Accelerating Bayesian inference with computationally intensive models, with application to Pine Island Glacier

Patrick R. Conrad (MIT), Patrick Heimbach (MIT), Youssef Marzouk (MIT), Natesh Pillai (Harvard), and Aaron Smith (Univ. of Ottawa)

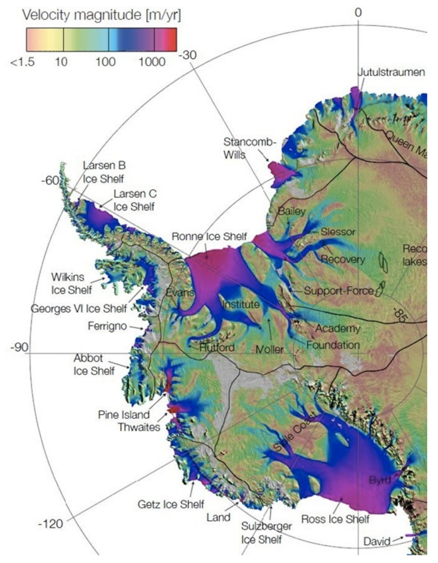
Massachusetts Institute of Technology

MIT AEROASTRO

Antarctica and climate change

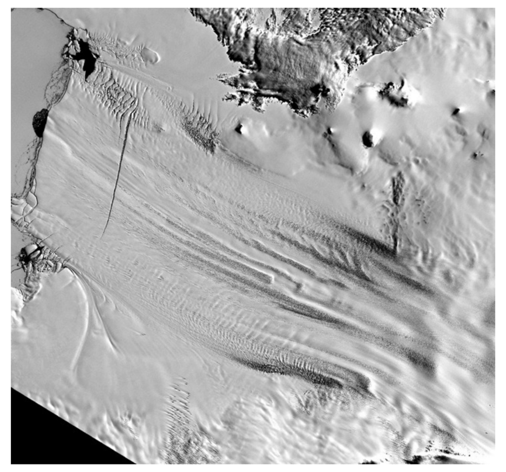
The Western Antarctic Ice Sheet has recently shown growing mass loss along the Amundsen coast

Western Antarctic Ice Sheet



[Rignot et al. 2011]

Pine Island Glacier



[NASA]

Vast uncertainty in ice-ocean dynamics

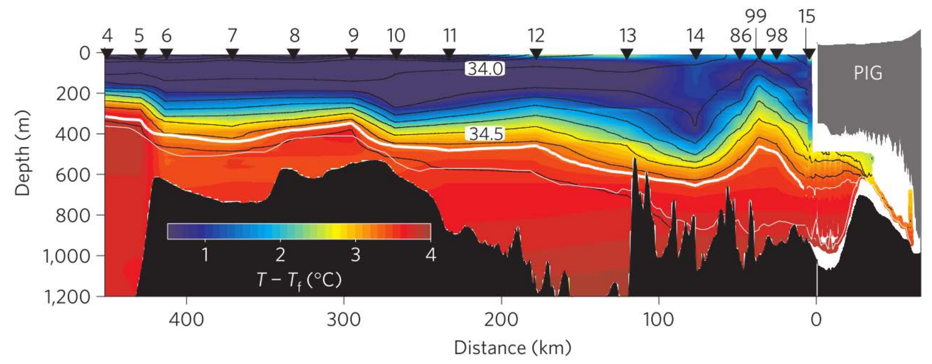
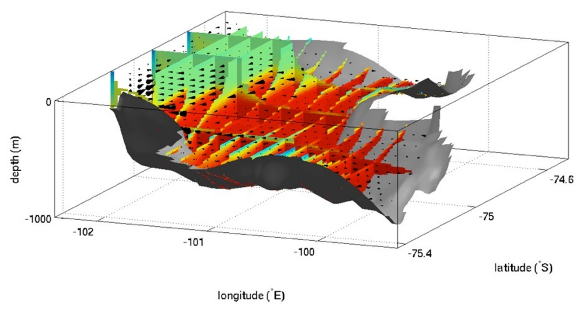


Figure: Temperature profile under Pine Island Glacier, Antarctica [Jacobs et al.]

