We first would like first to thank reviewer #2 for making constructive comments which will help to improve our manuscript. Please find our answer in blue in the text.

In this paper, Kittel and co-authors present a series of experiments in which they use the regional climate model MAR to simulate climate over Antarctica over the coming century with boundary conditions provided by four different earth system models (drawn from CMIP5 and CMIP6). They find a substantial difference in the surface mass balance of grounded versus floating ice in all cases, with the former acting as a net sink for sea level, and the latter acting as either a contributor or neutral, depending on the forcing. The authors also find that the integrated differences between MAR response to different ESMs is largely explained by differences in the timing and intensity of projected global warming. This allows the authors to develop simple polynomial functions relating near-surface temperature anomaly in the Antarctic region to SMB, snowfall, rainfall, and runoff anomalies. They then apply these relationships to the remainder of the CMIP5 and CMIP6 ensemble members to produce an approximation of model uncertainty in SMB anomaly for each CMIP scenario.

I find this paper to be a well-written and useful contribution to the community's understanding of variability in climate model predictions. It does a good job of laying out critical assumptions and also is careful to couch their results as model predictions (rather than a factual future). Besides a few requests for changes to the structure of the paper, and a few technical corrections, I suggest that the paper be published with little further review.

Major Points

I believe that Supplement S2 should be included in the main text, more or less in its entirety. The results section's primary points are devoted to summarizing its content and referencing its figures, so why not just include it in the manuscript?

As section S2 is only an evaluation of the MAR results on the present, we preferred to put it as supplementary material in order not to lengthen the manuscript and lose the main thread of the story. Nevertheless, following the reviewer's remark, we propose to move Fig. S3 and Tab. S2 into the main manuscript. We will also create a section before Section results based on previous Section S2 and that assesses the SMB downscalings over the present period (while SMB-component and near-surface-climate evaluation figures would remain in the supplement):

Our ESMs-based experiments closely reproduce the SMB and near-surface climate of MAR(ERA5) over the historical period (Supplement S2). The anomalies of the annual mean SMB modelled by MAR forced by each ESM compared to MAR(ERA5) are lower than the interannual variability (i.e, one standard deviation) of the SMB simulated by MAR(ERA5) over the historical period, suggesting that the biases are not significant. Overall, MAR(ACCESS1.3) has the best representation of the Antarctic SMB over the current climate (mean bias: -3 Gt yr−1, spatial rmse: 59 kg m−2 yr−1), while MAR(CESM2) is the least accurate (mean bias: -25 Gt yr−1 , spatial rmse: 90 kg m−2 yr−1). We refer to the Supplement Sect. S2 to more details about the evaluation of our experiments. The results of our experiments over the current climate are consistent with the ranking of the ESMs given by Agosta et al. (2015), Barthel et al. (2020), and Agosta et al. (in preparation). This *highlights the importance of selecting ESMs that correctly represent the historical climate around Antarctica as they strongly controls present biases independently of the capacity of the RCM to improve ESMs results.*

will become

3. Evaluation of MAR(ESM) simulations over present

Present biases might have a significant influence on the projections results as they could amplify in the future (Fettweis et al., 2013; Agosta et al., 2015; Krinner and Flanner 2018; Fettweis et al., 2020), highlighting the need for a thorough evaluation over the present climate. Since ESMs only simulate meteorological conditions representative of a certain climate, evaluating MAR ESM-forced simulations cannot be done using the observations directly. We then compared these simulations to the averaged MAR(ERA5) hereafter considered as a reference and evaluated in Sect.~S1.

MAR(ACCESS1.3) is the experiment that best compares with the reference MAR(ERA5) over the present climate. It displays the lowest integrated-SMB anomaly (Tab.~S2) and spatial RMSE and bias (Fig.~2). MAR(ACCESS1.3) underestimates SMB over Wilkes Land, Queen Mary Land and the Amundsen sector while it overestimates SMB over Queen Maud Land and the lee side of the Antarctic Peninsula. These negative anomalies are associated with the small underestimation of the summer and winter precipitable water in ACCESS1.3 (Agosta et al., 2015). This experiment also reveals mostly non-significant temperature biases in summer (Fig.~S3), except for a small negative bias over Ross and Rhone ice shelves, yielding very similar integrated melt values.

