Assessing model performance

James Watmough

Overview

- Holistic vs reductionist
  - keep in mind what your model is assuming and what it neglects
  - it's the things you didn't know you didn't know that stab you in the back
- model assessment is a testing of model assumptions
  - hypothesis testing involves model comparisons, not comparisons between models and *reality*
  - creative *visualizations* that best illustrate assessment
  - relevant statistics
  - avoid just listing and tallying scores
  - descriptive vs inferential statistics
- identify structural errors in models

Visualization

1. Descriptive statistics of model outputs (mean, variance, histograms)
   - is the model behaving as expected?
   - plausible predictions for latent variables
2. Graphical comparisons of model and data
   - predicted vs observed plot
   - *residuals* vs predicted (process *residuals* vs observation error)
   - residuals vs time/space/aspect
   - looking for outliers in data and structural *gaps* in predictions

Revising model after step one is OK, revising model after step 2 and *re-assessing against same data* is suspect!

Statistical Analyses

- Visual analysis are subject to our hard-wiring to pick out patterns, even when none are there
- Statistics can insert some objectivity into patterns identified through visual analyses
  - RMSE $\sqrt{\frac{1}{n}\left(\text{Predicted} - \text{Observed}\right)}$
  - correlation coefficients (deviations from Predicted $=$ Observed)
  - model variance vs observation variance (SD ratios)
  - Taylor Diagrams

Null Models, Hierarchical Models, Nested Models

- Remember we are not *ranking* models as good or bad, but assessing relative performance and identifying structural gaps

Calibration versus Validation

Analysis of Residuals

- residuals as observed − predicted
  - are residuals distributed as hypothesized?
  - look at residuals vs various aspects
    - predicted
    - time
    - cofactors
- residuals as likelihood of observed given model
  - quantiles of `observed` should be uniformly distributed