Jet Propulsion Laboratory
California Institute of Technology

# Gaussian Mixture Models for Inversion and Uncertainty Quantification

Otto Lamminpää*, Amy Braverman*, Maggie Johnson *, Jonathan Hobbs*, Jouni Susiluoto*, Mike Turmon*

February 28, 2024

* Uncertainty Quantification and Statistical Analysis Group
Jet Propulsion Laboratory, California Institute of Technology

# Overview

Introduction

Gaussian Mixture Inversion

Application to OCO-2

# Introduction

# Terrestrial Carbon Cycle

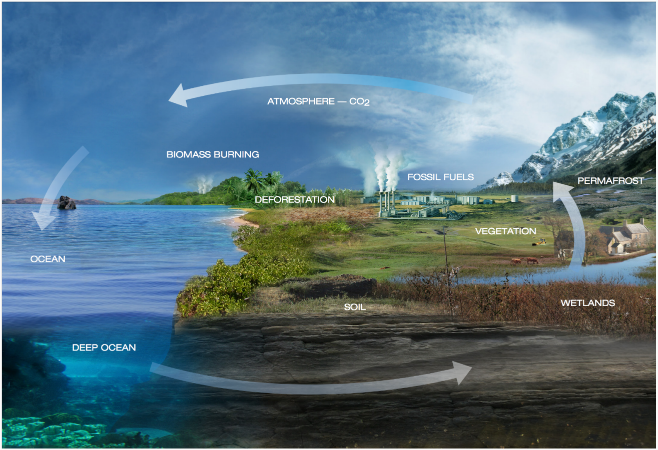



**Figure 1:** NASA/JPL

# Carbon Flux Mismatch

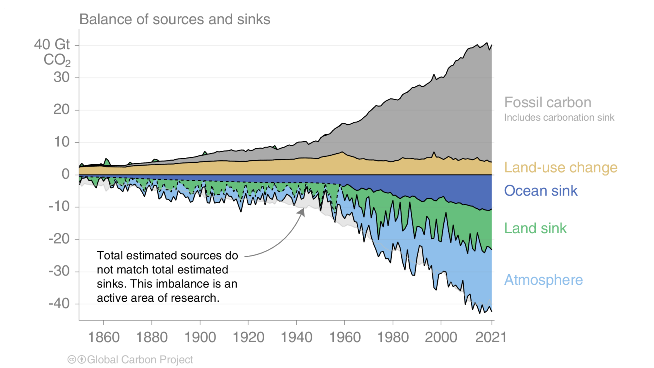



**Figure 2:** globalcarbonproject.org/carbonbudget/index.htm

# Carbon Flux Inversion

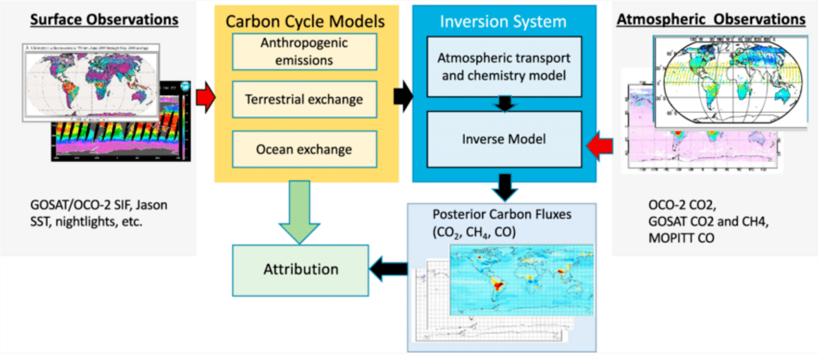



Figure 3: NASA/JPL

# The Orbiting Carbon Observatory 2 (OCO-2) Instrument

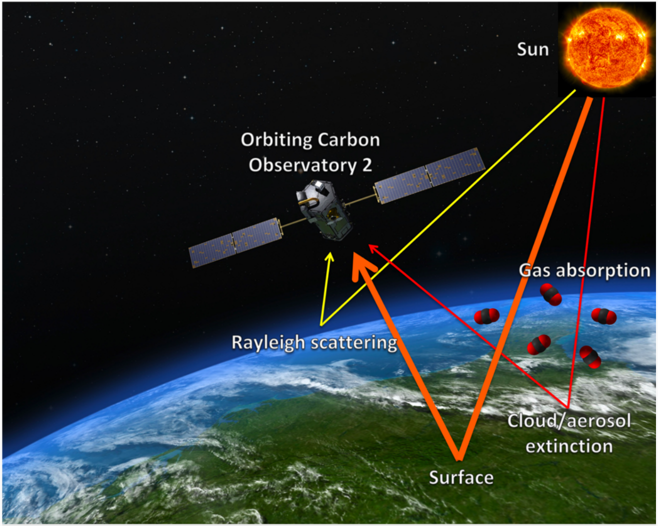
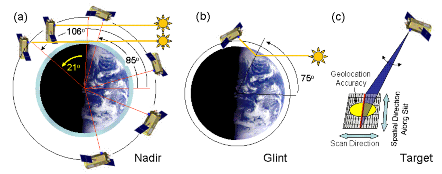
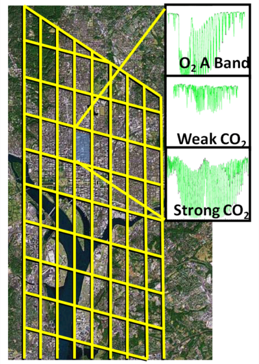


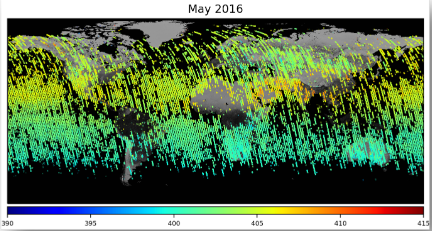

Figure 4: NASA/JPL

