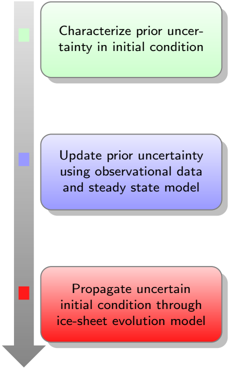
Predicting changes in sea level due to ice-sheet evolution in Greenland

Predicting Ice Sheet and Climate Evolution at Extreme Scales (PISCEES)

John D. Jakeman, Irina Teazur, Mauro Perego, Steve Price, Michael Eldred, Charles Jackson

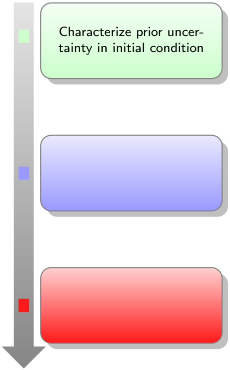
Quantifying Uncertainty in sea level change due to ice-sheet evolution



GOAL: Invert for unknown/uncertain ice sheet model parameters to find ice sheet initial state that

- matches observations (e.g. surface velocity, temperature, etc.)
- matches present-day geometry (elevation, thickness)
- is in "equilibrium" with climate forcings (SMB)
- Significantly reduce non physical transients without spin-up

Specifying an initial prior for uncertainty



Naive parameterization: Represent each degree of freedom on mesh be an uncertain variable

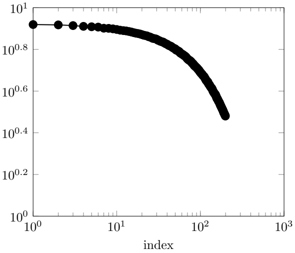
$$\beta(\boldsymbol{x}) = (z_1, \ldots, z_{n_{\mathrm{dof}}})$$

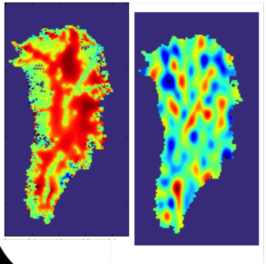
Assume level of smoothness; Petra, Stadler Ghattas square Laplace (Beltrami) operator, or use a KLE

$$\log(\beta(\boldsymbol{x})) = \log(\bar{\beta}) + \sum_{i=1}^{d} \sqrt{\lambda}_i \phi_i(\boldsymbol{x}) z_i$$

The latter results in significant dimension reduction, however there is choice of smoothness (correlation length). Moreover both approaches require an assumption of variance. What should this be?

Our prior





Log friction given by

$$\log(\beta(\boldsymbol{x})) = \log(\bar{\beta}) + \sum_{i=1}^{d} \sqrt{\lambda_i}\phi_i(\boldsymbol{x}) z_i$$

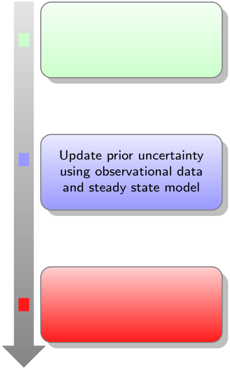
Covariance kernel

$$C_a(x_1, x_2) = 0.01 \exp\left[-\frac{(x_1 - x_2)^2}{0.001}\right]$$

Higher-dimension required for scientifically justified conclusions.

However this is a good test problem.

A data informed initial condition

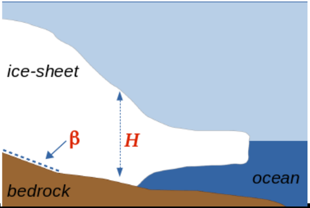


Available data/measurements:

- ice extension and surface topography
- surface velocity
- Surface Mass Balance (SMB)
- ice thickness H (sparse measurements)

Fields to be estimated:

- ice thickness H (allowed to vary but weighted by observational uncertainties)



Deterministic Inversion

Optimization problem: Find β and H that minimize

$$\begin{aligned}
\mathcal{J}(\beta, H) &= \int_\Gamma \frac{1}{\sigma_u^2} |\mathbf{u} - \mathbf{u}^{obs}|^2 \, ds && \text{surface velocity mismatch} \\
&+ \int_\Gamma \frac{1}{\sigma_\tau^2} |\mathrm{div}(\boldsymbol{U} H) - \tau_s|^2 \, ds && \text{SMB mismatch} \\
&+ \int_\Gamma \frac{1}{\sigma_H^2} |H - H^{obs}|^2 \, ds && \text{thickness mismatch} \\
&+ \mathcal{R}(\beta, H) && \text{regularization terms.}
\end{aligned}$$