MAR(NorESM1-M) presents mostly non-significant anomalies compared to MAR(ERA5) but overestimates the mean integrated annual SMB as a consequence of an overestimation of the snowfall and to, a lesser extent, a lower surface ablation (Tab.~S2). Higher snowfall values are modelled over Marie Byrd Land, the Peninsula, and the Brunt ice shelf whiler lower values compensate this overestimation over Queen Mary Land, Wilkes Land and the Amery ice shelf (Fig.~S4), which are strongly linked with the humidity anomalies in the forcing ESM (Agosta et al., 2015). NorESM1-M being too cold (with lower free atmosphere summer and ocean temperatures as well as higher sea ice concentration), MAR(NorESM1-M) displays a negative temperature anomaly up to 3°C over the plateau despite reducing the negative anomaly over half of the Antarctic ice sheet to non-significant differences in summer (Fig.~S3). This however leads to reduced surface melting (72 Gt yr-1).

MAR(CNRM-CM6-1) simulates nearly the same integrated snowfall amount as MAR(ERA5) but has a higher SMB RMSE due to a less accurate spatial representation of the precipitation. This results from an overestimation of the precipitable water combined to a higher mean sea level pressure in CNRM-CM6-1 potentially reducing cyclonic activity. MAR(CNRM-CM6-1) underestimates the SMB over the Ellsworth Land and the windward side of the Peninsula but overestimates it over Marie Byrd Land, Queen Maud Land and Victoria Land (Fig.~2). Agosta et al. (in preparation) revealed a strong negative temperature anomaly surrounding the ice sheet, yielding lower temperature in MAR(CNRM-CM6-1) *compared to MAR(ERA5) over the plateau. However, these differences are non significant over the margins, the Ronne ice shelf excepted (Fig.~S3).*

As it simulates lower snowfall amounts than MAR(ERA5), MAR(CESM2) slightly underestimates the mean integrated SMB. However, MAR(CESM2) represents a stronger accumulation over the area between the Peninsula, Queen Maud Land and Enderby Land (Fig.~2). This results from the significant overestimation of the precipitable water and the sea level pressure in CESM2 over this area. On the contrary, MAR(CESM2) simulates a lower accumulation over Wilkes Land and the Amundsen sector. CESM2 is colder than ERA5 but the difference is reduced in summer (Agosta et al., in preparation), leading to mostly non significant temperature anomalies in summer (Fig.~S3) .

In general, the SMB downscaled by MAR forced by the 4 ESMs is close to MAR(ERA5). The anomalies of the annual mean SMB are lower than the interannual variability of the SMB over the historical period. MAR(ACCESS1.3) has the best representation of the Antarctic SMB over the current climate (mean bias: -3 Gt yr-1, spatial rmse: 59 kg m-2 yr-1), while MAR(CESM2) is the least accurate (mean bias: -26 Gt yr-1, spatial rmse: 90 kg m-2 yr-1). The results of our experiments over the current climate are consistent with the ranking of the ESMs given by (Agosta et al., 2015, Barthel et al., 2020) and Agosta et al. (in preparation). This highlights the importance of selecting ESMs that correctly represent the historical climate (in particular the free atmosphere and the general circulation) around Antarctica as they induce biases in the downscaled near-surface climate independently of the capacity of the RCM to improve ESMs results. It is also important to note that the spatial and integrated anomalies are close to (or even lower than) the spread between several RCMs all forced by ERA-Interim (Mottram et al., 2020). This suggests a good ability of the different simulations to closely reproduce the SMB over the present climate and gives some confidence in the results of the future projections.

I also believe that it would be appropriate to cross-validate the quadratic fits with an independent dataset. For example, if one fits this polynomial to 3 of the 4 experiments, how well is the fourth predicted? Showing that it does a good job would go a long way towards ensuring that this surrogate model (which is what it is) is likely to be skillful at predicting the anomalies for other models.

We cross-validated our "regression model" fitted on 3 experiments compared to the last remaining one (See Fig. R1 to R4). Although the results remain good, the remaining error reflects the inherent error of the regression linked to the simplification, but also the inter-model variability for a same temperature threshold which cannot be captured by the other 3 models. For example, Figure R3B shows that the regression based on ACCESS1.3, NorESM1-M and CESM2 underestimates the negative anomaly of SMB on ice shelves. This is expected because the regression does not take into account the very negative anomalies that only MAR(CNRM-CM6-1) does. This also highlights the importance of inter-model variability for the projections and thus for maximising the projected warming.

