Ensembling geophysical models with Bayesian neural networks

Matt Amos, Ushinish Sengupta
Scott Hosking, Paul J. Young.
m.amos1@lancaster.ac.uk

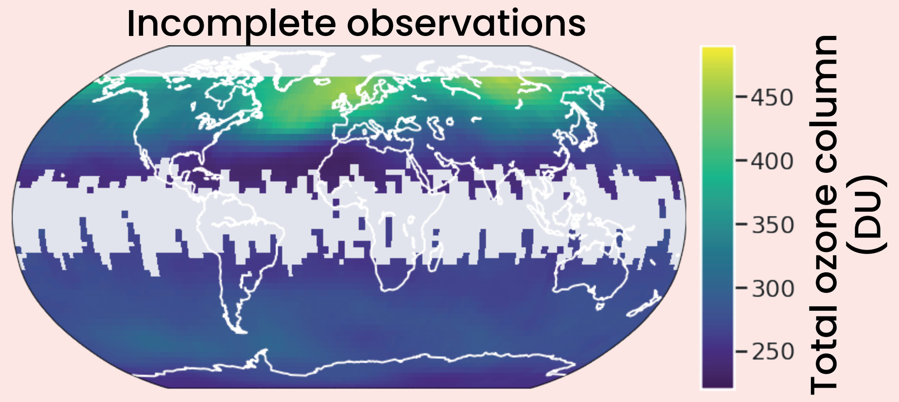
The basics

Two problems

Historic observations are incomplete. This is often due to suboptimal coverage or adverse meteorological conditions.

Methods to ensemble geophysical models are often basic and don't allow for variable model performance or similarities between models. Also the quantification of uncertainties is not rigorous.

Incomplete observations



Methods - The Bayesian neural network

We model observations as a linear combination of geophysical models:

$$\underbrace{y(\mathbf{x},t)}_{\text{Observations}} = \sum_{i=1}^{n} \underbrace{\alpha_i(\mathbf{x},t)}_{\text{Model weight}} \underbrace{M_i(\mathbf{x},t)}_{\text{Model prediction}} + \underbrace{\beta(\mathbf{x},t)}_{\text{Bias}} + \underbrace{\sigma(\mathbf{x},t)}_{\text{Heteroscedastic noise}}$$

We learn the model weights, bias, noise (as a function of space and time) with a **Bayesian neural network** which:

- Avoids overconfidence where there is no data
- Allows encoding of out prior beliefs that bias and noise should be small, and that a uniform weighting of geophysical models is a good starting point.

The ML details

The Bayesian neural network

Encoding domain knowledge in the prior

We map the coordinates from chemistry climate model output to make it 'ML ready'. Part of this is encoding our knowledge of the system into the data e.g. that January is adjacent to December or that 0º longitude is that same as 360º.

Time is warped onto a helix to enforce seasonal periodicity within the prior biases and weights.

$$t \rightarrow (\cos(2\pi t/T),\ \sin(2\pi t/T),\ t)$$

We convert to 3D Euclidean to make coordinates continuous across boundaries.

Scalable inference:

We use randomised MAP (maximum a posteriori) sampling to perform Bayesian inference. A number of neural networks are regularised (anchored) according to our prior, which is that all geophysical models are equally good. See the figure to the right for a depiction of this. The regularisation ensures that the variance is high for regions of sparse data as the individual neural network produce slightly different predictions.

The trained networks represent individual samples of the posterior, and from these we can compute the mean prediction and variance.

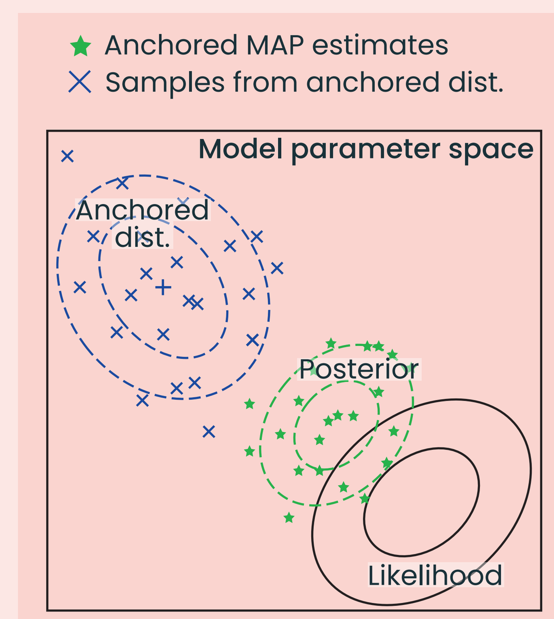
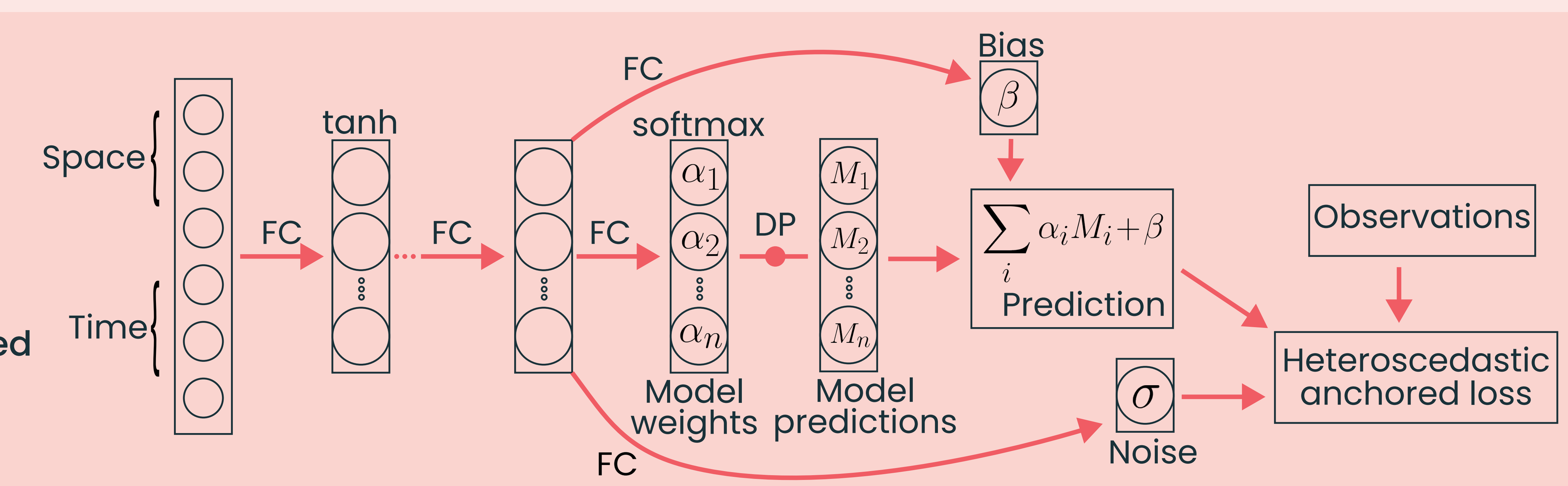


Figure adapted from Pearce et al. Uncertainty in neural networks: Bayesian ensembling (2020)

Neural network architecture:
Inputs are spatio-temporal coordinates

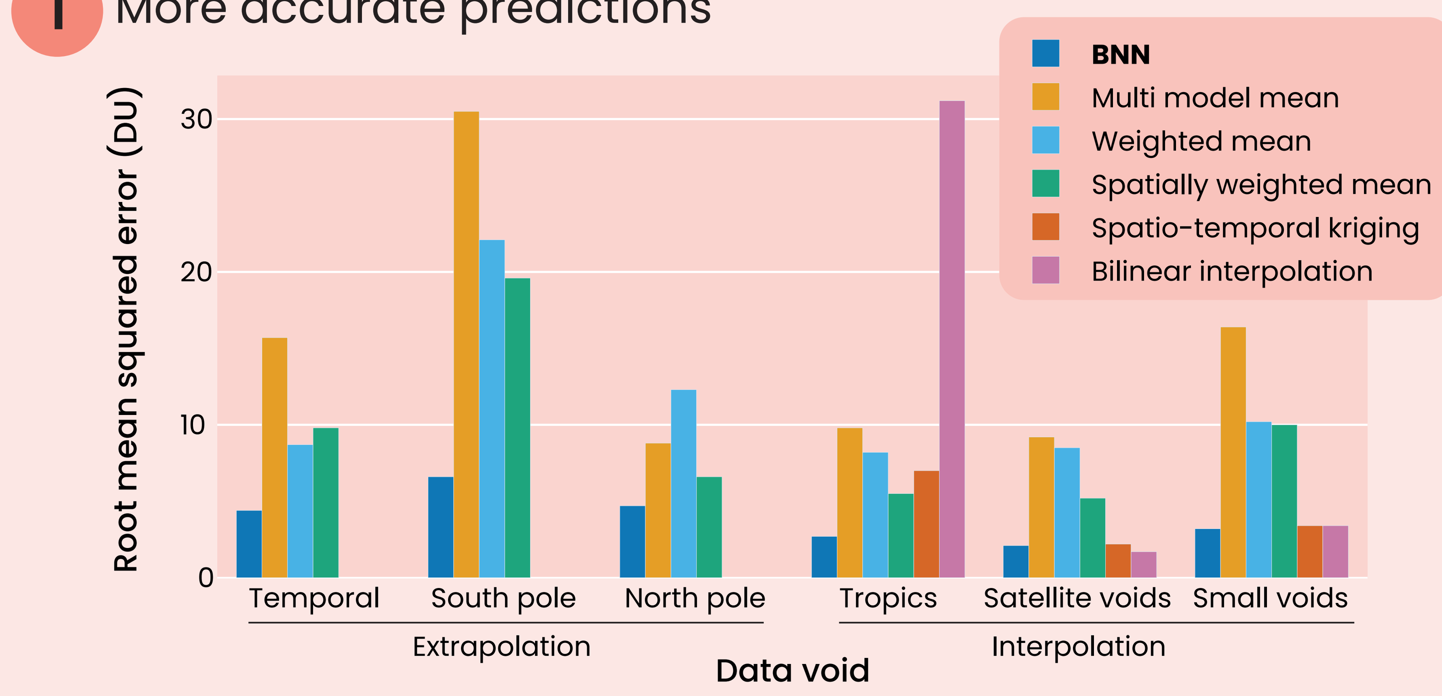
FC = Fully connected
DP = Dot product



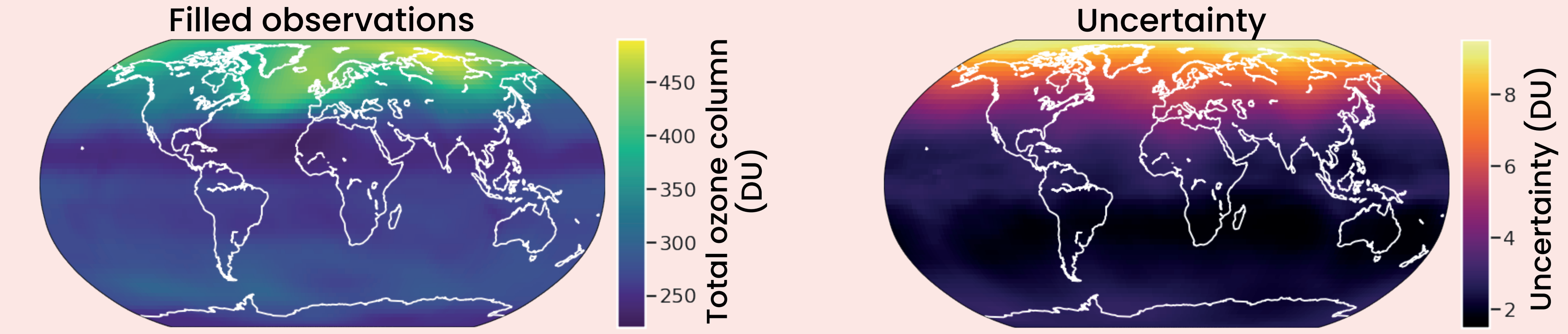
The results

Our Bayesian neural network (BNN) method:
1) outperforms existing ensembling methods
2) successfully fills observational gaps whilst quantifying the uncertainty of our predictions.
3) are interpretable

