Funding: NA14OAR4310136 NNX14AH02G

Optimal and Scalable Methods for Atmospheric Source Inversions

Nicolas Bousserez[1*]and Daven K. Henze[1]

[1] University of Colorado, Boulder, CO, USA; *Now at ECMWF, Reading, UK

INTRODUCTION: The increasing availability of satellite-based observations of atmospheric greenhouse gas (GHG) (e.g., GOSAT, OCO-2) offers an unprecedented opportunity to better monitor and understand the global carbon cycle. Current atmospheric source inversion systems using large satellite datasets are often limited by the computational cost of performing long-term inversions as well as quantifying the associated uncertainties. Previous approaches have relied on various dimension reduction techniques to make those computations more tractable, although they have often fallen short of combining both theoretical optimality and computational scalability. To address those limitations, we introduce a new approach for high-dimensional GHG source inversions, which combines cutting-edge randomization methods for large matrices decomposition and optimal dimension reduction techniques that maximize observational constraints. The resulting algorithm dramatically improves the computational scalability of the inversions. An additional and equally useful feature of this method is its ability to provide as a by-product of the optimization the spatiotemporal flux patterns that are independently and most constrained by the observations along with their associated posterior errors. Such diagnostics are crucial to better understand the information content of satellite-based observations used in current GHG source inversion systems and enable to answer critical questions such as: can current satellite observations provide seasonal to sub-seasonal constraints on the carbon cycle? At which spatial resolutions can those observations add significant information to the prior (bottom-up) estimates? Those aspects are illustrated with a pseudo-experiment based on a monthly methane source inversion over North America using GOSAT XCH4 column observations.

Methodology

Formalism and Notations

- <u>Bayesian framework:</u>

$$p(x|y) = \frac{\overset{\text{likelihood}}{p(y|x)}\overset{\text{prior}}{p(x)}}{\int p(y|x)p(x)dx}$$

- <u>Maximum a posteriori (normal distribution):</u>

$$\Rightarrow \mathbf{x}^a \equiv \operatorname{Arg}\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$$

$$= \operatorname{Arg}\min_{\mathbf{x}} J(\mathbf{x})$$

with: $$J(\mathbf{x}) \equiv \frac{1}{2}(\underset{\text{forward model}}{H}\mathbf{x}-\mathbf{y})^T\underset{\text{model-data error covariance}}{\mathbf{R}}^{-1}(H\mathbf{x}-\underset{\text{data}}{\mathbf{y}}) + \frac{1}{2}(\mathbf{x}-\underset{\text{prior}}{\mathbf{x}^b})^T\underset{\text{prior error covariance}}{\mathbf{B}}^{-1}(\mathbf{x}-\mathbf{x}^b)$$

<u>Dimensions:</u>

$\mathbf{B}$: nxn
$\mathbf{H}$: mxn
$\mathbf{R}$: mxm
$\mathbf{x}^b$: n
$\mathbf{y}$: m

Important Properties

- Diagonalization of the averaging kernel matrix:

$$\mathbf{A} = \mathbf{B}^{1/2}\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{B}^{-1/2}$$

$\mathbf{B}^{1/2}\mathbf{v}_i$ **independently constrained modes** → $d_i = \frac{\lambda_i}{1+\lambda_i}$ **relative contribution from observations**

- Observations are most informative along the first *p* eigenvectors $\{\mathbf{B}^{-1/2}\mathbf{v}_i\}_{i=1,..,p}$
- Minimization of averaged relative error:

$$\boxed{\mathbb{E}\|\mathbf{x}^a_{red}-\mathbf{x}^a\|^2_{(\mathbf{P}^a)^{-1}} = \min_{\Pi}\mathbb{E}\|\mathbf{x}^a_{\Pi}-\mathbf{x}^a\|^2_{(\mathbf{P}^a)^{-1}} = \sum_{i>k}\frac{\lambda_i}{1+\lambda_i}}$$

with $\|\mathbf{x}\|^2_{(\mathbf{P}^a)^{-1}} = \mathbf{x}^T(\mathbf{P}^a)^{-1}\mathbf{x}.$

