Probability as a Foundation for Data Uncertainty: Applications in Remote Sensing

Jon Hobbs

1 Jet Propulsion Laboratory, California Institute of Technology
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, $X$, 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 $\hat{X} - X$, given an estimate $\hat{X}$ based on observed data.

- Earth science data records are relying on increasingly complex methods for constructing estimates $\hat{X}$.
  - Remote sensing retrievals using satellite radiances and radiative transfer models (Rodgers, 2000)
  - Data assimilation using Earth system models and multiple data sources
• 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.
Multiple approaches for probabilistic assessment of observing systems

- **In situ validation**: Summarize the error distribution, $\hat{X} - X$, where substantially more accurate and precise observations of $X$ are available.

- **Simulation studies**: Monte Carlo experiments with the data-generating process, estimation procedure, and ensembles of user-specified true QOIs $X$. 

\[
N(\mu_X, \Sigma_X) \xrightarrow{\text{state}} X \xrightarrow{F(X, B) + \epsilon} Y \xrightarrow{R(Y, \beta)} \hat{X}
\]

- **Radiance**
- **Retrieval**
- **Algorithm inputs**
- **Error [ppm]**
- **Log Aerosol Optical Depth**
Error Distributions

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

Bias = −0.02
Std Dev = 0.74

CO₂ Retrieval Error

AOD Retrieval Error

5th = −0.065
50th = −0.003
95th = 0.110

July 18, 2018 Statistics and UQ 6 jpl.nasa.gov
Current Work

• Toward Unified Error Reporting (TUNER): International effort to provide validation-based error assessment for retrievals of comparable QOIs from different satellites. http://www.issibern.ch/teams/tuner/

• NASA AIST effort to develop tools for simulation-based UQ for retrievals (Hobbs et al., 2017)
  • Application to OCO-2 and AIRS Level 2 retrievals

• JPL internal initiative on UQ for Earth science applications
Acknowledgments

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

Questions?
Jonathan.M.Hobbs@jpl.nasa.gov

References