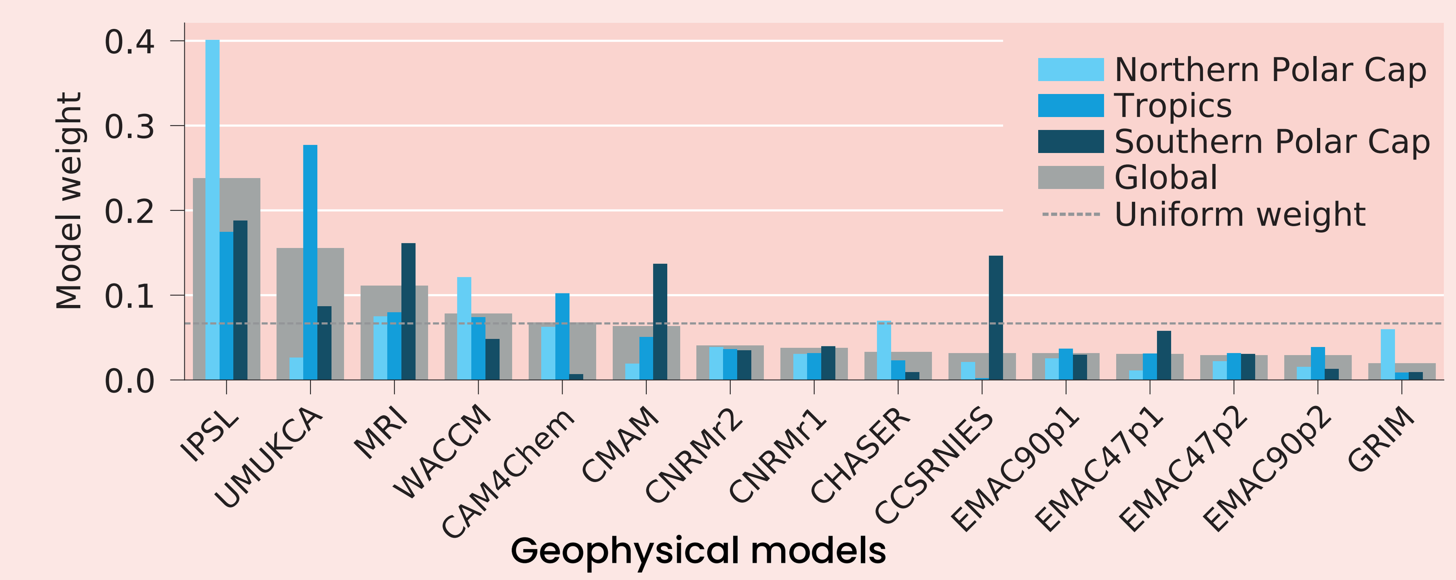
1 More accurate predictions



2 Predictions are uncertainty aware. Uncertainty is high over the north pole because of lack of observations.



3 We can quantify when and where a model contributes to the ensemble prediction



For even more information see our NeurIPS paper:
https://proceedings.neurips.cc//paper_files/paper/2020/hash/0d5501edb21a59a43435efa67f200828-Abstract.html