Weighted mean of Antarctic ozone from CCMI models

Matt Amos, Paul Young and Scott Hosking
Lancaster University - m.amos1@lancaster.ac.uk

What is process based weighting?

Here we present a methodology for creating a process based weighted mean, based upon model skill and model similarity. Models from the Chemistry-Climate Model Initiative (CCMI) [1] are weighted by their independence and performance for a range of chemistry-climate metrics. From this we can create an informed prediction of stratospheric ozone trends over Antarctica.

The end of model democracy?

The current standard is to take all the model output in an ensemble, assume they are all as correct and plausible, then take a multi model mean. Models are weighted equally; however, this assumes that the models are independent and equally good, which is not the case.

How do we define a process?

Using the current scientific understanding of a process or phenomenon we can identify the important drivers and factors on which the process is most dependent. For Antarctic stratospheric ozone, we are aware of major dependencies on temperature, photochemistry, ozone depleting substances and the polar vortex. From this we identify a series of metrics for which observations exist, allowing for direct comparison between the models.

Models and observations

The CCMI model ensemble is compared to observations (ERA-Interim, BDBP, MSU) using the refC1SD (specified dynamics) runs which are the model's best attempt at replicating the past between 1980 and 2010.

Seven metrics are chosen:
- Total ozone column
- Lower stratosphere temperature
- Ozone column gradient
- Stratospheric temperature gradient
- Ozone-temperature gradient
- Trend of the polar vortex breakdown
- Tendency of ozone and temperature trends

Later, to make projections we will use the refC2 model runs from CCMI, which are the model's projections based upon specific concentration pathways.

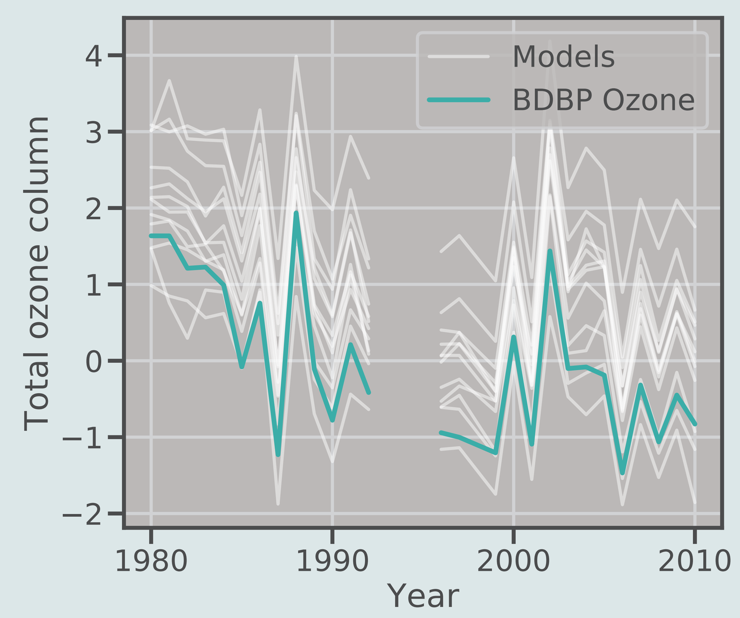

Figure 1
Total ozone column averaged over the southern polar cap for October and November for all refC1SD models. Normalised against the observations from the Bodeker Scientific total ozone column record.

Learning the weighting

Models are given a larger weighting when forming the ensemble mean if they perform better and are suitably independent to the other models in the ensemble. We have built on the weighting work of Knutti et al. [2] and calculate a weighting by comparing the past model runs (refC1SD) to the observations across all of the metrics.

