Reviewer #1

Dunmire et al. exhaustively evaluate the results of CESM2 over the Antarctic Ice Sheet against AWS observations and several often-used products such as RACMO, MAR, ERA5 or the reconstruction from Medley and Thomas (2019). The comparison is honest by highlighting both remaining biases and improvements against the previous version of the model. Furthermore, they discuss the future evolution of SMB using three different scenarios as it is a key variable for the ice sheet dynamics. One could argue than the topics is not particularly new, but this is an important study that would deserve a publication in the Cryosphere. CESM2 (one of the few polar-oriented ESMs) is often used either directly or downscaled to study the Antarctic climate and force ice sheet models. Knowing its biases is then important and I think that updates comparing to CESM1 (new emission scenarios, improved physics) are sufficient to justify a new study to present SMB projections. Furthermore, I found the paper well-structured, with clear figures and conclusions that are supported by results.

> We thank the reviewer for their positive and encouraging comments regarding our manuscript.

I have only minor comments listed hereafter that I hope will help the authors to improve their manuscript before its potential publication.

Minor comments

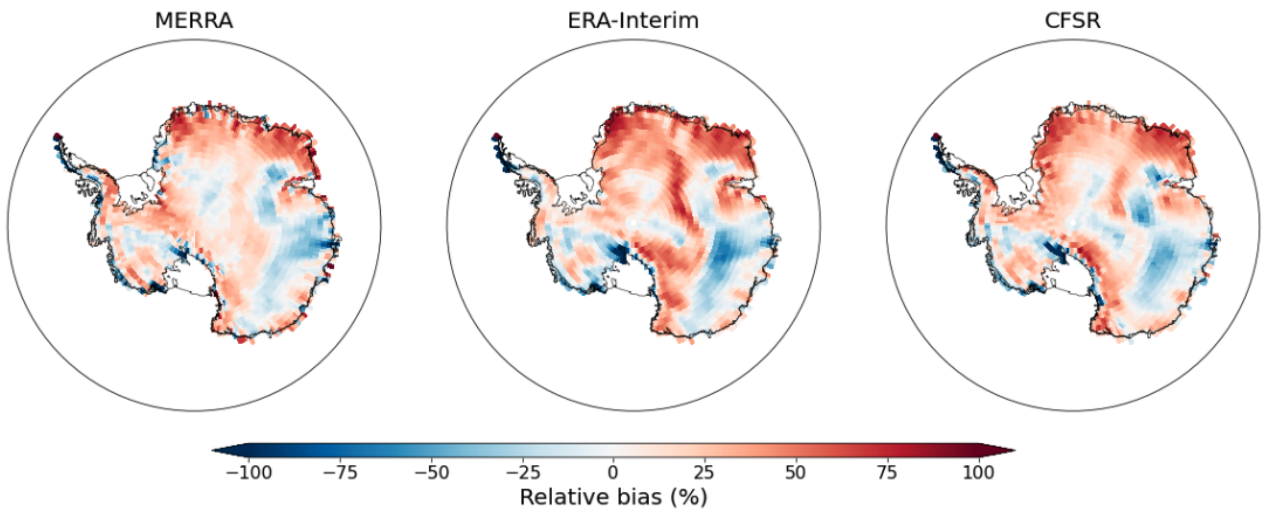
P4 Section 2.5: I recommend to specify the forcing of the RCMs (I think ERA-Interim for RACMO, ERA5 for MAR)

> Thanks for suggesting this. We will update lines 107-108 in Section 2.5 to read: "we compared CESM2 results to RCM output from the latest versions of RACMO2.3, which is forced with ERA-Interim (van Wessem et al., 2017), and MAR (version 3.11), which is forced with ERA5 (Kittel et al., 2021).

Furthermore, you used ERA5 as a reanalysis but the reconstruction based on MERRA2. Yet this choice is explained and fully justified by the better performance on the reconstruction with this specific reanalysis but I wonder if the different forcings (MERRA2, ERA-Interim, ERA5) could result in different trends and change your comparison with CESM2.

> With regard to the mean-state SMB for the grounded ice sheet, all three reconstructions from Medley and Thomas (2019) have insignificantly different ($p > 0.05$) average annual SMB and have similar spatial patterns of relative bias when compared to CESM2 (see below):

MERRA2 reconstruction annual SMB:	1788 Gt/yr
ERA-Interim reconstruction annual SMB:	1816 Gt/yr
CFSR reconstruction annual SMB:	1750 Gt/yr



With regard to SMB trends, all three reconstructions have an insignificant SMB trend from 1950 to 2000. We only use the reconstruction product to compare with the CESM2 mean-state SMB and trend in SMB over the historical period (not other variables such as temperature or precipitation). Thus, our choice in using the MERRA2 reconstruction product will not change the comparison with CESM2 and is fully justified in line 110: "In this study we used the MERRA-2 based SMB reconstruction as it most closely resembles observations (Medley and Thomas, 2019)."