# The Orbiting Carbon Observatory 2 (OCO-2) Instrument

Measured radiances are used to infer atmospheric $CO_2$ concentrations.



# Atmospheric Retrieval

The OCO-2 mission uses Optimal Estimation (Rodgers, 2000). The relationship between atmospheric state vector $x$ and radiance vector $y$ is modeled as

$$y = F(x, b) + \varepsilon, \quad x \sim N(x_a, S_a), \quad \varepsilon \sim N(0, S_y). \tag{1}$$

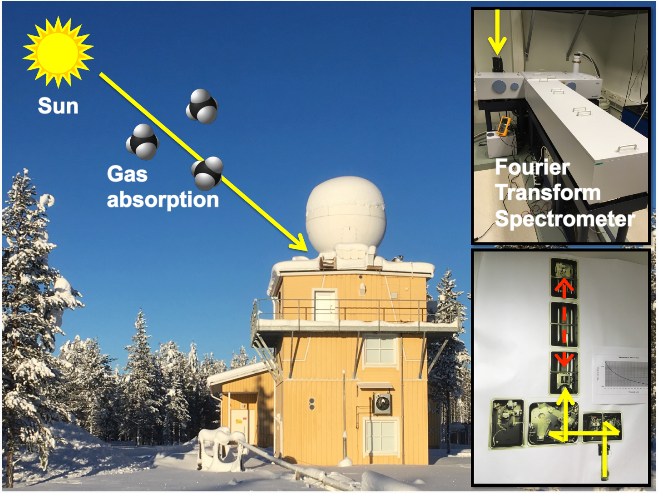
The atmospheric state is then inferred by solving the corresponding inverse problem:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \, [y - F(x, b)]^T S_y^{-1} [y - F(x, b)] + [x - x_a]^T S_a^{-1} [x - x_a] \,. \tag{2}$$

The operational retrieval also provides an estimate for the posterior covariance:

$$\hat{S} = \left(K^T S_y^{-1} K + S_a^{-1}\right)^{-1}. \tag{3}$$

# Ground Measurement: TCCON



# Outstanding Issues for OCO-2

-Retrievals take significant computational effort. (Re-)Processing the entire OCO-2 data record takes more than a year.

