# Post-hoc Uncertainty Quantification for Remote Sensing Observing Systems

Amy Braverman, Jon Hobbs, Joaquim Teixeira, and Michael Gunson
Jet Propulsion Laboratory, California Institute of Technology

January 31, 2020

# Outline

- Introduction
- Remote sensing observing systems
- Orbiting Carbon Observatory 2 (OCO-2) mission, science, and data
- Statistical model of the observing system and uncertainty
- Methodology
- Application to OCO-2
- Summary and discussion points

# Introduction

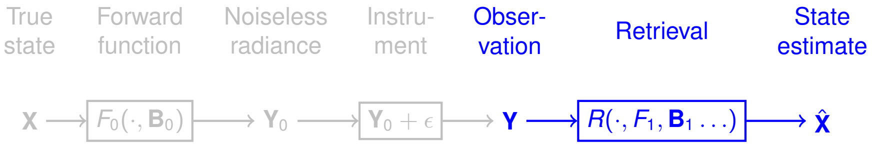
Remote sensing levels of data processing:

- Level 0: raw photon counts direct from satellite
- Level 1: georectified and calibrated radiances
- **Level 2: estimates of geophysical state**
- Level 3: “statistical summaries" of Level 2 on uniform space-time grid
- Level 4: output of models or data assimilation

*Level 2 “data" aren’t “data"; they are inferences!*

How do we calculate uncertainties of the point estimates provided by remote sensing observing systems?

## Remote sensing observing system:




Retrieval is an inference problem: estimate **X** when you only get to see **Y**.

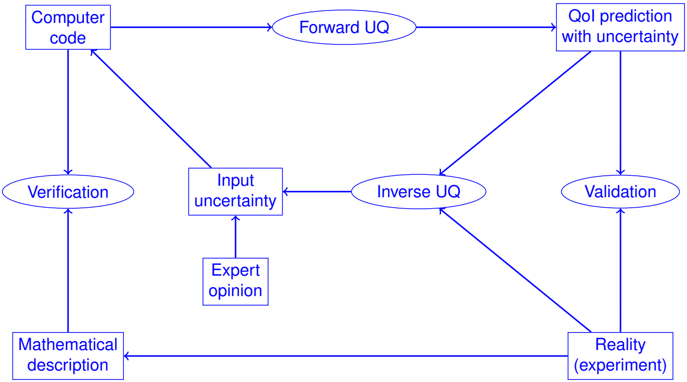
$F_0$ = nature's true forward function; $\mathbf{B}_0$ = other true quantities.

$F_1$ = forward model used in retrieval, $R$; $\mathbf{B}_1$ = other retrieval inputs.

$\epsilon$ = instrument measurement error.

$\ldots$ = other retrieval algorithm inputs.

## Uncertainty quantification:

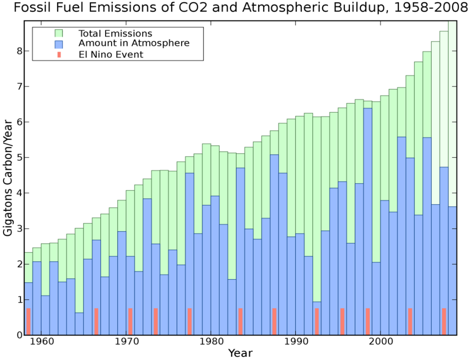



Adapted from Wu et al, (2018). DOI: 10.1016/j.nucengdes.2018.06.004.

# Introduction

- Uncertainty quantification (UQ) provides a formalism for understanding uncertainty in the output of computational models.
- Remote sensing data processing utilizes complex computational models.
- But:
    - we lack information about uncertainty of known inputs
    - there are many unknown unknowns (e.g., interaction effects)
    - computational artifacts must be taken into account
    - analysis must be performed after the fact and be computationally feasible

# OCO-2 mission and science




Graphic courtesy of Annmarie Eldering.

- Less than half of the carbon is staying in the atmosphere.
- Where are the sinks that are absorbing more than half of this $CO_2$?
- Why does $CO_2$ build-up vary from year to year with nearly uniform emission rates?
- How will $CO_2$ sinks respond to climate change?

# OCO-2 mission and science

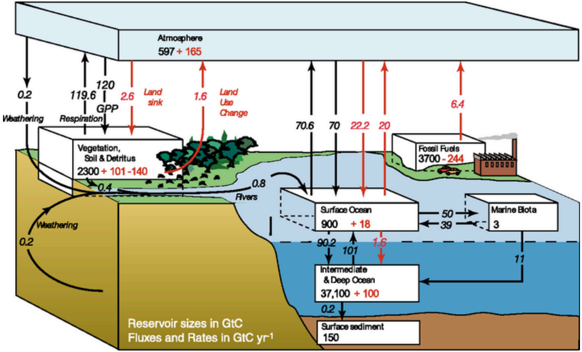
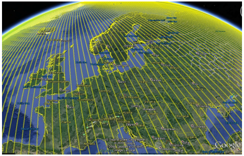



Figure 7.3 from *Climate Change 2007: Working Group I: The Physical Science Basis.*

- Benefit of space-based greenhouse-gas measurement is coverage and resolution.
- Challenge is the need for high-precision measurements to detect small changes in flux against large background variations.
- A primary objective of OCO-2 is to find and study the natural sinks.

