Jet Propulsion Laboratory
California Institute of Technology

Level 2 UQ: Enabling Simulation-Based Uncertainty Quantification for Remote Sensing Retrievals

Jon Hobbs[1]

Joint work with Ali Behrangi[1,2], Amy Braverman[1], Eric Fetzer[1], Kyo Lee[1], Hai Nguyen[1], and Joaquim Teixeira[1]

[1] Jet Propulsion Laboratory, California Institute of Technology

[2] University of Arizona

Objectives

- AIST-16-0030 project *Simulation-Based Uncertainty Quantification for Atmospheric Remote Sensing Retrievals*
- Project will "develop statistical methods and analysis software to facilitate uncertainty quantification (UQ) for Level-2 atmospheric remote sensing data products produced by operational retrieval algorithms."
    - Apply technology to understand sources of uncertainty in Atmospheric Infrared Sounder (AIRS) Level-2 retrieval algorithm
    - Use technology to characterize the feasibility of drought detection with AIRS on regional scales, and other applications that use AIRS data
- AIST program identifies this capability could support an atmospheric science analytic center.

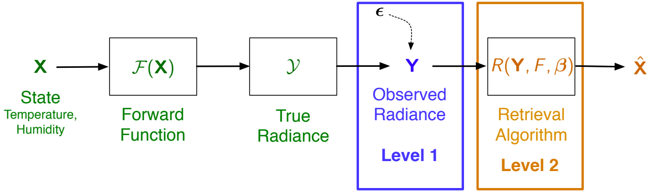
Data Uncertainty

- Data uncertainty represents lack of knowledge about a geophysical quantity of interest (QOI) *after observing relevant data*.
- The true value of the QOI, $\mathbf{X}$, is generally unknown, so plausible/likely values must be characterized.
- Probability offers a coherent framework for representing the distribution of the QOI, or the plausible error $\hat{\mathbf{X}} - \mathbf{X}$, given an estimate $\hat{\mathbf{X}}$ based on observed data.
- Earth science data records are relying on increasingly complex methods for constructing estimates $\hat{\mathbf{X}}$.
  - Remote sensing retrievals using satellite radiances and radiative transfer models
  - Data assimilation using Earth system models and multiple data sources

VVUQ

- National Research Council report (NRC, 2012) places uncertainty quantification (UQ) for complex physical systems in a probabilistic framework.
- UQ methodology seeks to identify the impact of sources, or contributors, to the distribution of the error for a QOI.
- A probabilistic framework benefits from representing the system as a data-generating process, with the QOI as an outcome.
- Monitoring the process includes describing the prediction error under a particular set of conditions, such as a particular version of a retrieval algorithm.
- Improving the process can result from improved understanding of error sources.
- UQ has a role in both monitoring and improvement.

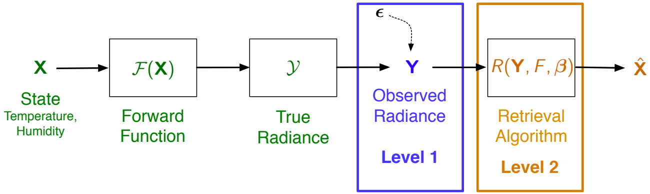
Observing System



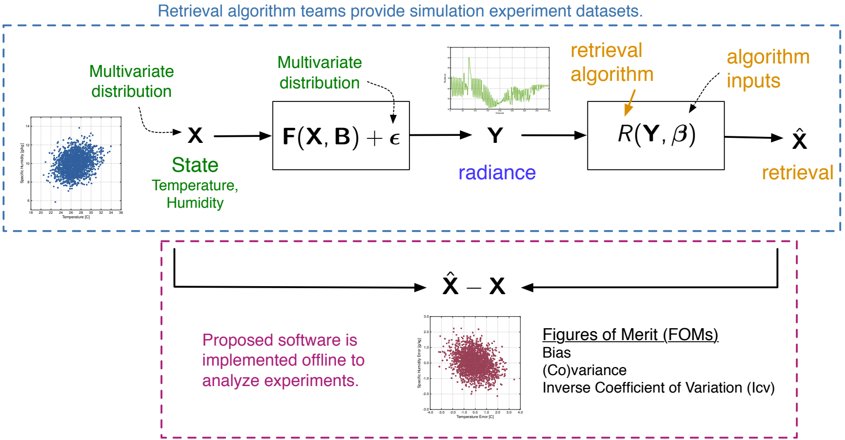
- Remote sensing observing system is a complex data-generating process with several key components.
    - True top-of-atmosphere radiance is a function of atmospheric state.
    - Instrument observes noisy radiance.
    - Retrieval algorithm produces estimate of state.
    - Science data system scales processing.
- Objective is inference on the state given the observed radiances, an *inverse problem*.

