Optimizing Confidence Intervals for Satellite-Based Carbon Flux Inversion

Michael (Mike) Stanley | Mikael Kuusela | April 13, 2022 | SIAM UQ

Thanks and acknowledgements

- Pratik Patil (CMU)
- Brendan Byrne (JPL)
- Junjie Liu (JPL)
- Peyman Tavallali (JPL)
- Daven Henze (UC Boulder)
- JPL UQ and CMU STAMPS

This work was supported by NSF grants DMS-2053804 and PHY-2020295 as well as JPL RSA No. 1670375.

Goal: provide UQ for carbon flux (CF) inversion

1. Provide a quick recap of CF inversion motivation,
2. Define our operational setup to test inversion methods,
3. Review a few other CF UQ approaches and some associated challenges,
4. Explain our methodological attempt to address these challenges,
5. Demonstrate our method on a toy problem,
6. And, elucidate our method’s application to CF UQ.

Identifying terrestrial carbon sources/sinks requires good UQ

- Insufficient understanding of natural carbon sinks creates a gap in our knowledge of the global carbon budget.
- Carbon fluxes inferred from satellite observations and a chemical transport model (CTM) come with uncertainty.
- We can use an Observing System Simulation Experiment (OSSE) to investigate these properties with some frequently used inversion elements.
  - Our OSSEs are defined over 8 months - Jan 1, 2010 to September 1, 2010

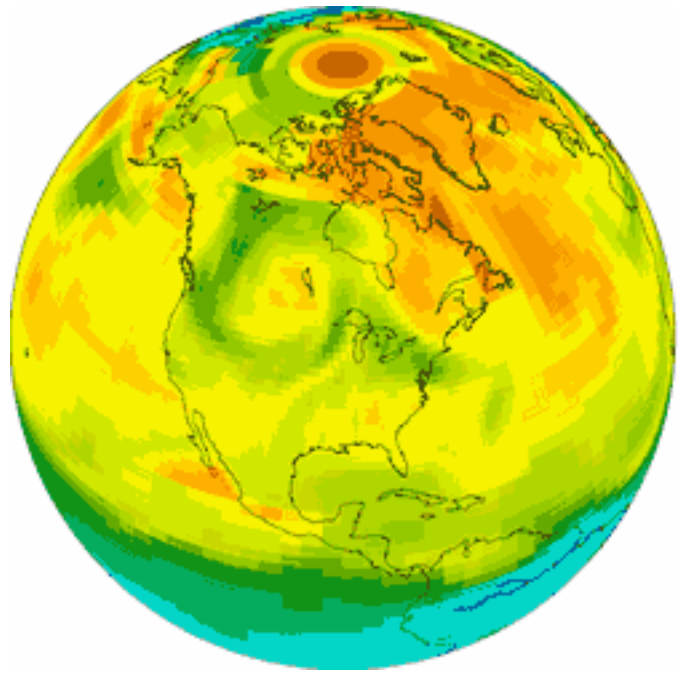
Inversion Elements

Simulator/Forward Model

Satellite Observations

True Fluxes*

Control Fluxes*



CarbonTracker

*: Net Ecosystem Exchange (NEE) Fluxes

4D-Var and GEOS-Chem Adjoint can provide flux estimation and UQ

$$J(\boldsymbol{c}) = \boxed{\frac{1}{2}(\boldsymbol{c} - \boldsymbol{c}_a)^\top \boldsymbol{S}_a^{-1}(\boldsymbol{c} - \boldsymbol{c}_a)} + \boxed{\frac{1}{2}(\boldsymbol{y} - \boldsymbol{K}_x\boldsymbol{c})^\top \boldsymbol{S}_O^{-1}(\boldsymbol{y} - \boldsymbol{K}_x\boldsymbol{c})}$$

- $\boldsymbol{c}$ - scaling factors to optimize (in $\mathbb{R}^{72\times46\times8}$)
- $\boldsymbol{c}_a$ - a priori scaling factors (set to unity)
- $\boldsymbol{S}_a, \boldsymbol{S}_O$ - a priori and observation covariance matrices
- $\boldsymbol{K}_x$ - forward model (GEOS-Chem + GOSAT XCO2 Observation Operator) with control flux x (we assume linear)
- $\boldsymbol{y}$ - satellite observations

Produces the MAP estimator - $\boldsymbol{c}_{MAP}(\boldsymbol{c}_a, \boldsymbol{y})$