-OE doesn't provide trustworthy uncertainty estimates. Posterior distribution might be non-Gaussian, overall estimate is generally too low. Error sources outside retrieval algorithm are not included.

-XCO2 estimates are biased. Bias correction needed after inital processing.

# Towards Solutions

-Direct retrievals using neural networks (David et al. [2021], Bréon et al [2022]) offer a fast means to going to XCO2 directly from input radiances.

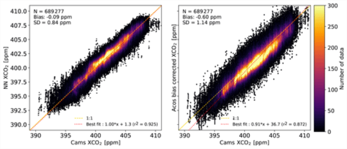



**Figure 5:** Image from David et al.

-MCMC aided by surrogate forward model. (Lamminpää et al. [2019] for exploring the non-Gaussian posterior.
-Simulation Based UQ (Braverman et al. [2021]) simultaneously tackles bias and non-Gaussianity.

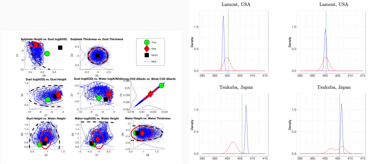
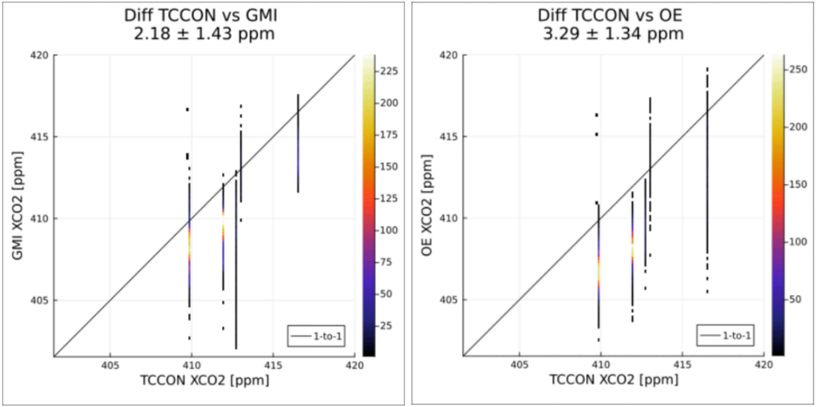



**Figure 6:** Left: Lamminpää et al., Right: Braverman et al.

# Results



Diff TCCON vs GMI
2.18 ± 1.43 ppm
GMI XCO2 [ppm]
TCCON XCO2 [ppm]
1-to-1
Diff TCCON vs OE
3.29 ± 1.34 ppm
OE XCO2 [ppm]
TCCON XCO2 [ppm]
1-to-1

# Gaussian Mixture Inversion

# Modeling a Non-Linear Function $u = f(w)$

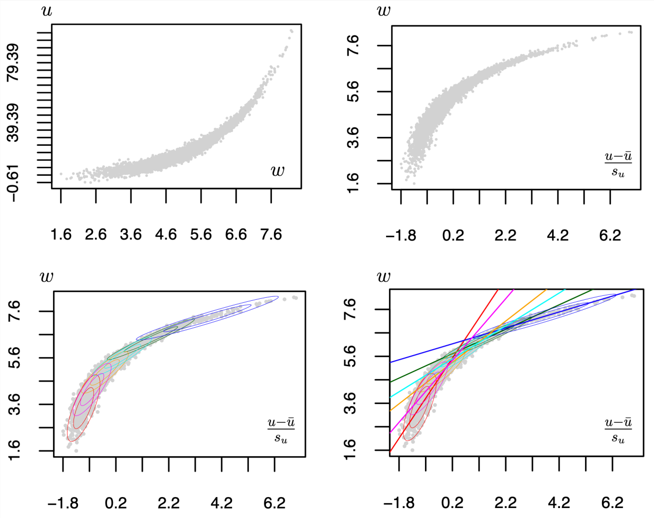
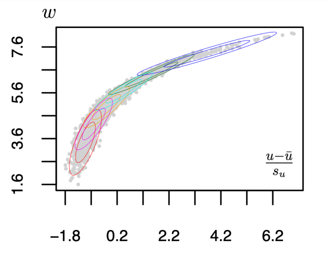



