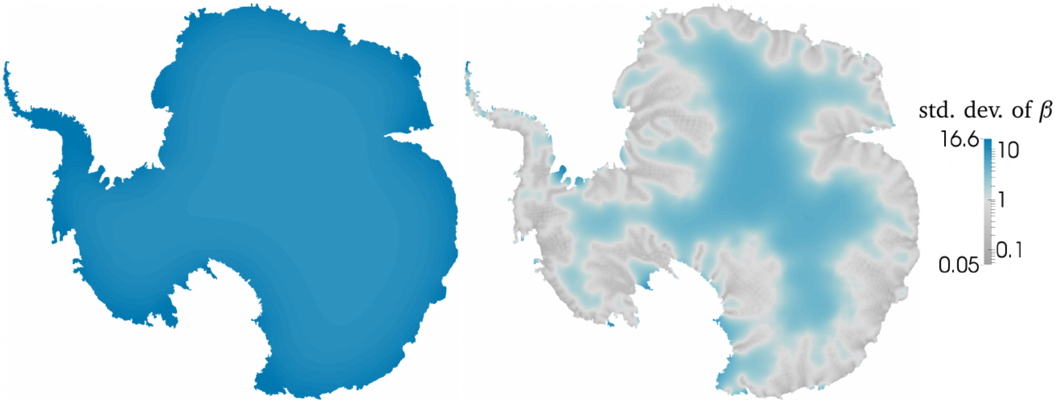


C_{post} Samples

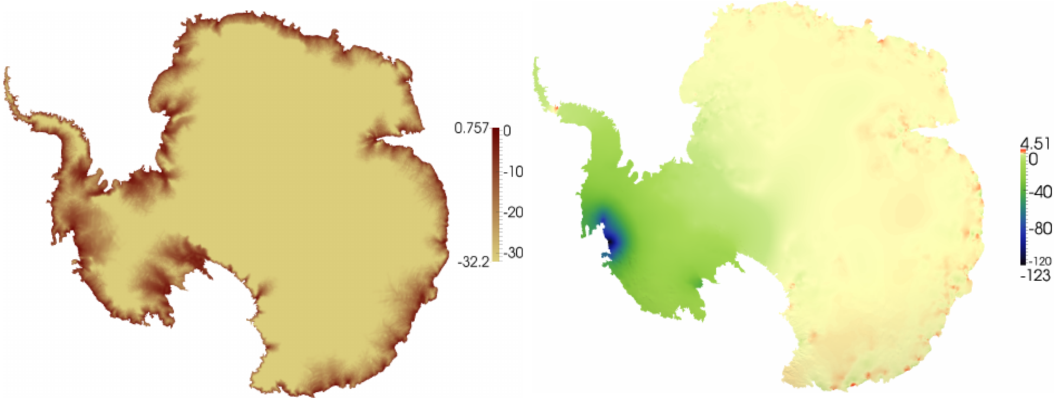
Antarctic Sliding Parameter



Reduction in Variance



q: Flux at Grounding Line



Left: gradient of q; right: joint maximizer of *sensitivity* and *variance*.

Conclusion

Theory says that for this problem, *the cost of the entire data-to-prediction framework, when measured in forward or adjoint Stokes solves, is a constant independent of the parameter or data dimension.*

We come very close, while also using optimal methods to keep the work per solve close to $\sim N/P$.

Acknowledgments

- U.S. Air Force Office of Scientific Research Computational Mathematics [FA9550- 12-1-0484]
- U.S. Department of Energy Office of Science Advanced Scientific Computing Research [DE-FG02-09ER25914, DE-FC02-13ER26128, and DE-SC0010518]
- U.S. National Science Foundation Cyber-Enabled Discovery and Innovation [CMS-1028889, OPP-0941678]
- Oak Ridge Leadership Facility at ORNL [DE-AC05-00OR22725]
- Texas Advanced Computing Center & XSEDE

References I

Burstedde, Wilcox, and Ghattas (2011). “p4est: Scalable Algorithms for Parallel Adaptive Mesh Refinement on Forests of Octrees”. DOI: 10.1137/100791634.

Creyts and Schoof (2009). “Drainage through subglacial water sheets”. DOI: 10.1029/2008JF00121.

Halko, Martinsson, and Tropp (2011). “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions”. DOI: 10.1137/090771806.

Rignot, Mouginot, and Scheuchl (2011). “MEaSUREs InSAR-based Antarctica ice velocity map”.

Stuart (2010). “Inverse problems: A Bayesian perspective”. DOI: 10.1017/S0962492910000061.