$\boldsymbol{U}$: computed depth averaged velocity
H: ice thickness
β: basal sliding friction coefficient
τ_s: SMB
$\mathcal{R}(\beta)$ regularization term

Inverting for friction (and thickness)

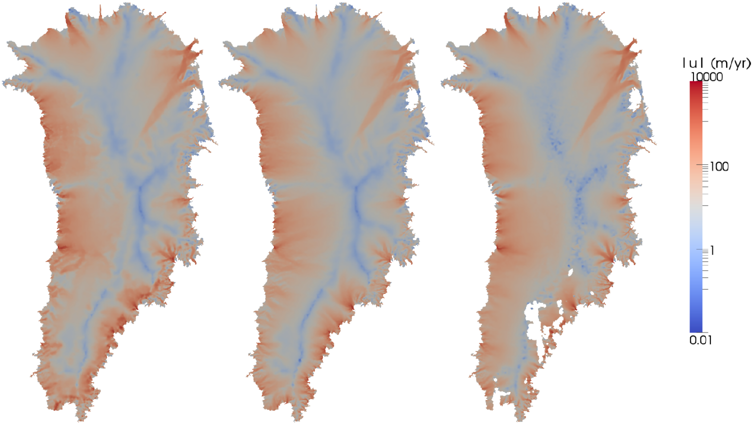


Figure: (left) β only surf. vel. (middle) β, H surf. vel. (right) Obs. surf. vel.

What is the effect of shocks on transient forward Model

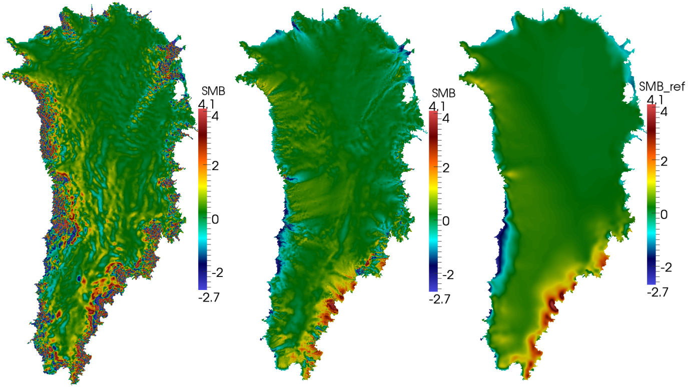
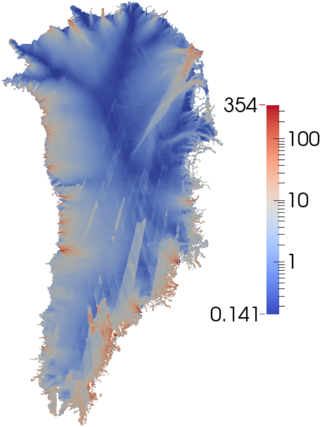
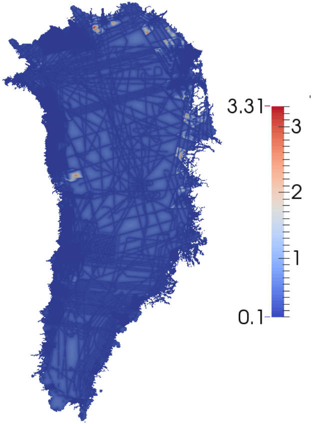


Figure: (left) β only SMB (middle) β, H SMB (right) SMB from climate model

Heavily processed errors in observational data



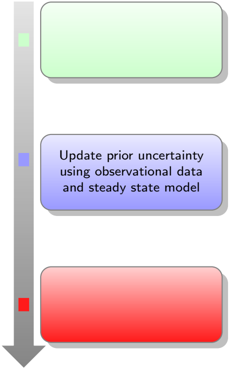
(a) Surf. vel. err. in m/yr



(b) Thickness err. in km

Uncertainty in SMB is not taken into account

Bayesian Inference



Additive Gaussian Noise Model

$$\boldsymbol{d} = \boldsymbol{f}(\boldsymbol{z}) + \epsilon, \quad \epsilon \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Gamma})$$

Separable I.I.D. Gaussian Prior

$$\pi_{\text{pr}}(\boldsymbol{z}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}\right)$$