Posterior Sampling

- The square-root of the optimal posterior error covariance approximation can be used to sample the posterior errors:

$$\delta\mathbf{w}_i = \mathbf{S}_{\text{red}}\xi_i,\ i=1,...,p \longrightarrow \text{Posterior samples}$$

with $$\begin{cases}\xi_i \sim \mathcal{N}(0,1) \\ \mathbf{S}_{\text{red}} = \mathbf{B}^{1/2}\sum_{i=1}^{p}(1+\lambda_i)^{-1/2}\mathbf{v}_i\mathbf{v}_i^T. \\ \mathbf{P}^a_{\text{red}} = \mathbf{S}_{\text{red}}\mathbf{S}^T_{\text{red}}\end{cases}$$

- Very cheap compared to ensemble of inversions (i.e., EDA).

Optimal Dimension Reduction

Prior-preconditioned Hessian

$$\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2} = \sum_i^k \lambda_i\mathbf{v}_i\mathbf{v}_i^T$$

Truncated Singular Value Decomposition (SVD)

Projection operator

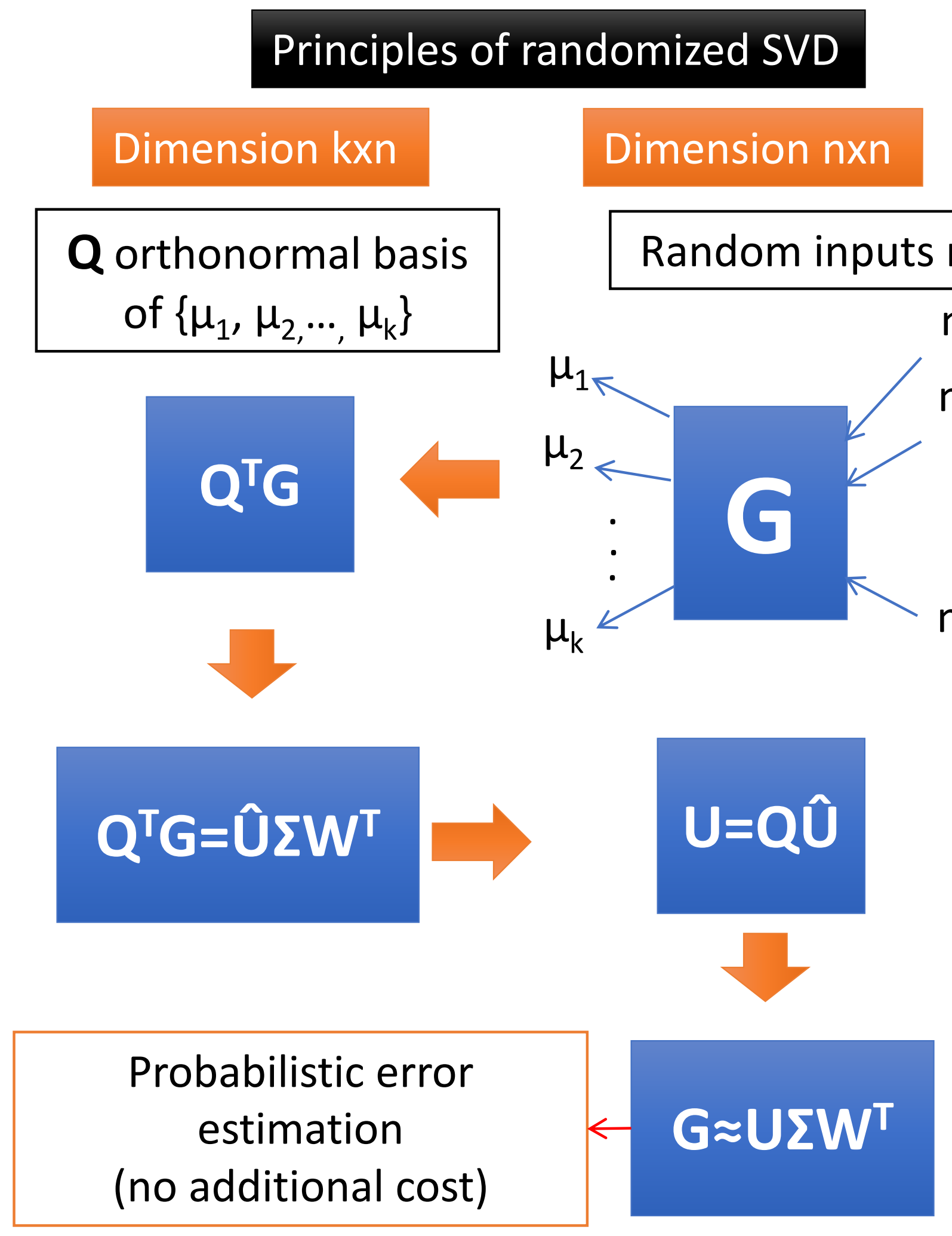
$$\mathbf{\Pi}_k = \mathbf{B}^{1/2}\sum_i^k\mathbf{v}_i\mathbf{v}_i^T\mathbf{B}^{-1/2}$$

