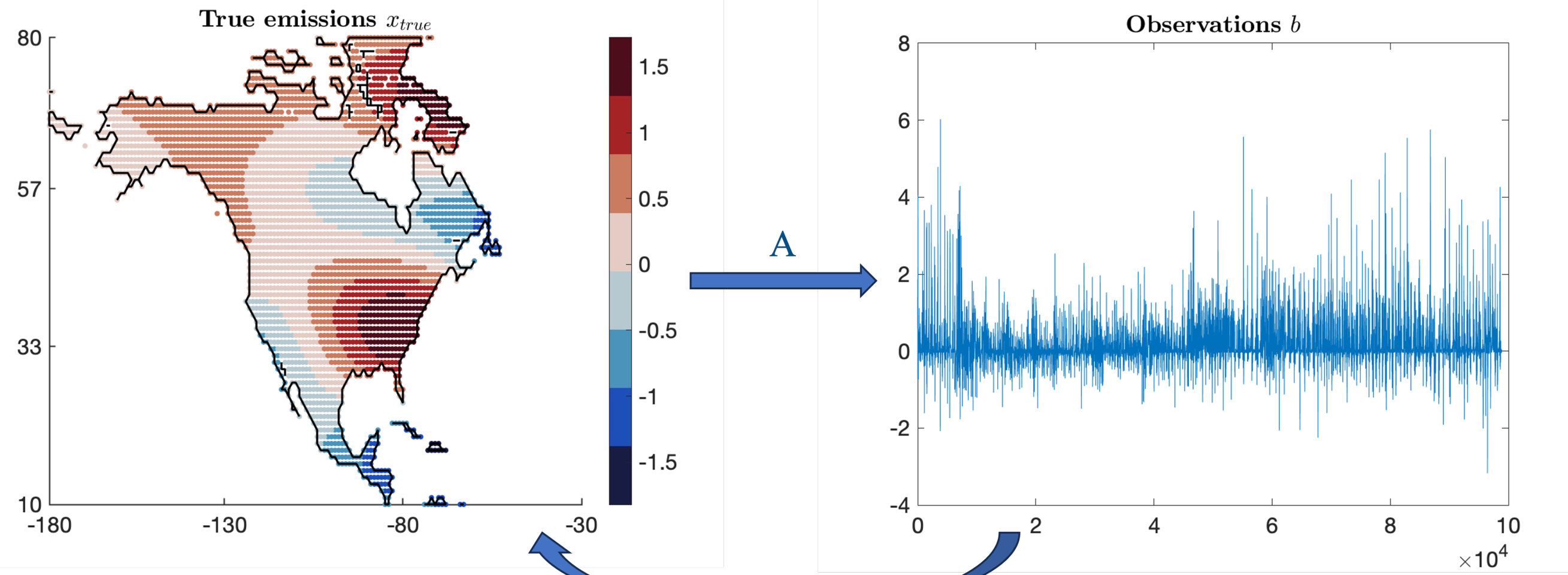




Motivation

- Inverse problems arise in many applications
 - Many unknown parameters
 - Dynamic problems have parameters that change over time
 - Example: Recovering CO₂ emissions from satellite data



Inverse Problem

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{e}$$

- $\mathbf{b} \in \mathbb{R}^m$ contains noisy observations
- $\mathbf{A} \in \mathbb{R}^{m \times n}$ represents the forward model
- $\mathbf{e} \in \mathbb{R}^m$ is noise
- $\mathbf{x}_{\text{true}} \in \mathbb{R}^n$ is the vectorized true image
- Uncertainty quantification (UQ) for large-scale problems is computationally challenging

Goal: Develop efficient sampling approaches that exploit generalized Golub-Kahan (genGK) methods

Hierarchical Bayesian Approach

Idea: Reformulate inverse problem as statistical inference where \mathbf{x} and \mathbf{e} are random variables:

$$\mathbf{x} \mid \delta \sim \mathcal{N}(\boldsymbol{\mu}, \delta^{-1}\mathbf{Q}) \quad \mathbf{e} \mid \lambda \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{R})$$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{0} \in \mathbb{R}^m$ are the mean vectors, assume $\boldsymbol{\mu} = \mathbf{0}$
- $\mathbf{Q} \in \mathbb{R}^{n \times n}$, $\mathbf{R} \in \mathbb{R}^{m \times m}$ symmetric positive definite
- $\delta > 0$, $\lambda > 0$ hyper-parameters with gamma hyperpriors

$$\pi(\lambda) \propto \lambda^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda) \quad \pi(\delta) \propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta)$$