Bayesian Interpretation

$$\boldsymbol{c} \sim \mathcal{N}(\boldsymbol{c}_a, d^2\boldsymbol{I})$$

$$\boldsymbol{y} \,|\, \boldsymbol{c} \sim \mathcal{N}(\boldsymbol{K}_x\boldsymbol{c}, \boldsymbol{S}_O)$$

$$\boldsymbol{c} \,|\, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$$

d^2 : Prior uncertainty

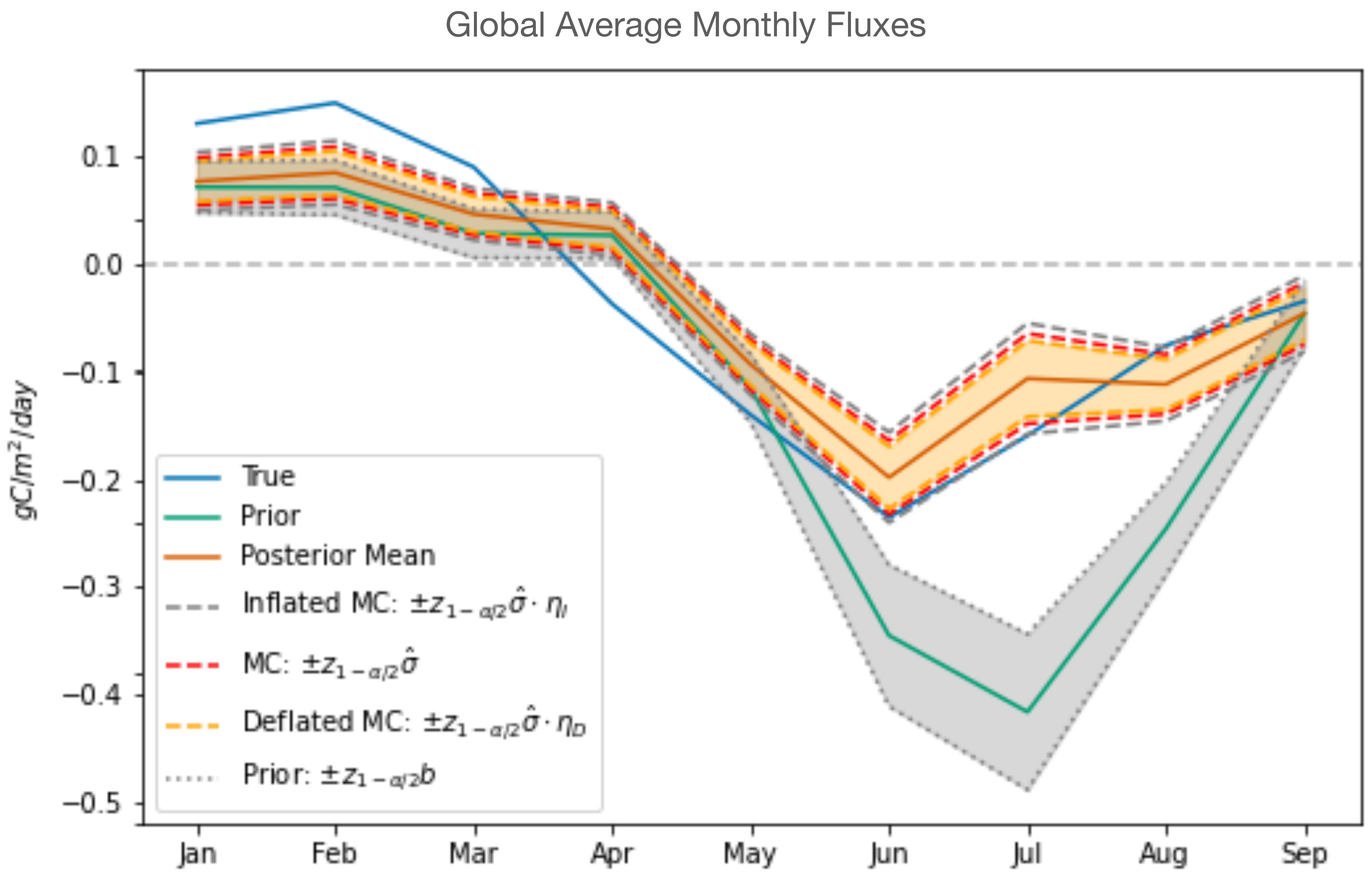
There are a variety of ways to approach Bayesian UQ for high-dimensional linear models

- **Monte Carlo-based** methods
    - Chevallier et al., 2007 | Bousserez et al., 2015 | Stanley et al., 2022
- **Low-Rank Hessian Approximation** methods
    - Flath et al., 2011 | Kalmikov and Heimbach, 2014 | Bousserez et al., 2015 | Bousserez and Henze, 2018
- **MCMC-based** approaches
    - WOMBAT, Zammit-Mangion, A., et al., 2022 | SN-MCMC Petra et al., 2007

A Monte Carlo method reveals a key challenge in the Bayesian formulation

- A misspecified prior distribution can make a problem well-posed at the cost of introducing a bias.

- With our OSSE and the MC method, we observe the effects of the prior's misspecification



We endeavor to eliminate the prior to avoid the misspecification effect

- We propose a prior-free approach based on [Patil et al., 2020] involving the direct computation of endpoints of a confidence interval with coverage guarantees

- With a linear forward model $\boldsymbol{K} \in \mathbb{R}^{m \times n}$, linear functional $\theta(\boldsymbol{x}) = \boldsymbol{h}^T \boldsymbol{x}$ and confidence level $1 - \alpha$,

Statistical Model

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{I})$$

CI Definition

$$\left[\underline{\theta}, \bar{\theta}\right] = \left[\min_{x \in C_\alpha} \theta(\boldsymbol{x}), \max_{x \in C_\alpha} \theta(\boldsymbol{x})\right]$$

Coverage Guarantee

where $\mathbb{P}\left(\theta^* \in [\underline{\theta}, \bar{\theta}]\right) \geq 1 - \alpha$, and

$\theta^* = \theta(x^*)$ and $\boldsymbol{x}^*$ is the true parameter value.