The RMSE of each cross-validation can be compared to the RMSE of the original "validation". The cross-reconstruction is better when compared to ACCESS1.3 (NorESM1-M) over the grounded ice sheet (over the ice shelves) and of the same order over the ice shelves (the grounded ice sheet). When compared to CNRM-CM6-1 or CESM2, the RMSE is larger. As these experiments simulate larger anomalies, the same relative error leads to a larger absolute error. Furthermore, as explained before, these “cross-regression” do not take into account all the inter-model (or inter-downscaling) variability inducing a bias. This is especially true for the strongest warmings, which are reached by only two models (CNRM-CM6-1 and CESM2). Removing one of these two during the crossfit therefore results in a large increase in uncertainty (and error) for the strongest anomalies caused by the strongest warming as it precludes the representation of model inter-variability. This again highlights the importance of using as many as ESM candidates as possible and that more downscalings are needed to reduce projected uncertainties.

We think that this cross-validation gives similar conclusions to the evaluation of our regression already present in additional material so we suggest to not add it.

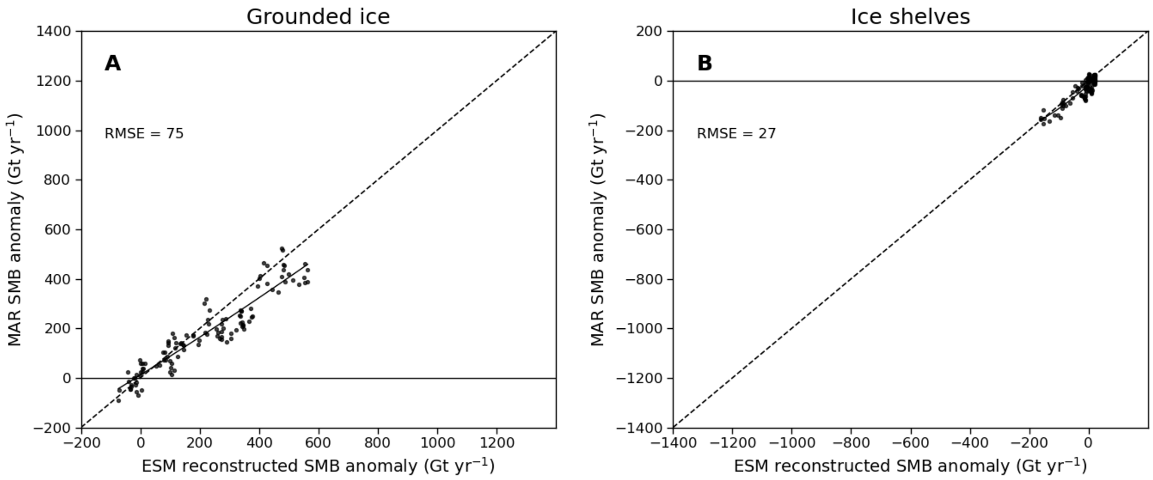

Fig R1: Evaluation of the MAR(ACCESS1.3) reconstructions based with regressions derived from the three other model anomalies ($f(x) = -3.2\ TAS^2 + 130.5\ TAS\ -17.3$ and $f(x) = -13.4\ TAS^2 + 37.4\ TAS - 3.8$) compared to the original MAR SMB anomalies over the grounded ice (A) and the ice shelves (B).