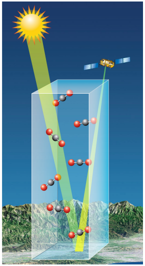
# OCO-2 mission and science



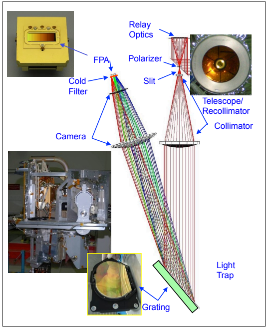
Ground tracks

OCO-2 in polar orbit at 705 km. 16-day ground-track repeat cycle (14.57 orbits per day).

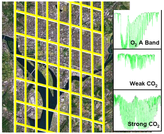
Longitude offset between orbits = 25 degrees; 1.5 degrees between tracks after 16 days.



Measurement principle




Measurement technology




Ground footprints

Eight 1.29 x 2.25 km footprints cross-track.

Three co-bore-sighted high-resolution grating spectrometers, observing in oxygen-A, weak CO2 and strong CO2 bands (3024 wavelengths).

- For each footprint, OCO-2 measures a $3048 \times 1$ vector of radiances, **Y** (Crisp, 2015).
- The retrieval uses **Y** to estimate the $45 \times 1$ atmospheric state vector, **X**, which includes CO2 dry air mole-fraction at 20 altitudes.

# OCO-2 mission and science

Greatest interest lies in the scalar quantities,

$$\hat{\mathrm{X}} \approx \mathrm{E}(\mathbf{h}'\hat{\mathbf{X}}_{1:20}|\mathbf{Y}), \quad \hat{\mathrm{S}} \approx \mathrm{var}(\mathbf{h}'\hat{\mathbf{X}}_{1:20}|\mathbf{Y}),$$

where $\mathbf{X}_{1:20}$ is the first 20 elements of the OCO-2 state vector (the CO2 vertical profile), and $\mathbf{h}$ is the 20-element pressure weighting function.

OCO-2 uses the Optimal Estimation (OE; Rodgers 2000) method to "retrieve" the maximum a posteriori estimate, $\hat{\mathrm{X}}$, and then computes $\hat{\mathrm{S}}$ as a linear function of $\hat{\mathrm{X}}$, all under Gaussian assumptions.

- $\hat{\mathrm{X}}$ is treated as the posterior mean total column mole-fraction of CO2.
- $\hat{\mathrm{S}}$ is treated as the posterior variance of total column mole-fraction of CO2.

# Statistical model

| True state | Forward function | Noiseless radiance | Instru-ment | Obser-vation | Retrieval | State estimate |
| --- | --- | --- | --- | --- | --- | --- |
| $\mathbf{X}$ | $F_0(\cdot, \mathbf{B}_0)$ | $\mathbf{Y}_0$ | $\mathbf{Y}_0 + \epsilon$ | $\mathbf{Y}$ | $R(\cdot, F_1, \mathbf{B}_1 \ldots)$ | $\hat{\mathbf{X}}$ |

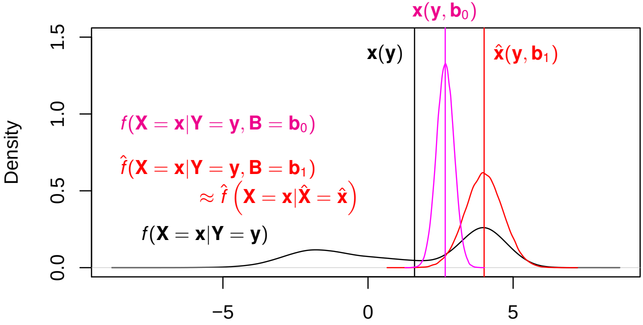
$$\mathbf{X} \longrightarrow \boxed{F_0(\cdot, \mathbf{B}_0)} \longrightarrow \mathbf{Y}_0 \longrightarrow \boxed{\mathbf{Y}_0 + \epsilon} \longrightarrow \mathbf{Y} \longrightarrow \boxed{R(\cdot, F_1, \mathbf{B}_1 \ldots)} \longrightarrow \hat{\mathbf{X}}$$

Uncertainty is quantified by conditional distributions.

- OE: $P(\mathbf{X}|\mathbf{Y})$; only source of uncertainty is $\epsilon$.
- Our view: $P(\mathbf{X}|\hat{\mathbf{X}})$; uncertainty induced by mismatches of $F_1$ to $F_0$, $\mathbf{B}_1$ to $\mathbf{B}_0$, non–Gaussianity, computational artifacts (e.g., discretization, definition of vertical grid, etc.), and unknown interactions among all these sources.
- How to proceed given that we can't even enumerate all sources of uncertainty?