Observing System

- General retrieval objective: infer unknown surface and/or atmosphere states from remote sensing observations.
- Typically heterogeneous collection of unknowns, such as surface and atmosphere characteristics.
- Simulation of the data-generating process provides UQ insights.
- Ideally UQ includes characterizing the joint distribution of $[\mathbf{X}, \hat{\mathbf{X}}]$.



OSUE



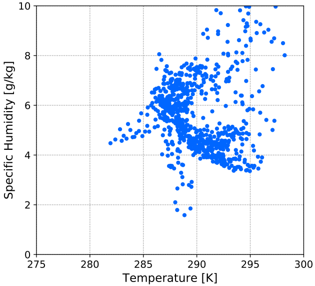
- Observing system uncertainty experiment

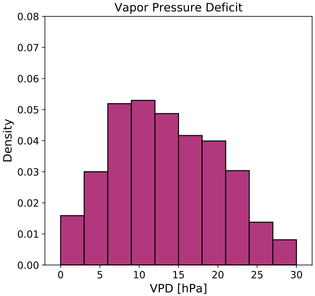
Figures of Merit

- Retrieval properties can be summarized with figures of merit (FOM) based on Monte Carlo experiment.
- FOM is a quantitative summary of the joint distribution $[\mathbf{X}, \mathbf{Y}, \hat{\mathbf{X}}]$. Examples:
    - Item-by-item average error, or bias
    - Item-by-item error standard deviation
    - Covariance matrix of retrieval errors
- Additional multivariate FOMs have been proposed for retrieval simulation experiments. (Hobbs et al., 2017; Cressie and Burden, 2015)
    - Normalized bias and error correlation for heterogeneous state vectors
    - Diagnosis of retrieval-based uncertainty estimates
- Project includes development of conditional FOMs based on $[\mathbf{X}|\mathbf{Y}]$

QOI

- Framework has flexibility for different retrievals R.
- Often interest in a functional QOI $\mathbf{g}(\mathbf{X})$ and retrieval $\mathbf{g}(\hat{\mathbf{X}})$.
    - Orbiting Carbon Observatory-2 (OCO-2) users focus on scalar X_{CO2}.
    - Some AIRS applications use vapor pressure deficit (VPD) as a primary QOI.

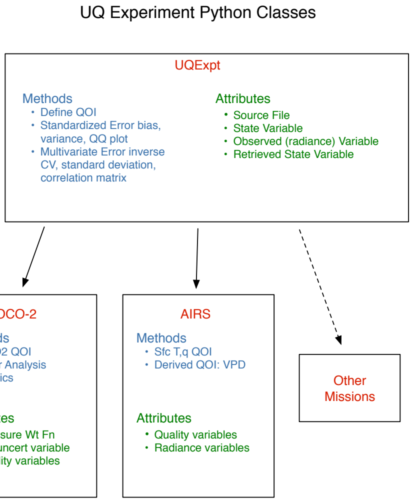




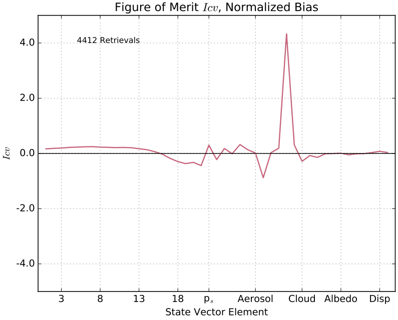
Project Objectives

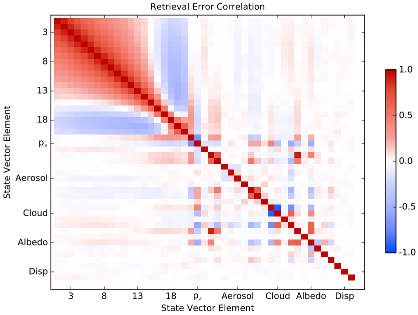
- Python module for analysis of OSUEs
  - Generic classes for figures of merit (FOM) that apply to various retrievals
  - Retrieval-specific classes: OCO-2, AIRS
- Implement OSUE for AIRS operational retrieval
  - Experiments for a variety of conditions, termed *geophysical templates*
  - Identify implications for AIRS data in applications

Module


UQ Experiment Python Classes
UQExpt
Methods
• Define QOI
• Standardized Error bias, variance, QQ plot
• Multivariate Error inverse CV, standard deviation, correlation matrix
Attributes
• Source File
• State Variable
• Observed (radiance) Variable
• Retrieved State Variable
OCO-2
Methods
• XCO2 QOI
• Error Analysis Metrics
Attributes
• Pressure Wt Fn
• OE uncert variable
• Quality variables
AIRS
Methods
• Sfc T,q QOI
• Derived QOI: VPD
Attributes
• Quality variables
• Radiance variables
Other Missions