The definition of the feasible set, C_α, depends on the desired coverage properties

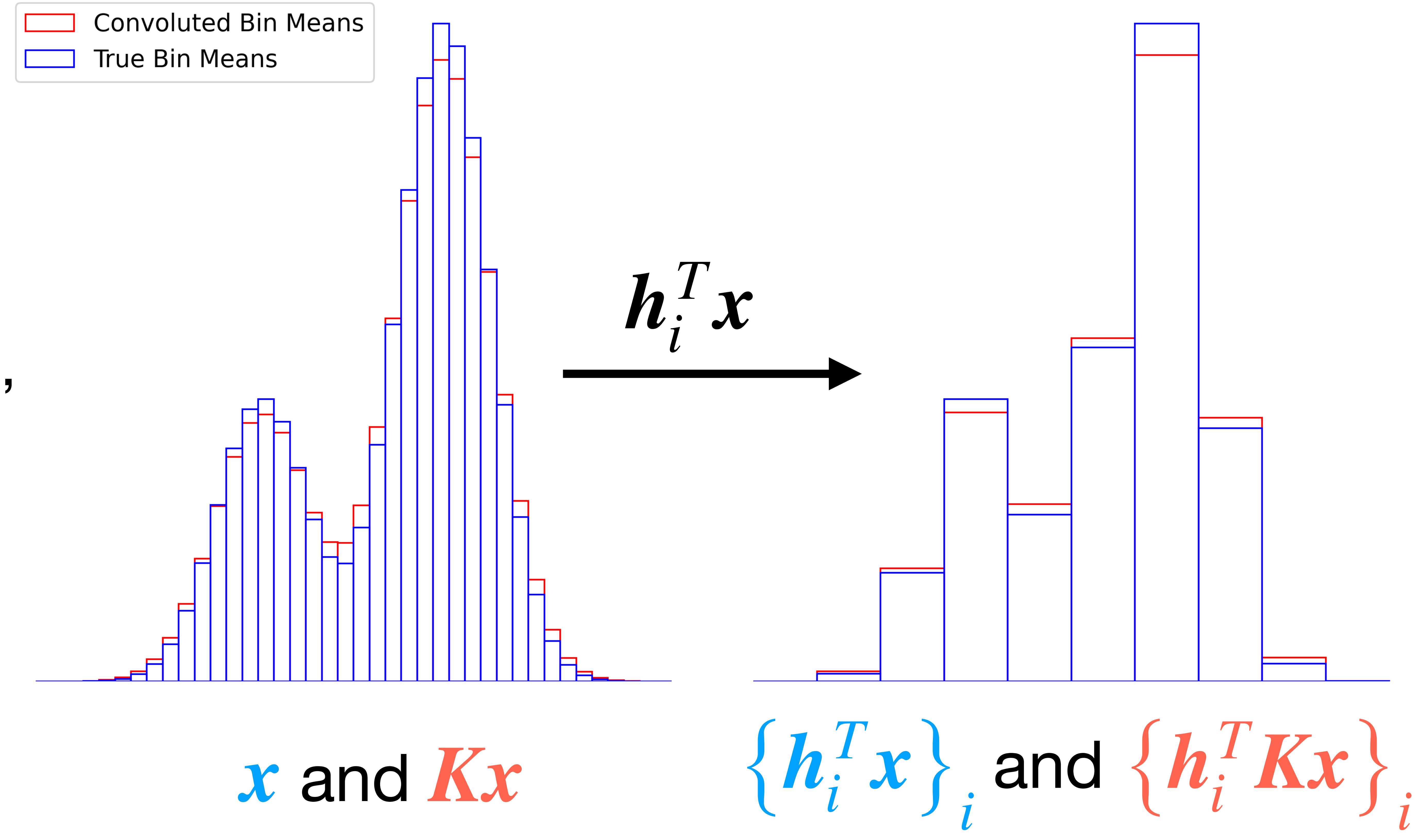
- Define $C_\alpha := \{\boldsymbol{x} \mid \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2 \leq \psi_\alpha^2$ and $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$
    - where $\boldsymbol{A}$ and $\boldsymbol{b}$ characterize physical constraints we might know about the problem.
- How we choose ψ_α^2 determines the coverage properties of the interval $\left[\underline{\theta}, \bar{\theta}\right]$
    - For ***simultaneous coverage*** : $\psi_\alpha^2 := \chi_\alpha^2(m)$ [Stark 1992]
    - For ***one-at-a-time coverage*** : $\psi_\alpha^2 := z_{1-\alpha/2}^2 + s^2$, where $s^2 = \min_{\boldsymbol{x}:\, \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}} \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|_2^2$ [Patil et al. 2020, Rust and O'Leary 1994]

We want to minimize interval length subject to a coverage guarantee

- The one-at-a-time ellipsoid constraint appears to have empirically correct coverage but is not yet provable [Patil et al. 2020].
- Based on the one-at-a-time intervals, in a decision theoretic sense, we can characterize interval endpoints as decision rules from a set $\mathscr{D}$, where all decision rules $\delta \in \mathscr{D}$ have desired coverage guarantee
  - Optimal decisions can then be characterized as those δ producing the shortest expected interval with respect to a prior —> "Prior-Optimized" (PO) Intervals
  - [Patil et al. 2020, Stanley et al. 2021].