# Statistical model

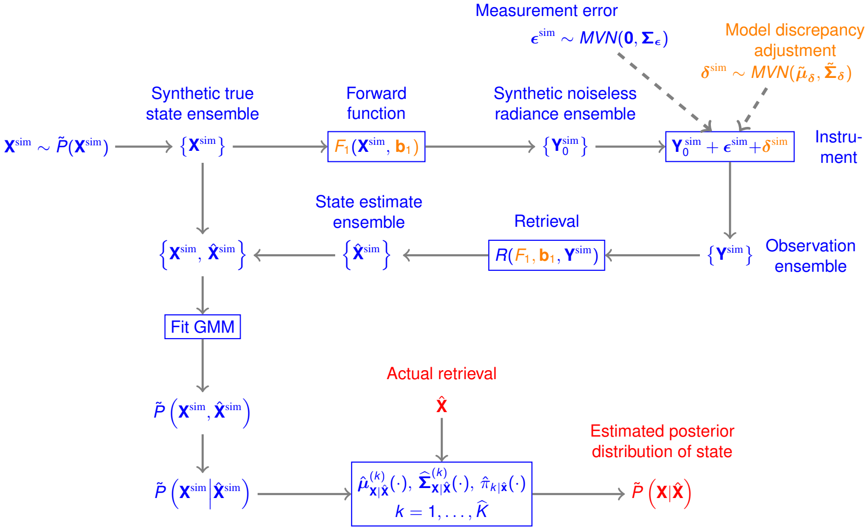
$$\mathbf{X} \sim P_{\mathbf{X}}(\mathbf{x};\theta), \quad \mathbf{Y}_0 = F_0(\mathbf{X}, \mathbf{b}_0), \quad \mathbf{Y} = \mathbf{Y}_0 + \epsilon, \quad \epsilon \sim MVN(\mathbf{0}, \mathbf{\Sigma}_\epsilon),$$
$$\theta_{\mathbf{X}} = \{\mu_{\mathbf{X}}, \mathbf{\Sigma}_{\mathbf{X}}, \ldots\}, \quad \text{and} \quad \hat{\mathbf{X}}(\mathbf{Y}, \mathbf{b}_1) = R(\mathbf{Y}, F_1, \mathbf{b}_1).$$



# Methodology

- "Top-down" approach: simulate the *entire* observing system, and compare retrieved states to synthetic truth over a representative ensemble of conditions.
- Treats the observing system/retrieval as an estimator and focuses on quantifying its mechanistic properties.
- Quanfity mechanistic performance via $P(\mathbf{X}|\hat{\mathbf{X}})$; this distribution is the most complete description of uncertainty in $\mathbf{X}$ after observing $\hat{\mathbf{X}}$.

# Methodology



# Methodology

- Model $P(\mathbf{X}|\hat{\mathbf{X}})$ as a Gaussian mixture model (GMM; McLachlan and Peel, 2000) derived from $P(\mathbf{X}, \hat{\mathbf{X}})$.
- Estimate the parameters of $P(\mathbf{X}|\hat{\mathbf{X}})$ from realistic, simulated data.
- Estimate the (conditional) GMM for a new (actual) $\hat{\mathbf{X}}$:
  - plug the new $\hat{\mathbf{X}}$ into the formulas for regression means and variances for all components in the GMM
  - simulate from the conditional GMM

# Methodology

The Gaussian mixture density for a multivariate random vector **V** is,

$$f_{\mathbf{V}}(\mathbf{v}) = \sum_{k=1}^{K} \pi_k \, \phi\left(\mathbf{v}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \quad \sum_{k=1}^{K} \pi_k = 1,$$

where:

$\phi\left(\mathbf{v}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$ = the multivariate normal density function with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$, evaluated at $\mathbf{v}$;

$\pi_k$ = the (mixing) weight of component $k$,

$K$ = the total number of components.

We abbreviate this density by,

$$\mathbf{V} \sim \mathrm{GMM}\left(K, \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k\}_{k=1}^{K}\right).$$

# Methodology

The parameters of this model are,

$$\left\{ \widehat{K}, \hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_{\widehat{K}}, \widehat{\boldsymbol{\Sigma}}_1, \ldots, \widehat{\boldsymbol{\Sigma}}_{\widehat{K}}, \hat{\pi}_1, \ldots, \hat{\pi}_{\widehat{K}} \right\}.$$

We use the R package mclust (Scrucca et al., 2016 ) to estimate them.

# Methodology

Partition $\mathbf{V} = (\mathbf{W}', \mathbf{U}')$.