**Figure 7:** Image from Braverman et al. [2021]

# Gaussian Mixture Model: Joint Distribution

$$p(y,x) = p(z) = \sum_{k=1}^{K} \pi^{(k)} \phi(z; \mu^{(k)}, \Sigma^{(k)}) \tag{4}$$

$$z = \begin{bmatrix} y \\ x \end{bmatrix}, \quad \Sigma^{(k)} = \begin{bmatrix} \Sigma_{yy}^{(k)} & \Sigma_{yx}^{(k)} \\ \Sigma_{xy}^{(k)} & \Sigma_{xx}^{(k)} \end{bmatrix} \tag{5}$$



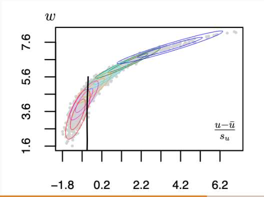
# Gaussian Mixture Model: Conditional Distribution

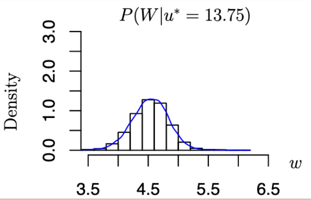
$$p(x|y) = \sum_{k=1}^{K} \pi_{x|y}^{(k)} \phi(x; \mu_{x|y}^{(k)}, \Sigma_{x|y}^{(k)}) \tag{6}$$

$$\pi_{x|y}^{(k)} = \frac{\pi^{(k)} \phi(y; \mu_y^{(k)}, \Sigma_{yy}^{(k)})}{\sum_{l=1}^{K} \pi^{(l)} \phi(y; \mu_y^{(l)}, \Sigma_{yy}^{(l)})} \tag{7}$$

$$\mu_{x|y}^{(k)} = \mu_x^{(k)} + \Sigma_{xy}^{(k)} \left(\Sigma_{yy}^{(k)}\right)^{-1} \left[y - \mu_y^{(k)}\right] \tag{8}$$

$$\Sigma_{x|y}^{(k)} = \Sigma_{xx}^{(k)} - \Sigma_{xy}^{(k)} \left(\Sigma_{yy}^{(k)}\right)^{-1} \Sigma_{yx}^{(k)} \tag{9}$$



# GMM Posterior As A Solution To Inverse Problem

A Bayesian solution to an non-linear inverse problem of solving for $x$ in $y = F(x) + \varepsilon$ is given by

$$p(x|y) \propto p(y|x)p(x). \tag{10}$$

Instead of gradient-based solutions, we propose *Gaussian Mixture Inversion* (GMI): given $x \sim p(x)$ and $y = F(x) + \varepsilon$, the posterior density will be approximated as

$$\widetilde{p}(x|y) = \sum_{k=1}^{K} \pi_{x|y}^{(k)} \phi(x; \mu_{x|y}^{(k)}, \Sigma_{x|y}^{(k)}) \tag{11}$$

defined as before.