Likelihood is

$$\hat{\pi}_{\text{lhood}}(\boldsymbol{z}) = \exp\left(-\frac{1}{2}(\boldsymbol{d} - \boldsymbol{f}(\boldsymbol{z}))^T \boldsymbol{\Gamma}_{\text{obs}}^{-1}(\boldsymbol{d} - \boldsymbol{f}(\boldsymbol{z}))\right)$$

Posterior is

$$\pi_{\text{pos}}(\boldsymbol{z}) = C_{\text{evid}}^{-1}\hat{\pi}_{\text{lhood}}(\boldsymbol{z})\pi_{\text{pr}}(\boldsymbol{z})$$

Approximate the misfit

Compute approximation of misfit

$$m(\boldsymbol{z}) = \frac{1}{2}(\boldsymbol{d} - \boldsymbol{f}(\boldsymbol{z}))^T \boldsymbol{\Gamma}_{\text{obs}}^{-1}(\boldsymbol{d} - \boldsymbol{f}(\boldsymbol{z}))$$

Approximate a function by another different but related function or fewer variables

Linear transformation of parameters

$$m(\boldsymbol{z}) \approx \boxed{g}\left(\boxed{\boldsymbol{W}_1^T}\boldsymbol{z}\right)$$

Function of few variables $s \leq d$

Data informed subpaces

Compute active subspace of misfit

$$m(\boldsymbol{z}) \approx \; g\left(\boldsymbol{W}_1^T \boldsymbol{z}\right)$$

Active subspace computed using

Eigenvalue decomposition

$$\int_{\mathbb{R}^d} \nabla_{\mathbf{x}} m(\boldsymbol{z}) \nabla_{\mathbf{x}} m(\boldsymbol{z})^T \mathrm{d}\rho(\boldsymbol{z}) = \boldsymbol{W} \boldsymbol{\Lambda} \boldsymbol{W}^T$$

Gradient of misfit

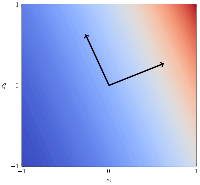
Averaged over prior

Partition $\boldsymbol{z}$ into inactive and active variables

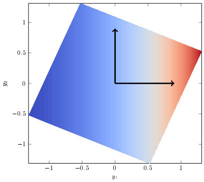
$$\boldsymbol{z} = \boldsymbol{W}_1^T \boldsymbol{z} + \boldsymbol{W}_2^T \boldsymbol{z}, \quad \boldsymbol{W} = [\boldsymbol{W}_1 \quad \boldsymbol{W}_2]$$

How Does Active subspaces work

$$m(\boldsymbol{z}) = \exp\left(0.7z_1 + 0.3z_2\right).$$



(c) Original function



(d) Rotated function

Sample gradient using MC

$$[\nabla m(\boldsymbol{z}^{(1)}), \cdots, \nabla m(\boldsymbol{z}^{(M)})]$$

Compute Gauss Newton approximation of Hessian averaged over prior

$$\boldsymbol{C} = \frac{1}{M}\sum_{i=1}^{M} \nabla m(\boldsymbol{z}^{(i)}) \nabla m(\boldsymbol{z}^{(i)})^T$$

Perform eigenvalue decomposition

$$\boldsymbol{C} = \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^T,$$

Partition $\boldsymbol{z}$ into inactive and active variables

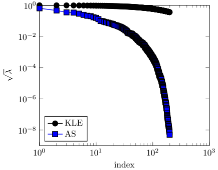
$$\boldsymbol{z} = \boldsymbol{W}_1^T\boldsymbol{z} + \boldsymbol{W}_2^T\boldsymbol{z}, \quad \boldsymbol{W} = [\boldsymbol{W}_1 \quad \boldsymbol{W}_2].$$

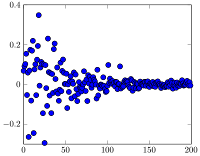
Dimension reduction for Greenland

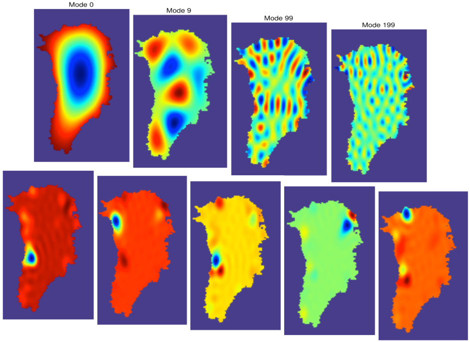
Gradients obtained using adjoint solve

$$m(\boldsymbol{z}) = \frac{1}{2}\nabla \boldsymbol{f}(\boldsymbol{z})^T \boldsymbol{\Gamma}_{\text{obs}}^{-1}(\boldsymbol{d} - \boldsymbol{f}(\boldsymbol{z}))$$