By Bayes theorem, we get the posterior, which is the solution of the inverse problem

$$\pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b}) \propto \pi(\mathbf{b} \mid \mathbf{x}, \lambda) \pi(\mathbf{x} \mid \delta) \pi(\lambda) \pi(\delta)$$

Challenges for large-scale UQ:

- Need efficient methods to explore the posterior
 - Maximum a Posteriori estimate: $\max_{\mathbf{x}, \lambda, \delta} \pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b})$
 - Estimating variances requires drawing samples, which can be expensive
- \mathbf{A} and \mathbf{Q} may only be accessible via matrix-vector multiplications
- Computing factorization $\mathbf{Q} = \mathbf{L}\mathbf{L}^\top$ is computationally infeasible

Hierarchical Gibbs Sampling:

- **Idea:** Sample from the conditional densities of \mathbf{x} , λ and δ :

$$\lambda \mid \mathbf{b}, \mathbf{x} \sim \Gamma(m/2 + \alpha_\lambda, \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{R}^{-1}}^2 + \beta_\lambda)$$

$$\delta \mid \mathbf{b}, \mathbf{x} \sim \Gamma(n/2 + \alpha_\delta, \frac{1}{2} \|\mathbf{x} - \boldsymbol{\mu}\|_{\mathbf{Q}^{-1}}^2 + \beta_\delta)$$

$$\mathbf{x} \mid \mathbf{b}, \lambda, \delta \sim \mathcal{N}\left((\lambda \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} + \delta \mathbf{Q}^{-1})^{-1} \lambda \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}, (\lambda \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} + \delta \mathbf{Q}^{-1})^{-1}\right)$$

where $\|\mathbf{x}\|_{\mathbf{M}} = \sqrt{\mathbf{x}^\top \mathbf{M} \mathbf{x}}$

```

1: Initialize  $(\lambda^0, \delta^0)$  and  $\hat{\mathbf{x}}^0 = (\lambda^0 \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} + \delta^0 \mathbf{Q}^{-1})^{-1} \lambda^0 \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}$ , set  $J$ 
2: for  $j = 1 : J$ 
3:   Compute  $\lambda^j \sim \Gamma(m/2 + \alpha_\lambda, \frac{1}{2} \|\mathbf{Ax}^{j-1} - \mathbf{b}\|_{\mathbf{R}^{-1}}^2 + \beta_\lambda)$ 
4:   Compute  $\delta^j \sim \Gamma(n/2 + \alpha_\delta, \frac{1}{2} \|\mathbf{x}^{j-1} - \boldsymbol{\mu}\|_{\mathbf{Q}^{-1}}^2 + \beta_\delta)$ 
5:   Compute  $\mathbf{x}^j \sim \mathcal{N}\left((\lambda^j \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} + \delta^j \mathbf{Q}^{-1})^{-1} \lambda^j \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}, (\lambda^j \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} + \delta^j \mathbf{Q}^{-1})^{-1}\right)$ 
6: end for
    
```

Sampling from a Gaussian Distribution with genGK

Note in line 5 of the hierarchical Gibbs sampling algorithm:

- λ, δ fixed
- $\mathbf{x} \mid \mathbf{b}, \lambda, \delta \sim \mathcal{N}(\hat{\mathbf{x}}, \boldsymbol{\Gamma})$ with

$$\boldsymbol{\Gamma} = (\lambda \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A} + \delta \mathbf{Q}^{-1})^{-1} \quad \text{and} \quad \hat{\mathbf{x}} = \lambda \boldsymbol{\Gamma} \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}$$

Compute samples, \mathbf{x}^j , as

$$\mathbf{x}^j = \hat{\mathbf{x}} + \boldsymbol{\Gamma}^{1/2} \boldsymbol{\xi}, \quad \text{with} \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