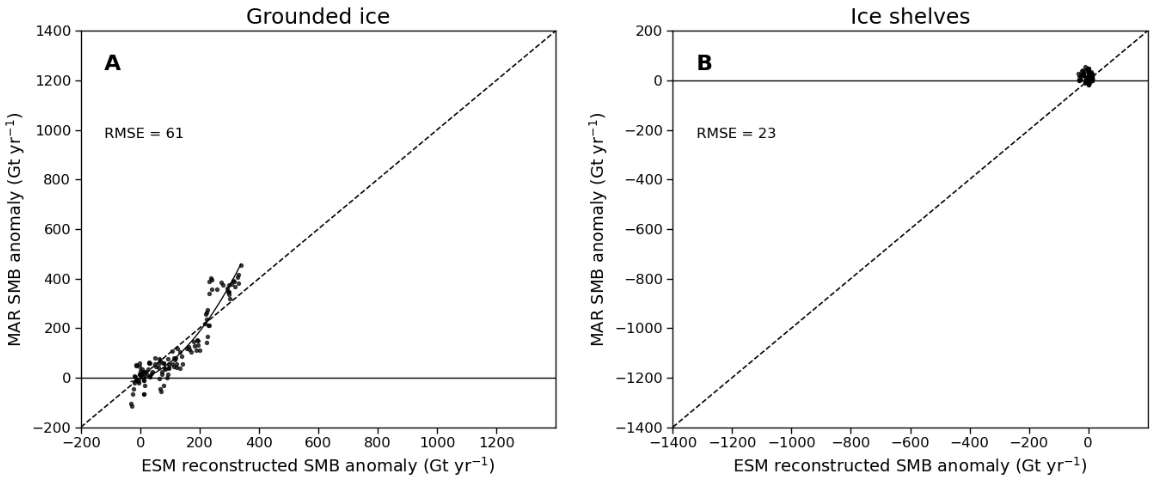

Fig R2: Evaluation of the MAR(NorESM1-M) reconstructions based with regressions only derived from the three other model anomalies ($f(x) = -0.7\ TAS^2 + 109.6\ TAS - 3.3$ and $f(x) = -12.2\ TAS^2 + 29.1\ TAS - 3.8$) compared to the original MAR SMB anomalies over the grounded ice (A) and the ice shelves (B).

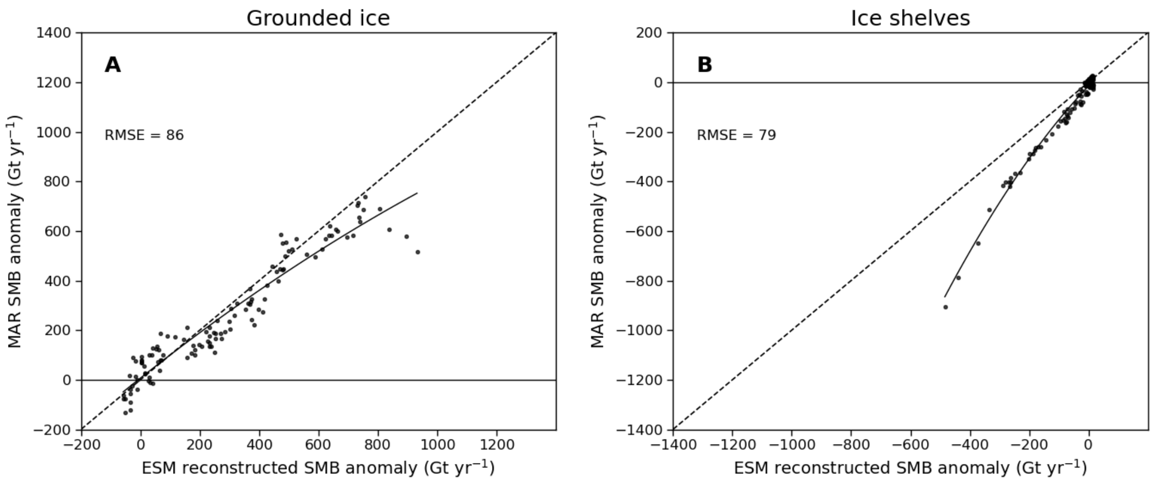

Fig R3: Evaluation of the MAR(CNRM-CM6-1) reconstructions based with regressions only derived from the three other model anomalies ($f(x) = -0.1\ TAS^2 + 115.2\ TAS -12.6$ and $f(x) = -10.3\ TAS^2 + 27.1\ TAS - 1.4$) compared to the original MAR SMB anomalies over the grounded ice (A) and the ice shelves (B).