The regression mean function for component $k$ evaluated at $\mathbf{U} = \mathbf{u}$ is,

$$\hat{\mu}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}) = \hat{\mu}^{(k)}_{\mathbf{W}} + \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WU}} \left( \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UU}} \right)^{-1} \left[ \mathbf{u} - \hat{\mu}^{(k)}_{\mathbf{U}} \right],$$

where $\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{V}}$ is the estimated covariance matrix for component $k$, which is partitioned as,

$$\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{V}} = \left[ \begin{array}{c|c} \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WW}} & \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WU}} \\ \hline \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UW}} & \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UU}} \end{array} \right].$$

# Methodology

The regression covariance function for component $k$ evaluated at $\mathbf{U} = \mathbf{u}$ is,

$$\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}) = \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WW}} - \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{WU}} \left(\widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UU}}\right)^{-1} \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{UW}}.$$

The (posterior) mixing weights are,

$$\hat{\pi}_{k|\mathbf{u}} = \tilde{P}(\kappa = k|\mathbf{U} = \mathbf{u}) = \frac{\hat{\pi}_k\, \phi\left(\mathbf{u}; \hat{\boldsymbol{\mu}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}), \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u})\right)}{\sum_{l=1}^{\widehat{K}} \hat{\pi}_l\, \phi\left(\mathbf{u}; \hat{\boldsymbol{\mu}}^{(l)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}), \widehat{\boldsymbol{\Sigma}}^{(l)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u})\right)},$$

for $k = 1, \ldots, \widehat{K}$.

$\kappa$ indicates component membership: the probability that $\mathbf{V}$ comes from component $k$ is $\hat{\pi}_k$ before observing $\mathbf{U}$, and $\hat{\pi}_{k|\mathbf{u}}$ after seeing $\mathbf{U} = \mathbf{u}$.

# Methodology

Conditional distribution of $\mathbf{W}$ given $\mathbf{U} = \mathbf{u}^*$:

$$[\mathbf{W}|\mathbf{U} = \mathbf{u}^*] \sim \mathrm{GMM}\left(\widehat{K}^{\mathrm{sim}}, \left\{\hat{\boldsymbol{\mu}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}^*), \widehat{\boldsymbol{\Sigma}}^{(k)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}^*), \hat{\pi}_{k|\mathbf{u}^*}\right\}_{k=1}^{K}\right).$$

Approximate this distribution by simulating from it $B$ times:

1. Let $\kappa_b$ be a univariate random variable taking values in the set $\{1, \dots, \widehat{K}^{\mathrm{sim}}\}$ with,

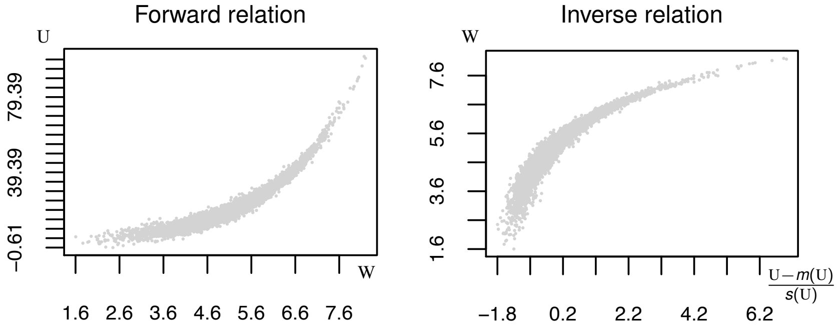
$$P(\kappa_b = k) = \hat{\pi}_{k|\mathbf{u}^*}, \quad k \in \{1, \dots, \widehat{K}^{\mathrm{sim}}\}.$$

2. Draw $B$ random variables,

$$\mathbf{W}^*_b \sim \mathrm{N}\left(\hat{\boldsymbol{\mu}}^{(\kappa_b)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}^*), \widehat{\boldsymbol{\Sigma}}^{(\kappa_b)}_{\mathbf{W}|\mathbf{U}}(\mathbf{u}^*)\right), \quad b = 1, \dots, B.$$

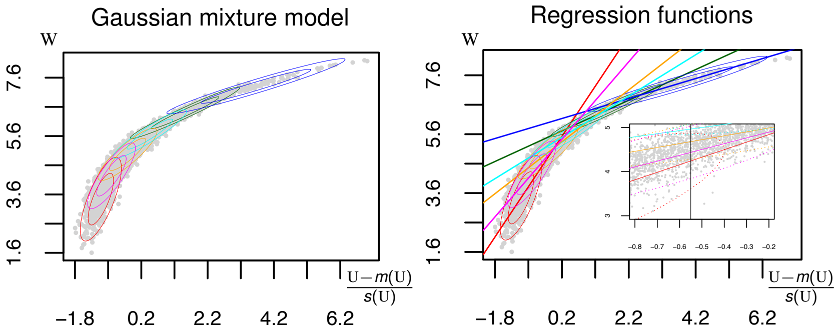
3. Fit kernel density estimate to $(\mathbf{W}^*_1, \dots, \mathbf{W}^*_B)$.

# Methodology



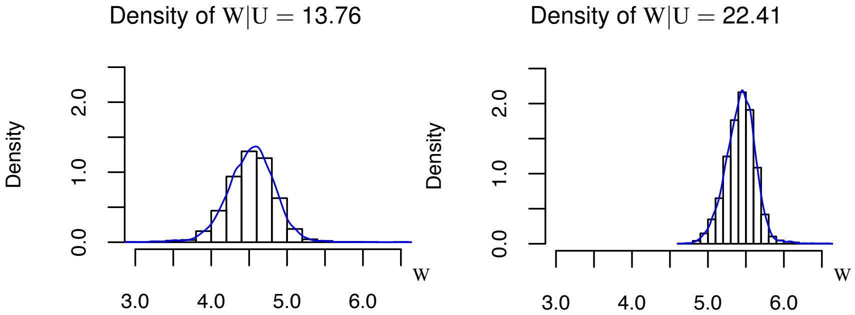

$$W \sim N(5,1), \quad U = (1.75)^{W} + \epsilon, \quad \epsilon \sim N(1,2).$$

# Methodology

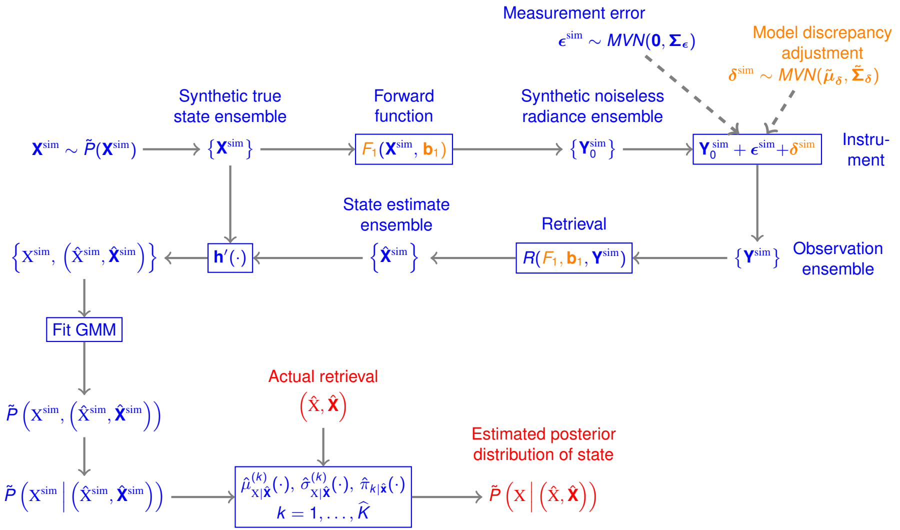



For any $\mathrm{U} = \mathrm{u}^*$, obtain mean, variance, and mixing weight for each component from its regression functions, and simulate.

# Methodology



# Application to OCO-2



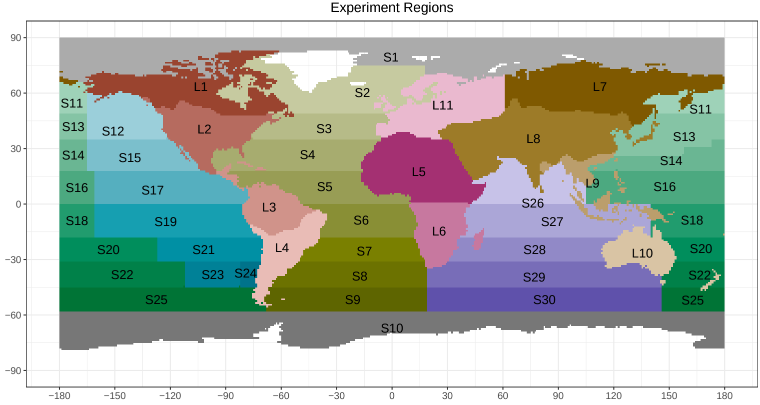
# Application to OCO-2

Synthetic true state ensemble, $\tilde{P}(\mathbf{X}^{\text{sim}})$:

- Ocean and land treated separately– land discussed here.
- Actual OCO-2 retrievals partitioned into 11 land regions by week-long time periods.
- Gaussian mixture model fit (by mclust) separately to each region-week ("template").
- Make 5000 iid draws from each template GMM.

The strategy is similar to a semi-parametric bootstrap.

# Application to OCO-2

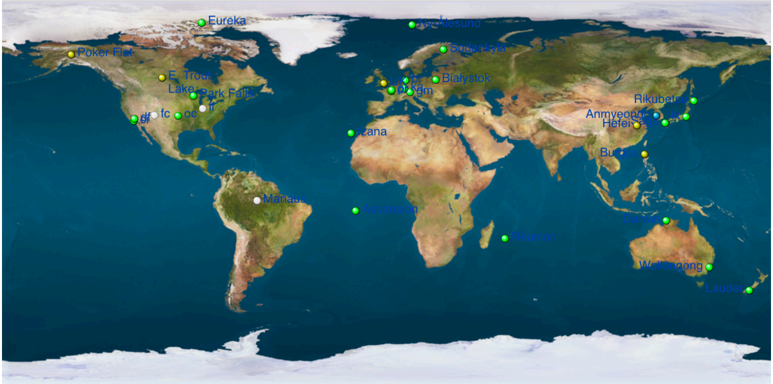


# Application to OCO-2

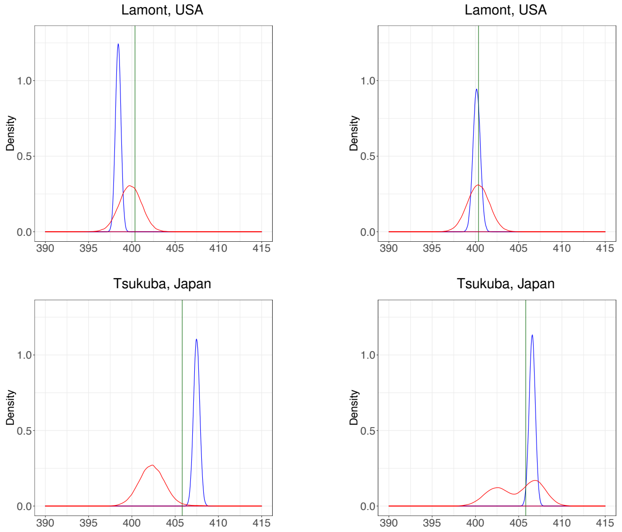
How well does this work?

- Evaluate simulation-based distributions using "ground-truth" (Total Carbon Column Observing Network; TCCON, Wunch et al., 2017) where it exists.
- Consider OCO-2 footprints that contain TCCON sites during four months: August 2015, November 2015, February 2016, and May 2016.
- Also evaluate operationally derived distributions: $\mathrm{N}(\hat{\mathrm{X}}, \hat{\mathrm{S}})$.
- Two metrics: centrality of the TCCON value in the reported distribution, and bias of the distribution mean relative to the TCCON value.

## Total Carbon Column Observing Network (TCCON)



# Application to OCO-2




TCCON value

Operational distribution

Simulated distribution

# Application to OCO-2

Metrics for evaluating simulated and operational distributions.

Metric 1:

$$G_{op}(\mathrm{X}_T) = P_{op}(\mathrm{X} \leq \mathrm{X}_T) \approx \Phi\left(\mathrm{X}_T; \hat{\mathrm{X}}, \hat{\mathrm{S}}\right),$$

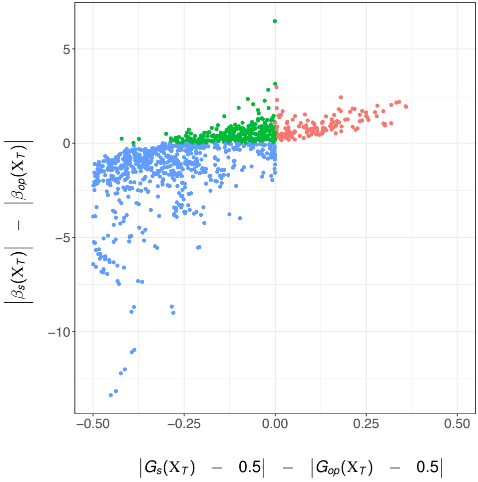
$$G_s(\mathrm{X}_T) = P_s(\mathrm{X} \leq \mathrm{X}_T) \approx \frac{1}{B}\sum_{b=1}^{B} \mathcal{I}\left(\mathrm{W}_b^* \leq \mathrm{X}_T\right).$$

Metric 2:

$$\beta_{op}(\mathrm{X}_T) = (\hat{\mathrm{X}} - \mathrm{X}_T) \quad \text{and} \quad \beta_s(\mathrm{X}_T) = \left(\frac{1}{B}\sum_{b=1}^{B} \mathrm{W}_b^*\right) - \mathrm{X}_T.$$

$\mathrm{X}_T$ denotes the TCCON value, subscropts *op* and *s* denote operational and simulated quantities, respectively.

# Application to OCO-2




$N = 944$

63.1%

25.7%

11.2%

# Application to OCO-2

## Checking interval coverage...

- Let $Q^s_{\alpha/2}$ and $Q^s_{1-\alpha/2}$ denote the lower and upper $\alpha/2$ quantiles of the simulated distribution.
- Let $Q^{op}_{\alpha/2}$ and $Q^{op}_{1-\alpha/2}$ denote the lower and upper $\alpha/2$ quantiles of the operational (Gaussian) distribution.
- Then $\left[Q^s_{\alpha/2}, Q^s_{1-\alpha/2}\right]$ and $\left[Q^{op}_{\alpha/2}, Q^{op}_{1-\alpha/2}\right]$ are $(1-\alpha)100\%$ confidence intervals.
- Compute

$$p^s_{1-\alpha} = \frac{1}{N_r}\sum_{n=1}^{N_r} \mathcal{I}\left(\mathrm{X}_T \in \left[Q^s_{\alpha/2}, Q^s_{1-\alpha/2}\right]\right), \text{ and}$$

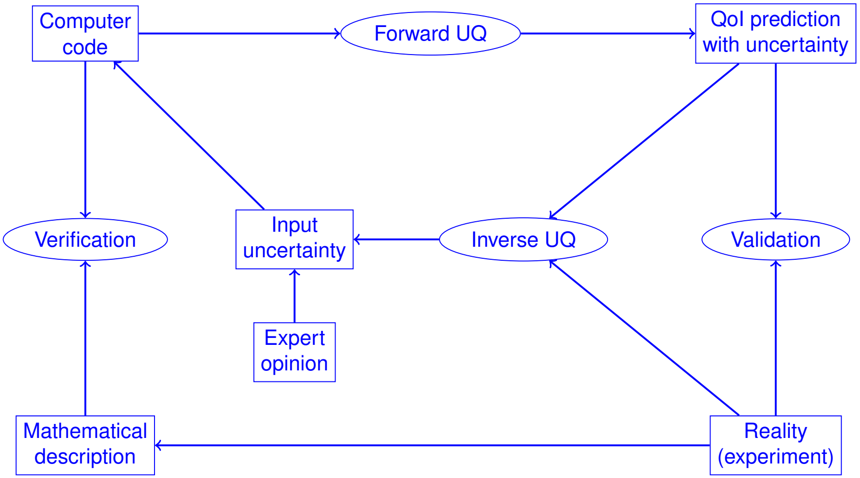
$$p^{op}_{1-\alpha} = \frac{1}{N_r}\sum_{n=1}^{N_r} \mathcal{I}\left(\mathrm{X}_T \in \left[Q^{op}_{\alpha/2}, Q^{op}_{1-\alpha/2}\right]\right),$$

where $N_r$ is the number of OCO-2 footprints.

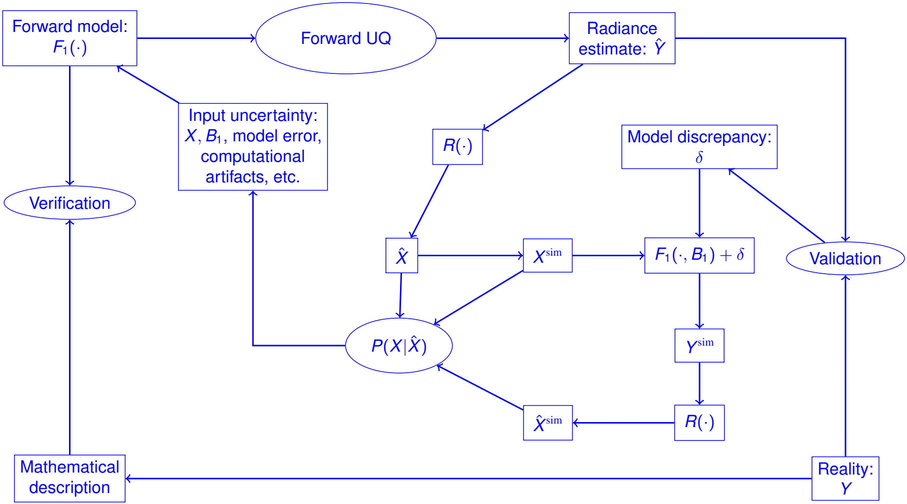
# Application to OCO-2

| Site | Week | $p^{op}_{.95}$ | $p^{op}_{.50}$ | $p^{s}_{.95}$ | $p^{s}_{.50}$ | $N_r$ |
|---|---|---|---|---|---|---|
| Bialystok, Poland | 2016-02-17 | 0.013 | 0.000 | 1.000 | 1.000 | 80 |
| Darwin, Australia | 2015-08-10 | 0.663 | 0.366 | 0.970 | 0.782 | 202 |
| Lamont, OK USA | 2015-11-02 | 0.288 | 0.143 | 1.000 | 0.909 | 132 |
| Lauder, New Zealand | 2016-02-29 | 0.449 | 0.170 | 0.932 | 0.441 | 118 |
| Orleans, France | 2015-11-02 | 0.627 | 0.322 | 0.746 | 0.661 | 59 |
| Sodankyla, Finland | 2015-08-20 | 0.533 | 0.133 | 0.600 | 0.267 | 15 |
| Tsukuba, Japan | 2016-05-13 | 0.390 | 0.169 | 0.974 | 0.818 | 77 |
| Wollongong, Australia | 2015-11-24 | 0.272 | 0.115 | 0.973 | 0.778 | 261 |

# Summary and discussion



# Summary and discussion



# Summary and discussion points

- The challenge was to develop a practical method to assign uncertainties to every estimate produced by a remote sensing obsreving system.
- Uncertainty is defined by the conditional distribution of the true state given the operationally produced state estimate.
- Our method is similar to the bootstrap bias correction (Davison and Hinkley, 1997), but goes further: we correct the entire distribution (not just the mean).
- Simulation-based confidence intervals are usually valid, but not always efficient.
- Operational confidence intervals are rarely valid.

# Summary and discussion points

- Method appears to perform relatively well in this study, but there are indications that we need to adjust the stratification so that the model is more globally representative.
- The most computationally intensive stage in building the model is the retrieval on the simulated radiances.
- Once the model is built, it is very fast to apply to new/actual retrieved estimates.

# References

Crisp, David (2015). Measuring atmospheric carbon dioxide from space with the Orbiting Carbon Observatory-2 (OCO-2), *Earth Observing Systems, Proceedings of the SPIE*, Volume 9607, doi: 10.1117/12.2187291.