Posterior mean of projected Bayesian problem

$$\mathbf{x}^a_{\Pi_k} = \mathbf{B}^{1/2}\left[\sum_i^k(1+\lambda_i)^{-1}\mathbf{v}_i\mathbf{v}_i^T\right]\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}^b))$$

Scalable Computation of the Optimal Projection

- Standard matrix-free SVD algorithms based on **Krylov subspace** approach:
  - Symmetric operator: Lanczos.
  - General operator: Arnoldi.
  - →Numerically **unstable**.
  - →**Sequential**: computationally expensive.
- **Randomized SVD** of large matrices (*Halko et al., 2011*):
  - **Excellent convergence** properties in general.
  - **Inherently stable**.
  - **Parallel implementation**.
  - Cheap probabilistic posterior estimate of SVD accuracy.



Principles of randomized SVD

Dimension kxn

Dimension nxn

Q orthonormal basis of {μ_1, μ_2,…, μ_k}

Random inputs η_i

$Q^TG=\hat{U}\Sigma W^T$

$U=Q\hat{U}$

$G\approx U\Sigma W^T$

Probabilistic error estimation (no additional cost)

<u>Method:</u>

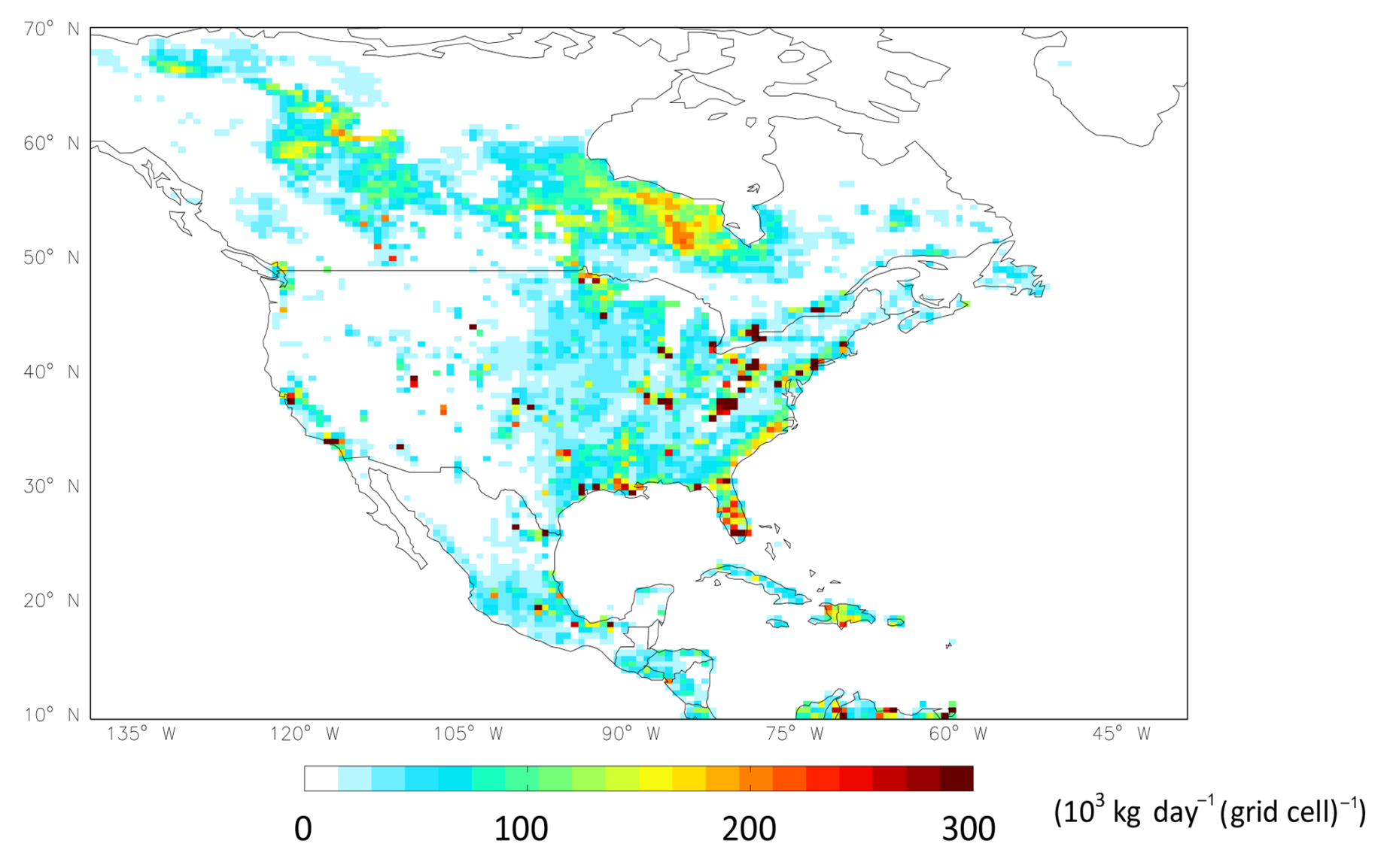
- A randomized SVD algorithm is used to compute a truncated eigendecomposition of the prior-preconditioned Hessian ($\mathbf{B}^{1/2}\mathbf{H}\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$).
- The posterior flux mode, posterior errors and the degree of freedom for signal (DOFs) for the reduced problem are computed analytically using the eigenvectors and eigenvalues of the prior-preconditioned Hessian.

<u>Reference</u>: Bousserez, N. and Henze, D.K., Optimal and Scalable Methods to Approximate the Solutions of Large-Scale Bayesian Problems: Theory and Application to Atmospheric Inversion and Data Assimilation. *Q.J.R. Meteorol. Soc.*, doi:10.1002/qj.3209

Satellite-based methane inversion

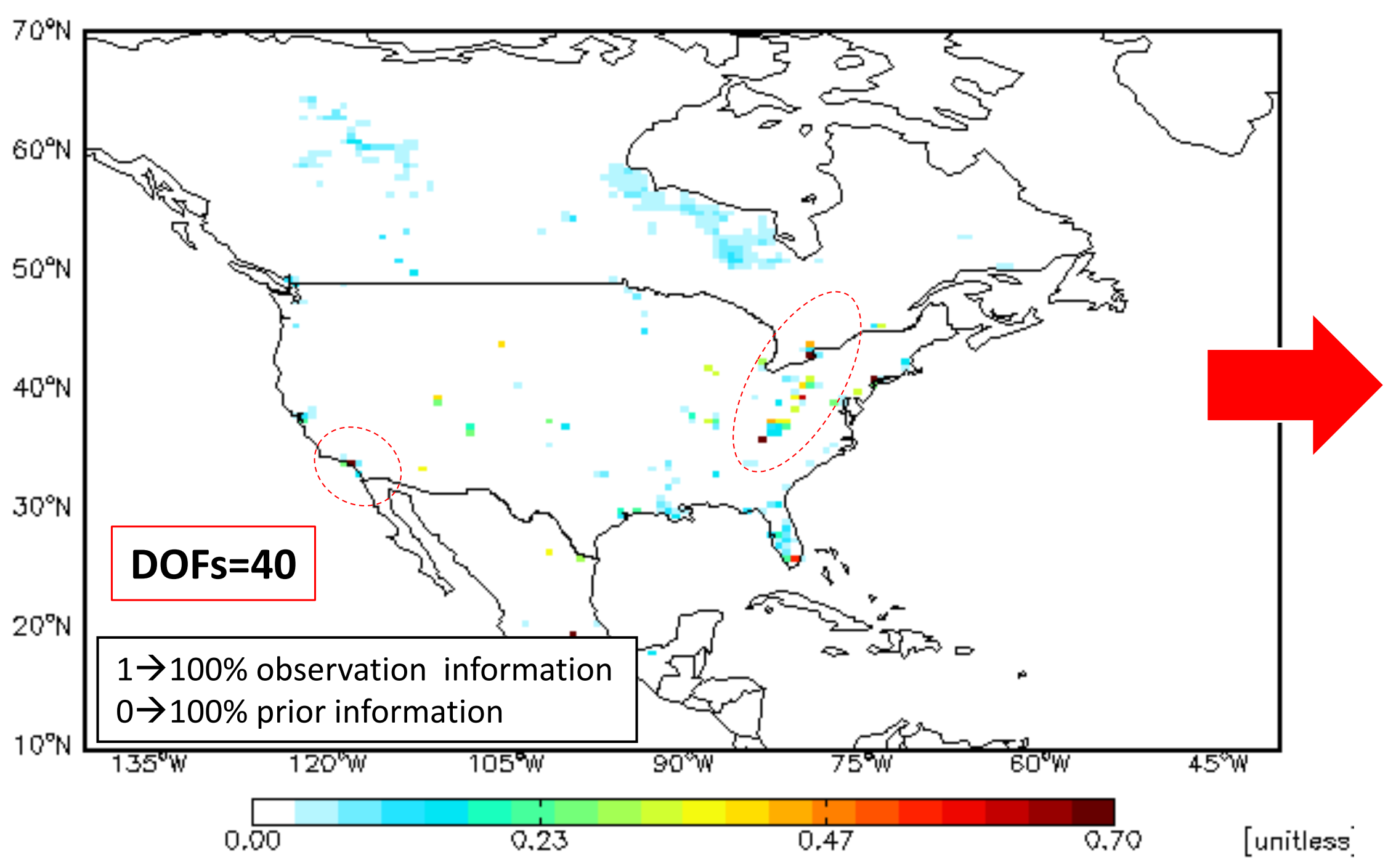
Observation Simulation System Experiment (OSSE)

- Methane fluxes inversion for July 2008.
- GEOS-Chem CTM forward and adjoint models.
- Nested North America domain at 0.5°x0.6°.
- Prior error of 40% of prior emissions.
- GOSAT XCH_4 columns data.
- **Control space dimension *n*=18271.**
- **Randomized SVD to compute optimal approximation.**



Total daily prior CH_4 emissions (EDGAR v4.2)