- Estimate $\hat{\mathbf{x}}$ with genGK iterative methods

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} -\log \pi(\mathbf{x} \mid \mathbf{b}, \lambda, \delta) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \frac{\lambda}{2} \|\mathbf{Ax} - \mathbf{b}\|_{\mathbf{R}^{-1}}^2 + \frac{\delta}{2} \|\mathbf{x}\|_{\mathbf{Q}^{-1}}^2$$

1. Introduce a change in variables $\mathbf{s} = \mathbf{Q}^{-1} \mathbf{x}$

$$\min_{\mathbf{s} \in \mathbb{R}^n} \frac{\lambda}{2} \|\mathbf{AQs} - \mathbf{b}\|_{\mathbf{R}^{-1}}^2 + \frac{\delta}{2} \|\mathbf{s}\|_{\mathbf{Q}}^2$$

2. Project transformed problem onto a Krylov subspace of increasing dimension
3. Using genGK process to generate a basis \mathbf{V}_k of subspace: Let $\gamma_1 = \|\mathbf{b}\|_{\mathbf{R}^{-1}}$

$$\mathbf{U}_{k+1} \gamma_1 \mathbf{e}_1 = \mathbf{b}, \quad \mathbf{AQV}_k = \mathbf{U}_{k+1} \mathbf{B}_k, \quad \text{and} \quad \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} = \mathbf{V}_k \mathbf{B}_k^\top + \alpha_{k+1} \mathbf{v}_{k+1} \mathbf{e}_{k+1}^\top$$

with

$$\mathbf{U}_{k+1}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} = \mathbf{I}_{k+1} \quad \text{and} \quad \mathbf{V}_k^\top \mathbf{Q} \mathbf{V}_k = \mathbf{I}_k$$

4. Let $\mathbf{s}_k = \mathbf{V}_k \mathbf{y}_k \implies$ projected problem

$$\min_{\mathbf{s}_k \in \mathcal{R}(\mathbf{V}_k)} \frac{\lambda}{2} \|\mathbf{AQs}_k - \mathbf{b}\|_{\mathbf{R}^{-1}}^2 + \frac{\delta}{2} \|\mathbf{s}_k\|_{\mathbf{Q}}^2 \iff \min_{\mathbf{y}_k \in \mathbb{R}^k} \frac{\lambda}{2} \|\mathbf{B}_k \mathbf{y}_k - \gamma_1 \mathbf{e}_1\|_2^2 + \frac{\delta}{2} \|\mathbf{y}_k\|_2^2$$

5. Project back: $\mathbf{x}_k = \mathbf{Qs}_k = \mathbf{QV}_k \mathbf{y}_k \approx \hat{\mathbf{x}}$

- Estimate $\boldsymbol{\Gamma}^{1/2}$ using genGK

$$\begin{aligned} \boldsymbol{\Gamma}^{1/2} &= \delta^{-1/2} \mathbf{Q}^{1/2} \left(\mathbf{I} + \frac{\lambda}{\delta} \mathbf{Q}^{1/2} \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{AQ}^{1/2} \right)^{-1/2} \\ &\approx \delta^{-1/2} \mathbf{Q}^{1/2} \left(\mathbf{I} + \frac{\lambda}{\delta} \underbrace{\mathbf{Q}^{1/2} \mathbf{V}_k \mathbf{B}_k^\top \mathbf{B}_k \mathbf{V}_k^\top \mathbf{Q}^{1/2}}_{\mathbf{Z}_k \boldsymbol{\Theta}_k \mathbf{Z}_k^\top} \right)^{-1/2} \\ &= \delta^{-1/2} \mathbf{Q}^{1/2} (\mathbf{I} - \mathbf{Z}_k \mathbf{D}_k \mathbf{Z}_k^\top) \end{aligned}$$