# Test cases

Benchmark solution: Adaptive Metropolis (AM). Sample points $x_{t+1}$ from a proposal distribution $N(x_t, C_t)$, where $C_t$ is the covariance matrix of the chain at time $t$:

$$C_t = cov([x_1, \ldots, x_t]), \tag{12}$$

accept new point with probability

$$\alpha(x_t, x_{t+1}) = min\left(1, \frac{\pi(x_{t+1})}{\pi(x_t)}\right) \tag{13}$$

Further, we observe

$$\mathbb{E}^Y\left[d_{Hell}\left(\pi(\cdot|Y), \pi_N(\cdot|Y)\right)^2\right] \leq 2\left(d_{Hell}(\pi(\cdot,\cdot), \pi(\cdot,\cdot)_N)^2 + d_{Hell}(\pi_Y, \pi_{Y_N})^2\right) \tag{14}$$

# 1D Toy Example

We validate our approach to posterior approximation by considering the following two synthetic example. Let

$$x \in \mathbb{R}, \quad y = f(x) + \varepsilon \in \mathbb{R}, \quad f(x) = \sin(2x) - \cos(3x) \qquad (15)$$

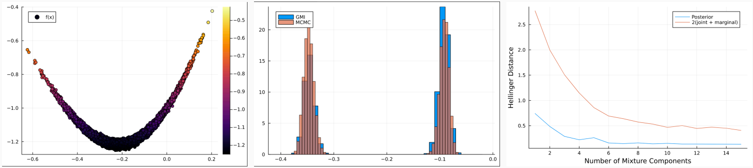



**Figure 8:** Test case where true x = -0.4

# 2D Toy Example

Next, consider

$$x \in \mathbb{R}^2, \quad y = f(x) + \varepsilon \in \mathbb{R}, \quad f(x) = \sin(x_1) - \cos(x_2) \tag{16}$$

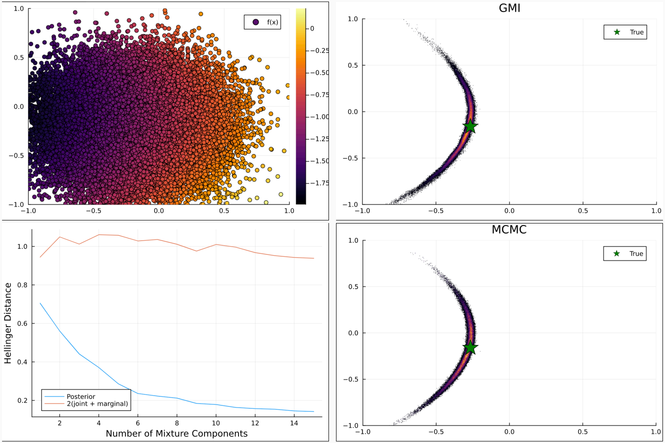



**Figure 9:** Test case where true x = (-0.2, -0.2)

# Application to $OCO$-2

# Example Using Simulated OCO-2 Data

- Retrieval for a single pixel. Quantity of interest: column averaged CO2 concentration, denoted XCO2.
- State vector *x* with realistic atmospheric and surface conditions.
- Simulated measurement: forward model evaluated at *x*, add synthetic measurement error.
- Radiance dimension reduction using PCA.
- Compare GMI and MCMC posteriors (with and without "model discrepancy", using a forward model emulator. **See talk by Jouni Susiluoto, MS157, on Thursday!**).

# Training data

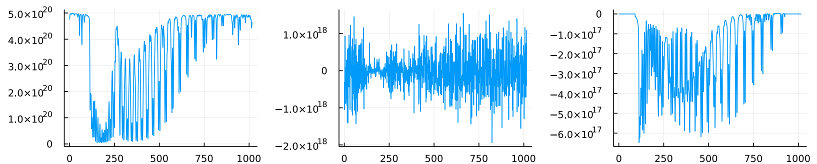
$$y = F(x, b) + \varepsilon + \delta \tag{17}$$

- $x$: operational prior, combination of priors, more comprehensive local / global distribution.
- $b$: forward model parameter uncertainty included by prescribing a distribution.
- $\varepsilon$: error model, possibility of off-diagonal elements in covariance, can be non-Gaussian.
- $\delta$: model discrepancy for including model misspecification, can include new ML bias-correction, other methods for accounting "Unknown Unknowns" and spectral residuals.