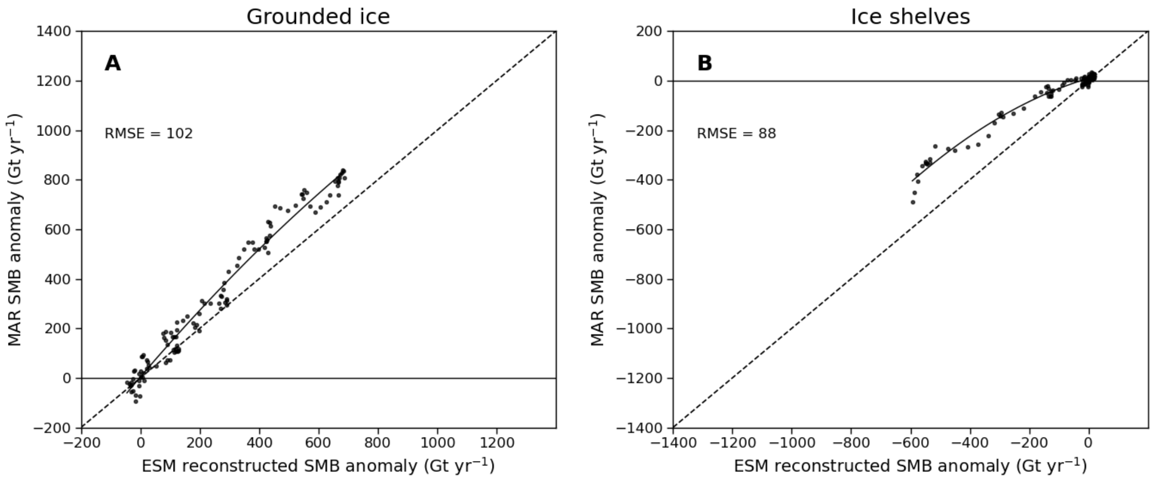

Fig R4: Evaluation of the MAR(CESM2) reconstructions based with regressions only derived from from the three other model anomalies (f(x) = -2.5 TAS² + 111.8 TAS -12.1 and f(x) = -15.9 TAS² + 40.1 TAS - 5.7) compared to the original MAR SMB anomalies over the grounded ice (A) and the ice shelves (B).

Minor Points

L13 Specify what a 'lower surface mass balance' means. More negative?

We suggest to change L13:
Over the ice shelves, the strong runoff increase associated with higher temperature is projected to lower the SMB with a stronger decrease in CMIP6-ssp585 compared to CMIP5-RCP8.5.

by:

Over the ice shelves, the strong runoff increase associated with higher temperature is projected to decrease the SMB (more strongly in CMIP6-ssp585 compared to CMIP5-RCP8.5).

Fig. 1 Would it be possible to include observed temperature anomaly in some form? maybe ERA5?

We added ERA5 over 1960 -- 2020 using the latest backward release of ERA5 prior to 1979. However, initial results suggest that the reanalysis is not as reliable as before 1979 than after (Hersbach et al., 2020). This is likely a combination of climatic isolation of Antarctica and poor observation coverage. The new Fig.1 now also illustrates that ESMs correctly reproduce the mean warming since 1960.

