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
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:

True state \( X \) $\rightarrow$ Forward function \( F_0(\cdot, B_0) \) $\rightarrow$ Noiseless radiance \( Y_0 \) $\rightarrow$ Instrument observation \( Y_0 + \epsilon \) $\rightarrow$ Retrieval \( R(\cdot, F_1, B_1 \ldots) \) $\rightarrow$ State estimate \( \hat{X} \)

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

\( F_0 \) = nature’s true forward function; \( B_0 \) = other true quantities.

\( F_1 \) = forward model used in retrieval, \( R \); \( B_1 \) = other retrieval inputs.

\( \epsilon \) = instrument measurement error.

\ldots = other retrieval algorithm inputs.
Uncertainty quantification:

- Computer code
- Forward UQ
- QoI prediction with uncertainty
- Verification
- Input uncertainty
- Inverse UQ
- Validation
- Mathematical description
- Expert opinion
- Reality (experiment)

Adapted from Wu et al, (2018). DOI: 10.1016/j.nucengdes.2018.06.004.
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
Less than half of the carbon is staying in the atmosphere.

Where are the sinks that are absorbing more than half of this CO2?

Why does CO2 build-up vary from year to year with nearly uniform emission rates?

How will CO2 sinks respond to climate change?
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

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.

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).
Greatest interest lies in the scalar quantities,

\[
\hat{X} \approx E(h'\hat{X}_{1:20}|Y), \quad \hat{S} \approx \text{var}(h'\hat{X}_{1:20}|Y),
\]

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

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

- \(\hat{X}\) is treated as the posterior mean total column mole-fraction of CO2.
- \(\hat{S}\) is treated as the posterior variance of total column mole-fraction of CO2.
Statistical model

True state → Forward function → Noiseless radiance → Instrument → Observation → Retrieval → State estimate

\[ X \xrightarrow{F_0(\cdot, B_0)} Y_0 \xrightarrow{Y_0 + \epsilon} Y \xrightarrow{R(\cdot, F_1, B_1 \ldots)} \hat{X} \]

Uncertainty is quantified by conditional distributions.

- OE: \( P(X|Y) \); only source of uncertainty is \( \epsilon \).

- Our view: \( P(X|\hat{X}) \); uncertainty induced by mismatches of \( F_1 \) to \( F_0 \), \( B_1 \) to \( B_0 \), 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

\[ X \sim P_X(x; \theta), \quad Y_0 = F_0(X, b_0), \quad Y = Y_0 + \epsilon, \quad \epsilon \sim \text{MVN}(0, \Sigma_\epsilon), \]
\[ \theta_X = \{\mu_X, \Sigma_X, \ldots\}, \quad \text{and} \quad \hat{X}(Y, b_1) = R(Y, F_1, b_1). \]
“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.

- Quantity mechanistic performance via $P(X|\hat{X})$; this distribution is the most complete description of uncertainty in $X$ after observing $\hat{X}$.
Methodology

Synthetic true state ensemble: $X_{\text{sim}} \sim \tilde{P}(X_{\text{sim}})$

Forward function: $F_1(X_{\text{sim}}, b_1)$

Synthetic noiseless radiance ensemble: $Y_{\text{sim}} = F_1(X_{\text{sim}}, b_1)$

Measurement error: $\epsilon_{\text{sim}} \sim \text{MVN}(0, \Sigma_\epsilon)$

Model discrepancy adjustment: $\delta_{\text{sim}} \sim \text{MVN}(\tilde{\mu}_\delta, \tilde{\Sigma}_\delta)$

Instrument

State estimate ensemble

Retrieval

Observation ensemble

Fit GMM

Actual retrieval

Estimated posterior distribution of state

$\tilde{P}(X_{\text{sim}}, \hat{X}_{\text{sim}})$

$\tilde{P}(X_{\text{sim}} | \hat{X}_{\text{sim}})$

$\tilde{P}(X_{\text{sim}} | \hat{X})$

$\tilde{P}(X | \hat{X})$
Methodology

- Model $P(X|\hat{X})$ as a Gaussian mixture model (GMM; McLachlan and Peel, 2000) derived from $P(X, \hat{X})$.