Gradients obtained Albany are w.r.t mesh DOF, must use multivariate chain rule to express w.r.t KLE vars







Active subspaces for inference

$$\pi_{\text{pos}}(\boldsymbol{z}) \approx C_{\text{evid}}^{-1} \underbrace{\exp(-\hat{m}_s(\boldsymbol{W}_1^T \boldsymbol{z}))\pi_{\text{pr}}(\boldsymbol{W}_1^T \boldsymbol{z})}_{\text{Approximate posterior only in directions informed by data}} \overbrace{\pi_{\text{pr}}(\boldsymbol{W}_2^T \boldsymbol{z})}^{\text{Revert to prior in uninformed directions}}$$

Various levels of approximation can be employed

- Reduce dimension but no surrogate of misfit
    - Perform MCMC in active subspace to improve mixing
- Surrogate of misfit with rotation but no dimension reduction
    - Leverage increased sparsity induced by rotation,
- Surrogate of misfit and reduce dimension
    - Build surrogate of subspace and perform MCMC on the surrgoate

Compare errors as accuracy of inversion increases

1. Dimension reduction via KLE
2. Dimension reduction via AS
3. Quadratic PCE over active variables
4. Laplace posterior at MAP

1. Dimension reduction via KLE
2. Dimension reduction via AS
3. High-order PCE
4. Stochastic Newton on PCE

1. Dimension reduction via KLE
2. Dimension reduction via AS
3. No surrogate of steady-state model
4. Stochastic Newton on steady-state simulation model

Error in posterior

Lessons can be learned by avoiding the use of highest fidelity model.

Laplace approximation

Given Linear Model $G(\boldsymbol{z}) = \boldsymbol{G}\boldsymbol{z}$ we have

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{z} + \epsilon$$

The resulting posterior is Gaussian

$$\boldsymbol{\mu}_{\text{pos}} = \boldsymbol{\Gamma}_{pos}\boldsymbol{\Gamma}_{\text{pr}}\boldsymbol{\mu}_{\text{pr}} + \boldsymbol{G}^T\boldsymbol{\Gamma}_{\text{obs}}^{-1}\boldsymbol{y}, \quad \text{and} \quad \boldsymbol{\Gamma}_{\text{pos}} = (\boldsymbol{H} + \boldsymbol{\Gamma}_{\text{pr}}^{-1})^{-1}$$

$$\boldsymbol{H} = \boldsymbol{G}^T\boldsymbol{\Gamma}_{\text{obs}}^{-1}\boldsymbol{G}$$

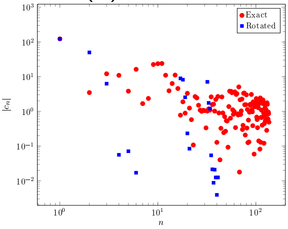
Approximate misfit over active variables using a quadratic function obtained via compressed sensing. Using $M = 733$ samples and a PCE with 20301 terms

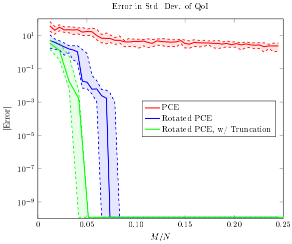
$$\frac{\|m(\boldsymbol{z}) - \hat{m}(\boldsymbol{z})\|_{\ell^2_\rho}}{\|m(\boldsymbol{z}) - \sum_{i=1}^{M} m(\boldsymbol{z}^{(i)})\|_{\ell^2_\rho}} \approx 0.981$$

Quadratic gradient enhanced PCE does not fit in desktop memory. Need large-scale compressed sensing tools.

