Statistical Challenges in Uncertainty Quantification

Murali Haran

Department of Statistics, Pennsylvania State University

PCHES

October 2019.

Talk Goal

- Provide a sketch for statistical approach to uncertainty quantification
- Will not discuss details of our new research/methodology
- Idea is to share thoughts about what we mean by uncertainty and how we account for it
- Start a discussion about how this may be related to other projects, and other notions of uncertainty

Projections

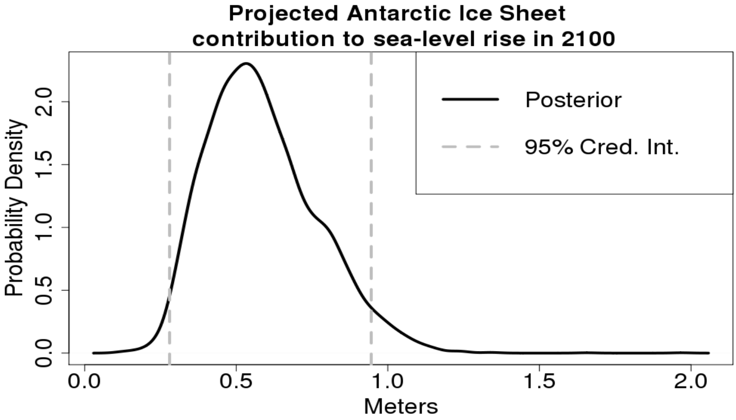
- How do we project sea level rise? Coastal flooding/extreme storm surge events?
- Approaches:
  - **Purely statistical models**: use past data to understand the statistical distribution of, say, extreme storm surge events. Make projections based on estimate of how the distribution is evolving over time
  - **Physics-based models**: make projections based on models that describe dynamics. These projections also use observational data (“constraints”) to
- Both types of projections are highly uncertain

Projections and Statistics

- Basic principle: predictions should be reported as probability distributions
- Rigorous approach for sound decision-making
  Only "coherent" approach is to use probability (see work by de Finetti, Savage and others)
- For assessing risk of a decision: risk is an expected value (probability-weighted average) of an outcome

It is easy to dismiss our work if we do not write down uncertainties

Example of End Result

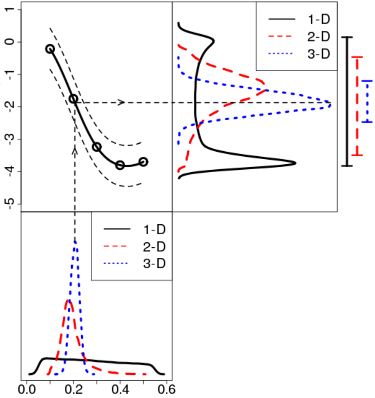


(Lee et al., 2019)

Important: this probability distribution accounts for many sources of uncertainty...

Translating Uncertainties into Projections

E.g. parametric uncertainty gets translated into projection uncertainty



(Chang et al., 2016)

Uncertainties

1. Models can never fully describe a complex system. Structural uncertainty: Which model features to prioritize?
2. Internal/natural climate variability: interactions between components of the climate system can result in variability in the climate system that is “internal” to the system.
3. Boundary or initial condition uncertainty.
4. “Forcings” uncertainty, e.g. uncertainty about emissions.
5. **Parameter uncertainty**: parameters (“dials” in the computer model) may be uncertain.
6. Observations are never perfect: measurement error, interpolation uncertainty

Quantifying Uncertainties

Stats/UQ group (Ben Lee, Vivek Srikrishnan, Kelsey Ruckert, Klaus, Murali + many grads/postdocs) has focused on addressing a few different kinds of uncertainty:

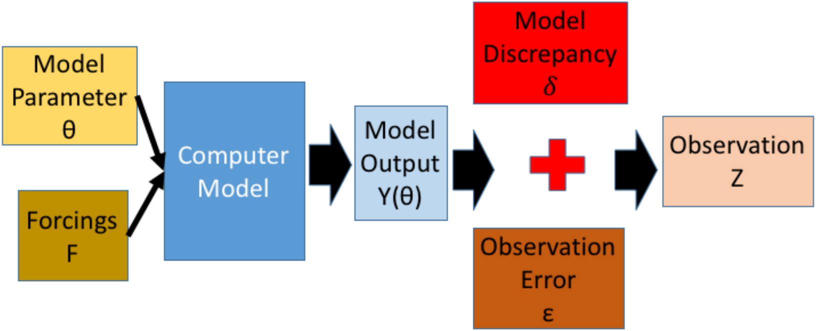
Structural uncertainty via Bayesian model averaging (BMA)

- Each model is given a probability (weight) based on match to observations, p_i for model M_i, $i = 1, .., k$
- Projections = weighted average of model projections (M_i^*),

$$\text{BMA projection} = \sum_i p_i M_i^*$$

Quantifying Uncertainties: Parametric + Observational

An outline of how physical model is combined with observations to learn about parameter θ. Result: *probability distribution* on θ



(Lee et al., 2019; Chang et al., 2016)

Bayesian Approach for UQ

How do we translate that diagram into a distribution on θ?

- Write down probability distributions (statistical models) for model ouptut, data, observational error: each of the boxes and how the relate to each other
- Let ξ be all remaining parameters in the models
- Specify **prior** distribution $p(\boldsymbol{\theta}, \xi)$
- Results based on posterior distribution: the probability distribution of θ, ξ *given* everything we have observed, $\pi(\boldsymbol{\theta}, \xi \mid \mathbf{Z}, \mathbf{Y}) \propto \mathcal{L}(\boldsymbol{\theta}, \xi) p(\boldsymbol{\theta}, \xi)$

Uncertainties (some of them) are now quantified!

Research Challenges?

- Modeling: How to formulate the flow chart + probability models in the right way
- Computing: How to obtain posterior distribution on θ?
  - Expensive: requires that we run the computer model at lots of settings
  - Data sets can be large and complicated (e.g. satellite data)
  - Emulation approach (Chang et al, 2016, 2014): replace computer model with approximation
  - New heavily parallelized particle Monte Carlo algorithm (Lee et al., 2019)
- Others?