

Discussing some ideas (HoliSoils)

Lorenzo Menichetti

2023-03-20

Uncertainties to include

- ▶ Structure of SOC decay model
- ▶ Parameterization of SOC decay model
- ▶ Initialization
- ▶ Climate scaling
- ▶ Input estimation

The Bayesian model (to repeat for each SOC decomposition structure)

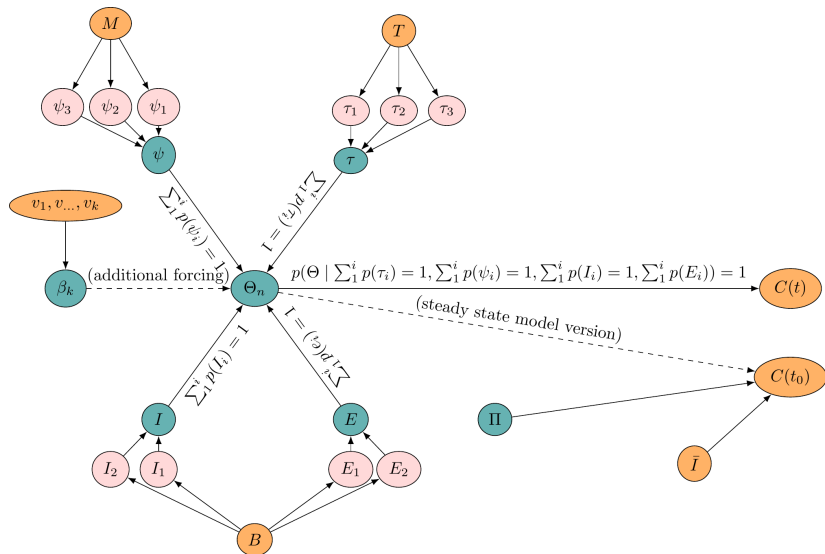


Figure 1: Orange = data, pink = multiple choices, teal = standard node

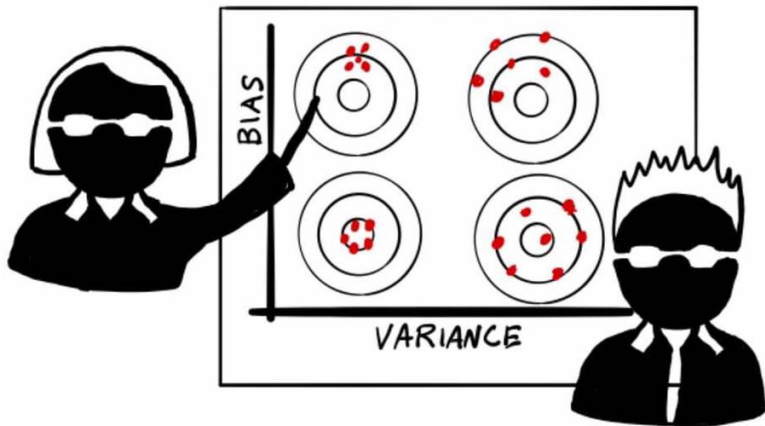
Considering the model structural uncertainty

The model structural uncertainty is left outside the Bayesian calibration in order to treat it more in detail.

More specifically, I want to optimize the final ensemble for robustness with model selection.

Optmizing for fitness alone does not cut it, **because overfitting**.

The bias/variance trade off



Even with parametric models, the training data is never fully representative. A model too compliant with the data will have good fit, but can carry over **bias**. The worst for robust extrapolation in a distant future.

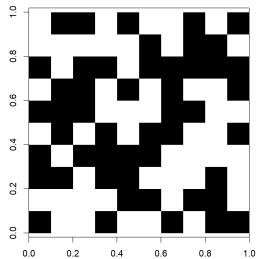
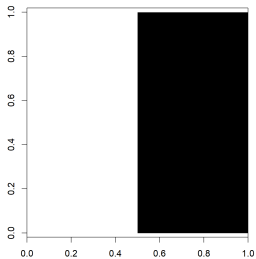
Model complexity definitions

My aim is minimizing ensemble bias, still retaining a decent fitness. A less complex model tend to have less bias (and less fitness), and I define complexity around this aim

Graph/parameters based definitions

Number of parameters, or better number of connections in a DAG.

Entropy based definitions



Entropy based definitions: residuals

Residuals:

In the literature it is commonly said that entropy in the residuals means less bias, since it means that the information left in the error is less organized, and more probably random. **More entropy = better**

Entropy and equifinality?

Will models with high equifinality, more complex, have higher entropy of their posteriors? More interactions means multimodal distributions, so there seems to me to be a mid-frequency spectrum where **more entropy = worse**

Possible entropy definition: Shannon H

We can consider the MCMC a discrete distribution and apply Shannon H calculation to it

$$H = \sum_{i=1}^n -p(x_i) \cdot \log_2 p(x_i) \quad (1)$$

Where x is our likelihood and i is each realization of the MCMC

Other possible entropy definitions?

Is anyone confident with signal processing? Like Fourier and so on?
If we could analyze the distributions in the frequency spectrum,
something interesting might appear, maybe?