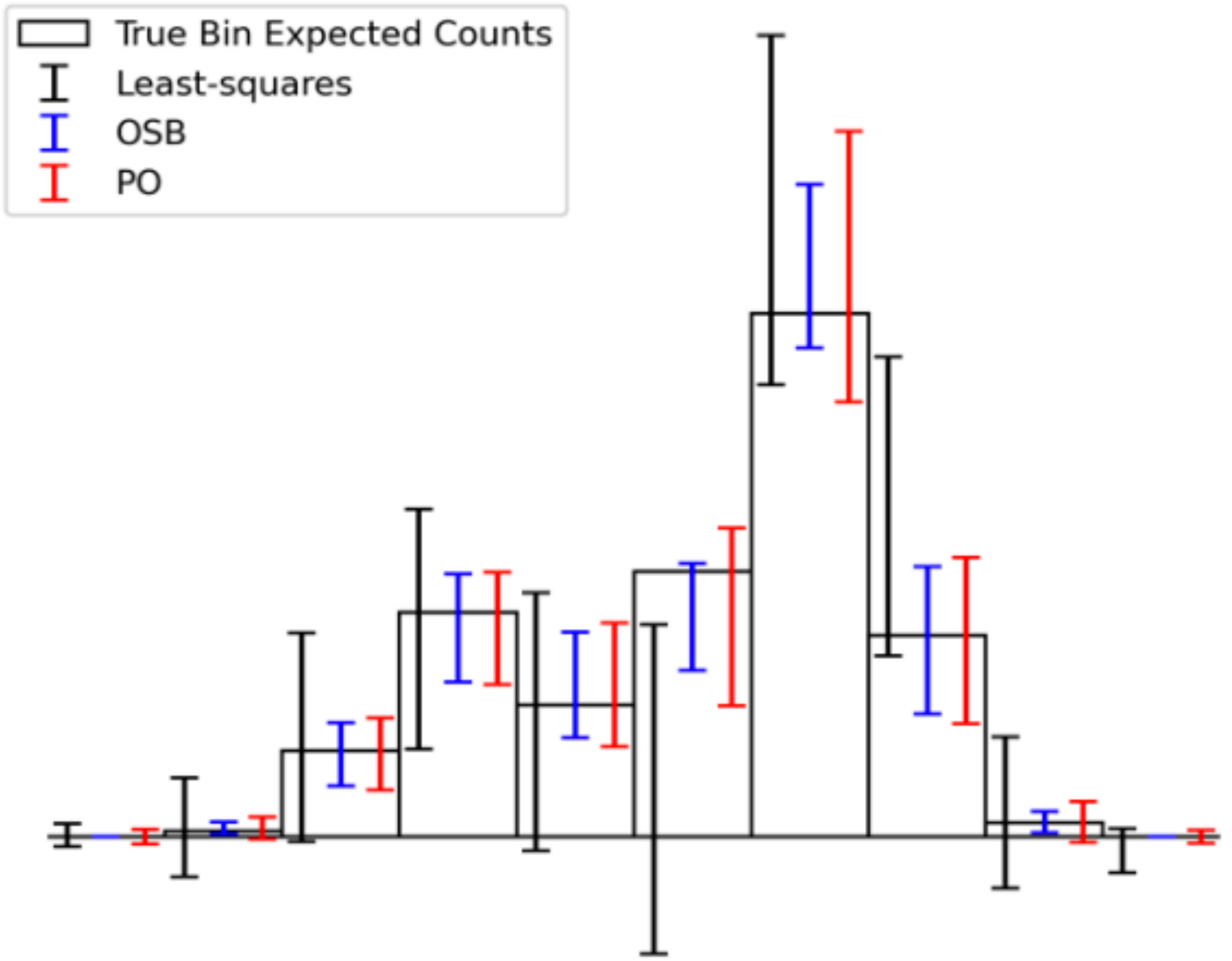
Method exposition on a toy-model: density deconvolution

- Using the same linear model
  - $\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}),$
- Information that each bin mean must be non-negative
  - $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$ where $\boldsymbol{A} = -\boldsymbol{I}$ and $\boldsymbol{b} = \boldsymbol{0}$,
- a collection of functionals $\{\boldsymbol{h}_i\}_{i=1}^{10}$,
- and a confidence level $1 - \alpha$, e.g. 95%



