Jet Propulsion Laboratory
California Institute of Technology

# A Simulation-Based Perspective on the Joint Probability Distribution of Atmospheric States and AIRS Retrievals

Jon Hobbs[1]
Joint work with Ali Behrangi[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

- Advanced Information Systems Technology (AIST) program within NASA ESTO
- 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 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

# Data Uncertainty

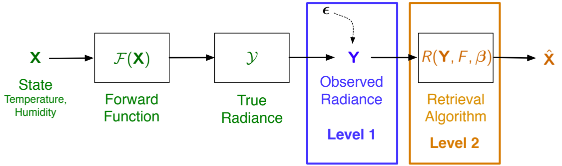
- Data uncertainty represents lack of knowledge about a geophysical quantity of interest (QOI) *after observing relevant data*.
- The true value of the QOI, $\mathbf{X}$, is generally unknown, so plausible/likely values must be characterized.
- Probability offers a coherent framework for representing 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

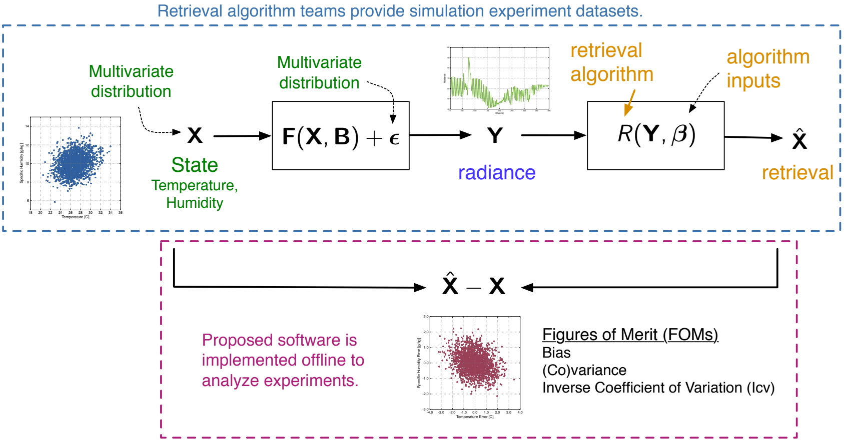
- National Research Council report (NRC, 2012) places uncertainty quantification (UQ) for complex physical systems in a probabilistic framework.
- UQ methodology seeks to identify the impact of sources, or contributors, to the distribution of the error for a QOI.
- A probabilistic framework benefits 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

- 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 $[\mathbf{X}, \mathbf{Y}, \hat{\mathbf{X}}]$

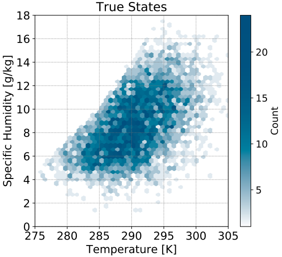
| | |
|---|---|
| $\mathbf{b} = E\left(\hat{\mathbf{X}} - \mathbf{X}\right)$ | Bias |
| $\mathbf{V} = Cov\left(\hat{\mathbf{X}} - \mathbf{X}\right)$ | Covariance |
| $\mathbf{D} = (diag(\mathbf{V}))^{1/2}$ | Std Dev |

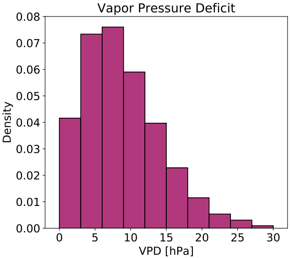
- Multivariate FOMs have been proposed for retrieval simulation experiments. (Hobbs et al., 2017; Cressie and Burden, 2015)

| | |
|---|---|
| $\mathbf{P} = \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}$ | Correlation |
| $\mathbf{z} = \mathbf{D}^{-1}\mathbf{b}$ | Icv |

# QOI

- Framework has flexibility for different retrievals *R*.
- Additional FOMs can diagnose reported retrieval uncertainties.
  - Role of nonlinearity
  - Prediction interval (region) coverage
- Often interest in a functional QOI $\mathbf{g}(\mathbf{X})$ and retrieval $\mathbf{g}(\hat{\mathbf{X}})$. Example: vapor pressure deficit (VPD)