# Model Discrepancy Adjustment

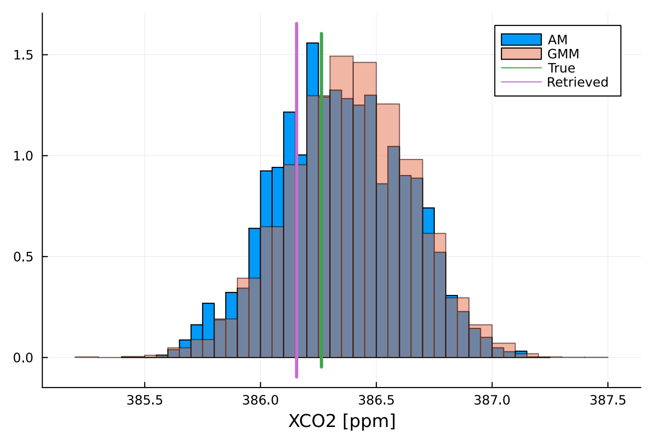
To test performance with more realistic data, we perturb modeled radiances with model discrepancy adjustment:

$$y = F_1(x, b_1) + \varepsilon + \delta, \quad \delta = F_0(x, b_0) - F_1(x, b_1). \tag{18}$$



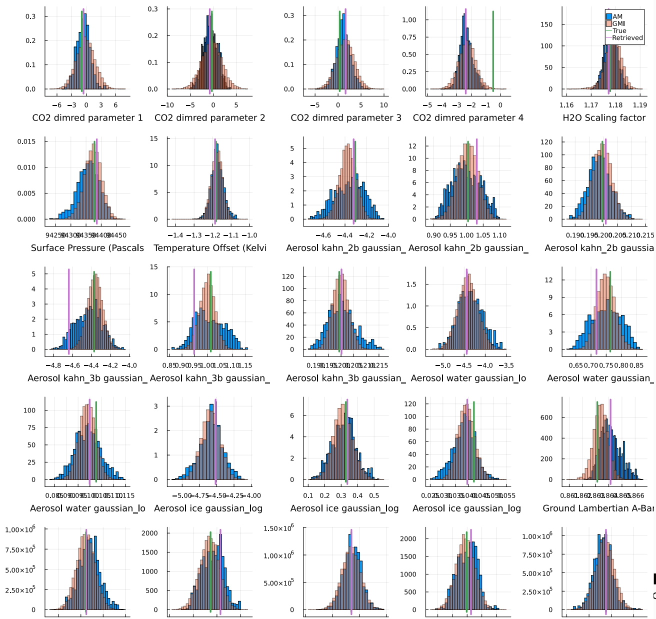

Left: O2-A Band radiance. Middle: a realization of measurement error $\varepsilon \sim \mathcal{N}(0, S_y)$. Right: a realization of model discrepancy $\delta \sim \mathcal{N}(\mu_\delta, S_\delta)$.

# Example Using Simulated OCO-2 Data

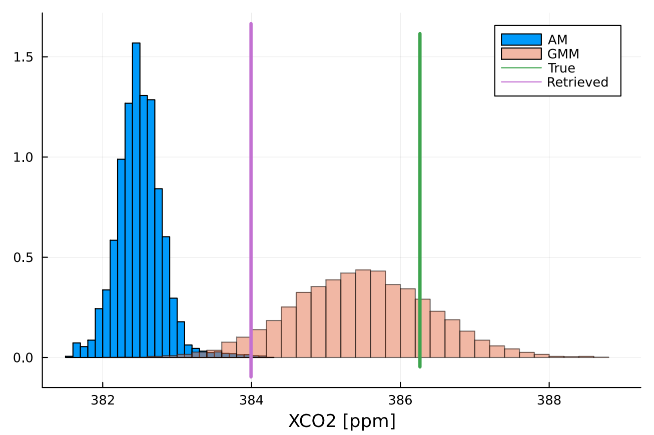



(Without model discrepancy: $y_{obs} = F(x) + \varepsilon$)