Performance is judged by how well a model can replicate the behaviour seen in the observations. Similarity, the measure of independence, is found by comparing between model runs. These two values for skill and similarity are combined to produce the final weight per model. σ_D and σ_S are parameters which allow tuning of the weighting.

$$\underbrace{w_i}_{\text{Weight of a model}} = \left(\underbrace{\sum_{k=1}^{M}}_{\text{Sum over all metrics}} \exp\left(- \overbrace{\frac{D_{ik}^2}{n_{ik}\sigma_D^2}}^{\text{Difference between obs and model}} \right) \right) \Bigg/ \left(M + \sum_{k=1}^{M} \underbrace{\sum_{j \neq i}^{N}}_{\text{Sum over all models}} \exp\left(- \overbrace{\frac{S_{ijk}^2}{n_{ik}\sigma_S^2}}^{\text{Difference between model and model}} \right) \right)$$

The final weighting is found by fusing all the hindcast model runs with observations whilst simultaneous tuning the σ_D and σ_S parameters using out of sample testing. This ensures the weighted mean is neither constructed from a small handfuls of models (over-fitting), nor converging to a multi model mean.

Is the weighted mean better?

To test the efficacy of process weighting schemes there are two methods available to us:

Out of sample testing. The ozone record extends further than the refC1SD model runs, which allows comparison between the created weighted average (formed from refC2 runs) and the most recent part of the ozone record.

Perfect model testing. We can run the same weighting process but instead use the model runs as pseudo-observations. This allows testing between the between the weighted average and the 'truth' which is in this case the refC2 run. We can test like this for all the models and we find that there is an overall benefit to using a weighted mean as opposed to a multi model mean.

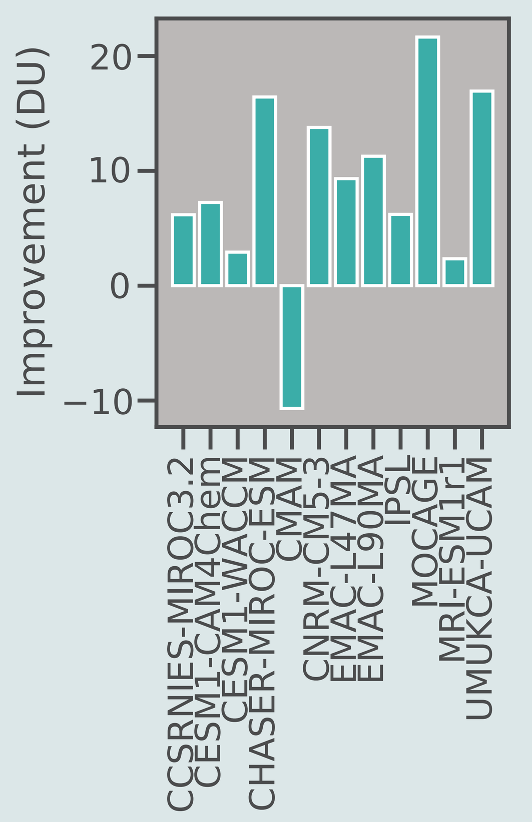

Figure 2
The mean difference between the weighted average and multi model mean when treating each model on the x axis as the 'pseudo truth'. This shows the weighted average found using this methodology is better than the multi model mean for all but one model.

Weighted ozone projection

The weighted projection is found by applying the learnt weightings onto the refC2 projections and combining these all together. We find a difference of around 10 DU for the first half of the 21st century, representing a greater ozone depletion than projected by the multi model mean. Correspondingly there is a later ozone recovery projected using the weighted average instead of the multi model mean. Towards the end of the century the two projection methods converge. For comparison, Dhomse et al. [3] found the return dates (to 1980 values) to be 2060 (2055-2066 within 1σ). We find a very similar recovery date of 2059.

