Statistical Foundations for Representing Uncertainty in Earth Science Data Records

Jon Hobbs [1]

[1] Jet Propulsion Laboratory, California Institute of Technology

January 11, 2018

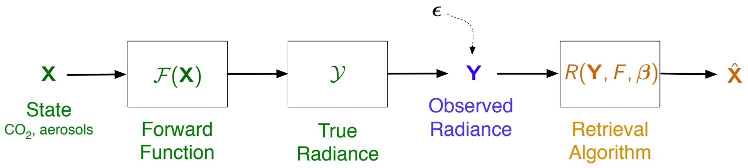
Data Uncertainty

- 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 (Rodgers, 2000)
  - 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 quantity of interest (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

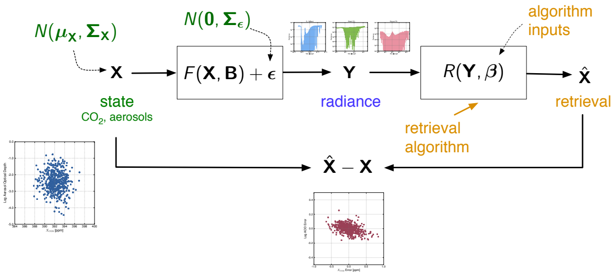


- Remote sensing observing system is a complex data-generating process with several key components.
  - True top-of-atmosphere radiance is a function of atmospheric state.
  - Instrument observes noisy radiance.
  - Retrieval algorithm produces estimate of state.
  - Science data system scales processing.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

Approaches

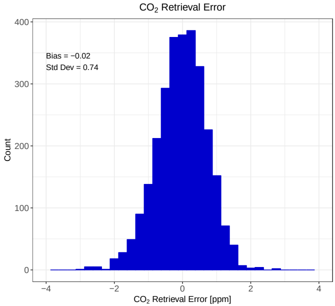
Multiple approaches for probabilistic assessment of observing systems

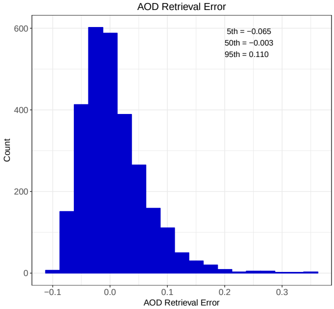
- In situ validation: Summarize the error distribution, $\hat{\mathbf{X}} - \mathbf{X}$, where substantially more accurate and precise observations of $\mathbf{X}$ are available.
- Simulation studies: Monte Carlo experiments with the data-generating process, estimation procedure, and ensembles of user-specified true QOIs $\mathbf{X}$.



Error Distributions

- How should uncertainty be summarized?
  - Bias, variance may be sufficient for a symmetric error distribution.
  - Quantiles may be more appropriate for skewed, mutli-modal distributions.





Current Work

- Toward Unified Error Reporting (TUNER): International effort to provide validation-based error assessment for retrievals of comparable QOIs from different satellites.
- NASA AIST effort to develop tools for simulation-based UQ for retrievals (Hobbs et al., 2017)
- JPL internal initiative on UQ for Earth science applications

Acknowledgments

- Contributions from Amy Braverman, Mike Gunson, David Moroni, Hai Nguyen, and Mike Turmon are appreciated.

Questions?

Jonathan.M.Hobbs@jpl.nasa.gov

References

Hobbs, J., Braverman, A., Cressie, N., Granat, R., and Gunson, M. (2017). Uncertainty quantification for retrieving atmospheric CO2 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.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding*. World Scientific, Hackensack, NJ.