# Full posterior

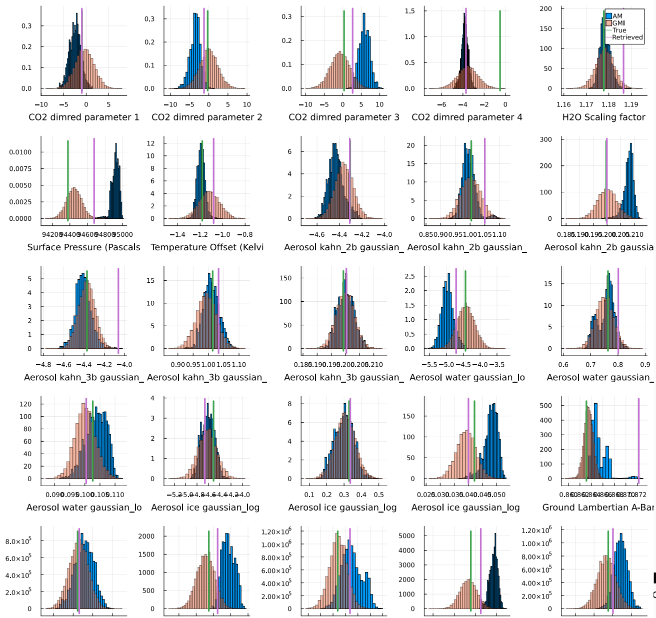


# Example Using Simulated OCO-2 Data




(With Model Discrepancy: $y_{obs} = F(x) + \varepsilon + \delta$)

# Full posterior, with model discrepancy



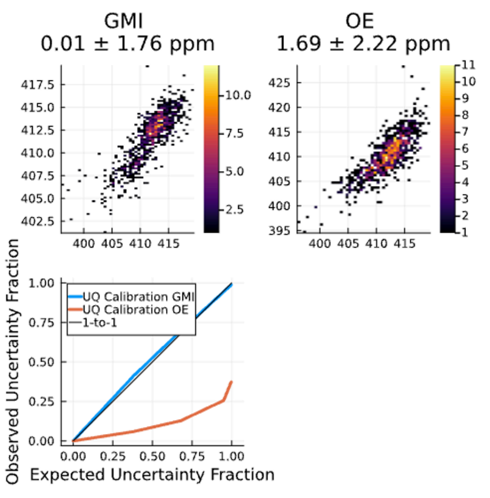
# Timing

- 10000 state vectors sampled from prior; input to forward model (emulator); added noise.
  Time: 104s (parallel in DGX station with 20 threads).
- Traned GMM with Julia's GaussianMixtures package: 100 iterations for learning, 20 mixture components.
  Time: 44s.
- Condition on measured radiance; sample 10000 realizations from conditional distribution.
  Time: 0.014s.
- Adaptive Metropolis ran for 1000000 iterations.
  Time: approx. 17000s, or 4.5h (using emulator).

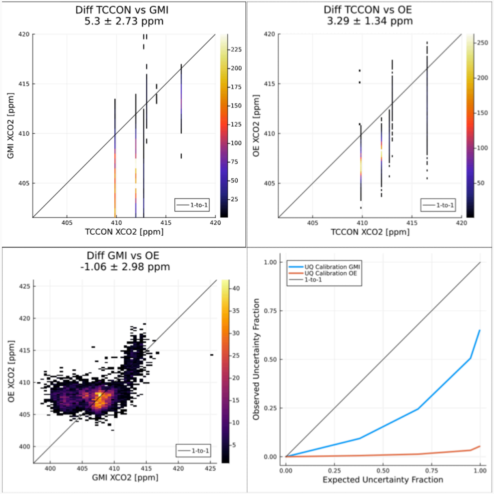
# Example Using Real Data

- Training data: partition the Globe into clusters according to real OCO-2 measurements via self-organizing maps. For each cluster, derive marginal distribution on $x$ and $\delta$. Obtain $y = F(x, b) + \varepsilon + \delta$.
- Evaluate model performance against operational retrieval by left-out simulated data, and TCCON co-located real world retrievals of XCO2.

