Geophysical Research Abstracts
Vol. 21, EGU2019-5053, 2019
EGU General Assembly 2019

Advanced Data Clustering Methods for Climate Model Intercomparison

Ryan Hossaini, Amber Leeson, and Richard Hyde
Lancaster University, Lancaster Environment Centre, Lancaster, United Kingdom (r.hossaini@lancaster.ac.uk)

Clustering – the automated grouping of similar data – can provide powerful and unique insight into large and complex data sets, in a fast and computationally-efficient manner. While clustering has been used in a variety of fields (from image processing to economics), its application within atmospheric science has been fairly limited to date, and the potential benefits of the application of advanced clustering techniques to climate data (both model output and observations) may yet to be fully realised. Here, we explore the specific application of clustering to the calculation of multi-model means from climate model output.

A standard rudimentary approach to multi-model mean (MMM) calculation simply involves taking the arithmetic mean of all models in a given ensemble, over a particular space/time domain (a ‘one model one vote’ approach). We hypothesise that clustering can provide a useful data-driven method of (a) excluding ‘poor’ model data from MMM calculations, on a grid-cell basis, thus (b) maximising retention of ‘good’ data, and avoiding the blanket exclusion of models, where appropriate. We focus our analysis on chemistry-climate model (CCM) output of tropospheric ozone – an important greenhouse gas – from the recent Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP).

Cluster-based MMM fields of tropospheric column ozone were generated from the ACCMIP ensemble using the Data Density based Clustering (DDC) algorithm. The cluster-based MMM was compared to the simple arithmetic MMM (one model one vote approach) and each MMM was evaluated against an observed satellite-based tropospheric ozone climatology, as used in the original ACCMIP study. As a proof of concept, we show the proposed clustering technique can offer improvement in terms of reducing the absolute bias between the MMMs and observations. For example, the global mean absolute bias from the cluster-based MMM is reduced in all months, up to $\sim$15%, compared to the simple arithmetic MMM. On a grid-cell basis, the bias is reduced at more than 60% of all locations. Some locations are found to be unaffected by the clustering process, while in others the bias increases, albeit slightly. This and other caveats of the clustering techniques are discussed.

Finally, while we have focused on tropospheric ozone, the principles underlying the cluster-based MMMs are applicable to other prognostic variables from climate models. We further demonstrate that clustering can provide a viable and useful framework in which to assess and visualise model spread, offering insight into geographical areas of agreement between models and a qualitative measure of diversity across an ensemble. Future directions of the work, including incorporating performance-based weighting of models, will be discussed.