Davison, A.C. and Hinkley, D.V. (1997). Bootstrap Methods and Their Application, Cambridge University Press, Cambridge, UK,

McLachlan, Geoffrey and Peel, David (2000). Finite Mixture Models, Wiley Series in Probability and Statistics, New York.

Rodgers, Clive D. (2000). Inverse Methods for Atmospheric Sounding, Theory and Practice, World Scientific, Singapore.

# References

Scrucca, Luca and Fop, Michael and Murphy, Thomas Brendan and Raftery, Adrian E. (2016). Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models, *The R Journal*, Volume 8, Number 1, pages 205–233.

Wunch, Debra and co-authors (2017). Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON, *Atmospheric Measurement Techniques*, Volume 10, pages 2209-?2238, doi: 10.5194/amt-10-2209-2017.

# Acknowledgements

We would like to thank the Orbiting Carbon Observatory-2 Project for sponsoring this work.

# Backup slides

# Model discrepancy

- Model discrepancy is $\delta^{\text{sim}} = F_0(\mathbf{X}, \mathbf{b}_0) - F_1(\mathbf{X}, \mathbf{b}_1)$.
- We would like to simulate from the distribuiton of $\delta^{\text{sim}} \sim \mathrm{N}\left(\tilde{\boldsymbol{\mu}}_\delta, \tilde{\boldsymbol{\Sigma}}_\delta\right)$.
- Assume this distribution is Gaussian with mean $\mathrm{E}(\delta^{\text{sim}})$ and covariance matrix $\mathrm{cov}(\delta^{\text{sim}})$.
- But we only have access to $F_1(\hat{\mathbf{X}}, \mathbf{b}_1)$, which motivates the approximation,

$$\delta^{\text{sim}} \approx F_0(\mathbf{X}, \mathbf{b}_0) - F_1(\hat{\mathbf{X}}, \mathbf{b}_1) - \left[F_1(\mathbf{X}^{\text{sim}}, \mathbf{b}_1) - F_1(\hat{\mathbf{X}}^{\text{sim}}, \mathbf{b}_1)\right].$$

# Model discrepancy

- Let $\mathbf{Y} \equiv F_0(\mathbf{X}, \mathbf{b}_0) + \epsilon$, and $\hat{\mathbf{Y}} \equiv F_1(\hat{\mathbf{X}}, \mathbf{b}_1)$. Then,

$$\delta^{\mathrm{sim}} \approx F_0(\mathbf{X}, \mathbf{b}_0) - F_1(\hat{\mathbf{X}}, \mathbf{b}_1) - \left[F_1(\mathbf{X}^{\mathrm{sim}}, \mathbf{b}_1) - F_1(\hat{\mathbf{X}}^{\mathrm{sim}}, \mathbf{b}_1)\right].$$

$$= \left(\mathbf{Y} - \epsilon - \hat{\mathbf{Y}}\right) - \left(\mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right),$$

$$\delta^{\mathrm{sim}} + \epsilon \approx \left(\mathbf{Y} - \hat{\mathbf{Y}}\right) - \left(\mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right).$$

- Expected value:

$$\mathrm{E}(\delta^{\mathrm{sim}} + \epsilon) \approx \mathrm{E}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) - \mathrm{E}\left(\mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right),$$

$$\tilde{\mu}_\delta + \mathbf{0} \approx \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{Y}_n - \hat{\mathbf{Y}}_n\right) - \frac{1}{M}\sum_{m=1}^{M}\left(\mathbf{Y}_{0,m}^{\mathrm{sim}} - \hat{\mathbf{Y}}_m^{\mathrm{sim}}\right),$$

where $n = 1, \ldots, N$ indexes actual OCO-2 retrievals, and $m = 1, \ldots, M$ indexes trials of the simulation.

# Model discrepancy

- Covariance:

$$\mathrm{cov}(\boldsymbol{\delta}^{\mathrm{sim}} + \boldsymbol{\epsilon}) \approx \mathrm{cov}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \mathrm{cov}\left(\mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right) - 2\,\mathrm{cov}\left(\mathbf{Y} - \hat{\mathbf{Y}}, \mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right)$$

$$\tilde{\boldsymbol{\Sigma}}_\delta = \mathrm{cov}\left(\boldsymbol{\delta}^{\mathrm{sim}}\right) \leq \mathrm{cov}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \mathrm{cov}\left(\mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right) - \mathrm{cov}(\boldsymbol{\epsilon}),$$

$$\approx \widehat{\mathrm{cov}}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) + \widehat{\mathrm{cov}}\left(\mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right) - \mathrm{cov}(\boldsymbol{\epsilon}),$$

assuming $\mathrm{cov}\left(\mathbf{Y} - \hat{\mathbf{Y}}, \mathbf{Y}_0^{\mathrm{sim}} - \hat{\mathbf{Y}}^{\mathrm{sim}}\right) \geq 0$, and $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}^{\mathrm{sim}}$ are independent.