Using physical constraint information provides significant interval length improvement

- Seeing the data y, we perform inference on the functionals of the true bin counts by computing our CIs, $\left[\underline{\theta}, \bar{\theta}\right]_i$
- Interval Types
  - OSB == "One-at-a-time Strict Bounds" using $\psi_\alpha^2 := z_{1-\alpha/2}^2 + s^2$
  - PO == "Prior Optimized" (i.e., the decision theoretic framework)
  - Least-squares is equivalent to OSB with no $\boldsymbol{Ax} \leq \boldsymbol{b}$ constraint [Patil et al. 2020]



[Stanley et al. 2021]

Applying the method to CF inversion

Defining the ***lower endpoint*** optimization

- First, we make the reasonable assumption that the forward model, $\boldsymbol{K}$, is linear
- We use the $\psi_\alpha^2 := \chi_\alpha^2(m)$ ellipsoid constraint

$$\underset{\boldsymbol{c}}{\text{minimize}} \quad \boldsymbol{h}^T\boldsymbol{c}$$

$$\text{subject to} \quad (\boldsymbol{y} - \boldsymbol{K}_x\boldsymbol{c})^T\boldsymbol{S}_o^{-1}(\boldsymbol{y} - \boldsymbol{K}_x\boldsymbol{c}) \leq \chi_\alpha^2(m)$$

$$\boldsymbol{A}\boldsymbol{c} \leq \boldsymbol{b}$$

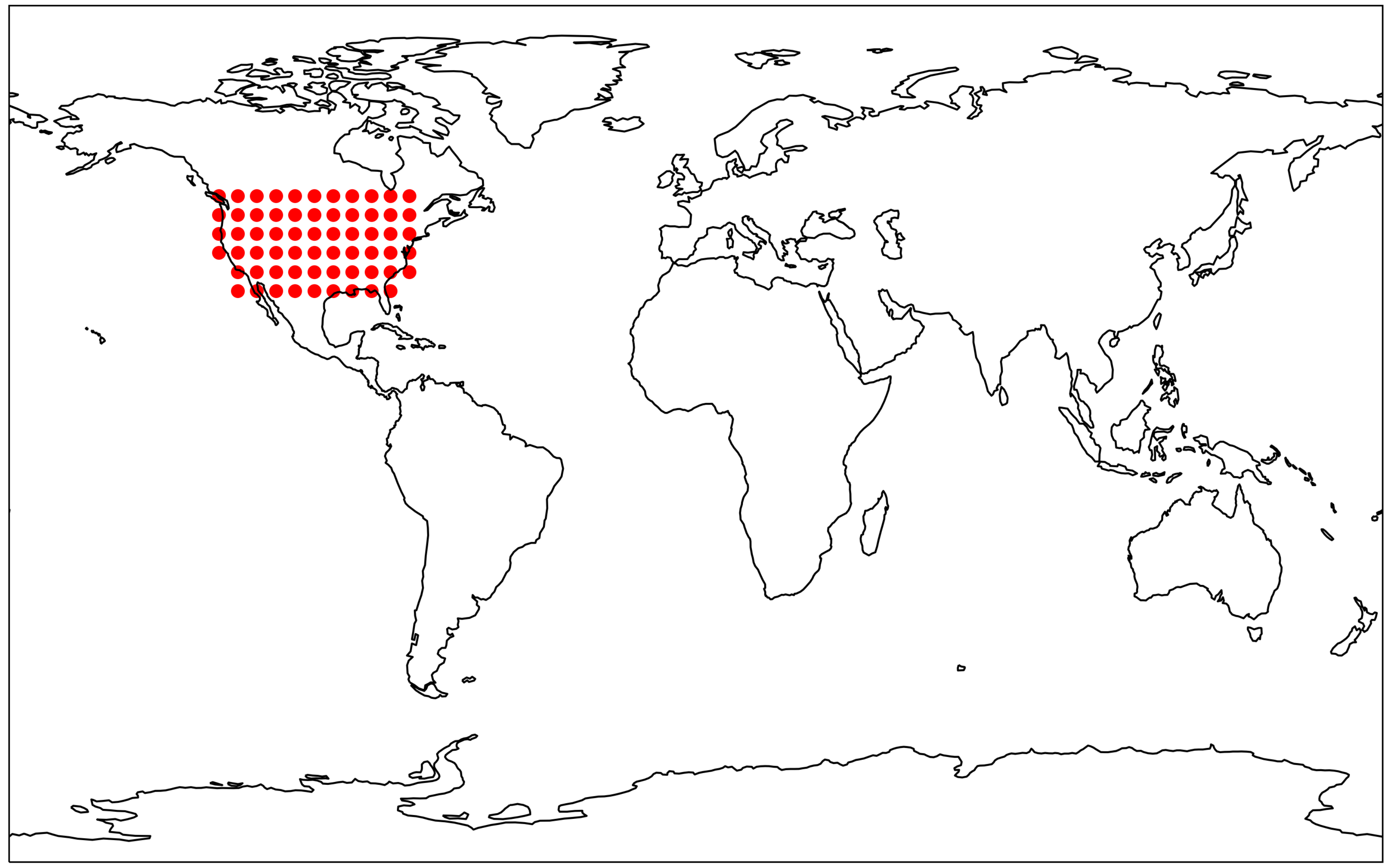
$\boldsymbol{h} \in \mathbb{R}^n$: functional of interest
$\boldsymbol{x} \in \mathbb{R}^{n'}$: control flux
$\boldsymbol{K}_x \in \mathbb{R}^{n \times m}$: forward operator with control flux x
$\boldsymbol{c} \in \mathbb{R}^n$: monthly scaling factors
$\boldsymbol{y} \in \mathbb{R}^m$: GOSAT XCO2 observations
$\boldsymbol{S}_O \in \mathbb{R}^{m \times m}$: observation error covariance
$\boldsymbol{A} \in \mathbb{R}^{s \times n}, \boldsymbol{b} \in \mathbb{R}^s$: physical constraint matrix/vector

Applying the method to CF inversion

Creating a functional

- A scientific question can motivate a functional's definition
  - e.g., what is the average June 2010 flux over the continental US?
- To define this functional $\boldsymbol{h}$,
  - let $h_i = \overline{NEE}_i \cdot a_i$, $i = 1,\ldots,n$, where
    - a_i's are the region area weights and
    - $\overline{NEE}_i$'s are the monthly average control fluxes

Spatial Locations of non-zero functional values



Applying the method to CF inversion

Creating the constraints

- Based on the idea that net ecosystem exchange (NEE) is decomposable into ecosystem respiration (R_e) and gross primary product (GPP) we have
  - $NEE = R_e - GPP$ where $R_e \geq 0$ and $GPP \geq 0$ [Byrne 2018]
- To incorporate this decomposition with our monthly scaling factors, for each month t and spatiotemporal index i we have,
  - $c_i \cdot \overline{NEE}_{t,i} + \overline{GPP}_{t,i} \geq 0 \implies c_i \geq -\dfrac{\overline{GPP}_{t,i}}{\overline{NEE}_{t,i}}$
  - $\implies \boldsymbol{A} = \boldsymbol{I}$ and $b_i = -\dfrac{\overline{GPP}_{t,i}}{\overline{NEE}_{t,i}}$

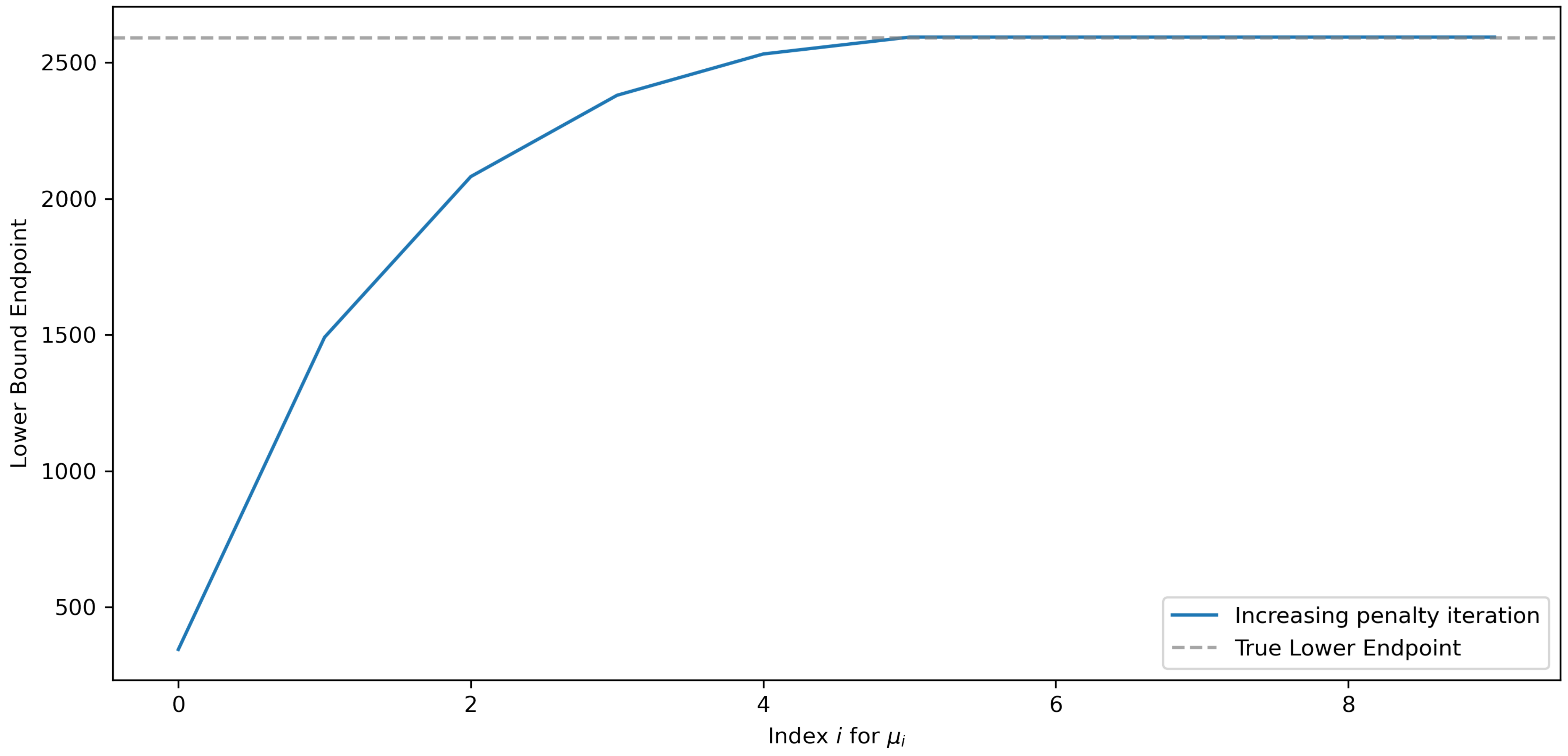
Solving this optimization comes with some challenges

- The forward model is only accessible via runs of GEOS-Chem on a supercomputer (i.e., each iteration is costly)
- From the GEOS-Chem Adjoint model, we are limited to using the gradient of the 4D-Var objective function
- The following optimization approaches the original as positive $\mu \to \infty$

$$\underset{\boldsymbol{c} \geq \boldsymbol{b}}{\text{minimize}} \quad \boldsymbol{h}^T \boldsymbol{c} + \mu \max \left\{ 0, (\boldsymbol{y} - \boldsymbol{K}_x \boldsymbol{c})^T \boldsymbol{S}_o^{-1} (\boldsymbol{y} - \boldsymbol{K}_x \boldsymbol{c}) - \chi^2_\alpha(m) \right\}$$

Lower bound optimization approaches from below in deconvolution example

- Starting with $\mu = 1$ and for each iteration using the update $\mu \leftarrow 2\mu$, we display the following convergence for one of the deconvolution interval lower bounds,



Summary of key points

- Current UQ approaches for CF inversion are exposed to prior misspecification
- We propose the application of a prior-free methodology for directly computing flux confidence intervals with coverage guarantees
    - The linear functional form is well suited for ill-posed inverse problems because of the implicit regularization
    - The variational form is well-suited to the computational complexity of problems involving complex simulators
- For CF inversion, decomposition of NEE flux motivates the affine inequality constraints

Special thanks to Cheyenne at NCAR for hosting our computations!

NCAR

References

(1) Bousserez, N., et al. (2015). Improved analysis-error covariance matrix for high-dimensional variational inversions: application to source estimation using a 3D atmospheric transport model. *Quarterly Journal of the Royal Meteorological Society*.

(2) Byrne, B. (2018). Monitoring the carbon cycle: Evaluation of terrestrial biosphere models and anthropogenic greenhouse gas emissions with atmospheric observations. *PhD Thesis*.

(3) Chevallier, F., Breon, F.-M., and Raynor, P. J. (2007). Contribution of the orbiting carbon observatory to the estimation of CO2 sources and sinks: theoretical study in a variational data assimilation framework. *Journal of Geophysical Research.* 112.

(4) Flath, H. P. et al. (2011). Fast algorithms for bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximation. *SIAM Journal on Scientific Computing*. 33.

(5) Kalmikov, A. G., and Heimbach, P. (2014). A hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM Journal on Scientific Computing*. 36.

(6) Patil, P., Kuusela, M., Hobbs, J. (2020). Objective frequentist uncertainty quantification for atmospheric CO2 retrievals. To appear in the *SIAM / ASA Journal on Uncertainty Quantification*. Preprint on arXiv: https://arxiv.org/abs/2007.14975.

(7) Petra, N., et al. (2007). A computational framework for infinite-dimensional bayesian inverse problems, part II: stochastic newton mcmc with application to ice sheet flow inverse problems. *SIAM Journal on Scientific Computing*. 36.

(8) B. W. Rust and D. P. O'Leary (1994). Confidence intervals for discrete approximations to ill-posed problems. *Journal of Computational and Graphical Statistics*. 3.

(9) Stanley, M., Kuusela, M., Byrne, B., Liu, J. (2022). Estimating Posterior Uncertainty via a Monte Carlo Procedure Specialized for Data Assimilation. In preparation.

(10) Stanley, M., Patil, P., Kuusela, M. (2021). Uncertainty quantification for wide-bin unfolding: one-at-a-time strict bounds and prior-optimized confidence intervals. Preprint on arXiv: https://arxiv.org/pdf/2111.01091.pdf.

