Uncertainty Quantification for Remote Sensing Retrievals: Monte Carlo Experiments for the Orbiting Carbon Observatory-2

Jon Hobbs (398L)
Joint work with
Amy Braverman, Noel Cressie,
Robert Granat, and Michael Gunson

Jet Propulsion Laboratory
California Institute of Technology

November 11, 2015
Uncertainty Quantification

- Contributions are not perfectly estimated, and plausible values are reflected as probability distributions.
- Uncertainty arises because the climate system is nonlinear with many feedback mechanisms.
- Uncertainty quantification (UQ) targets identifying and reducing uncertainties for quantities of interest associated with complex systems (Smith, 2014).

Figure 8.16 from IPCC AR5 Working Group I report
Carbon cycle scientists combine data on carbon dioxide (CO$_2$) concentrations with process models to infer carbon sources and sinks.

Estimates of CO$_2$ from satellites such as the Orbiting Carbon Observatory-2 (OCO-2) provide substantial spatial and temporal coverage.

Satellite observations are indirect so UQ is challenging.

Reported uncertainties, or standard errors, dictate relative weight of data in flux estimation.

http://oco.jpl.nasa.gov
Key state variables, $X$, for OCO-2 include:

- CO$_2$ concentration
- Pressure
- Scattering particles
- Surface albedo

Observed radiance, $Y$, is a function of the state.
OCO-2 Measurement

- OCO-2 observation \( Y \) includes radiances at 1016 wavelengths in each of three spectral bands.
- Objective is to estimate \( X_{\text{CO}_2} \), the total column \( \text{CO}_2 \) concentration, given observed radiances \( Y \). The estimate’s uncertainty should also be quantified.
- Inference utilizes a forward model, a mathematical representation of the relationship between \( X \) and \( Y \).
Uncertainty quantification (UQ) relies on a probabilistic treatment of sources of uncertainty (Smith, 2014).

- Inherent variability in the state or measurement process
- Lack of complete knowledge about fixed parameters or the physical model of the process

Monte Carlo simulation is a tool for propagation of uncertainty through a model.

Alternatives can be necessary depending on simulation scope and computational expense.
Randomly generated ensemble of $X_{CO2}$ and log aerosol optical depth.

For illustration, suppose $X_{CO2}$ and log aerosol optical depth have a bivariate Gaussian distribution,

$$X \sim N(\mu_X, \Sigma_X)$$

Monte Carlo investigation begins by randomly generating state vectors $X$.

The forward model is evaluated for each state, yielding synthetic radiances $Y$. 
The atmospheric state is the quantity of interest, *inverse inference* is necessary.

A *retrieval algorithm* to produce an *estimate* is required.

The simulation framework can interrogate the propagation of uncertainty to the retrieval.
- Quantify the impact of retrieval choices that are subject to uncertainty on the overall bias and variance of the retrieval errors, $\hat{X} - X$.
- The OCO-2 retrieval problem is underdetermined and prior information is utilized in a Bayesian setting (Rodgers, 2000).
- The retrieval algorithm produces an uncertainty estimate $\hat{S}$.
- The retrieval algorithm requires an a priori mean $\mu_a$ be input as a guess for the true ensemble mean.
Full set-up:

\[ \Delta = (\hat{X} - X) \]

\[ Y = F(X, B) + \epsilon \]

\[ N(\mu_a, \Sigma_a) \]

\[ N(\mu_e, \Sigma_e) \]

\[ R(Y, \mu_a, \Sigma_a, \hat{F}, \hat{B}, \mu_e, \Sigma_e) \]

\[ [\hat{X}, \hat{S}] \]
Surrogate Model Experiment

- Quantify the impact of a misspecified, uncertain a priori mean $\mu_a$ on retrieval bias and error covariance.

- Implement with a computationally efficient surrogate model.
Error variance in $X_{CO_2}$ grows with increasing uncertainty in the parameter $\mu_a$.

Both systematic and stochastic components interact to produce largest bias in M2V2.

Error distribution for $X_{CO_2}$ under different experimental conditions. Points depict the mean and error bars enclose the center 95% of the distribution.
Experiment Summary

- Ratio of error standard deviation to reported retrieval standard deviation

\[
\frac{\sqrt{\text{Var}(\Delta_{\text{XCO}_2})}}{\sqrt{E(\hat{\text{Var}}_{\text{XCO}_2})}}
\]

<table>
<thead>
<tr>
<th></th>
<th>M0V0</th>
<th>M0V1</th>
<th>M0V2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.025</td>
<td>1.045</td>
<td>1.234</td>
</tr>
<tr>
<td>M1V0</td>
<td>1.021</td>
<td>1.026</td>
<td>1.223</td>
</tr>
<tr>
<td>M2V0</td>
<td>0.993</td>
<td>1.002</td>
<td>1.648</td>
</tr>
</tbody>
</table>

- As uncertainty in $\mu_a$ increases, variability in realized errors surpasses the reported uncertainty.
- Bias is also largest for M2V2 treatment.
A geographically and seasonally comprehensive set of experiments is forthcoming. Spatially and temporally varying marginal distributions \((\mu_X, \Sigma_X)\) are required. Methodology is general and can be potentially applied for retrievals based on physical or empirical models. The approach can be used to provide an estimate of the retrieval error variance when an operational algorithm does not routinely produce an uncertainty estimate. Other applications of UQ in inverse problems can incorporate a Monte Carlo approach.

- Atmospheric data assimilation
- Hydrologic model calibration
Suggestions and contributions from Jenný Brynjarsdóttir, Brian Connor, Dejian Fu, James McDuffie, Vijay Natraj, Hai Nguyen, Chris O’Dell, Joaquim Teixeira, and Mike Turmon are appreciated.

Questions?

Copyright 2015 California Institute of Technology, all rights reserved. Government sponsorship acknowledged.
References