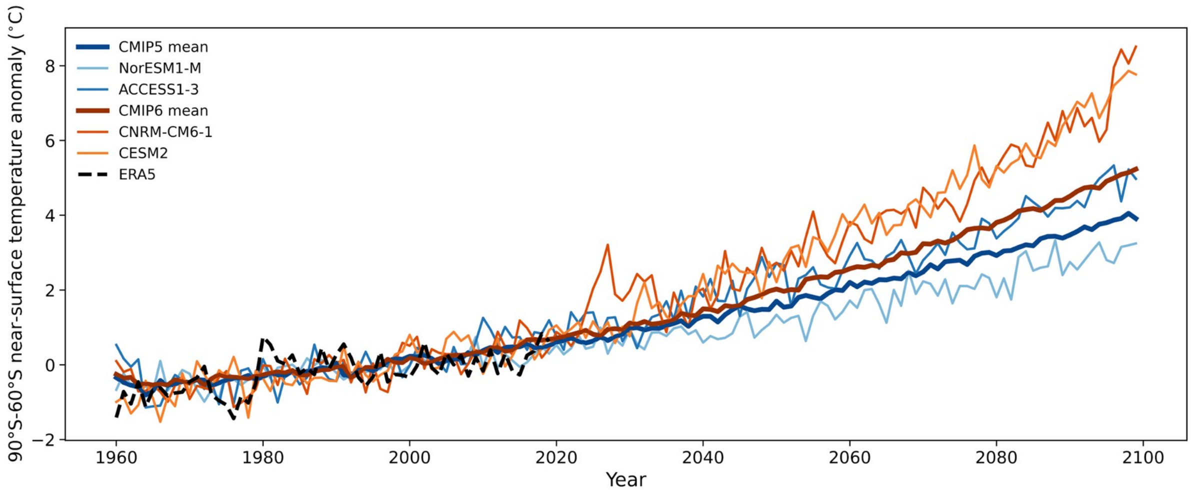

Fig1.: Time series of the 90°S–60°S annual near-surface temperature anomaly (°C) between compared to the reference period (1981–2010) from the ERA5 reanalysis and ESMs using the extreme high-emission scenarios RCP8.5 and ssp585 after their historical period (2004 for CMIP5 and 2014 for CMIP6). The thick blue and red lines represent the mean annual warming from 28 CMIP5 and 34 CMIP6 ESMs. Thinner orange and blue lines are for ESMs selected as boundary conditions for our regional climate model MAR: CNRM-CM6-1 and CESM2 (CMIP6, ssp585) and NorESM1-M and ACCESS1-3 (CMIP5, RCP85). The dashed black line is the ERA5 reanalysis (1960--2020).

L128 It’s okay to leave the details to the references, but it would be helpful to at least qualitatively describe the methodology for comparing ESMs to ERA5 that are used here.

This aspect of the paper had been less developed as it was based entirely on the method defined in Agosta et al, 2015 and Barthel et al, 2020. In addition, the extended discussion including CMIP6 of this selection will be the subject of a paper (Agosta et al., in preparation) with in part the same authors of our manuscript. It will include the rankings of each mode including new CMIP6 models (more models are now available than when we had to make our selection) and a discussion on the influence of the choice of metrics and other parameters not included in the score (resolution too low, potential importance of some metrics in relation to others) and on the strategies to be followed when selecting models (i.e. take only the best on the present climate, diversify warming or exclude models with too large a ECS even if this means excluding some good models) for both the Greenland and Antarctic ice sheets. We felt that this whole subject deserved to be discussed separately as it could have an important influence on the projections that will be made and the possible impacts of these results on policy makers.

However, we have extended the section presenting the selection procedure (L121-128) following the reviewer’s remarks. Furthermore we have inverted the two first paragraphs of the selection section .

We changed:

The selection of ESMs that were dynamically downscaled by MAR was based on their ability to 1) represent the current climate (air temperature and humidity, sea surface conditions, *and large-scale circulation) around the AIS and 2) diversify the projected changes during the 21st century. These criteria ensure on one hand, that the ESM biases will not have a prejudicial effect on the projections since the present state determines future biases (Agosta et 2015, Krinner and Flanner 2018) and on the other hand that we assess the AIS response to a wide range of projected temperature increases for a better quantification of the future uncertainties. We therefore selected ESMs by comparing them to the ECMWF reanalysis ERA5 (Hersbach et al., 2020) over the recent "historical" period (1980--2004) following the method defined in Agosta et al., 2015 and Barthel et al. (2020) for CMIP5, extended here to CMIP6 and applied only to the Antarctic atmosphere.*

Large-scale forcing models were chosen among the CMIP5 and CMIP6 ESMs. CMIP6 models rely on an improved and more sophisticated representation of the global climate system than CMIP5. They incorporate better coupling between the different components of the Earth system, improved present- and better-constrained future concentrations scenarios of long-lived greenhouse gases and aerosols (Eyring et al., 2016,O'Neill et al., 2016). Additionally, most CMIP6 ESMs are also run on a higher spatial resolution. First analyses of the CMIP6 results revealed higher equilibrium climate sensitivity in this new-generation models (Mauritsen et al., 2019, Voldoire et al., 2019, Zelinka et al., 2020, Meehl et al., 2020, Wyser et al., 2020), suggesting warmer future climates, while based on similar future scenarios in terms of global radiative forcing. However, this higher climate sensitivity is potentially not supported by paleo-climate records (Zhu et al., 2020). We therefore also included models from the CMIP5 dataset, some of which show a good agreement with reanalyses over the current Antarctic climate (Agosta et al., 2015, Palerme et al., 2017). We only chose the scenarios of large greenhouse gas emissions from CMIP5 (RCP8.5) and its updated version in CMIP6 (ssp585) in order to obtain stronger warming signals. These two scenarios have an equivalent global radiative forcing of +8.5 W m-2 by 2100, but differ in how the anthropogenic forcing is split between individual drivers of global warming (O'Neill et al., 2016).