OCO-2 Multivariate



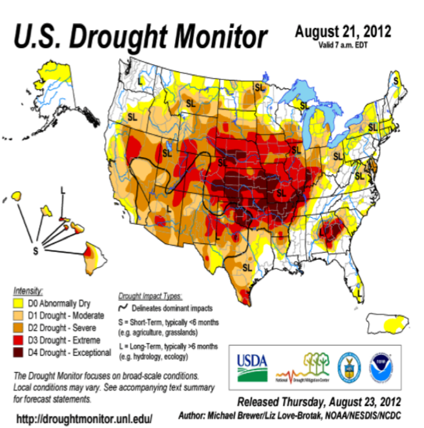


Multivariate retrieval error distribution for a single OCO-2 experiment

- Normalized bias, or *Icv*, summarizes bias relative to error variability for heterogenous state vector components.
- Retrieval error correlation matrix depicts magnitude of association among retrieval errors for state vector components.

Templates

- AIRS data have demonstrated utility, through derived quantities, to detect drought onset (Behrangi et al., 2016)
- Uncertainty in retrieved temperature, humidity propagate to derived drought indices
- Template based on 2012 Midwest US drought
- Ensemble of true states assembled from MERRA2



http://droughtmonitor.unl.edu

AIRS OSUE Workflow

Step I
Subset MERRA2 data (3-hourly & instantaneous)

1. 3-D variables
   - Air temperature
   - Specific humidity
   - Geopotential height
   - Cloud fraction
2. 2-D variables
   - Surface temperature
   - Surface pressure
3. Constants
   - Surface altitude
   - Land fraction

Step II
Prepare input files to run SARTA

1. Subset along AIRS tracks
   - Day (21 UTC)
   - Night (09 UTC)

2. Vertical interpolation
   - Air temperature
   - Water vapor molecular density

Step III
Run SARTA

1. Pick the closest AIRS granule to the MERRA grid
   - Copy viewing geometries (landfrac, solar zenith angle, etc.) from AIRS L1B into SARTA rtp file
   - Assume that within a 3x3 L1B grid, the only variable that varies is cloud fraction.

Step IV
Add radiance measurement errors

1. Errors are assumed to be zero mean and independent across channels
2. Standard deviation are given by the AIRS radiance errors.

Step IV
Run AIRS retrieval

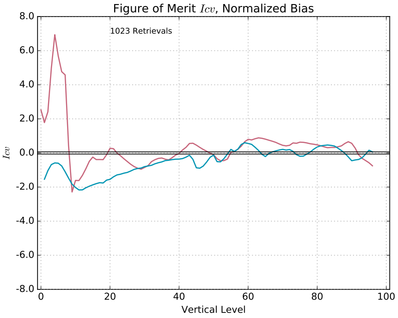
1. Copy the SARTA radiance back onto the AIRS L1B file

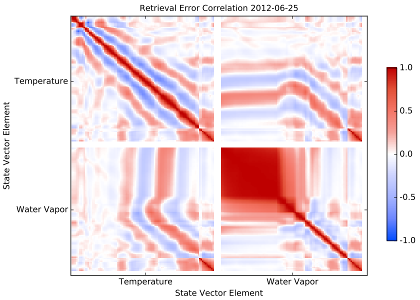
2. Retrievals are done using the standard AIRS algorithm.

3. Cloud clearing is included as part of the retrieval.

- AIRS OSUE workflow

AIRS Multivariate





Multivariate retrieval error distribution for a single AIRS experiment

- Error correlation matrix demonstrates vertical dependence for both variables, plus cross-dependence

Discussion

- Upcoming activities
  - Experimental design for AIRS simulations with clouds
  - Identify and process additional geophysical templates
  - Potential incorporation to Level 3 products
  - Python module examples and documentation
- Interaction with AIRS project and science teams
  - Synergy with other activities: validation, data fusion
  - Long term: potential contribution to uncertainty information in products
- Offline analysis with OSUEs executed by other project teams
  - Assessment of algorithm design as part of ReFRACtor (PI James McDuffie, JPL)

Questions?

Jonathan.M.Hobbs@jpl.nasa.gov

Jet Propulsion Laboratory
California Institute of Technology

References

References

Behrangi, A., Fetzer, E. J., and Granger, S. L. (2016). Early detection of drought onset using near surface temperature and humidity observed from space. *International Journal of Remote Sensing*, 37:3911–3923.

Cressie, N. and Burden, S. (2015). Figures of merit for simultaneous inference and comparisons in simulation experiments. *Stat*, 4:196–211.

Hobbs, J., Braverman, A., Cressie, N., Granat, R., and Gunson, M. (2017). Uncertainty quantification for retrieving atmospheric CO_2 from satellite data. *SIAM/ASA J. Uncertainty Quantification*, 5:956–985.

NRC (2012). *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*. The National Academies Press, Washington, DC.