Since you also used a collection of products that could all give you the same information, is there any reason why you selected ERA5? For instance, precipitation or temperature could be also compared with RACMO or MAR. Try to justify why you selected ERA5 for these comparisons.

The near-surface precipitation and temperature trends from ERA5, RACMO2.3, and MAR are very similar (see figures below). We will add the figures below to the appendix and justify our decision to use ERA for the comparison with CESM2 by adding the following to our methods section 2.2.3:

"We also compared the CESM2 trend in near-surface temperature and precipitation from 1979-2015 with that from ERA5. We used ERA5 for this comparison because (a) it is the latest reanalysis product, with updated model physics and the highest horizontal resolution, and (b) has similar near-surface temperature and precipitation trends to the RCMS used in this study (Fig. A1, A2). The ERA5 near-surface temperature trend is also consistent with observations (Zhu et al., 2021)"

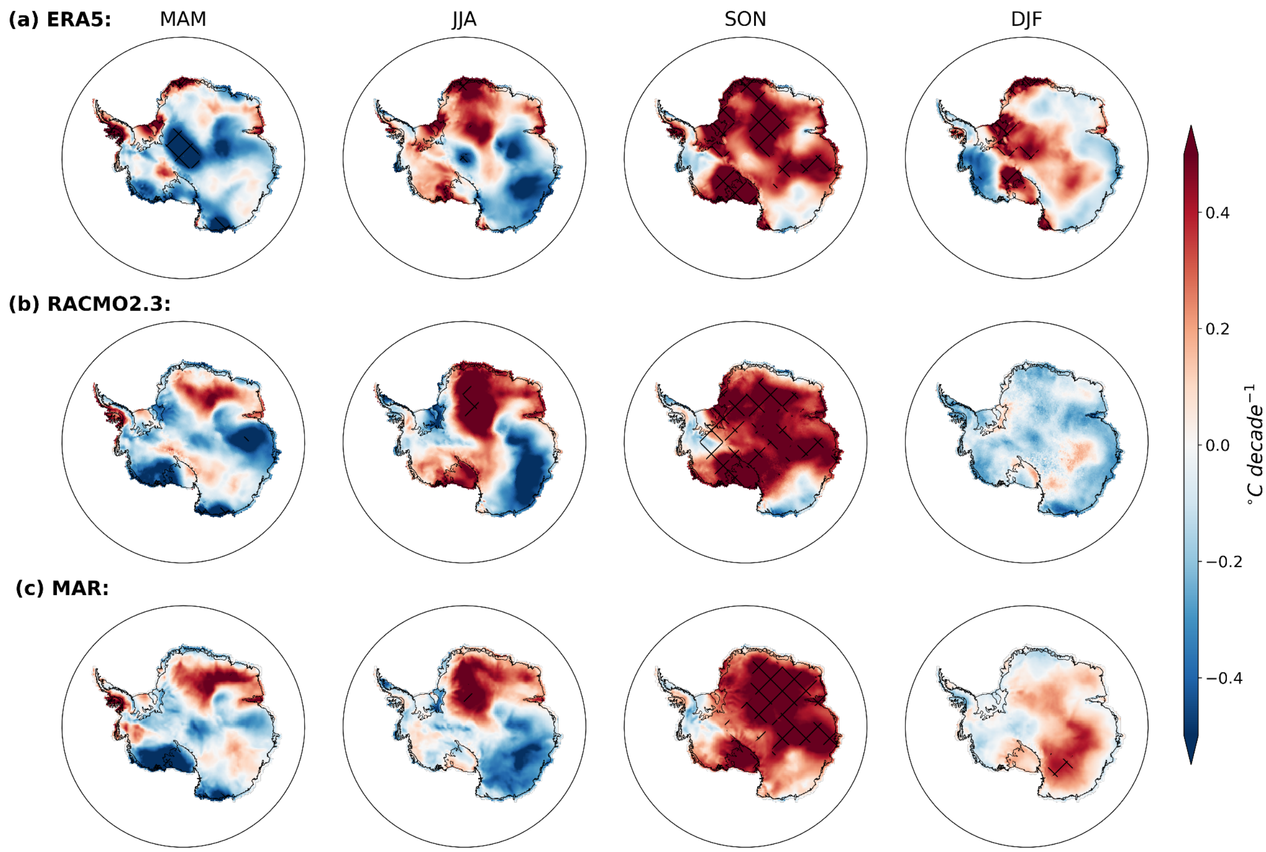


Fig. A1. 1979-2015 seasonal trend in near-surface temperature from ERA5 (a), RACMO2.3 (b), and MAR(c).