by

Large-scale forcing models were chosen among the CMIP5 and CMIP6 ESMs. CMIP6 models rely on an improved and more sophisticated representation of the global climate system than CMIP5. They incorporate better coupling between the different components of the Earth system, improved present- and better-constrained future concentrations scenarios of long-lived greenhouse gases and aerosols (Eyring et al., 2016,O'Neill et al., 2016). Additionally, most CMIP6 ESMs are also run at a higher spatial resolution. First analyses of the CMIP6 results revealed higher equilibrium climate sensitivity in this new-generation models (Mauritsen et al., 2019, Voldoire et al., 2019, Zelinka et al., 2020, Meehl et al., 2020, Wyser et al., 2020), suggesting warmer future climates, while based on similar future scenarios in terms of global radiative forcing. However, this higher climate sensitivity is potentially not supported by paleo-climate records (Zhu et al., 2020). We therefore also included models from the CMIP5 dataset, some of which show a good comparison with reanalyses over the current Antarctic climate (Agosta et al., 2015, Palerme et al., 2017). We only chose the scenarios of the largest greenhouse gas emissions from CMIP5 (RCP8.5) and its updated version in CMIP6 (ssp585) in order to obtain stronger warming signals and then SMB sensitivities. These two scenarios have an equivalent global radiative forcing of +8.5 W m-2 by 2100, but differ in how the anthropogenic forcing is split between individual drivers of global warming (O'Neill et al., 2016).

The selection of ESMs that were dynamically downscaled by MAR was based on their ability to i) represent the current climate (air temperature and humidity, sea surface conditions, and large-scale circulation) around the AIS and ii) diversify the projected changes during the 21st century. These criteria ensure on one hand, that the ESM biases will not have a prejudicial effect on the projections since the present state determines future biases (Agosta et al., 2015, Krinner and Flanner 2018) and on the other hand that we assess the AIS response to a wide range of projected temperature increases for a better quantification of the future uncertainties for a same scenario. We therefore firstly ranked ESMs by comparing them to the ECMWF reanalysis ERA5 (Hersbach et al., 2020) over the recent "historical" period (1980--2004) following the method defined in Agosta et al. (2015) and Barthel et al. (2020) for CMIP5, extended here to CMIP6 and applied only to the Antarctic atmosphere. The method firstly computes the root mean square error (RMSE) compared to ERA5 for several climate variables (mean air temperature at 850 hPa, annual precipitable water, annual sea level pressure, summer sea surface temperature and winter sea ice extent over 1980--2004) that are supposed to determine the SMB (Agosta et al. , 2015). The score of each ESM is then obtained by averaging its RMSEs that were previously normalized with regards to the multi-model median and interquartile range. This enables the combination of several metrics using the same weight for each of the metrics. Once the models were ranked on the basis of their score against ERA5, the final selection was made to diversify the changes expected at the end of the century and the availability of 6-hourly outputs in the CMIP5/CMIP6 database at the end of 2019 when we started our experiments.

L170 SIC: ‘Suplement’
L227 SIC: ‘questionning’,
SIC: ‘in pace’
Thanks for highlighting these mistakes! They were corrected in the revised version of the manuscript.

Sec 4.1 When reporting the bounds on SLE in this section, is the error in the surrogate model accounted for? The fit to the simulations isn’t perfect, so there should be some extra variance injected to account for potential mismatch between Eq. 2/3 and the true model predictions, rather than just the spread in the predictions themselves

The regression error has not been included in the shading of Figures 8 and 9 or the bounds since we only aimed to illustrate the uncertainty caused by inter-model variability. We recognize that the total future uncertainty should be greater if we add to the inter-model uncertainty the uncertainty due to the intrinsic errors of the regression, in a way that is difficult to quantify.

Nevertheless, the RMSEs between the original and reconstructed anomalies (ie, 68 kg/m^2yr and 38 kg/m^2yr) are 1) lower (a little more larger) for the grounded (ice shelves) than the present SMB variability 2) lower than the future projected changes in our simulations or in the CMIP reconstructions 3) lower than the interannual variability in MAR projected anomalies 4) lower than the inter-model variability within a same scenario.

The regression error thus appears to be of lower importance than present and future interannual variability, projected changes and inter-model variability, suggesting that it introduces at least second-order uncertainty with respect to all these indicators.

We will add after L338:

Note that the uncertainties associated to mean reconstituted anomalies are only based on the intermodel variability over both the grounded ice sheet and the ice shelves but the uncertainties would have been larger if the biases of MAR (over current climate) and Eq.~2 and Eq. ~3 had been taken into account.