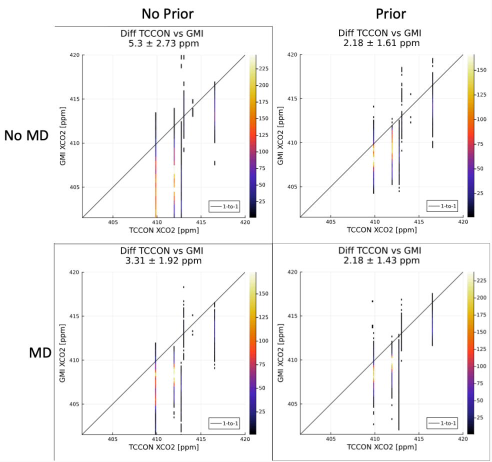
# Comparison with Left-Out Data



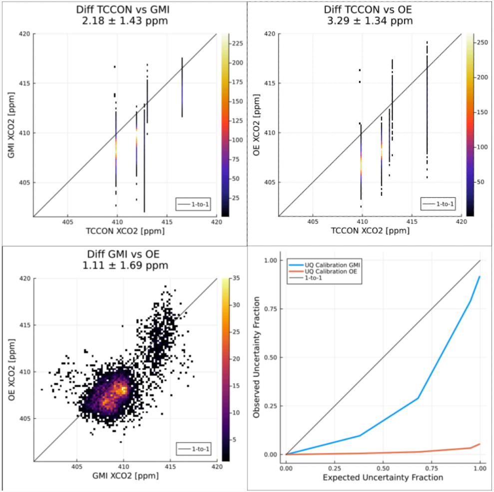
# Comparison with Real Data



# Include Model Discrepancy and Prior Mean In Training



# Comparison with Real Data (Revisited)



# Remark: Averaging Kernels?

$$
\begin{aligned}
A &= (S_a^{-1} + K^T S_\varepsilon^{-1} K)^{-1} K^T S_\varepsilon^{-1} K \\
&= \widehat{S} K^T S_\varepsilon^{-1} K \\
&= \widehat{S}(K^T S_\varepsilon^{-1} K + S_a^{-1} - S_a^{-1}) \\
&= \widehat{S}(\widehat{S}^{-1} - S_a^{-1}) \\
&= \widehat{S}\widehat{S}^{-1} - \widehat{S} S_a^{-1} \\
&= I - \widehat{S} S_a^{-1}
\end{aligned}
\tag{19}
$$

# Further work

- Wasserstein loss and cross-validation to fit GMM.
- Fitting/learning the mixture model with e.g. mixture density networks. Different basis functions, e.g. Cauchy.
- Dimension reduction for data using Autoencoders, UMAP.
- Other methods with similar capabilities: Deep Ensemble methods, VAEs, GAN-FLOW.

# Thank you!

References:


1. Braverman et al. *Post hoc Uncertainty Quantification for Remote Sensing Observing Systems*, 2021 https://epubs.siam.org/doi/10.1137/19M1304283.
2. David et al: XCO2 estimates from the OCO-2 measurements using a neural network approach, 2021 https://doi.org/10.5194/amt-14-117-2021
3. Bréon et al.: On the potential of a neural-network-based approach for estimating XCO2 from OCO-2 measurements, 2022 https://doi.org/10.5194/amt-15-5219-2022
4. Lamminpää et al. Accelerated MCMC for Satellite-Based Measurements of Atmospheric CO2, 2019 https://doi.org/10.3390/rs11172061


This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged