Uncertainty quantification for OCO-2 remote sensing retrievals via Monte Carlo simulation

Amy Braverman, Jon Hobbs, and Mike Gunson
Jet Propulsion Laboratory, California Institute of Technology

June 6, 2017

Outline

- Goals
- Approach
- Methods
- Results
- Conclusions

Goals

- Obtain uncertainties (biases and standard errors) for OCO-2's estimates of the posterior mean and posterior variance of XCO2 over specified regions and time periods.
- Non-linear error analysis: do so without assuming Gaussianity or linearity.
- Computation must be tractable.

Approach

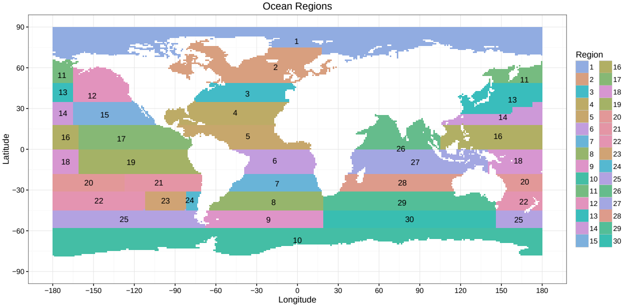
$$
\begin{array}{ccccccccccccccc}
X_1 & \leftarrow & \mathbf{X}_1 & \rightarrow & \boxed{F_0(\cdot,\mathbf{B}_0)} & \rightarrow & \boxed{\mathbf{Y}_{01}+\epsilon_1} & \rightarrow & \mathbf{Y}_1 & \rightarrow & \boxed{R(\cdot,F,\ldots,\mathbf{B})} & \rightarrow & \hat{\mathbf{X}}_1,\hat{\mathbf{S}}_1 & \longrightarrow & (\hat{X}_1,\hat{S}_1) \\
X_2 & \leftarrow & \mathbf{X}_2 & \rightarrow & \boxed{F_0(\cdot,\mathbf{B}_0)} & \rightarrow & \boxed{\mathbf{Y}_{02}+\epsilon_2} & \rightarrow & \mathbf{Y}_2 & \rightarrow & \boxed{R(\cdot,F,\ldots,\mathbf{B})} & \rightarrow & \hat{\mathbf{X}}_2,\hat{\mathbf{S}}_2 & \longrightarrow & (\hat{X}_2,\hat{S}_2) \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
X_N & \leftarrow & \mathbf{X}_N & \rightarrow & \boxed{F_0(\cdot,\mathbf{B}_0)} & \rightarrow & \boxed{\mathbf{Y}_{0N}+\epsilon_N} & \rightarrow & \mathbf{Y}_N & \rightarrow & \boxed{R(\cdot,F,\ldots,\mathbf{B})} & \rightarrow & \hat{\mathbf{X}}_N,\hat{\mathbf{S}}_N & \longrightarrow & (\hat{X}_N,\hat{S}_N)
\end{array}
$$

$$
\boxed{P(\mathbf{X},\mathbf{Y})} \propto P(\mathbf{X}|\mathbf{Y}) \Rightarrow \left(\mu_{\mathbf{X}|\mathbf{Y}}, \Sigma_{\mathbf{X}|\mathbf{Y}}\right) \Rightarrow \left(\mu_{X|\mathbf{Y}}, \sigma^2_{X|\mathbf{Y}}\right).
$$

Figure: *Simulation framework for assessing the statistical properties of OCO-2's retrieved XCO2 posterior mean ($\hat{X}$) and retrieved posterior variance ($\hat{S}$).*

Synthetic “true" states

Define ensembles of synthetic true state vectors, $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N\}$, by membership in Andy Jacobson’s flux ocean regions over one-week periods:



Draw synthetic true states from Gaussian mixture models fit to OCO-2 retrieved state vectors, with aerosol parameters replaced by MERRA daily averages, for each region-week.

Forward function and model discrepancy

$$X_1 \leftarrow \mathbf{X}_1 \rightarrow \boxed{F(\cdot, \mathbf{B})} \rightarrow \boxed{\mathbf{Y}_{01} + \epsilon_1 + \delta_1} \rightarrow \mathbf{Y}_1 \rightarrow \boxed{R(\cdot, F, \ldots, \mathbf{B})} \rightarrow \hat{\mathbf{X}}_1, \hat{\mathbf{S}}_1 \rightarrow (\hat{X}_1, \hat{S}_1)$$

$$X_2 \leftarrow \mathbf{X}_2 \rightarrow \boxed{F(\cdot, \mathbf{B})} \rightarrow \boxed{\mathbf{Y}_{02} + \epsilon_2 + \delta_2} \rightarrow \mathbf{Y}_2 \rightarrow \boxed{R(\cdot, F, \ldots, \mathbf{B})} \rightarrow \hat{\mathbf{X}}_2, \hat{\mathbf{S}}_2 \rightarrow (\hat{X}_2, \hat{S}_2)$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$X_N \leftarrow \mathbf{X}_N \rightarrow \boxed{F(\cdot, \mathbf{B})} \rightarrow \boxed{\mathbf{Y}_{0N} + \epsilon_N + \delta_N} \rightarrow \mathbf{Y}_N \rightarrow \boxed{R(\cdot, F, \ldots, \mathbf{B})} \rightarrow \hat{\mathbf{X}}_N, \hat{\mathbf{S}}_N \rightarrow (\hat{X}_N, \hat{S}_N)$$

Since we use F (and $\mathbf{B}$) to generate radiances as well as to retrieve them, we add an additional error term to the radiances: δ_n.

δ_n is a random draw from $N(\mu_\delta, \boldsymbol{\Sigma}_\delta)$. μ_δ and $\boldsymbol{\Sigma}_\delta$ estimated from actual and simulated residuals.

Simulation experiment

$$X_1 \leftarrow \boxed{\mathbf{X}_1} \rightarrow \boxed{F(\cdot, \mathbf{B})} \rightarrow \boxed{\mathbf{Y}_{01} + \epsilon_1 + \delta_1} \rightarrow \boxed{\mathbf{Y}_1} \rightarrow \boxed{R(\cdot, F, \dots, \mathbf{B})} \rightarrow \hat{\mathbf{X}}_1, \hat{\mathbf{S}}_1 \rightarrow (\hat{X}_1, \hat{S}_1)$$

$$X_2 \leftarrow \boxed{\mathbf{X}_2} \rightarrow \boxed{F(\cdot, \mathbf{B})} \rightarrow \boxed{\mathbf{Y}_{02} + \epsilon_2 + \delta_2} \rightarrow \boxed{\mathbf{Y}_2} \rightarrow \boxed{R(\cdot, F, \dots, \mathbf{B})} \rightarrow \hat{\mathbf{X}}_2, \hat{\mathbf{S}}_2 \rightarrow (\hat{X}_2, \hat{S}_2)$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

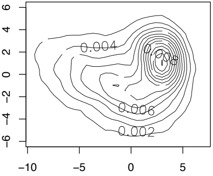
$$X_N \leftarrow \boxed{\mathbf{X}_N} \rightarrow \boxed{F(\cdot, \mathbf{B})} \rightarrow \boxed{\mathbf{Y}_{0N} + \epsilon_N + \delta_N} \rightarrow \boxed{\mathbf{Y}_N} \rightarrow \boxed{R(\cdot, F, \dots, \mathbf{B})} \rightarrow \hat{\mathbf{X}}_N, \hat{\mathbf{S}}_N \rightarrow (\hat{X}_N, \hat{S}_N)$$

Gaussian Mixture Model

$$(\mathbf{X}_n, \mathbf{Y}_n) \sim \sum_{k=1}^{K} 1(C_n = k) N\left(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\Sigma}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{Y}}, \boldsymbol{\rho}_{\mathbf{X},\mathbf{Y}}\right)$$

$$P(C_n = k) = p_k$$

superposition of K multivariate Gaussians



via wtd avg of K multivariate regressions

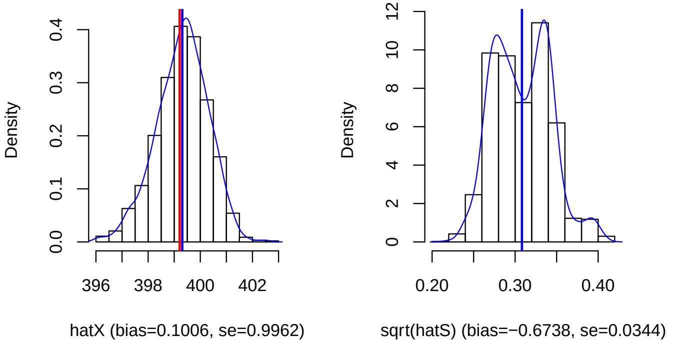
$$\left(\boldsymbol{\mu}_{\mathbf{X}_1|\mathbf{Y}_1}, \boldsymbol{\Sigma}_{\mathbf{X}_1|\mathbf{Y}_1}\right)$$
$$\left(\boldsymbol{\mu}_{\mathbf{X}_2|\mathbf{Y}_2}, \boldsymbol{\Sigma}_{\mathbf{X}_2|\mathbf{Y}_2}\right)$$
$$\vdots$$
$$\left(\boldsymbol{\mu}_{\mathbf{X}_N|\mathbf{Y}_N}, \boldsymbol{\Sigma}_{\mathbf{X}_N|\mathbf{Y}_N}\right)$$

$$\rightarrow$$

$$\left(\mu_{X_1|\mathbf{Y}_1}, \sigma^2_{X_1|\mathbf{Y}_1}\right)$$
$$\left(\mu_{X_2|\mathbf{Y}_2}, \sigma^2_{X_2|\mathbf{Y}_2}\right)$$
$$\vdots$$
$$\left(\mu_{X_N|\mathbf{Y}_N}, \sigma^2_{X_N|\mathbf{Y}_N}\right)$$

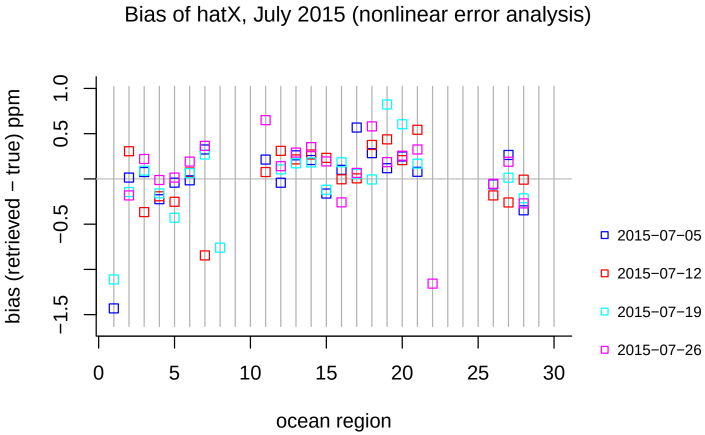
Example

Region 16 (east of Papua/New Guinea), week of July 5, 2015:



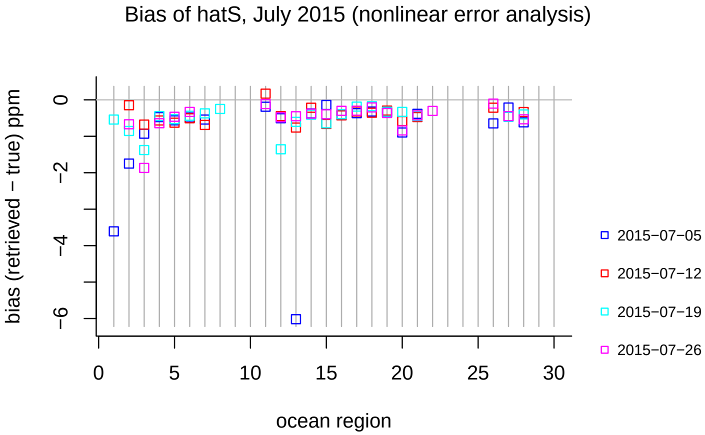
Example

Summary of results for all regions, all weeks of July 2015:



Example

Summary of results for all regions, all weeks of July 2015:



Conclusions

- Monte Carlo uncertainty experiment aims to quantify the uncertainty (both bias and variance) of retrieval algorithm output quantities as estimates of their true counterparts.
- Experiment performed on a regional basis, not sounding-by-sounding, in order to achieve required replication.
- Multivariate Gaussian mixture model (vs. multivariate unimodal Gaussian used in Rodgers' OE framework) allows for heterogeneity within region.
- Gaussian mixture approach also provides a theoretical mechanism for obtaining sounding-by-sounding bias estimates based on radiance vectors (forthcoming).
- Monte Carlo framework can be applied to any retrieval, not just optimal estimation retrievals.

Questions? Contact Amy.Braverman@jpl.nasa.gov.

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.