Earth System
Dynamics

Discussions

Interactive comment on "Selecting a climate model subset to optimise key ensemble properties" *by* Nadja Herger et al.

Anonymous Referee #1

Received and published: 16 May 2017

This is an interesting investigation into methods of selecting subsets of the ensemble. I think it's a useful contribution that reaches a number of conclusions that were not a priori obvious.

One weakness (which is shared with many papers) is the limited discussion of principles underlying the selection of the sub-ensemble. Having a good ensemble mean is one possible property that we might like an ensemble to have, but it's not clear whether/when/why it is important. To illustrate, if we consider a simple one-dimensional case where truth is known to take the value 5, is it better to use an ensemble of two models which take the values 3.5 and 4, or a pair with the values 0 and 9, or yet another pair with the values -10 and +20? The ensemble mean improves (relative to truth) across these three sets, but the models themselves are getting worse, which may be a concern. Another distinction between these ensembles is that both members of the

first pair share a bias in sign whereas the other two ensembles bound reality which is close to (at) the 50th percentile. I don't think these questions are easily answered but they do seem fundamental to the whole concept of how and why we use ensembles, so I think they ought to be discussed a bit more fully in the manuscript. Do the authors actually have a good argument why they would like to find an ensemble with a good mean? The analysis does also consider the issue of ensemble spread (both in model selection and assessment of the predictions) to some extent but this isn't really placed in any coherent mathematical framework. For example, the extended cost function on page 9 provides one route to distinguishing more clearly between the three different types of sub-ensembles in my example, but there does not seem to be any structured reasoning behind any particular choice.

The method of ordering by model performance seems to have some superficial similarities with Bayesian Model Averaging principles, albeit with 0-1 rather than continuous weights (and implicitly a uniform prior even when initial condition ensembles are present). It might be worth mentioning the link though I don't suppose the conclusions drawn here will be directly applicable to BMA due to the methodological differences. In particular the implied uniform prior even when IC ensembles are present would probably be considered inappropriate for any more formal implementation of BMA. On the other hand this similarity does highlight the major issue with the method, which is why the RMSE of the ensemble mean is considered to be an appropriate optimisation target in the first place. For BMA (which, at least in many artificial idealised cases, is basically the correct solution to ensemble calibration and weighting) the ensemble mean is not optimised in any meaningful sense even though it will tend to be moved towards observations in the posterior.

Despite the limitations noted above, there is clearly value in investigating the performance of different methods of ensemble selection, so I don't have any hesitation in recommending publication.

"binary": this word appears 3 times, it might be worth explaining this more fully at

the outset as meaning weights of 0 or 1 (and why this restricted choice is significant/beneficial). Actually, I believe the issue is not so much the contrast of binary (or even discrete) weights with continuous, but rather more precisely the number of zero weights, since this is what allows some models to be discarded, thereby reducing computational effort. See for example the lasso approach to regression which might have been a plausible alternative to the 0/1 methods used here. However I'm not suggesting that the authors need to investigate this as part of this piece of work.

Fig 1: The red triangles are not explained in the caption, though presumably they represent the optima from the black triangle cases.

C3