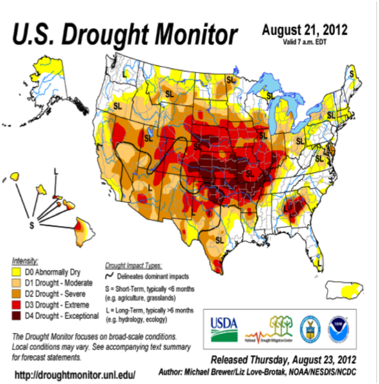



# Project Objectives

- 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

# Templates

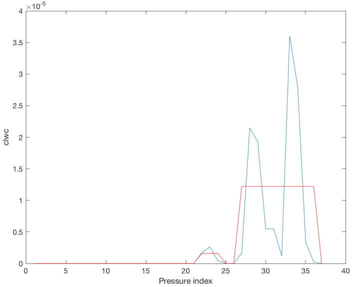
- A simulation experiment is executed with reference to a specific set of atmospheric and observing conditions, which constitute a *geophysical template*.
    - Range of times/locations
    - Reference data (reanalysis, in situ data)
- Project's AIRS templates motivated by applications
    - Drought detection (Behrangi et al., 2016)
    - Validation with MAGIC campaign (Zhou et al., 2015)

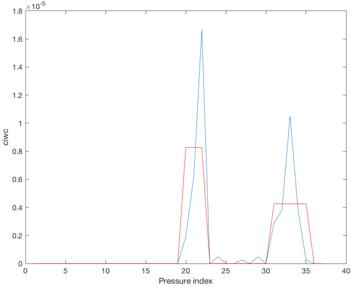



http://droughtmonitor.unl.edu

# Forward Model

- SARTA two-slab forward model (DeSouza-Machado et al., 2018)
- Construct cloud slab state from reanalysis cloud water/ice content



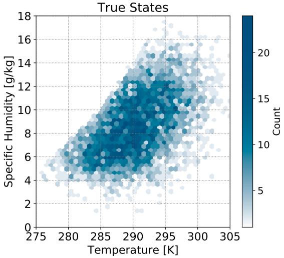



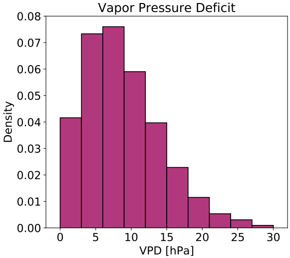

Example cloud slab definitions

# State Vector

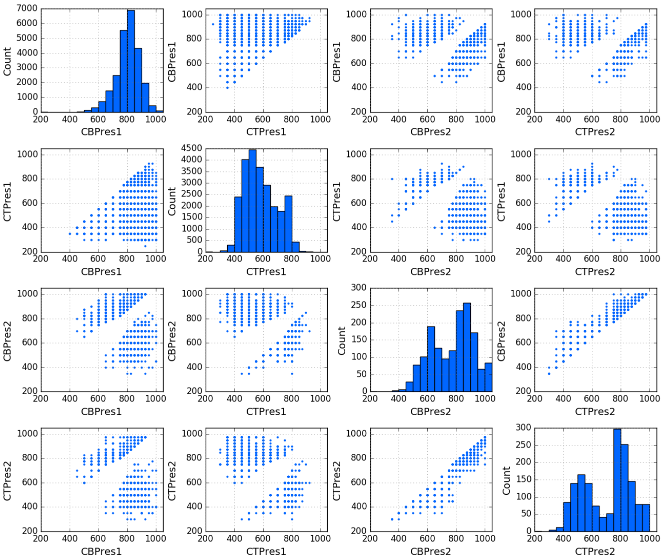
State contains atmospheric vertical profiles plus cloud properties for $\leq$ two slabs

Temperature vertical profile
RH vertical profile
Cloud fraction (each FOV)
Cloud type
Cloud temperature
Cloud top pressure
Cloud bottom pressure
Cloud particle size
Cloud non-gas water
Surface pressure, temp, altitude



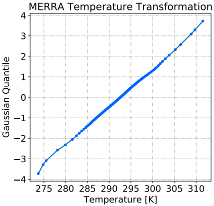



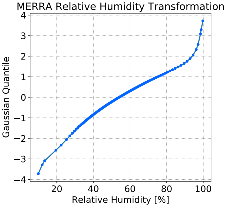
# Cloud States

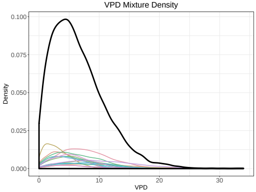


# Probability Model

- State vector ensembles can be informed by reanalysis, model nature run, and actual retrievals.
- Develop probabilistic representation using mixture modeling.
- Apply quantile transformation to preserve physical constraints.
- Synthetic states randomly generated from fitted model.

