PECOS

Predictive Engineering and Computational Sciences

Research Challenges in VUQ

Robert Moser

The University of Texas at Austin

March 2012

Justifying Extrapolitive Predictions

Need a credible process to make predictions of unobserved QoI's using imperfect models and imperfect data that accounts for the imperfections.

"Prediction is difficult, especially about the future"–Niels Bohr

Modeling Uncertainties

- Prior information—this may be all you have
  - Ignorance representations (e.g. for max. entropy)
  - Qualitative information (e.g. expert opinion)
  - Physics constraints
  - Inconsistent legacy data
  - Correlations
- Uncertainty in data
  - Complete characterization of all uncertainties
  - Correlation and dependencies with other data and possibly with prior information
- Model inadequacy
  - Physics constraints
  - Spatial/temporal structure for functions & fields
  - Calibration & priors for uncertainty model parameters

A Processes for Predictive Validation

1.Inform models τ^m and ϵ_{mod}

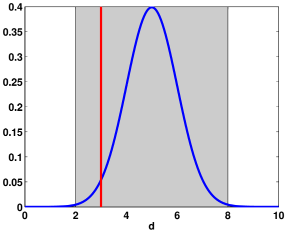
- Use data for observables D_c from scenarios ξ_c
- Bayesian inference to calibrate θ_i for models τ_i^m and meta-parameters for $\epsilon_{\mathrm{mod}\,i}$ for model classes i
- Bayesian model selection among model classes i

2. Challenge selected models

- Use data for observables D_v from scenarios ξ_v (include D_c from ξ_c)
- Are physics $+$ uncertainty models consistent with observations?
- Uncertainty models **must** account for all discrepancies between physics models and observations

Consistency Assessment

- Physical model + probabilistic statements consistent with data?
- Are all available data plausible plausible results of the physics and uncertainty models?
- What measures are appropriate?
    - Credibility intervals, area metric, p-values
    - As stated, this does not appear to be a Bayesian quesiton, is there a Bayesian formulation?



A Processes for Predictive Validation

3. Asses validity of predictions

- Does prediction scenario exercise embedded models outside the conditions for which it has been challenged? (requires characterization of relevant “conditions”)
- Are prediction quantities q sensitive to uncertainties to which observed quantities are not?
- Are prediction uncertainties in q too large for decision maker

Entitled to make predictions **only** if answers to questions in (3) are “**no**”

Predictive Assessments

Relevant Scenario Parameters for Embedded Models

- For some cases, this is clear (e.g. T and P in chemical kinetics models)
- When it is not, how to determine scenario parameterization?
- If it's another modeling assertion, needs to be "validated"
- Characterize when an unreliable embedded model is being used extrapolatively, rather than interpolatively.

Have Dominant Uncertainties Been Well Characterized

- Predictions & uncertainties should be dominated by well-known and well-calibrated components of the model.
- How can this be assessed rigorously?
  - Like a signal to noise ratio

Data Uncertainty and Model Inadequacy

$$\begin{aligned} 0 &= \mathcal{R}(u, \tau(\theta) + \epsilon_{\mathrm{mod}}; \xi) \\ D &= d(u, \tau(\theta) + \epsilon_{\mathrm{mod}}; \xi) + \epsilon_{\mathrm{exp}} \\ Q &= q(u, \tau(\theta) + \epsilon_{\mathrm{mod}}; \xi) \end{aligned}$$

Predictive Uncertainty

$$\begin{aligned} 0 &= \mathcal{R}(u, \tau(\theta); \xi) \\ D &= d(u, \tau(\theta); \xi) + \epsilon_{\mathrm{exp}} + \tilde{\epsilon}_{\mathrm{mod}} \\ Q &= ? \end{aligned}$$

Kennedy & O'Hagen

Uncertainty in predictions, **q**, arise from uncertain parameters, $\boldsymbol{\theta}$, $\boldsymbol{\xi}$
AND model inadequacy ϵ_{mod}

Some Caveats:

- all sources of uncertainty have been identified
- the data are accurate with well characterized uncertainties
- computational models are reliable (verified)
- numerical solutions are well resolved

Algorithms for Inference With Stochastic Models

- Example, when $\mathcal{R}$ is a PDE, inadequacy model for τ makes it a stochastic PDE
  - Also have to calibrate inadequacy model
- Likelihood evaluation involves solution of stochastic PDE
- Naive sampling algorithms lead to nested MCMC/MC sampling
- Need effective algorithms to avoid this calculation or make it tractable