- Estimate the parameters of $P(X|\hat{X})$ from realistic, simulated data.

- Estimate the (conditional) GMM for a new (actual) $\hat{X}$:
  - plug the new $\hat{X}$ into the formulas for regression means and variances for all components in the GMM
  - simulate from the conditional GMM
The Gaussian mixture density for a multivariate random vector \( \mathbf{V} \) is,

\[
    f_{\mathbf{V}}(\mathbf{v}) = \sum_{k=1}^{K} \pi_k \phi(\mathbf{v}; \mu_k, \Sigma_k), \quad \sum_{k=1}^{K} \pi_k = 1,
\]

where:

\[
    \phi(\mathbf{v}; \mu_k, \Sigma_k) = \text{the multivariate normal density function with mean vector } \mu_k \text{ and covariance matrix } \Sigma_k, \text{ evaluated at } \mathbf{v};
\]

\[
    \pi_k = \text{the (mixing) weight of component } k,
\]

\[
    K = \text{the total number of components}.
\]

We abbreviate this density by,

\[
    \mathbf{V} \sim \text{GMM } (K, \{\mu_k, \Sigma_k, \pi_k\}_{k=1}^{K}).
\]
The parameters of this model are,

\[ \{ \hat{K}, \hat{\mu}_1, \ldots, \hat{\mu}_K, \hat{\Sigma}_1, \ldots, \hat{\Sigma}_K, \hat{\pi}_1, \ldots, \hat{\pi}_K \} . \]

We use the R package mclust (Scrucca et al., 2016) to estimate them.
Partition $V = (W', U')$.

The regression mean function for component $k$ evaluated at $U = u$ is,

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

where $\hat{\Sigma}_V^{(k)}$ is the estimated covariance matrix for component $k$, which is partitioned as,

$$
\hat{\Sigma}_V^{(k)} = \begin{bmatrix}
\hat{\Sigma}_{WW}^{(k)} & \hat{\Sigma}_{WU}^{(k)} \\
\hat{\Sigma}_{UW}^{(k)} & \hat{\Sigma}_{UU}^{(k)}
\end{bmatrix}.
$$
The regression covariance function for component $k$ evaluated at $U = u$ is,

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

The (posterior) mixing weights are,

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

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

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

Conditional distribution of $W$ given $U = u^*$:

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

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

1. Let $\kappa_b$ be a univariate random variable taking values in the set $\{1, \ldots, \hat{K}^{\text{sim}}\}$ with,

$$P(\kappa_b = k) = \hat{\pi}_{k|U^*}, \quad k \in \{1, \ldots, \hat{K}^{\text{sim}}\}.$$

2. Draw $B$ random variables,

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

3. Fit kernel density estimate to $(W^*_1, \ldots, W^*_B)$. 
\[ W \sim N(5, 1), \quad U = (1.75)^W + \epsilon, \quad \epsilon \sim N(1, 2). \]
For any $U = u^*$, obtain mean, variance, and mixing weight for each component from its regression functions, and simulate.
Density of $W|U = 13.76$

Density of $W|U = 22.41$
Application to OCO-2

Synthetic true state ensemble
\( \mathbf{X}^{\text{sim}} \sim \tilde{\mathcal{P}}(\mathbf{X}^{\text{sim}}) \) → \( \{ \mathbf{X}^{\text{sim}} \} \)

Forward function
\( F_1(\mathbf{X}^{\text{sim}}, \mathbf{b}_1) \) → \( \{ \mathbf{Y}_0^{\text{sim}} \} \)

Synthetic noiseless radiance ensemble

Measurement error
\( \mathbf{\epsilon}^{\text{sim}} \sim \text{MVN}(\mathbf{0}, \Sigma_{\mathbf{\epsilon}}) \)

Model discrepancy adjustment
\( \mathbf{\delta}^{\text{sim}} \sim \text{MVN}(\tilde{\mathbf{\mu}}_{\delta}, \tilde{\Sigma}_{\delta}) \)

Instrument

Observation ensemble
\( \{ \mathbf{Y}^{\text{sim}} \} \)

State estimate ensemble
\( \{ \mathbf{\hat{X}}^{\text{sim}} \} \)

Retrieval
\( R(F_1, \mathbf{b}_1, \mathbf{Y}^{\text{sim}}) \) → \( \{ \mathbf{Y}^{\text{sim}} \} \)

Actual retrieval
\( (\mathbf{\hat{X}}, \mathbf{\hat{X}}) \)

Estimated posterior distribution of state
\( \tilde{\mathcal{P}}(\mathbf{X} | (\mathbf{\hat{X}}, \mathbf{\hat{X}})) \)

Fit GMM

Fit GMM
\( \mathcal{P}(\mathbf{X}^{\text{sim}}, (\mathbf{\hat{X}}^{\text{sim}}, \mathbf{\hat{X}}^{\text{sim}})) \)

Actual retrieval
\( (\mathbf{\hat{X}}, \mathbf{\hat{X}}) \)

Estimated posterior distribution of state
\( \tilde{\mathcal{P}}(\mathbf{X} | (\mathbf{\hat{X}}, \mathbf{\hat{X}})) \)

24
Application to OCO-2

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

- 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.
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: $N(\hat{X}, \hat{S})$.

- Two metrics: centrality of the TCCON value in the reported distribution, and bias of the distribution mean relative to the TCCON value.
Implementation and results

Total Carbon Column Observing Network (TCCON)
Metrics for evaluating simulated and operational distributions.

Metric 1:

\[
G_{op}(X_T) = P_{op}(X \leq X_T) \approx \Phi \left( X_T; \hat{X}, \hat{S} \right),
\]

\[
G_{s}(X_T) = P_{s}(X \leq X_T) \approx \frac{1}{B} \sum_{b=1}^{B} \mathcal{I}(W_b^* \leq X_T).
\]

Metric 2:

\[
\beta_{op}(X_T) = (\hat{X} - X_T) \quad \text{and} \quad \beta_{s}(X_T) = \left( \frac{1}{B} \sum_{b=1}^{B} W_b^* \right) - X_T.
\]

\(X_T\) denotes the TCCON value, subscripts \(op\) 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 $Q_{\alpha/2}^{op}$ and $Q_{1-\alpha/2}^{op}$ denote the lower and upper $\alpha/2$ quantiles of the operational (Gaussian) distribution.

- Then $[Q_s^{\alpha/2}, Q_s^{1-\alpha/2}]$ and $[Q_{\alpha/2}^{op}, Q_{1-\alpha/2}^{op}]$ are $(1 - \alpha)100\%$ confidence intervals.

- Compute

$$p_{1-\alpha}^s = \frac{1}{N_r} \sum_{n=1}^{N_r} \mathbb{I} \left( X_T \in [Q_s^{\alpha/2}, Q_s^{1-\alpha/2}] \right),$$

$$p_{1-\alpha}^{op} = \frac{1}{N_r} \sum_{n=1}^{N_r} \mathbb{I} \left( X_T \in [Q_{\alpha/2}^{op}, Q_{1-\alpha/2}^{op}] \right),$$

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

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

Computer code  \rightarrow  Forward UQ  \rightarrow  QoI prediction with uncertainty

Verification  \rightarrow  Input uncertainty  \rightarrow  Inverse UQ  \rightarrow  Validation

Mathematical description  \rightarrow  Expert opinion

Reality (experiment)
Summary and discussion

Forward model: \( F_1(\cdot) \)

Forward UQ

Radiance estimate: \( \hat{Y} \)

Verification

Input uncertainty: \( X, B_1, \) model error, computational artifacts, etc.