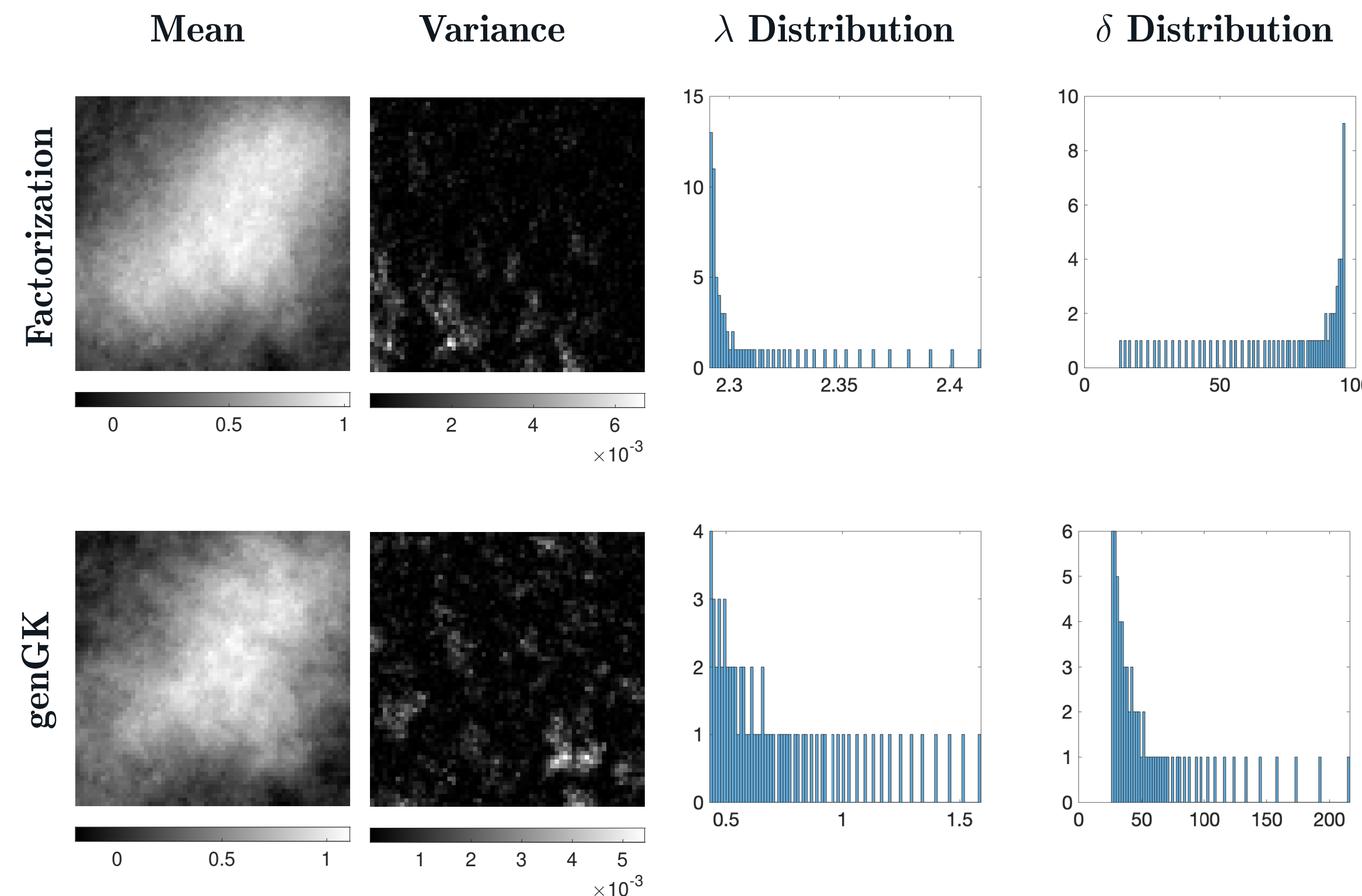
where $\mathbf{D}_k \equiv \mathbf{I}_k - (\mathbf{I}_k + \frac{\lambda}{\delta} \boldsymbol{\Theta}_k)^{-1/2}$

- Draw the sample

$$\mathbf{x}^j = \mathbf{x}_k + \delta^{-1/2} \mathbf{Q}^{1/2} (\mathbf{I} - \mathbf{Z}_k \mathbf{D}_k \mathbf{Z}_k^\top) \boldsymbol{\xi}$$

Seismic Tomography Results

- \mathbf{x}_{true} represents a vectorized 64×64 image of the Earth's interior
- observations $\mathbf{b} \in \mathbb{R}^{3200}$ contains simulated measurements with 2% noise
- compare to hierarchical Gibbs with full factorization of \mathbf{Q} : Matérn Kernel with smoothness $\nu = 0.5$ and correlation $\ell = 0.25$

