The Effect of Including Observational Uncertainty in the ILAMB Scoring System

Nathan Collier, Forrest Hoffman (ORNL)
Dave Lawrence (NCAR)

RUBISCO Science Friday
January 2021

Introduction

For some time we have been developing the methodology to include observational uncertainty in our scoring. There are still only few datasets which provide an estimate, but the list is growing:

- precipitation, runoff, water storage, radiation, sensible heat, latent heat, and ground heat (Conserving Land-Atmosphere Synthesis Suite (CLASS), doi:10.25914/5c872258dc183))
- nitrogen fixation (Davies-Barnard, et al., doi:10.1029/2019GB006387)
- boreal forest biomass (Thurner, et al., doi:10.1111/geb.12125) *

In this talk, I will present our ideas and make the case that this methodology should be used when uncertainties are present by default.

Bias Uncertainty Score

Bias score from (*Collier, 2018*), where the overbar reflects a time mean,

$$\begin{aligned} bias(x) &= \overline{v_{\mathrm{mod}}}(x) - \overline{v_{\mathrm{ref}}}(x) \\ \varepsilon_{bias}(x) &= bias(x)/std(v_{\mathrm{ref}}(x)) \\ s_{bias}(x) &= exp(-\varepsilon_{bias}(x)) \end{aligned}$$

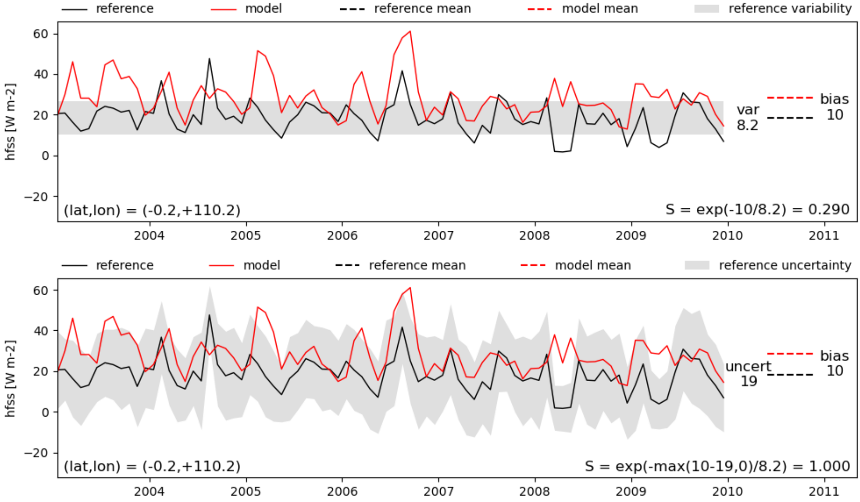
which reflects that the error in bias is relative to the variability at a given location. If a variable's uncertainty is given as $\delta v(t, x)$, then we can formulate a relative error which only penalizes bias beyond the uncertainty as,

$$\begin{aligned} \varepsilon_{bias}^{uncert}(x) &= max(|bias(x)| - \overline{\delta v}(x), 0)/std(v_{\mathrm{ref}}(x)) \\ s_{bias}^{uncert}(x) &= exp(-\varepsilon_{bias}^{uncert}(x)) \end{aligned}$$

What does this look like at a point?

Top: standard method

Bottom: method with uncertainty



RMSE Uncertainty Score

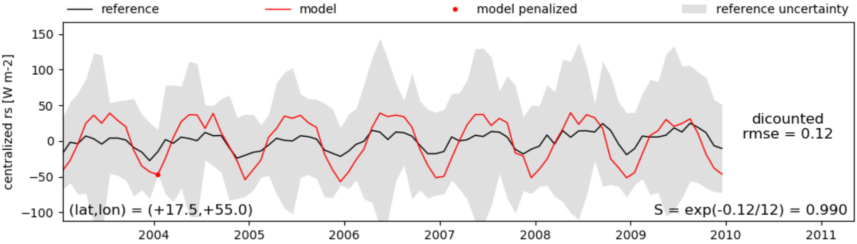
In the same spirit as the bias, we formulate a RMSE relative error only penalizing points which go beyond the uncertainty at any point in space and time.

$$\varepsilon_{rmse}^{uncert}(\mathrm{x}) = \frac{\sqrt{\frac{1}{t_f - t_0}\int_{t_0}^{t_f} max\left(\left|v_{\mathrm{mod}}^c(t,\mathrm{x}) - v_{\mathrm{ref}}^c(t,\mathrm{x})\right| - \delta u(t,\mathrm{x}), 0\right)^2 \; dt}}{std(v_{\mathrm{ref}}(\mathrm{x}))}$$

What does this look like at a point?

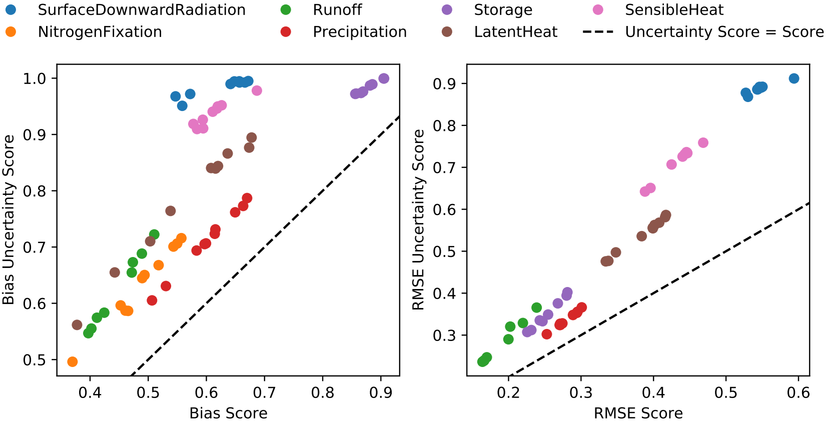
Top: standard method

Bottom: method with uncertainty



Comparison of Scores with Uncertainty

- Each dot is a pair of scores of a model in the CMIP6 archive.
- All pairs are in the top-left half, reflecting that uncertainty scores are strictly $\geq$ than the regular scores.
- Linear structure of each variable reflects that models retain the rank relative to others.

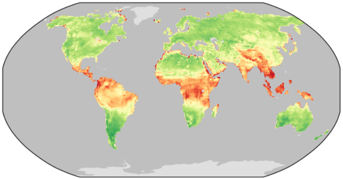


RMSE Score Map - Radiation

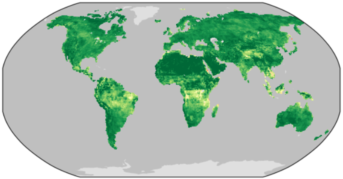
In the case of incoming radiation from CLASS, including uncertainty reveals that most models are perfect globally (CESM2 shown).

- may be distateful that our comparison does not distinguish among models
- does it rather mean that models are doing well enough given our certainty in the data
- suggests models should spend effort elsewhere and that we need more certain data

RMSE Score



RMSE Uncertainty Score



Remarks

- That there is no major reshuffling of model ranking is a nice result, this implies that we have not biased our notion of which models are 'good' by exlcuding uncertainty
- This is likely due to greater variance across models in areas with larger uncertainty
- This result holds across 7 variables from 2 sources, but will need more verification
- While this methodology in some sense blunts our tool, it also keeps us from making strong statements where we are not certain
- It makes score maps more useful for identifying areas of improvement
- Including it our main methodology will only affect comparisons where observational uncertainty is present