The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-240-RC2, 2020

The Cryosphere Discussions

Interactive comment on "Scoring Antarctic surface mass balance in climate models to refine future projections" *by* Tessa Gorte et al.

Anonymous Referee #2

Received and published: 16 January 2020

This discussion paper assess surface mass balance (SMB) across CMIP5 and CMIP6 based on a specific set of metrics. A subset of four CMIP5 models is selected to refine 21st century Antarctic mean AIS SMB projections. They show weaker sensitivity and a smaller inter-model spread than the full ensemble. This is an important topic due to the implications for Antarctica's contribution to future global sea level change.

Unfortunately I do not feel that this discussion paper in its present form is suitable for full publication in The Cryosphere. There are a number of major points on the approach, method and presentation that need addressing. Indeed since some parts of the description were not clear, it was difficult to fully evaluate the method. Major points are listed below along with suggestions on how to address them. Given the number of major general comments, at this stage I have not included minor technical correc-

tions to wording or Figure presentation. Overall, important aspects that are required include (among other things) utilizing the CMIP6 HighResMIP experiments to assess resolution-related aspects, incorporating multiple ensemble members to assess the role of internal variability and a more in-depth explanation, motivation and development (i.e. relative to other literature) of the scoring method. Indeed one possibility would be to re-formulate the manuscript with a focus on comparing scores across different resolutions in the CMIP6 HighResMIP experiments and less of a focus on projections.

General comments: 1. Overall model evaluation approach The overall approach to model evaluation presented here differs to that of most other similar studies that I am aware of. For example Agosta et al. and Barthel et al. evaluate CMIP5 models based on selecting those most appropriate for driving regional models and not explicitly on the CMIP5-simulated SMB / surface climate. The main reason for this is that the low-resolution GCMs are not able to realistically capture the correct processes and features associated with the steep orography near the Antarctic coast. This is the job of the regional model and the global model provides the bigger-picture climate responses. In this discussion paper the authors take the approach of explicitly evaluating global standard-resolution models (CMIP5 and CMIP6) directly on their representation of SMB instead of the broader approach taken in the above mentioned studies. They then present projections of SMB change directly from the global models. This presents a number of issues, which are listed below. In addition there are questions on the description and choices made in the methods used, which are also outlined below. Overall, key recommendations are to utilize the CMIP6 HighResMIP experiments to assess resolution-related aspects, to incorporate multiple ensemble members to assess the role of internal variability and explain and motivate the scoring method (and relate to a wider range of existing literature).

1.1 Comparing GCMs with reconstructions A major issue with comparing standard resolution GCMs and observations/reconstructions, is that full GCMs are not able to reproduce the detail required in regions of high precipitation. Therefore a standard-resolution

GCM that reproduces observed/reconstructed Antarctic-wide time-mean SMB is quite possibly doing so for the wrong reasons. This therefore may not be the most appropriate model for projections. The authors should utilize the HighResMIP dataset to determine the resolution dependence of participating models and the potential implications this might have on model selection. This is relevant to all 5 of the criteria used (mean SMB, SMB variability, SMB trends, modes of variability (EOF analysis) and variance explained by the modes). With regard to the EOF analysis, from Figure 5 seems to suggest highly regionalized nature of patterns from the reconstructions. Indeed, an assessment of natural variability is again crucial here in identifying uncertainty in comparing observations and models.

1.2 A lack of mechanistic explanation for why each of the 5 criteria are relevant for improving reliability of projections Firstly the authors should outline the rationale for inclusion of each of the criteria and how they may potentially improve reliability of projections. It is important to discuss this in the context of existing literatures. For example, Krinner et al. (2014) found that future change in SMB was more associated with thermodynamic, rather than dynamic, factors. Secondly the authors should consider the possibility of leave-one-out cross validation, whereby the real world is can be replaced by each member of the model ensemble in turn to see whether evaluation against that model can help improve predictions from that model. This can help to identify which criteria are most relevant in terms of future projections.