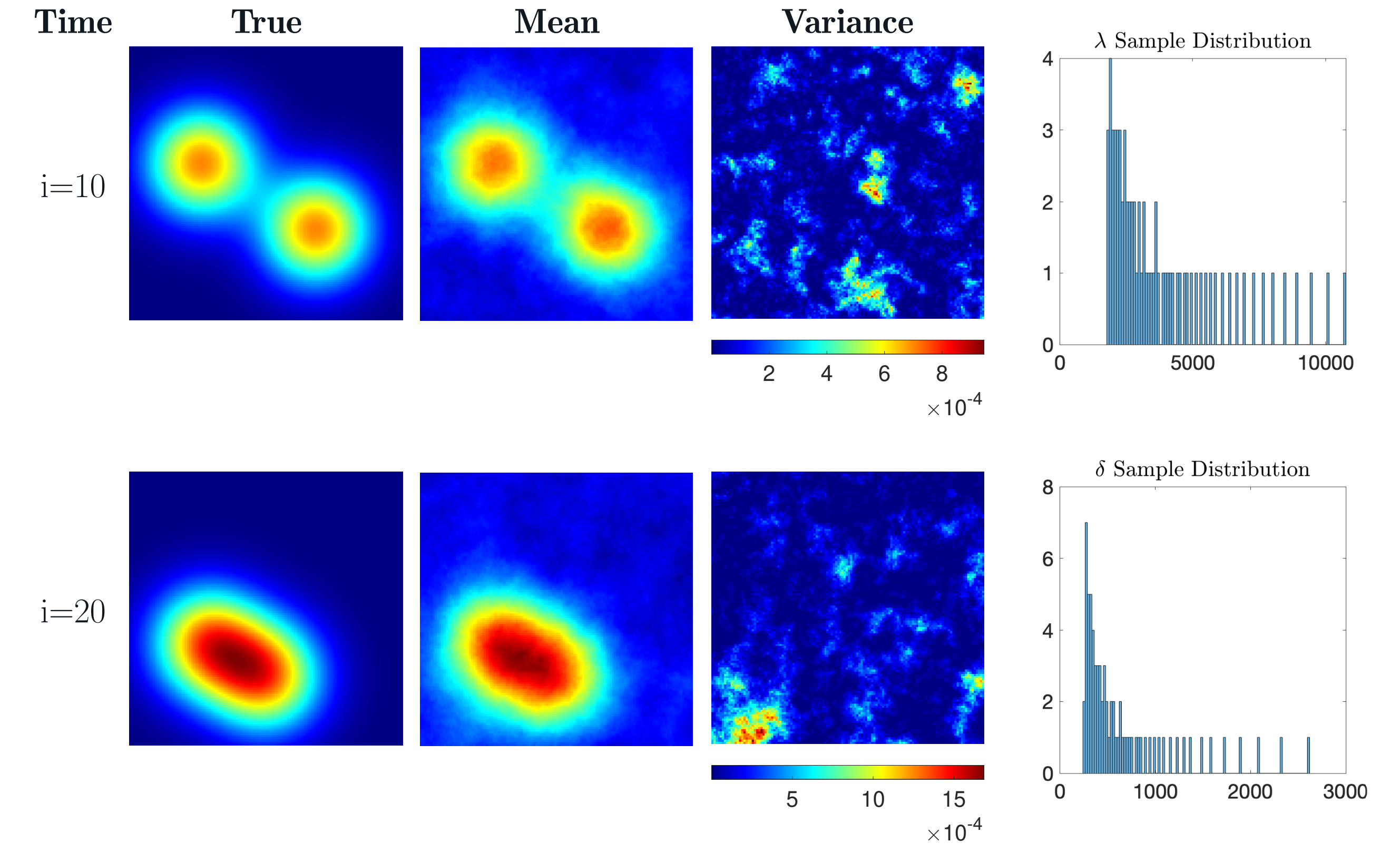


Dynamic Photoacoustic Tomography Results

- $\mathbf{x}^{(i)} \in \mathbb{R}^{16,384}$ is a vectorized 128×128 image
- forward model $\mathbf{A} \in \mathbb{R}^{65,160 \times 327,680}$ is spherical projection with 18 angles
- observations $\mathbf{b} \in \mathbb{R}^{65,160}$ contain projection data with 2% noise