- How readily is heat absorbed by the ice?
- How much mixing occurs near the ice-ocean interface?
- Ultimately, can we predict melt rates and the stability of the glacier?

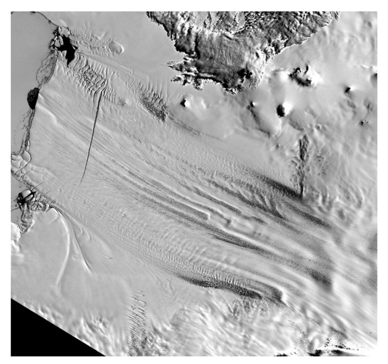
Forward model of ice-ocean coupling

- MIT General Circulation Model, configured for Pine Island
- Realistic geometry on coarse scale (4 km × 4 km × 20 m) or fine scale (1 km × 1 km × 20 m) models
- Several input parameters are unknown

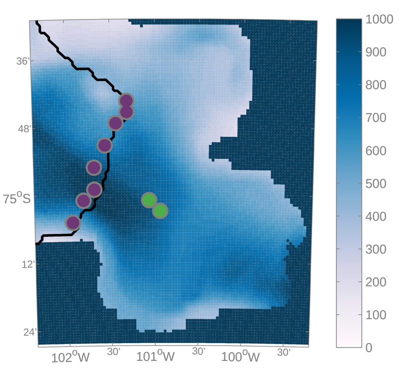


Constructing an inference problem

Satellite image



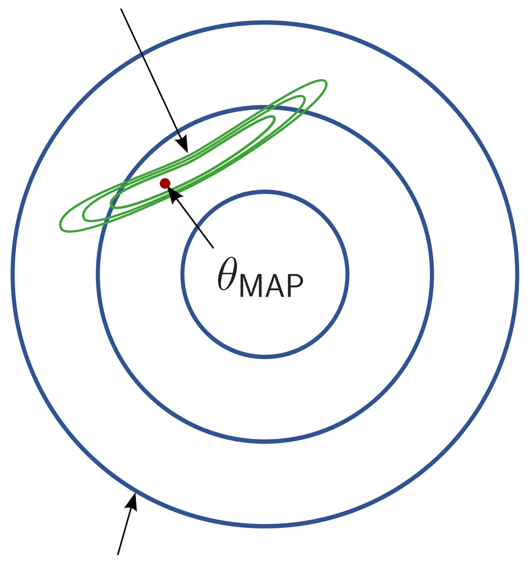
Bathymetry and sample locations



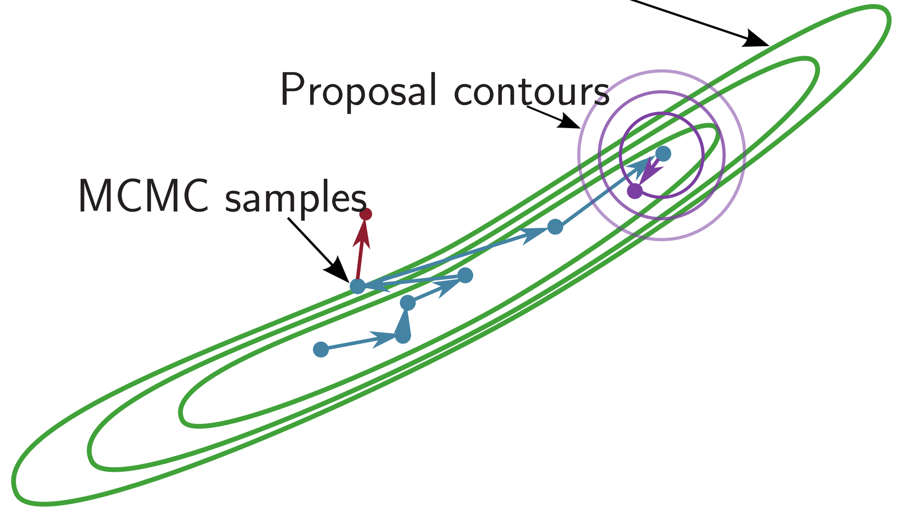
- Representative locations for temperature and salinity observations

Bayesian inference illustration

- Bayesian inference expresses our prior beliefs over parameters $\theta \in \mathbb{R}^n$, with a probability density,
$$p(\theta),$$
and constructs a posterior probability density,
$$p(\theta|\mathbf{d}) \propto \mathcal{L}(\theta|\mathbf{d}, \mathbf{f}(\theta))p(\theta)$$
expressing our beliefs after comparing the data $\mathbf{d} \in \mathbb{R}^d$, to the computationally expensive forward model $\mathbf{f}(\theta)$.
- Well suited to limited data and complex models



Markov chain Monte Carlo (MCMC)



- Significant literature discusses proposals that "mix" quickly, i.e., that generate nearly independent samples
- Evaluates forward model N times
- Run-time can be dominated by cost of $\mathbf{f}$
- Standard MCMC links cost of understanding $p(\theta|\mathbf{d})$ and $\mathbf{f}(\theta)$

MCMC with Local Approximations

Given X_0, initialize $\mathcal{S}_0$, then simulate chain $\{X_t\}_{t \leq N}$ with kernel:

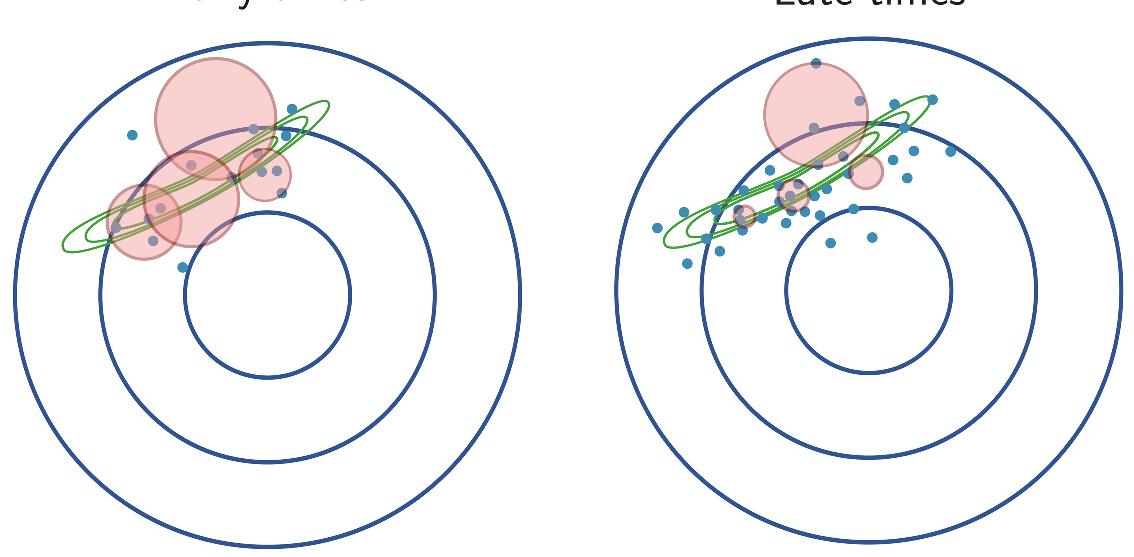
MH Kernel $K_t(x, \cdot)$

1. Given X_t, draw $q_t \sim Q(X_t, \cdot)$ from kernel Q with (symmetric) translation invariant density $q(x, \cdot)$
2. Compute acceptance ratio
$$\alpha = \min\left(1, \frac{\mathcal{L}(\theta|\mathbf{d}, \tilde{\mathbf{f}}_t(q_t))p(q_t)}{\mathcal{L}(\theta|\mathbf{d}, \tilde{\mathbf{f}}_t(X_t))p(X_t)}\right)$$
3. As needed, select new samples near q_t or X_t, yielding $\mathcal{S}_t \subseteq \mathcal{S}_{t+1}$. Refine $\tilde{\mathbf{f}}_t \to \tilde{\mathbf{f}}_{t+1}$.
4. Draw $u \sim \mathcal{U}(0, 1)$. If $u < \alpha$, let $X_{t+1} = q_t$, otherwise $X_{t+1} = X_t$.

Local approximations

- To compute $\tilde{\mathbf{f}}(\theta)$, construct a model over ball $\mathcal{B}_R(\theta)$
- Use samples $\theta_i \in \mathcal{S}$ at distance $r = \|\theta - \theta_i\| < R$
- Approximation converges locally under loose conditions [Cleveland]
- For example, quadratic approximations over $\mathcal{B}_R(\theta)$ [Conn *et al.*]:
$$\|\mathbf{f} - \mathcal{Q}_R\mathbf{f}\| \leq \|\mathbf{f}\|\kappa\lambda R^3$$

Local approximation illustration



Models are refined using new points chosen when model quality appears poor

Ergodicity and exactness of approximate samplers

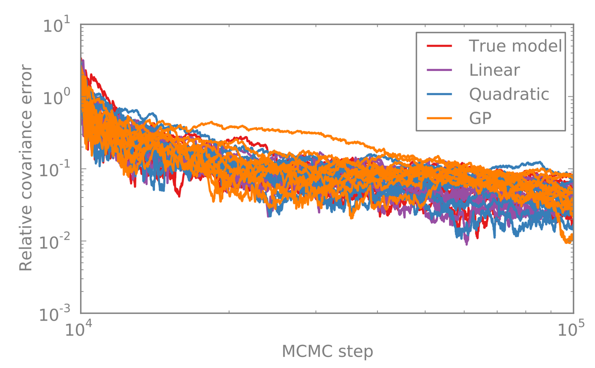
Assume the log-posterior is approximated with local quadratic models and $\theta \in \mathcal{X} \subseteq \mathbb{R}^n$ for *compact* $\mathcal{X}$, or $p(\theta|\mathbf{d})$ obeys a *Gaussian envelope*:
$$\lim_{r\to\infty} \sup_{|\theta|=r} |\log(p(\theta|\mathbf{d})) - \log(p_\infty(\theta))| = 0$$
for some quadratic form $\log(p_\infty)$ with negative definite coefficient matrix.

Then under standard regularity assumptions for geometrically ergodic kernel K_∞ and posterior $p(\theta|\mathbf{d})$, the chain X_t is **ergodic** and asymptotically samples from the **exact posterior**:
$$\lim_{t\to\infty} \|\mathbb{P}(X_t) - p(\theta|\mathbf{d})\|_{TV} = 0$$

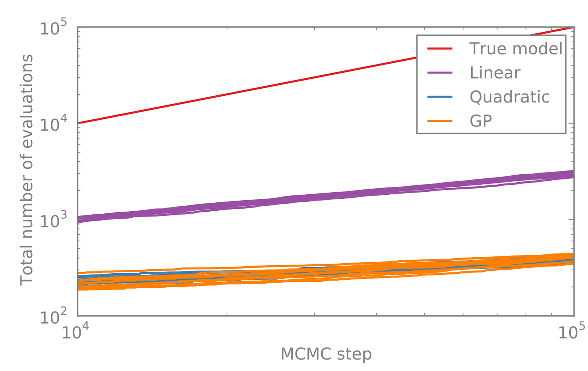
Example: Elliptic permeability inversion

Infer parameters of k given observations of u in the PDE:
$$\nabla_{\mathbf{s}} \cdot (k(\mathbf{s}, \theta)\nabla_{\mathbf{s}} u(\mathbf{s}, \theta)) = 0,$$

Accuracy of chains



Cost of chains



Prior and likelihood selection

- Priors are log-normal with expert-chosen mean and width
- Likelihoods are i.i.d. Gaussian with variance suggested by *in situ* experimental data

Parameter	Nominal value, μ'	Prior "width" σ'
Drag coefficients	1.5E-3	1.5E-3
Heat & Salt transfer	1.0E-4	0.5E-4
Prandtl Number	13.8	1.
Schmidt Number	2432.	200.
Horizontal Diffusion	5.0E-5	5.0E-5
ZetaN	5.2E-2	0.5E-3
Temperature	–	0.04
Salinity	–	0.1

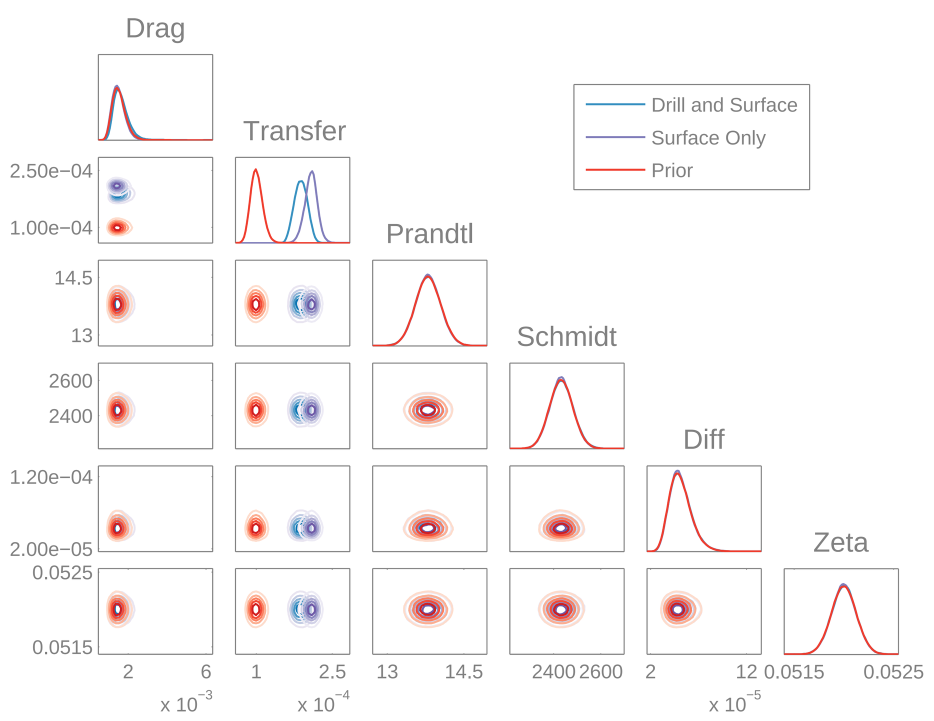
Computational details and results

- Compute synthetic data using the fine scale model, try to infer them using the coarse scale
- Constructed 30 parallel chains with shared evaluations
- Chains run for approximately two weeks
- Results shown after burn-in is removed

Inference cost summary

	Samples	Model runs	Savings
Drill and surface	225,000	53,000	$\geq$ 4.2x
Surface only	450,000	52,000	$\geq$ 8.6x

Prior and posterior marginals



Contributions

- Introduce a novel framework for using **local approximations** within MCMC; prove that the framework produces **asymptotically exact samples**.
- Demonstrate strong numerical performance on canonical inference problems.
- Construct a realistic, synthetic inference problem for ice-ocean coupling near Pine Island Glacier.
- Apply local approximation methods to reduce computational cost of inference in the Pine Island Glacier setting.

This work is supported by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award Number DE-SC0007099, part of the SciDAC Institute for the Quantification of Uncertainty in Extreme-Scale Computations (QUEST).

prconrad@mit.edu, ymarz@mit.edu