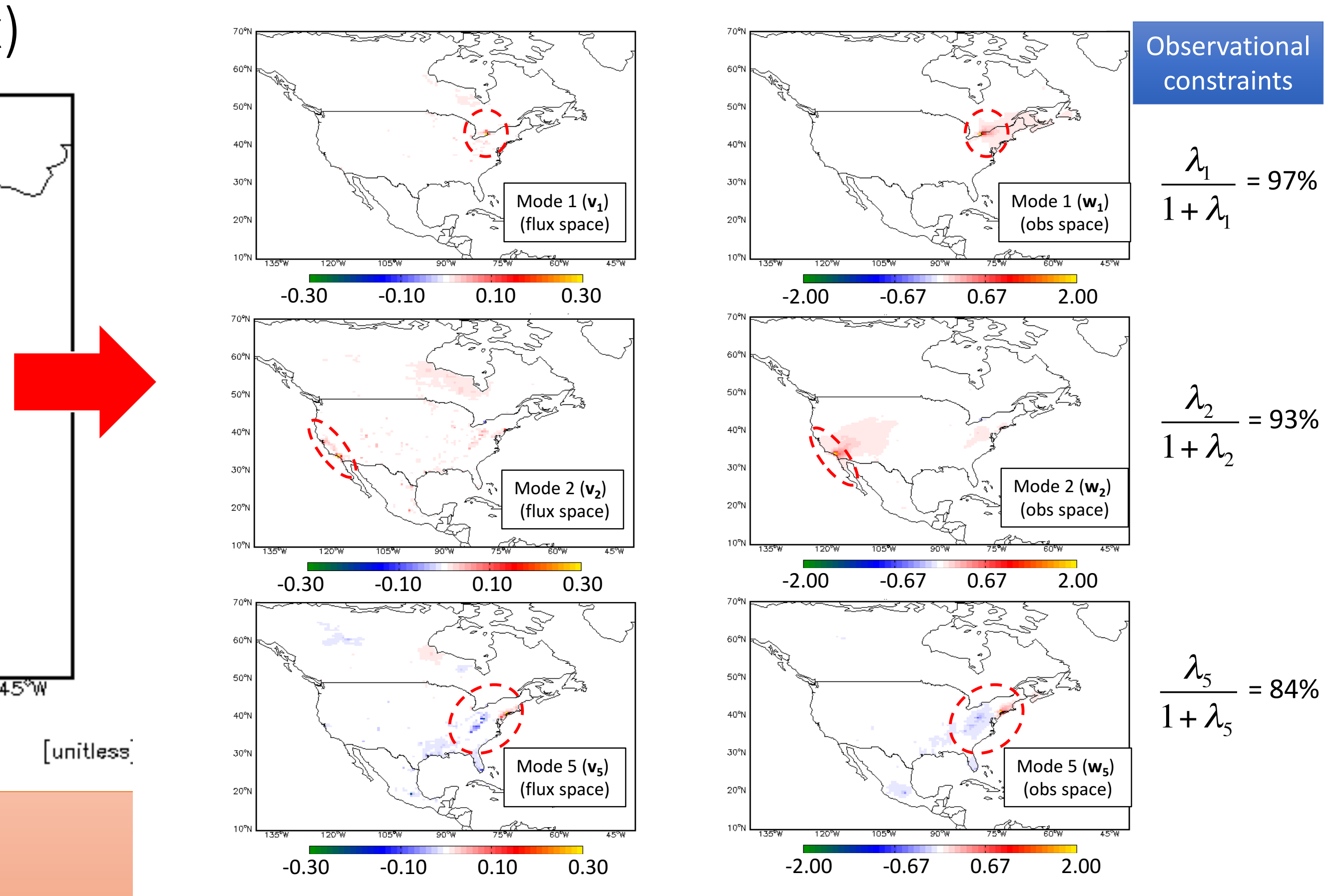
Resolution of Inversion (diagonal of averaging kernel matrix)



