The Cryosphere
Discussions

Open Access

EGU

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

Anonymous Referee #1

Received and published: 14 January 2020

In this article, the authors use a reconstruction of the Antarctic surface mass balance, annually resolved, from ice cores, to evaluate CMIP5 and CMIP6 global climate models and to constrain projections based on their evaluation criteria.

I think this is an important subject that deserve to be published in TC, I found the paper generally well written with nicely shaped figures. However I think the paper need major revision before being published.

Major

Methodology

"To score the time series magnitude, we assigned a score, x, for how many x-times the reconstruction uncertainty was requiredÂăfor the entire time series to be within the

reconstruction uncertainty."
I think you should reformulate this sentence in a more mathematical framework. What did you code? What is the minimum value of your score, 0 or 1?
if I understand well, you did max(abs(Model - obs))/(reconstruction uncertainty)? So you scaled the maximum difference of model to obs with the reconstruction uncertainty? Why not using the RMSE scaled by the reconstruction uncertainty?

"involves finding what spatial SMB patterns explain the highest variance in the AIS-integrated SMB time."
are you sure this what EOF do? Is is not the variance of space-time SMB variability?
time series (typo?)

"To avoid manually sorting the top three modes of variability for all 53 models, we generated difference maps between each of the top three reconstructed modes and each of the top three modes for each model:"
why do you do this only for the top 3 modes of each model and not e.g. the top 10?

"We then sorted the top modes of variability for each model based on smallest difference"
what do you call "the smallest difference"? Do you average absolute differences over the map? Do you compute a RMSE?

"After compiling scores for all five of the aforementioned scoring criteria, we normalized each set of scores to be on a scale from one to ten to ensure that each criterion was equally weighted."
So, if I understand well, you divide each criteria by the max of the criteria? This scaling is extremely sensitive to outliers. You should consider scaling by the interquantile range or by the standard deviation of each of you criteria.

"To refine the scope of what we predict for SMB in the future, we used a subsetÂăof models that had a final score in the top 10th percentile of CMIP5 and compared them

to the entire scope of CMIP5"
I am not sure it is a correct method. How much is your method sensitive to the number of models you keep? Why do you use this "10th percentile" criteria? I think that 4 models is too little to compute a robust statistic. Is it statistically correct to compare 4 members to 30 members? You should consider e.g. ensemble regression based on models' scores (Bracegirdle and Stephenson, 2012, doi: 10.1007/s00382-012-1330-3)

Results

Figure 1.
when I see the spatial pattern of trends in 1B and 1C, I wonder why you use a criteria for SMB-integrated values instead of comparing spatial maps of trends? I think using spatial maps of trends would be more relevant.

"Looking at multiple time "slices" allows us to investigate if models capture the reconstructed SMB trends for the whole time series compared to more recent decades. Here, we looked at three time slices: the entire over-lapping time series from 1850-2000, the last century from 1900-2000, and the last 50 years from 1950-2000."
I understand that simulating correctly the trends for 1950-2000 may be useful because it quantifies if the global climate models are able to simulate correctly the response to anthropogenic forcing. However I don't think that scoring the trends over the century is useful for your purpose. Your uncertainty on century-scale trends is very small and I wonder if it is not underestimated. It seems difficult to estimate century-scale internal variability from a 200 year reconstruction in fact.

"All CMIP5 and CMIP6 models overestimate SMB variability. ĂăThe CMIP5 and CMIP6 models range from overestimates of 144% to 261% and 151% to 217% of the reconstruction standard deviation, respectively"
A strong warning here. I have doubts on the reliability of the reconstruction for inter-annual variability. How does the reconstruction interannual variability compare with the reanalyses variability for the common period? I suspect that the annual accumulation

signal extracted from ice cores is dampened.

"This dipole corresponds to variability in precipitation generated by variations in the track and strength of the Amundsen Sea Low. The Amundsen Sea Low, which represents the pole of circulation variability in Antarctica (Turner et al., 2013), is marked by high precipitation around the coast of the Antarctic Peninsula (Grieger et al., 2016)."
All this sentence is strange. It is more a discussion than a result. "The Amundsen Sea Low, which represents the pole of circulation variability in Antarctica"? What is a pole of circulation variability?

"The second mode of variability represents high variability in West Antarctica and the Antarctic Peninsula. This could be caused by the topography in these regions which can induce large amounts of snowfall."
I am not sure that you interpret the EOFs correctly. The spatial pattern of an EOF associated to its time series explains to a certain amount of the space-time variability, but it does not mean that where the EOF spatial pattern is high there is a high variability. " This could be caused by the topography in these regions which can induce large amounts of snowfall." I don't understand why?

"By comparison, one of the better scoring models for the EOF map criterion, CMCC CM, also shows a dipole between the Antarctic Peninsula and the Ross Sea region for the top mode as well as strong variance signal around the Antarctic Peninsula for mode 2 and a quadrupolar pattern for mode 3."
When looking at Fig. 5, EOF modes from the two climate models do not ressemble the reconstruction EOF modes, even for the best performing model (row B). Maybe showing the patterns with the same sign as for the reconstruction modes will help (multiply by -1 the climate model patterns). But still, they will remain very different. E.g. in row B there is no high spot near Davis for EOF 3, and there is a large dipole in WAIS. Are you sure of your computation? If yes, are you sure your analysis is relevant?

* What are the biases of the best scoring models for the large scale circulation fields

(e.g. sea level pressure over Southern Ocean) over the last 40 years?

Fig 9 and associated text :
The climate sensitivity for SMB must be shown in % K-1, because SMB varies exponentially with temperature. You should revise the end of section 4 with regard to climate sensitivities computed in % K-1.