Enhancing sparsity via rotation

$m(\boldsymbol{z}) = \boldsymbol{z}^T \boldsymbol{A} \boldsymbol{z}$





$\text{rank}(A) = 5$, $d = 15$, $z_i \sim N(1/2, 1/5)$

Approximate misfit with quadratic in rotated $d = 200$ space

$$\frac{\|m(\boldsymbol{z}) - \hat{m}(W^T \boldsymbol{z})\|_{\ell^2_\rho}}{\|m(\boldsymbol{z}) - \sum_{i=1}^{M} m(\boldsymbol{z}^{(i)})\|_{\ell^2_\rho}} \approx 0.190$$

Approximate misfit with quadratic in rotated and truncated $d = 73$ space

$$\frac{\|m(\boldsymbol{z}) - \hat{m}_{s=73}(W_1^T \boldsymbol{z})\|_{\ell^2_\rho}}{\|m(\boldsymbol{z}) - \sum_{i=1}^{M} m(\boldsymbol{z}^{(i)})\|_{\ell^2_\rho}} \approx 0.136$$

Draw sample $\boldsymbol{z}'_k$ from proposal Compute the approximate misfit at proposal sample

$$m(\boldsymbol{z}'_k) = \boxed{\sum_{j=1}^{N_{\mathrm{mc}}} m(\boldsymbol{W}_1 \boldsymbol{y}_k + \boldsymbol{W}_2 \boldsymbol{x}_j)}$$

Compute acceptance ratio (assuming symmetric proposal)

$$\gamma = \frac{\exp(-g(\boxed{\boldsymbol{y}'_k}))\pi_{\mathrm{pr}}(\boxed{\boldsymbol{y}'_k})}{\exp(-g(\boldsymbol{y}_k))\pi_{\mathrm{pr}}(\boldsymbol{y}_k)}$$

Determine acceptance - draw $u \sim U[0,1]$

$$\boldsymbol{y}_{k+1} = \begin{cases} \boldsymbol{y}'_k & \text{if } \gamma \geq u \\ \boldsymbol{y}_k & \text{otherwise} \end{cases}$$

If accepted

$$\boldsymbol{z}_{kl} = \boldsymbol{W}_1 \boldsymbol{y}_k + \boxed{\boldsymbol{W}_2 \boldsymbol{x}_{kl}}, \quad \boldsymbol{x}_{kl} \sim \pi_{\mathrm{pr}}(\boldsymbol{x})$$

Similar to Likelihood informed subspace, but we do not have Hessian information

Given a Gaussian prior with mean $\bar{\boldsymbol{z}}$ and covariance $\boldsymbol{\Sigma}_{\text{prior}}$ the negative log likelihood objective gradient and inverse Hessian are

$$m(\boldsymbol{z}) = \frac{1}{2}\boldsymbol{r}^T\boldsymbol{\Sigma}_{\text{noise}}^{-1}\boldsymbol{r} + \frac{1}{2}(\boldsymbol{z}-\bar{\boldsymbol{z}})^T\boldsymbol{\Sigma}_{\text{prior}}^{-1}(\boldsymbol{z}-\bar{\boldsymbol{z}})$$

$$\boldsymbol{g} = \nabla m(\boldsymbol{z}) = \boldsymbol{r}^T\boldsymbol{\Sigma}_{\text{noise}}^{-1}\nabla\boldsymbol{r} + \boldsymbol{\Sigma}_{\text{prior}}^{-1}(\boldsymbol{z}-\bar{\boldsymbol{z}})$$

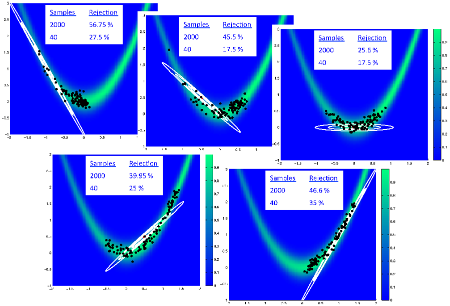
$$\boldsymbol{H}^{-1} = \boldsymbol{L}\{\boldsymbol{V}[(\boldsymbol{\Lambda}+\boldsymbol{I})^{-1}-\boldsymbol{I}]\boldsymbol{V}^T+\boldsymbol{I}\}\boldsymbol{L}^T$$

where $\boldsymbol{L}^T\boldsymbol{H}_{\text{misfit}}\boldsymbol{L} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$

Drawing samples from proposal using

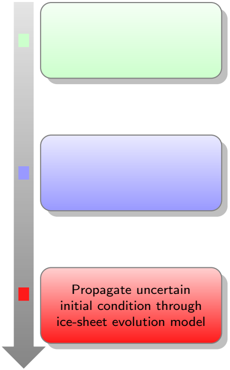
$$\boldsymbol{z} = \boldsymbol{z}^{\star} - \boldsymbol{H}^{-1}\boldsymbol{g} + \boldsymbol{H}^{-1/2}\boldsymbol{y}$$

