Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2017-28-RC2, 2017

Earth System Dynamics Discussions Open Access

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

Anonymous Referee #2

Received and published: 28 June 2017

General comments

The paper presents a new method to choose an 'optimal' ensemble from a multi-model ensemble based on model performance and interdependence. The paper certainly contains many interesting aspects and presents in principle, as the authors say, important work as little has been published on this subject so far to improve policy and user relevant information from an ensemble of opportunity like CMIP.

The paper is generally well written and fits within the scope of ESD. However, the authors present this method as something that is simple to calculate and generally applicable which is by no means the case. In fact, the authors lack to clearly highlight

the aspects of their work that go beyond what has already been published. The example given as an application of their method does not seem well suited as a proof of concept to select an optimal ensemble for climate applications as it is too simple. A demonstration of how their method can be applied to multi-variable problems using multiple metrics as it would typically be needed for climate analyses would be more helpful. Another important point that is not discussed sufficiently is how to account for observational uncertainties, which is of key importance when ranking and benchmarking models. Also, even though the term 'model interdependence' is repeatedly used, no attempt is made to define model interdependence or discuss the relevant aspects for determining an optimal ensemble. Further work is required to clarify what we can learn from this study and in which cases this method can be applied, before I can recommend publication in ESD, see details below.

Specific Comments

I have the following major concerns which I am hoping the authors can address:

1. What is the aim of this study? Is the aim to
   (a) present a new method: then please what is new, what are the differences and advantages compared to the other methods that have recently been published (e.g., [Knutti et al., 2017; Sanderson et al., 2015a; b]? Quantitative comparisons would be required.
   (b) to present a method that is only slightly different but to provide a demonstration that this method can be used for impact studies and other climate applications? The paper fails to convincingly show that this method can be applied for concrete applications, see further comments below. The example given in the manuscript is too simple to provide any helpful insights beyond of what has already been published (see references above).

Currently a mixture of both is presented.

2. The paper could expand on recommendations of pre-selection in an ensemble. The statement on p6, l.34 that similar improvements can be made if closely related model runs are a priori removed from the ensemble to start off with a more independent ensemble could be such a recommendation.

3. It is quite confusing that within a short time this is the forth (?) recommendation for a method that should be applied for model weighting considering both model performance and interdependence (with two of the authors of this paper being also authors on all the previous papers). Yet the authors do not show the differences between this newly presented method and the previous ones. Neither they give a recommendation whether this method now supersedes the previous ones nor do they provide a sophisticated comparison of the published methods for a concrete example. For example, how would the results on sea ice extent weighting from Knutti et al. [2017] change if this method instead of the Knutti et al. [2017] method was applied and what are the policy and stakeholder relevant implications when analyzing model ensembles?

4. Related to the above: if the authors can't convincingly show what is different to the above methods, then it is also not clear what is new.

5. Climate change is not a single, but a multi-variable problem. Using RMSE as only metric does not always seem appropriate, more comprehensive metrics are available (see for example Xu et al. [2016]). The authors show that the optimal ensemble is performing best if the bias of the model subset average should be minimized - essentially indicating that the solver is working as anticipated (p6, l24). However, if a bias correction with climatological mean temperature would be the answer for an optimal ensemble, one could for example tune the models accordingly. There are good reasons why one might not want to do so (see for example Mauritsen et al. [2012]). Why would an ensemble that captures mean

temperature be better than another one? The multi-variable issue is mentioned on p7,l29 but it would be good if the authors could expand their analysis to explore this further and if possible give advice to the reader.

6. The physical consistency is mentioned yet the authors are not evaluating the optimal ensemble whether it captures other important climate features including modes of variability. This strongly limits the applications of this method and generalizations of the application like the one on p4,l10 ('We argue optimally selecting ensemble members for a set of criteria of known importance to a given problem is likely to lead to more robust projections') should be avoided.

7. Related to the above: what about model tuning? A model could be tuned towards a correct present-day temperature climatology but it might still not be the best model to project climate? What about climate sensitivity?

8. Can process-oriented diagnostics be used? This might be an interesting option to avoid selecting models that get the right results for the wrong reasons.

9. The study is motivated by the need of the impact and user community who need concrete guidance on how to use the large zoo of model output available in the CMIP ensemble (e.g. first sentence in abstract). While this is true, the paper needs to improve on giving concrete guidance. It either needs to provide real-world examples or avoid generalizations of the applicability of the method. It mathematically works fine, but whether or not it should be applied depends on whether the diagnostics chosen for the benchmark are actually relevant for the specific application. Finding these diagnostics remains a challenge.

10. The authors show that different observational products lead to different ensembles (Figure 1 and S1). But given there is observational uncertainty, some choices would need to be made. It would be good if the authors could expand

on this topic and give a recommendation how observational uncertainty can be considered in the method, the formulas presented in section 4.1 and the code.

11. Section 4.2 applies the method to the future, keeping the limited sample of weighting the ensemble based on temperature means / trends. A model could simulate a correct present-day climatology but why would it be a good model to project future climate? One of the authors convincingly shows that there is hardly any correlation between present-day and future temperature patterns [Knutti et al., 2010]. Climate change is non-linear. Could the authors choose a multi-variate and preferably process-oriented diagnostic approach? Otherwise, please limit general statements for the applicability of this method to improve projections (see above).

Minor Comments

There seems to be a mistake how papers are cited as they are missing 'et al.'

References

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl (2010), Challenges in Combining Projections from Multiple Climate Models, J Climate, 23(10), 2739-2758, doi:doi:10.1175/2009JCLI3361.1.

Knutti, R., J. Sedláček, B. M. Sanderson, R. Lorenz, E. Fischer, and V. Eyring (2017), A climate model projection weighting scheme accounting for performance and interdependence, Geophys Res Lett, n/a-n/a, doi:10.1002/2016GL072012.

Mauritsen, T., et al. (2012), Tuning the climate of a global model, Journal of Advances

in Modeling Earth Systems, 4, doi:Artn M00a01, Doi 10.1029/2012ms000154.

Sanderson, B. M., R. Knutti, and P. Caldwell (2015a), Addressing Interdependency in a Multimodel Ensemble by Interpolation of Model Properties, J Climate, 28(13), 5150-5170, doi:10.1175/Jcli-D-14-00361.1.

Sanderson, B. M., R. Knutti, and P. Caldwell (2015b), A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble, J Climate, 28(13), 5171-5194, doi:10.1175/Jcli-D-14-00362.1.

Xu, Z., Z. Hou, Y. Han, and W. Guo (2016), A diagram for evaluating multiple aspects of model performance in simulating vector fields, Geosci. Model Dev., 9(12), 4365-4380, doi:10.5194/gmd-9-4365-2016.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-28, 2017.