\( \hat{X} \)

\( R(\cdot) \)

Model discrepancy: \( \delta \)

\( \hat{X}^{\text{sim}} \)

\( X^{\text{sim}} \)

\( F_1(\cdot, B_1) + \delta \)

Validation

Mathematical description

Reality: \( Y \)

\( \hat{Y} \)

\( Y^{\text{sim}} \)

\( \hat{X}^{\text{sim}} \)

\( R(\cdot) \)
The challenge was to develop a practical method to assign uncertainties to every estimate produced by a remote sensing observing 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.


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

Backup slides
Model discrepancy

- Model discrepancy is $\delta^{\text{sim}} = F_0(X, b_0) - F_1(X, b_1)$.

- We would like to simulate from the distribution of $\delta^{\text{sim}} \sim N\left(\tilde{\mu}_\delta, \tilde{\Sigma}_\delta\right)$.

- Assume this distribution is Gaussian with mean $E(\delta^{\text{sim}})$ and covariance matrix $\text{cov}(\delta^{\text{sim}})$.

- But we only have access to $F_1(\hat{X}, b_1)$, which motivates the approximation,

$$\delta^{\text{sim}} \approx F_0(X, b_0) - F_1(\hat{X}, b_1) - \left[F_1(X^{\text{sim}}, b_1) - F_1(\hat{X}^{\text{sim}}, b_1)\right].$$
Let $Y \equiv F_0(X, b_0) + \epsilon$, and $\hat{Y} \equiv F_1(\hat{X}, b_1)$. Then,

$$\delta^{\text{sim}} \approx F_0(X, b_0) - F_1(\hat{X}, b_1) - \left[ F_1(X^{\text{sim}}, b_1) - F_1(\hat{X}^{\text{sim}}, b_1) \right].$$

$$= \left( Y - \epsilon - \hat{Y} \right) - \left( Y_0^{\text{sim}} - \hat{Y}^{\text{sim}} \right),$$

$$\delta^{\text{sim}} + \epsilon \approx \left( Y - \hat{Y} \right) - \left( Y_0^{\text{sim}} - \hat{Y}^{\text{sim}} \right).$$

Expected value:

$$\mathbb{E}(\delta^{\text{sim}} + \epsilon) \approx \mathbb{E}\left( Y - \hat{Y} \right) - \mathbb{E}\left( Y_0^{\text{sim}} - \hat{Y}^{\text{sim}} \right),$$

$$\tilde{\mu}_\delta + 0 \approx \frac{1}{N} \sum_{n=1}^{N} (Y_n - \hat{Y}_n) - \frac{1}{M} \sum_{m=1}^{M} (Y_0^{\text{sim}} - \hat{Y}_m^{\text{sim}}),$$

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

- Covariance:

\[
\text{cov}(\delta_{\text{sim}} + \epsilon) \approx \text{cov}(Y - \hat{Y}) + \text{cov}(Y_{0} - \hat{Y}_{\text{sim}}) - 2 \text{cov}(Y - \hat{Y}, Y_{0} - \hat{Y}_{\text{sim}}) \\
\]

\[
\hat{\Sigma}_{\delta} = \text{cov}(\delta_{\text{sim}}) \leq \text{cov}(Y - \hat{Y}) + \text{cov}(Y_{0} - \hat{Y}_{\text{sim}}) - \text{cov}(\epsilon), \\
\approx \widehat{\text{cov}}(Y - \hat{Y}) + \widehat{\text{cov}}(Y_{0} - \hat{Y}_{\text{sim}}) - \text{cov}(\epsilon),
\]

assuming \( \text{cov}(Y - \hat{Y}, Y_{0} - \hat{Y}_{\text{sim}}) \geq 0 \), and \( \epsilon \) and \( \delta_{\text{sim}} \) are independent.