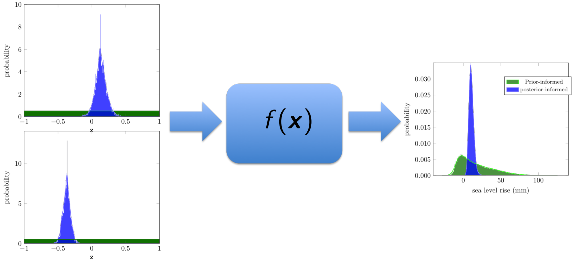
to increase acceptance rate in high-dimensions



We need to implement large-scale Lanczos eigenvalue solver to extend this technique from use surrogates to operate with Albany directly

Forward Propagation of Uncertainty





66 Models rusn required 1.25M CPU hours on 4km grid. NOT the finest resolution!

Compare errors as accuracy of forward propagation increases

- Linear forward map
- Gaussian posterior push forward

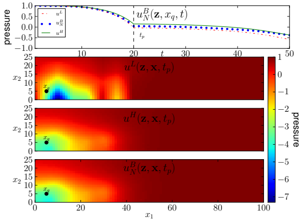
- Low rank multifidelity (MF) surrogate
- Push MCMC samples through MF surrogate

- No surrogate of transient model
- Push MCMC samples through transient simulation model

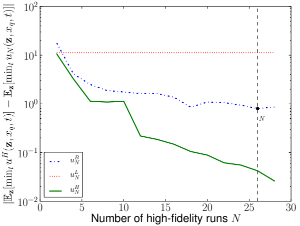
Error in posterior push forward

Lessons can be learned by avoiding the use of highest fidelity model.

$$u_N^B(\boldsymbol{\xi}, \mathbf{x}, t) = \sum_{n=1}^{N} c(u^L(\boldsymbol{\xi}, \mathbf{x}, t))\, u^H(\boldsymbol{\xi}_n^\gamma, \mathbf{x}, t)$$

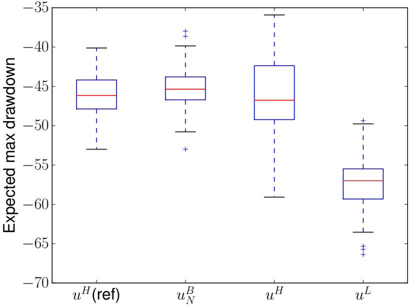


u^L is very inaccurate, however the error in u_N^B decreases as N increases.



Multifidelity algorithm can also be used for inverse problem.

- Reference u^H(ref): use 1000 HF runs to compute mean max drawdown.
- Low-fidelity u^L: use 2300 LF runs compute mean.
- Bi-fidelity u_N^B: use 1000 LF runs and $N = 26$ HF runs to compute mean.
- High-fidelity u^H: use 226 HF runs to compute mean.
- Cost to compute u_N^B, u^H and u^L are the same because 1 HF runs cost 50 time more than 1 LF run.

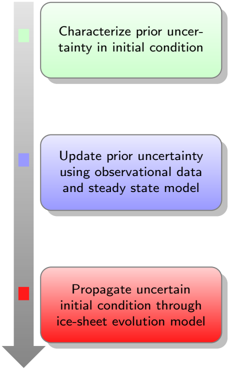


Uncertainty in predictions of sea-level change

Dimension reduction via KLE	Dimension reduction via KLE	Dimension reduction via KLE
Dimension reduction via AS	Dimension reduction via AS	No further dimension reduction
Quadratic PCE over active variables	High-order PCE using gradients	No surrogate of steady-state model
Laplace posterior at MAP	Stochastic Newton on PCE	Stochastic Newton on steady-state simulation model
Linear forward map	Low rank multifidelity (MF) surrogate	No surrogate of transient model
Gaussian posterior push forward	Push MCMC samples through MF surrogate	Push MCMC samples through transient simulation model

Error in uncertainty estimates for prediction of sea-level change

Forward Propagation of Uncertainty



- Large-scale KLE construction (using Anasazi from Trillinos)
- High-dimensional compressed sensing
- Compressed sensing for Gaussian random variables
- Low-rank multifidelity algorithm (including large scale Cholesky decomposition)
- Active subspace dimension reduction - Implement active subspace MCMC, Likelihood informed MCMC
- Stochastic Newton MCMC - Implement Large scale low-rank Hessian approximation