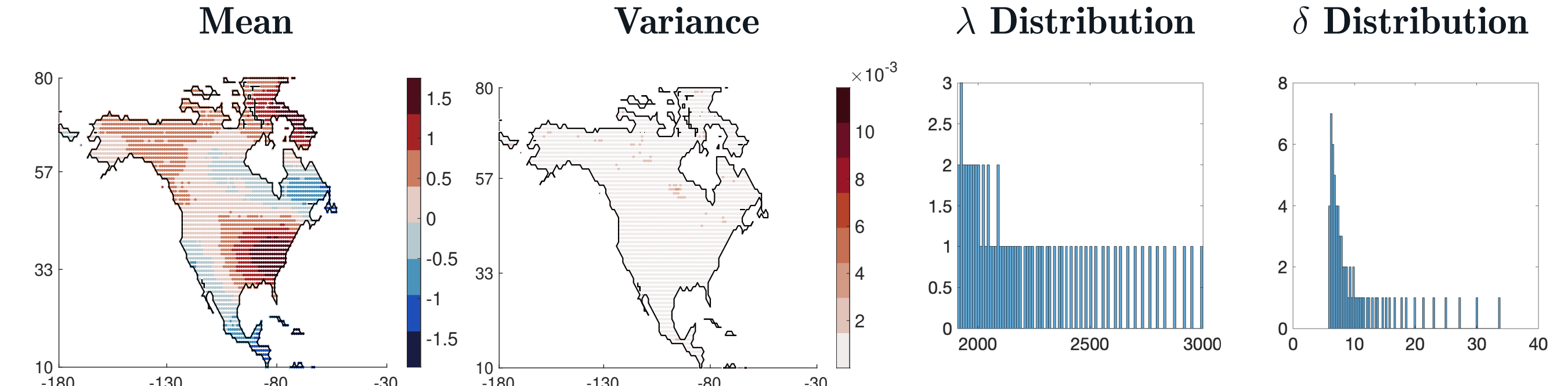
$$\mathbf{x}_{\text{true}} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \vdots \\ \mathbf{x}^{(20)} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{b}^{(1)} \\ \vdots \\ \mathbf{b}^{(20)} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}^{(1)} & & \\ & \ddots & \\ & & \mathbf{A}^{(20)} \end{bmatrix}, \quad \mathbf{Q} = \mathbf{Q}_t \otimes \mathbf{Q}_s$$

- $\mathbf{Q}_s, \mathbf{Q}_t$ Matérn Kernel with smoothness $\nu = 0.5, \nu = 2.5$ and correlations $\ell = 0.25, \ell = 0.1$



Atmospheric Inverse Modeling Results

- Solution $\mathbf{x} \in \mathbb{R}^{11,900}$ is vectorized 2D image of average CO₂ emissions over North America
- observations $\mathbf{b} \in \mathbb{R}^{98,880}$ contains satellite data with 4% noise
- \mathbf{Q} : Matérn Kernel with smoothness $\nu = 2.5$ and correlation $\ell = 0.05$



Conclusions

- UQ for large-scale inverse problems can be challenging
- For Gaussian posterior distributions, generalized Krylov methods can be used to efficiently
 - compute point estimates
 - perform UQ (e.g., draw samples)
- For non-Gaussian posterior distributions, hierarchical Bayesian approaches are needed
 - Sampling can be computationally infeasible and slow
 - Generalized Krylov methods can be exploited within hierarchical Bayesian approaches

References

- Acknowledgments:** This work was partially supported by the National Science Foundation ATD program under grants DMS-2026841, 2026830, and 2026835. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
- [1] Julianne Chung and Arvind K. Saibaba. “Generalized Hybrid Iterative Methods for Large-Scale Bayesian Inverse Problems”. In: *SIAM Journal on Scientific Computing* 39.5 (2017), S24–S46.
 - [2] Silvia Gazzola, Per Christian Hansen, and James G. Nagy. “IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems”. In: *Numerical Algorithms* 81.3 (2019), pp. 773–811.
 - [3] Arvind Saibaba, Julianne Chung, and Katrina Petroske. “Efficient Krylov subspace methods for uncertainty quantification in large Bayesian linear inverse problems”. In: *Numerical Linear Algebra with Applications* 27 (Aug. 2020).