DOFs=40

1→100% observation information
0→100% prior information

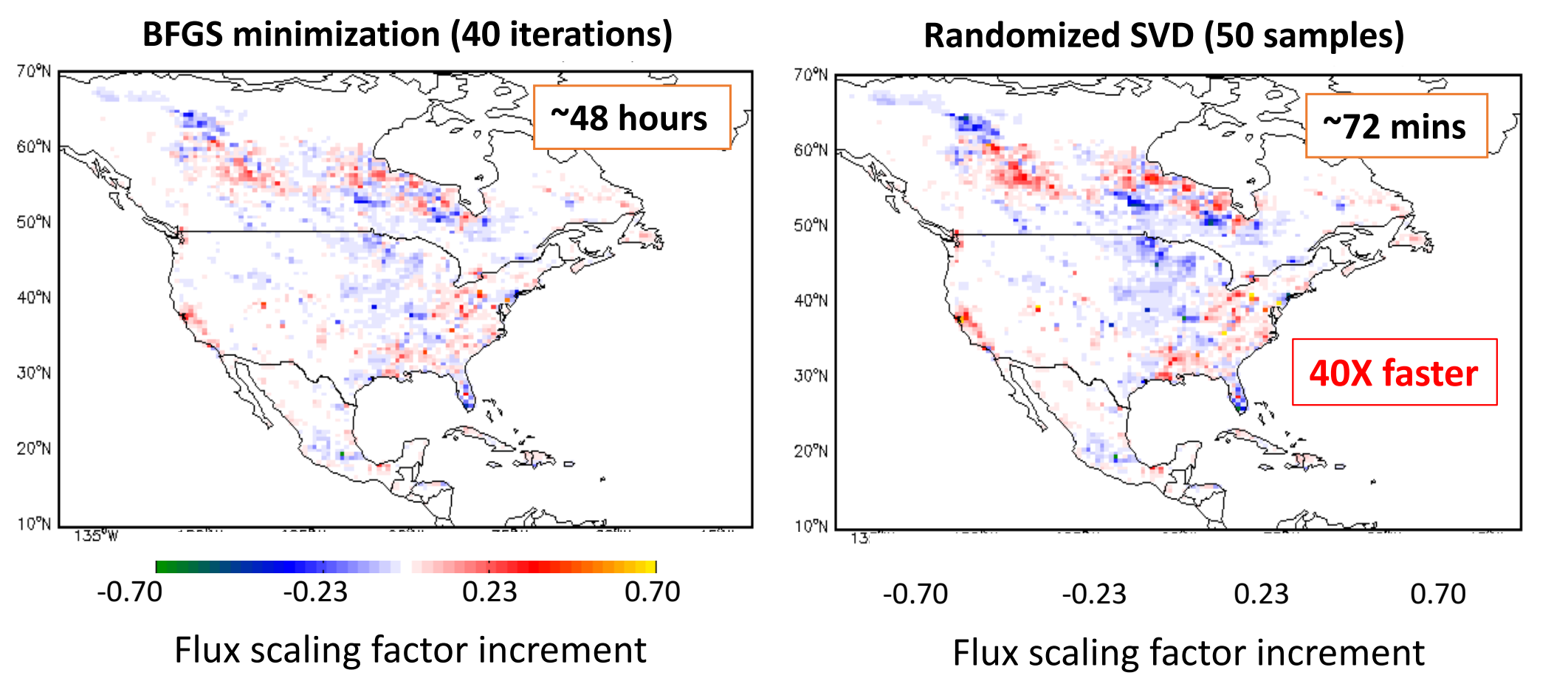
- Only a few methane fluxes can be resolved by the observations at grid-scale resolution.
- What are the spatial modes resolved by the inversion?



<u>Information content analysis:</u>

- Singular vectors in flux space represent independently constrained patterns.
- Singular vectors in observation space represent associated observational patterns.
- Singular values quantify observational constraints.

Computational Performance



Flux scaling factor increment

Flux scaling factor increment

<u>Optimization:</u>

- BFGS requires 40 iterations⇔40 sequential forward and adjoint runs (walltime~ 2 days).
- Randomization requires 50 samples⇔ 50 forward and adjoint runs in parallel (walltime~ 72 mins).

Current & Future work

- The method has been be extended to non-linear inverse problems such as incremental 4D-Var by replacing the iterative conjugate-gradient minimization by a randomized SVD approach.
- The optimal projection framework can be applied to error tuning approaches (e.g., Desroziers and Ivanov (2001)) to provide fast and spatially resolved prior (i.e., bottom-up) error estimates.