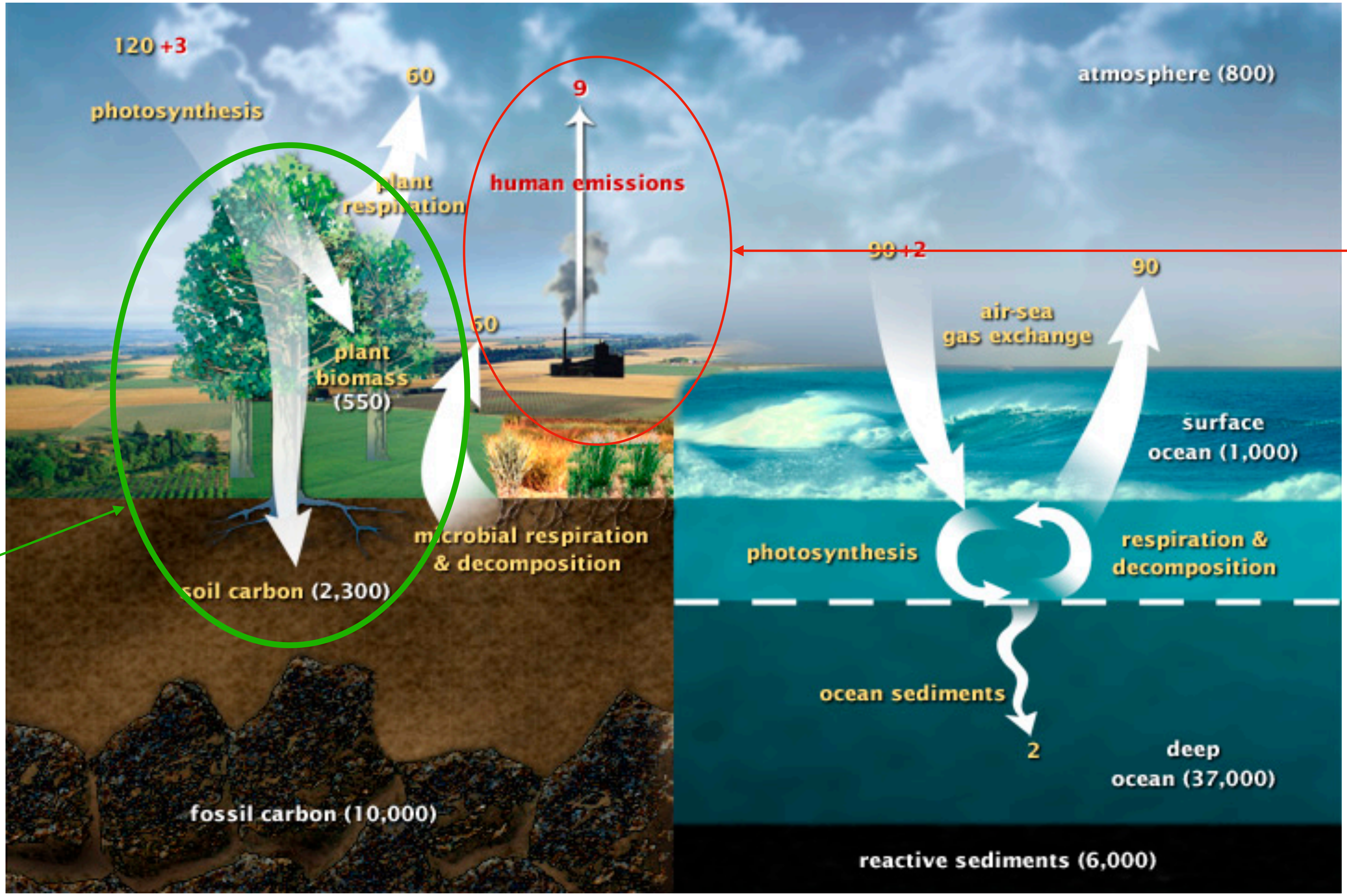
(11) P. B. Stark (1992). Inference in infinite-dimensional inverse problems: Discretization and duality. *Journal of Geophysical Research*. 97.

(12) Zammit-Mangion, A., et al. (2022). WOMBAT v1.0: a fully Bayesian global flux-inversion framework. *Geoscientific Model Development*. 15.

Thank you!

Appendix

The carbon cycle - sources and sinks



Carbon **Source***

Carbon **Sink***

Return

Prior-Optimized Confidence Intervals

[Stanley et al. 2021]

Prior distribution on x

Decision rule parameterized by $(\underline{w}, \underline{c}, \bar{w}, \bar{c})$

Bayes risk where loss is defined by interval length

$$\underset{\underline{\boldsymbol{w}},\underline{\boldsymbol{c}},\bar{\boldsymbol{w}},\bar{\boldsymbol{c}}}{\text{minimize}} \quad r(\pi_x, \delta)$$

$$\text{subject to} \quad \boldsymbol{h} + \boldsymbol{A}^T\underline{\boldsymbol{c}} - \boldsymbol{K}^T\underline{\boldsymbol{w}} = \boldsymbol{0}$$

$$\boldsymbol{h} - \boldsymbol{A}^T\bar{\boldsymbol{c}} - \boldsymbol{K}^T\bar{\boldsymbol{w}} = \boldsymbol{0}$$

$$\underline{\boldsymbol{c}}, \bar{\boldsymbol{c}} \geq \boldsymbol{0}$$

Constraints defining a set of decision rules with guaranteed coverage

Return