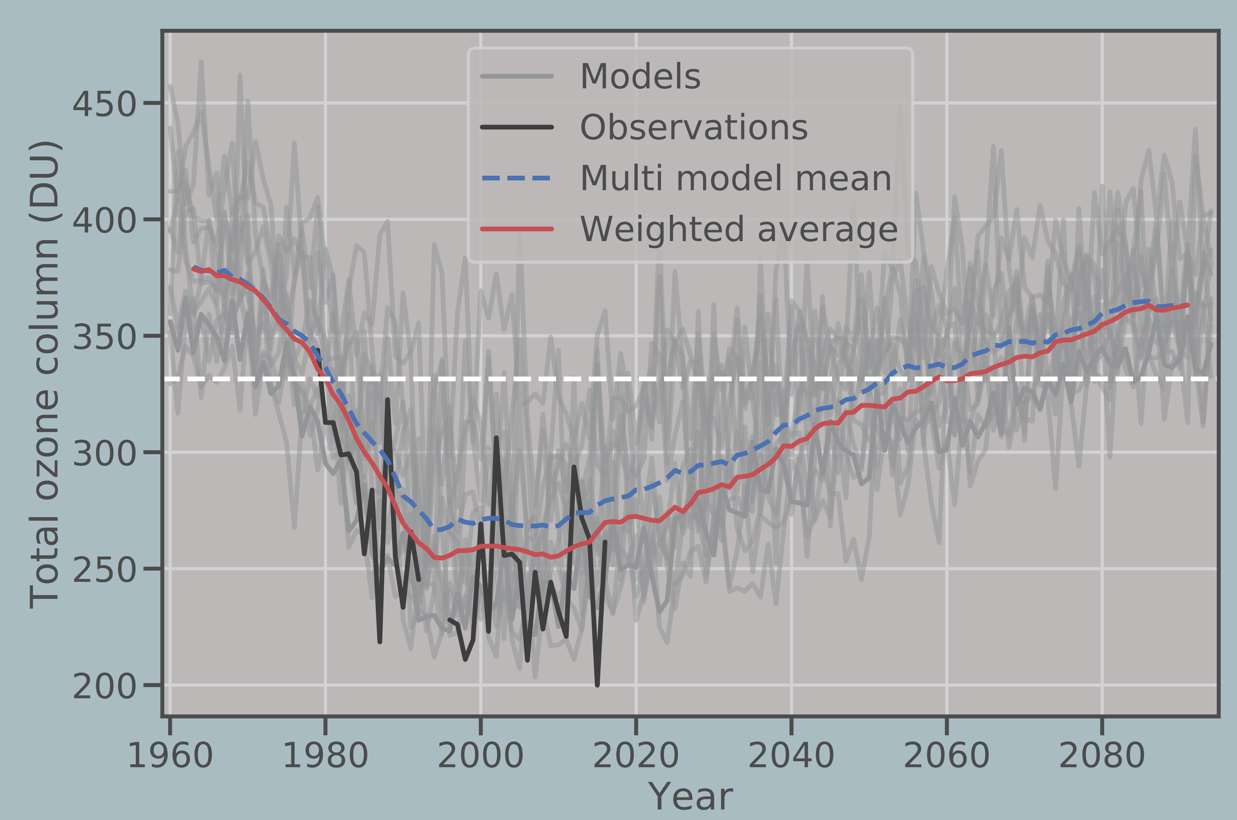

Figure 5
The weighted projection of total column ozone for springtime over the southern polar cap, both with a seven year boxcar smoothing . This is compared to the Bodeker ozone record and the multi model mean. RefC2 model runs were only included for models which performed refC1SD also.

What is learnt about the models?

By performing the weighting analysis over all the metrics we can learn about overall model performance and independence alongside creating a weighted projection. The summation over all metrics provides an insight into the entire process of evolving polar stratospheric ozone.