1.3 The methodological framework for model weighting In addition to the criteria selected, the rationale for the methodology on model weighting needs to be carefully introduced and motivated. Indeed it is common for a model weighting method to be developed initially in a separate paper and then applied to model output in subsequent papers. Specific suggestions are: Firstly the authors need to bring in more of the previous substantial literature on model weighting. Agosta et al. (2015) use a Climate Prediction Index approach which, as I understand it, draws from probability theory and the probability that observations and models may agree (this goes back to Murphy et

al., 2014). There are also detection and attribution approaches, which use past trends to scale future projections and should be mentioned. What is the advantage of the approach used in this discussion paper? Secondly, the authors should consider the implications of situations where the reconstruction uncertainty is small. In the extreme case where it approaches zero, in general models would be many multiples of this uncertainty range away from the reconstruction. How is/would this be handled in terms of relative weighting across different criteria? Thirdly, the method needs to be described more clearly and is in fact difficult to fully evaluate. The whole section needs to be improved and I have just identified one example starting on line 109. Specifically the text: "if a model time series was fully captured within $2\times$ the reconstruction uncertainty, the model would receive a score of 2". I could not find a clear definition of "reconstruction uncertainty". This exact term is only referred to once in the preceding text on line 69. Is it the same as the "total uncertainty" mentioned on lines 72/73? If so, how does the spatial and temporal information map of total uncertainty map onto the AIS-integrated SMB? In the same paragraph it is not clear what is meant by "model time series fully captured"? Does this mean that even extreme years in the model time series are considered? My recommendation is to write out these score criteria as equations to make it easier for the reader to understand and assess them.

1.4 The role of internal climate variability in trend and spatial EOF analysis The potential role of internal climate variability in evaluating trends is not mentioned, but could be very important. This could be very important for 50-year trends and the spatial EOF patterns. The authors should test the possible role of internal variability by assessing climate models with multiple ensemble members of their historical runs.

1.5 Final model selection The final selection of 4 CMIP5 models for projections should be compared and contrasted with related studies, Agosta et al., (2015) and Barthel et al. (2020). The reasons for, and implications of, differences should be discussed. What is the significance of the smaller spread across these four models. They come from only two model centers (GISS and MPI). Such close links calls into question the

statistical significance of spread across models from just two groups. This could be small or large by chance.

1.6 Impacts of wider factors on projections (e.g. conditions over the Southern Ocean). Another major caveat with the SMB-focused model evaluation is that wider model biases that are known to be important for projections, such sea-surface conditions surrounding Antarctica and hemispheric-scale atmospheric circulate biases, could have an effect on projections (e.g. Krinner et al., 2014; Kittel et al., 2018). The authors do acknowledge this, but don't make implications of differences clear. Could it be that the results of this study should be interpreted alongside other studies?

1.7 Inter-annual variability in GCMs and regional models. On line 79 it is stated that "Global climate models tend to show higher skill at representing interannual variability compared to regional climate models (Medley and Thomas, 2019)". It is not clear to me why this should be since regional models derive their variability from global models. It is also then notable that all CMIP5/6 models over-estimate SMB variability by so much (line 197). An explanation needs to be provided for this, or at least a discussion of the point.

References Agosta, C., Fettweis, X., & Datta, R. (2015). Evaluation of the CMIP5 models in the aim of regional modelling of the Antarctic surface mass balance. The Cryosphere Discussions, 9(3), 3113–3136. https://doi.org/10.5194/tcd-9-3113-2015

Kittel, C., Amory, C., Agosta, C., Delhasse, A., Doutreloup, S., Huot, P.-V., Wyard, C., Fichefet, T., and Fettweis, X.: Sensitivity of the current Antarctic surface mass balance to sea surface conditions using MAR, The Cryosphere, 12, 3827–3839, https://doi.org/10.5194/tc-12-3827-2018, 2018.

Krinner, G., Largeron, C., Ménégoz, M., Agosta, C., & Brutel-Vuilmet, C. (2014). Oceanic Forcing of Antarctic Climate Change: A Study Using a Stretched-Grid Atmospheric General Circulation Model. Journal of Climate, 27(15), 5786–5800. https://doi.org/10.1175/JCLI-D-13-00367.1

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768–772, doi:10.1038/nature02771, 2004.

Genthon, C., Krinner, C., & Castebrunet, H. (2009). Antarctic precipitation and climate-change predictions: Horizontal resolution and margin vs plateau issues. Annals of Glaciology, 50(50), 55–60. https://doi.org/10.3189/172756409787769681.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-240, 2019.