* Given the issues on the scoring and the relevance of selecting four models, the new version of the manuscript might give different results.

Minor

"Integrated over the grounded Antarctic ice sheet (AIS), the blowing snow and runoff terms are negligibly small (Lenaerts et al., 2012a)."
Drifting snow sublimation is still not well modeled and evaluated. You should reformulate, e.g. something like "we neglect blowing snow and runoff and estimate SMB as precipitation minus sublimation"

"Over longer time scales"
Which ones?

"The strong regional variability suggests an important impact of variations in synoptic-scale patterns around the AIS (Fyke et al. (2017); Marshall et al. (2017))."
It is known that synoptic scale patters drive the accumulation variability, reformulate, e.g. "Synoptic-scale variability induces a strong regional variability of the SMB"

"Additionally, as the atmosphere has been warming over large parts of the AIS and can hold more moisture per the Clausius-Clapeyron relation, SMB is expected to show an overall increase"
Previdi and Polvani (2016, https://iopscience.iop.org/article/10.1088/1748-9326/11/9/094001) state that "the forced SMB increase due to global warming in recent decades is unlikely tobe detectable as a result of large natural SMB variability". Your sentence is unclear and potentially wrong for the last decades. Modify and add

references.

"but many of those models tend to overestimate annual precipitation values due to poor representation of coastal topography"
Are you sure it is because of the poor representation of coastal topography?

"This allows the atmospheric moisture to penetrate too far inland and leads to excessive precipitation on much of the grounded AIS, while underestimating precipitation nearby the coasts (Lenaerts et al. (2012b))."
I did not read again this article, but it is about "Modeling drifting snow in Antarctica with a regional climate model: 1. Methods and model evaluation", so I am not sure it is the right paper to cite here? Do you have other references to show that resolution is the most important factor for modelling Antarctic precipitation?

"Barthel et al. (2019) investigated the Ice Sheet Model Intercomparison Project version 6 to determine a recommendation of which models to use for ice sheet model forcings based on best captured current Antarctic climate relative to observations and their ability to project certain metrics into the future"
It's "Ice Sheet Model Intercomparison Project *for CMIP6*" and not "version 6" (in fact it's version 1).
Barthel et al. (2019) evaluate the global climat models based on their ability to capture the large scale circulation around ice sheets compared to reanalyses . It is not "very similar" to your study because the "observation" they use is well evaluated (reanalyses large scale fields after 1979) and they don't use this criteria to constrain future projections.

"To improve upon model estimates, several groups have combined ice core data with models to create spatio-temporally robust SMB data sets (Monaghan et al. (2006), Thomas et al. (2017), Medley and Thomas (2019))."
this sentence should be in the Method section

"In this work, we leverage the availability of that new avenue for climate model evalua-

tion of AIS SMB, and compare the suite of CMIP5 and CMIP6 climate models to that new SMB reconstruction."
repetition of the sentence P2 L50-52, merge the two.

"they weighted each ice core spatially to generate the 200-year data set"
give the period

"they calculated spatial sampling uncertainty is basedÂăon the RMSE"
"they calculated spatial sampling uncertainty basedÂăon the RMSE"

"Global climate models tend to show higher skill at representing interannual variability compared to regional climate models (Medley and Thomas, 2019)."
it is not what is said in Medley and Thomas, 2019. They say "Because of their afore-mentioned ability to reproduce the interannual variability[17], which strengthens the weighting scheme, we used *global atmospheric reanalyses* over regional climate models.". So this statement is for *reanalyses* compared to RCM only, and is based on [17] Medley, B. et al. Airborne-radar and ice-core observations of annual snow accumulation over Thwaites Glacier, West Antarctica confirm the spatiotemporal vari-ability of global and regional atmospheric models. Geophys. Res. Lett. 40, 3649–3654 (2013).

"To get a comprehensive look at how well global climate models capture SMB, we compared the suite of CMIP5 models to the reconstruction."
and CMIP6?

P4 L90-95
I am not sure the detail of conversion of kg m-2 s-1 in Gt yr-1 is useful. Just saying that it is computed on the original GCM grid is enough.

P4 L99-100: remove parentheses

P4 L107: "the magnitude of the SMB time series"
do you mean "the SMB mean value"? If yes it seems clearer for me to replace "magni-

tude" by "mean value" everywhere.

"To achieve this goal, we analyzed trends from 1850-2000, 1900-2000, and 1850-2000."
typo
how do you combine the 3 periods?

"To score the time series variability, we detrended and normalized each time series to separate the SMB trend from its absolute magnitude using:"
I don't understand "to separate the SMB trend from its absolute magnitude"

"To do so, we performed an empirical orthogonal function (EOF) analysis"
on annual data over 1850-2005(?)

"By breaking this criterion down into two main factors, we were able to determine the models' abilities to accurately capture the modes of variability as well as how much variance each mode explained."
what are the two main factors you are talking about?

P5 L169 "All four of best scoring models are captured within the reconstructed uncer-tainty for the entire 150 year time series."
After reading further I understood that the best scoring models are for the combination of criteria. I think you should begin your result section by presenting the best scoring models (currently presented P10 and in the Figures' legends)

"We weighted all scores from the five scoring criteria equally on a scale from 1 to 10 with lower scores indicating better performance.ÂăThe final score, then, is the sum of all the individual scores, which is renormalized on a scale of 1 to 10 with lower scores still indicating better performance."
repetition of P5 L141-143

" The reconstructed AIS SMB averaged from 1801-2000 shows higher SMB values around the coastal areas, particularly in the Antarctic Peninsula and West Antarctic

regions (Fig. 1A)."
This is really the most basic feature of Antarctic SMB, this sentence is not useful.

C9