The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2020-86-RC2, 2020

The Cryosphere Discussions

Interactive comment on "Improved Multimodel Superensemble Forecast for Sea Ice Thickness using Global Climate Models" *by* Wang Yangjun et al.

Anonymous Referee #2

Received and published: 22 September 2020

General comments: The paper describes a new method for weighting models in a multi-model ensemble for Arctic sea ice thickness prediction. This is an important topic given the need to provide reliable predictions of changing Arctic sea ice. The authors test and compare this new method to multiple other methods which have been used previously. The various methods are demonstrated and applied to CMIP5 climate model projections with metrics for a single year (2018) used to compare them. They are then applied to projections of sea ice thickness for the 2020s, 2030s, and 2040s. Some of the metrics and analysis used would likely be relevant for initialized predictions of sea ice. However, I believe that there are some problems with this methodology and the metrics used in the paper particularly for the uninitialized climate projections to which they

are applied. These are further discussed below. To address these issues will require some major reworking of the manuscript, including a better consideration/discussion of model dependence and better metrics for assessment of the different techniques in the context of climate model projections. Given these concerns, I am recommending that the paper be rejected in its current form.

Specific comments:

1. Much of the material in the introduction is related to weather or seasonal prediction in which initialized forecasts are use. However, climate projections on longer timescales (such as those used here from the CMIP5 models) have some different considerations. In particular, the non-initialized climate model simulations (such as those in CMIP5) are not meant to predict the conditions from a single year, in which internal variability plays a large role, but instead the ensemble mean conditions represent the forced response of the system, which is relevant to longer timescale predictions. Given this, I am uncertain why metrics for a single year (2018) are used to assess the different ensemble techniques as applied to CMIP5 simulations. Could you please further explain, discuss this? Alternatively, you may want to consider assessing the methods using a perfect model setup in which you treat a model projection as the truth (for example, see Karpechko et al., 2013, doi: 10.1175/JAS-D-13-071.1) and note that in doing this, that model independence needs to be considered (see more on independence below).

2. A primary rationale for using multi-model ensemble predictions for climate is the thought that different models will have different biases and so considering multiple models allows us to characterize the uncertainty associated with imperfect models and to "average out" some of that uncertainty in our projections of the forced climate changes. Because of this, having models that are independent samples is an important consideration. Ensemble members from individual models are not independent since they are produced by the same model and ideally should only differ due to the internal variability that is not predictable on the timescales of consideration in uninitialized simulations. Because of this, in producing a simple multi-model ensemble mean (for example in

the IPCC), often only a single realization from each individual model is used. Additionally, there is considerable interdependence of models, especially from the same centers (for example see Knutti et al.s 2013 paper on climate model genealogy), which can be a problem for producing an ensemble average. It appears to me that for some of the ensemble techniques (like the simple ensemble mean), you have treated every simulation including ensemble members from an individual model as independent simulations and averaged them all together. Is this correct? If so, those methods will be heavily weighted towards the few models that have numerous members, which should be avoided. In general, the issue of model independence and the challenges inherent in the climate model ensembles of opportunity should be discussed in the manuscript and considered in the different ensemble techniques.

3. There are numerous papers discussing ensemble averaging techniques for climate projections for many different properties but these are not generally mentioned in the introduction. Many of these papers discuss challenges with model weighting in the CMIP ensembles of opportunity and some strategies behind devising reasonable weights. Please expand the introduction to include some of this background material (for a recent sea ice example see – Knutti et al., 2017, doi: 10.1002/2016GL072012).

4. Please enhance the description of the metrics used in the assessment of the different methods. Are you comparing the single time-varying multi-model ensemble values from each different method with the observations for 2018? It was not always clear whether things like the correlation coefficients were computed in space (with appropriate weighting for non-equal area grid cells) or computed in time (just over the months in 2018) or something else. Are all of the metrics just computed for 2018 using monthly data?

5. Why does the conventional multimodel superensemble method provide such bad results?

6. The observed training data that is used is only about 10 years long. On this

timescale, internal variability may have a large imprint on the trends and variability. Given this, uninitialized climate model simulation disagreement with observations may not mean that the model is deficient but could just be a consequence of internal variability. The weights will not necessarily reflect this and so “good” models could be down-weighted in the projections. Please discuss these limitations in the paper and the challenges in using a short observational record in comparison to the free-running climate simulations which have their own time-evolution of internal variability.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2020-86, 2020.