Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-194-RC2, 2020

Geoscientific Model Development Discussions

Open Access

Interactive comment on "Towards an objective assessment of climate multi-model ensembles. A case study in the Senegalo-Mauritanian upwelling region" *by* Juliette Mignot et al.

Anonymous Referee #2

Received and published: 3 March 2020

Mignot et al. use self organising maps to evaluate the behaviour of a multi-model ensemble in the Senegalo-Mauritanian upwelling region, with the aim of producing accurate projections of future climate changes in the region. Their algorithm aims to select models that yield a specific desired quantity - in this case, a multi model mean. They then project the selected models through the future to assess changes in the region.

There is clearly a great deal of potential in the technical work in this paper. The idea of using Self Organising Maps as a dimension reduction and interpretation technique is a good one, and appears to work very well. It can clearly add a great deal of value

Printer-friendly version

Discussion paper

to the analysis of a large multi-model ensemble in this region. However, I feel a degree of restructuring, clarification of the aims of the paper, and editing for overall clarity is required before the scientific content can be properly assessed.

I feel the paper would most benefit from restructuring so that the objectives, details and then method of assessment of the algorithm were more clearly laid out earlier in the paper, and the reader were more carefully led through that process. As it stands, intense technical detail follows very broad overview statements, and important details about the analysis are left until later in the paper, so the reader is left confused and searching for appropriate context into which to place technical detail. Some choices in the analysis feel arbitrary, and it is unclear whether this is because they are indeed arbitrary, or that they are inadequately described.

The most obvious candidate for restructuring is the start of section 3, describing the methods used for classification of the models. This section dives straight into a detailed description of self organising maps (SOMs), without a discussion of precisely what the algorithm aims to achieve, how that can be assessed, and why SOMs were chosen as opposed to any other dimension reduction technique. As it is, the paper reads as “we decided to use SOMS and this is what you can do with them”, rather than “we are trying to solve a specific problem and here is how SOMs can help”. One suggestion would be to take the description of the methods from the start of section 6 (Discussion and Conclusion), expand upon it and place it at the start of section 3.

The SOMs appear to successfully cluster the model field into regions with different dynamics. This seems useful and interesting. How does it help solve the specific problem? I think it would be useful to set out near the beginning of the paper the exact strategy that will be used, and how to tell if it is successful or not. For example, It seems clear that the assessment algorithm (starting line 232) can be used to rank the models in terms of their closeness to observations and dynamics in particular regions. One downside however, is that it does not give the modeller an intuition into how far the model is from “good” behaviour in absolute terms. We simply get an averaged “skill score" from 28% to 79%, but without an idea of how this might relate to more traditional measures of skill. So how close is the best model and how far is the worst model from reality? We have only a score (useful as that is) to guide us.

The paper makes the claim that it offers an objective method for the assessment of the behaviour of models with regards historical observations. I struggle to accept this, given the number of subjective choices made with regards to the way the analysis is conducted. Subjective judgements will always need to be made in the analysis of climate model output - this is inevitable, and perfectly reasonable as long as labelled as such. The paper only examines a subset of model fields for example, and a subjective choice as to which of those fields to select has been made.

A core problem that needs to be addressed in the paper can be illustrated by considering the section starting on line 285:

"As indicated in the introduction, the main objective of the methodology is to select an ensemble of models that represents at best the upwelling behavior with respect to the observations and to use this ensemble to predict the impact of climate change in the Senegalo-Mauritanian upwelling with some confidence. The problem is now to determine a subset of models that can adequately represent the observations, as the number of models is small enough we choose to cluster them by HAC according to their projections onto the seven axes provided by the MCA, and select the optimal jump in the hierarchical tree (Jain and Dubes, 1998)."

I cannot see a description of what it means for a subset of models to "adequately represent the observations". I also cannot see an adequate description for what the "optimal jump in the heirarchical tree" of Jain and Dubes (1998) is, or what it might mean for the ensemble members. The clustering of the models in figure 4 looks reasonable by eye, but there are a large number of other ways that the models could be clustered that might be equally as reasonable.

The authors claim that their algorithm selects a number of ensemble members that best represent an ensemble mean. I don't believe that they provide sufficient justification for why the ensemble mean should be selected for, or that the ensemble members their algorithm selects members in a way that is superior to a subjective selection. This is presented as a "model weighting" paper, and while that might be possible with this algorithm, I do not believe that is where the strength of the analysis lies. The paper would be better re-cast as a model analysis paper, using an interesting and useful algorithm to explore the dynamical deficiencies of the models in the region, and informing climate modellers of those deficiencies. I think if the authors wish it to be a model weighting paper, then more emphasis needs to be given to the meaning and justification of the weighting scheme. Further, the authors should develop placing the weighting scheme in the context of established work on the meaning of multi-model ensembles.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-194, 2019.