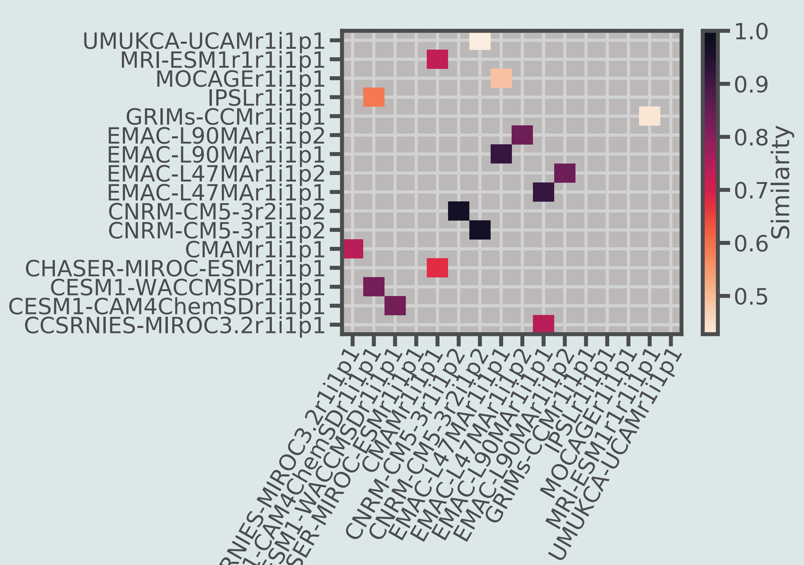

Figure 3
Reading across from the models on the y axis the graph shows the most similar model calculated across all the metrics. The colour represents the amount of similarity with one being the highest. Obvious significant pairings exist between like models as we would expect.

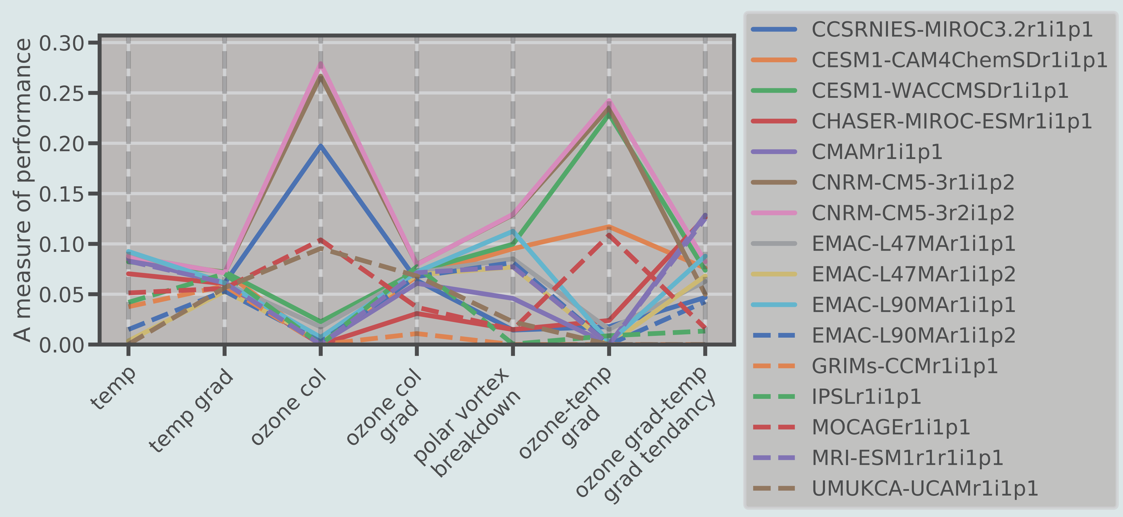

Figure 4
The models' skill (or performance) for each metric. This represents how well each model replicated the observations for each process based metric.

Key Message

Using a process based weighting method we have identified both model skill and similarity within an ensemble. This methodology was used to construct a weighted projection of Antarctic stratospheric ozone using the CCMI ensemble. We also showed that a weighted mean is better than a multi model mean.

References

[1] Morgenstern, O. et al. (2017) Review of the global models used within phase 1 of the Chemistry-Climate Model Initiative (CCMI). GMD 10, pp.639-671.
[2] Knutti, R. et al. (2017) . A climate model projection weighting scheme accounting for performance and interdependence. GRL. 44, pp.1909–1918.
[3] Dhomse, S. et al. (2018). Estimates of Ozone Return Dates from Chemistry-Climate Model Initiative Simulations. ACPD, pp.1-40.