Uncertainty Quantification and Climate Science

Amy Braverman

Jet Propulsion Laboratory, California Institute of Technology

July 29, 2012
Climate models are deterministic, mathematical descriptions of the physics of climate.

They provide a laboratory for experiments that we can’t carry out in the real world. Also used to make predictions about future climate.

Should we believe them? Can we quantify how uncertain their predictions are?

SAMSI program: UQ as a discipline and UQ as practiced re: climate models.

How can modern observational data sources be brought to bear?
UQ as a Discipline

- NRC study on verification, validation, and uncertainty quantification (2012).
- Verification = How well does computer code solve equations of the mathematical model?
- Validation = How well does the mathematical model represent the true physical system?
- UQ = “The process of quantifying uncertainties associated with model calculations of true, physical quantities of interest with the goals of accounting for all sources... and quantifying the contributions of specific sources.”
- Climate modeling case study.
Verification: two broad classes: code verification and solution verification.

Validation:

- How good is the underlying mathematical model? Leads to "structural" uncertainty.
- Model calibration: ancillary unknowns must be estimated from data. Leads "parametric" uncertainty.
- Interpolative predictions of primary unknowns ↔ simulations of past and present (we have data).
- Extrapolative predictions of primary unknowns ↔ predictions of the future (we have no data).

UQ:

- Quantify uncertainties in model inputs.
- Propagate these uncertainties through calculation.
- Quantify variability of the true quantities of interest.
- Aggregate uncertainties from different sources.
Quantity of interest: time for a bowling ball to fall from 100m tower.

Mathematical model: a function of \( g \) (gravitational constant). Assume it is imperfectly known and must be estimated.

Data: experimental drop times from 10, 20, 30, 40, 50m.

Validation experiment: drop time from 60m.

Sources of uncertainty: mathematical model, true value of \( g \), data on drop times used to estimate \( g \), form of the estimate of \( g \), extrapolation to 100m.
Quantity of interest: (say) annual precipitation in the western US in 30 years.

Mathematical model: equations reflecting best physics understanding instantiated on a coarse grid in space and time. Many poorly understood feedbacks. Exchangeability of present and future; effects of different forcings/scenarios.

Data: many heterogeneous sources (e.g., remote sensing, surface stations, radiosondes, aircraft) but massive yet incomplete, and themselves uncertain; often the result of inference.

Validation experiment: CMIP5 decadal “experiments" and observations?

Sources of uncertainty: process understanding over different spatial and temporal scales, model resolution, initial conditions, unknown parameters, inherent variability of (true) precipitation, and so on.
2011-12 Program on Uncertainty Quantification: Climate Modeling

The area of climate modeling is a quintessential field of application for UQ. In fact, a sizable part of the research done in the quantification of uncertainty in computer models has been driven by the pressing needs of climate modelers. Climate models are computer codes based on physical principles that simulate the complex interactions between the many parts of the Earth system such as atmosphere and oceans. A significant issue of interest is for instance the problem of regional climate as the global models cannot on their own give enough information at the regional/local scales. Various strategies (downscaling and upscaling) can be considered whereby various combinations of global and regional models are combined to gain information on model uncertainty. Other issues include: development of reliable atmospheric and ocean models and interface between the two, regional and local risk assessment from models, paleoclimate models and how to best deal with models that have huge uncertainties and biases.

Organizers: Amy Braverman (JPL, California Institute of Technology, and UCLA), Xabier Garaizar (LLNL), Dave Higdon (LANL), Gardar Johannesson (LLNL), Donald Lucas (LLNL)
Opening Workshop: review current landscape and state-of-the-art approaches.

Working Groups: conduct research on specific topics of interest to participants.

Observations Workshop: focus specifically on the role of observational data in UQ for climate models.

What’s new about this: data (observations) are becoming more important!
Two major roles:

- Estimate pdf’s (or parameters of those pdf’s) of processes that are not well enough understood to model directly. (Model “calibration" in UQ-speak, “parameterizations" in climate-model-speak.)

- Compare against climate model simulations to characterize uncertainties in model output. (Model “validation" in UQ-speak.)

Issues:

- *Uncertainty in the observations needs to be quantified and accounted for in any analysis that uses them.* How best to capture/communicate this? (Asheville: provide an ensemble of observations from which to sample?)

- Observational data sets are massive and heterogeneous (e.g., different support, measurement error, sampling).
**Observations**

**Inferred** (e.g., remote sensing)
- radiance measurement error
- underlying forward model
- method of inference ("retrieval")
- spatial and temporal aggregation (satellites)
- massive size

**Directly measured** (e.g., surface stations, aircraft)
- measurement error
- "point-level" spatial support
- spatially dense in places

- distributed storage
- sampling relative to "true" field
- spatially coarse, but often dense
- often temporally sparse
- relatively short record

- temporally dense in places
- records as long as hundreds of years
UQ for climate observations (remote sensing):

- Footprint radiance vector
- "Retrieval"
- Atmospheric state estimate

\[
\hat{Y} = F(X) + \epsilon
\]

\[
\hat{X} = E(X|Y)
\]

How to “validate” \( \hat{X} \)? How to quantify its uncertainty?

Bottom-up: propagate input uncertainties through retrieval.

Top-down: compare \( \hat{X} \) to “truth”. (“Truth”? Really?)
Data “fusion" and “homogenization":

- **Fusion**: combine multiple heterogeneous observational data sets, in a way that exploits their complementary strengths, to estimate true geophysical fields with minimum uncertainty.

- **Homogenization**: remove biases (or other artifacts) due to non-climate sources, from climate data.

- These are related- they are both about inferring the true field (or getting closer to it).
Uncertainty quantification for climate observations is important in both model “calibration”/parameterization and model “validation”.

Uncertainty quantification for climate observations is important upstream: observations used to formulate hypotheses ultimately instantiated in climate models.

Is UQ the same for all these purposes?

UQ for simulations of past/present vs. UQ for predictions of the future?
Comments?

Contact Amy.Braverman@jpl.nasa.gov.

Support for this work is provided by NASA’s Earth Science Data Records Uncertainty Analysis Program.

This work was performed at the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged. Copyright 2012, California Institute of Technology. All rights reserved.