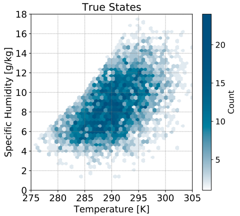


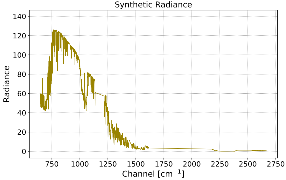



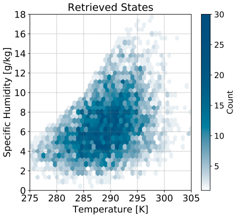



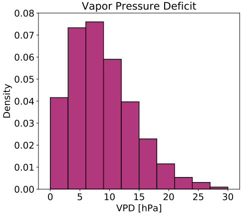

Probability model: Gaussian mixture with quantile transformation

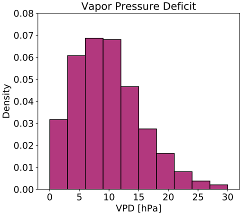
# Experiment


True States
Specific Humidity [g/kg]
Temperature [K]
Count

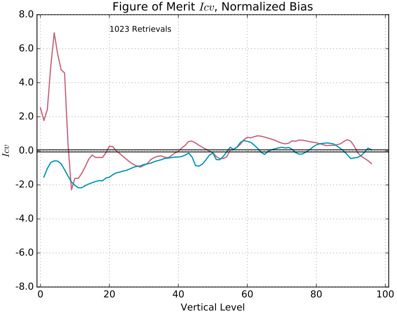

Synthetic Radiance
Radiance
Channel [cm$^{-1}$]

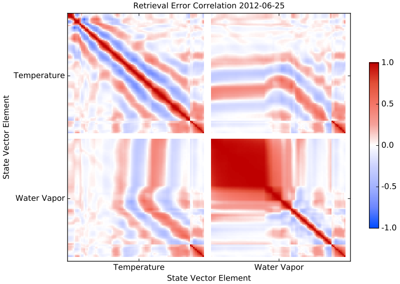

Retrieved States
Specific Humidity [g/kg]
Temperature [K]
Count


Vapor Pressure Deficit
Density
VPD [hPa]


Vapor Pressure Deficit
Density
VPD [hPa]

# AIRS Multivariate

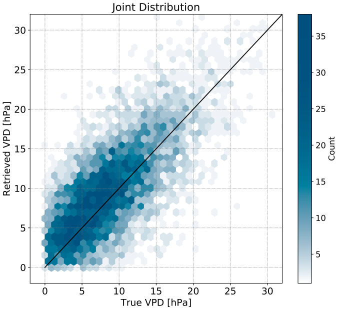


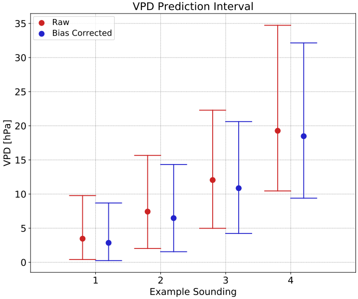




Multivariate retrieval error distribution for a single AIRS experiment

# Inference for VPD

- Experiment yields joint distribution for true $\mathbf{g}(\mathbf{X})$ and retrieved $\mathbf{g}(\hat{\mathbf{X}})$ QOI.
- Inference may focus on conditional distribution, $\left[\mathbf{g}(\mathbf{X})|\mathbf{g}(\hat{\mathbf{X}})\right]$
- Construct single-sounding prediction intervals, possibly bias-corrected



# Discussion

- Upcoming activities
  - MAGIC templates: Provide state vector ensembles to data fusion team for UQ pilot study
  - 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

- Suggestions and contributions from Bill Irion, Sergio DeSouza-Machado, Brian Kahn, Susan Kulawik, Maya Shen, and Ben Smith are appreciated.

Questions?

Jonathan.M.Hobbs@jpl.caltech.edu

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.

DeSouza-Machado, S., Strow, L. L., Tangborn, A., Huang, X., Chen, X., Liu, X., Wu, W., and Yang, Q. (2018). Single-footprint retrievals for AIRS using a fast twoslab cloud-representation model and the SARTA all-sky infrared radiative transfer algorithm. *Atmos. Meas. Tech.*, 11:529–550. doi:10.5194/amt-11-529-2018.

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.