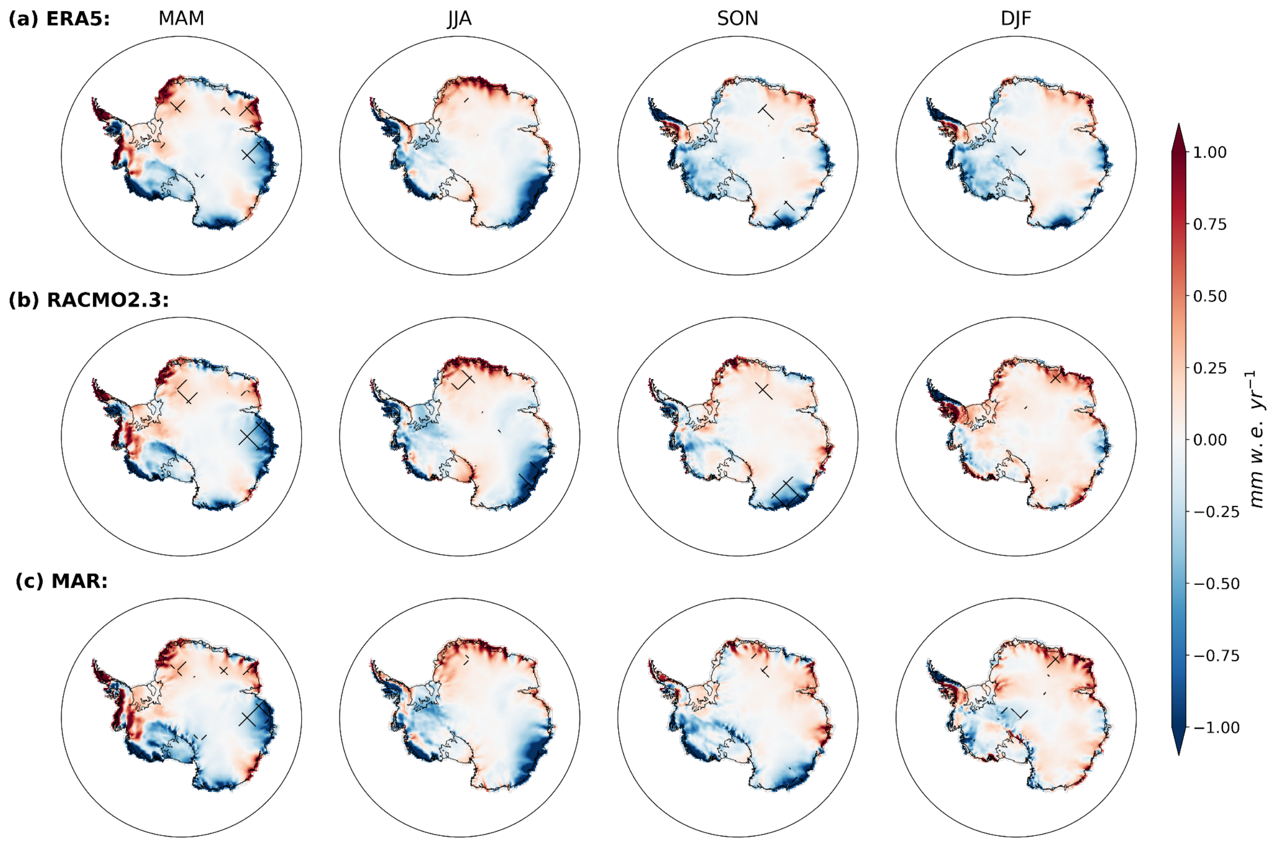


Fig. A2. 1979-2015 seasonal trend in precipitation from ERA5 (a), RACMO2.3 (b), and MAR(c).

P11 L205: According to Figure 6a, I guess the values are for the grounded ice sheet. What is the SMB of CESM2 over the ice shelves? There seems to have no evaluation/comparison over the ice shelves while other products than "the reconstruction" could be used there. Since ice shelves are particularly important for the Antarctic mass balance, this should be corrected.

> Figure 6 and lines 204-208 only include SMB over the grounded ice sheet because the reconstruction is only for the grounded ice sheet. We will clarify this in the text by changing lines ?? to "In CESM2, the annual average **grounded** surface mass balance (SMB) between 1979 and 2015 is…". We will also specify in the caption for Figure 6 that the timeseries (panel a) only includes the grounded ice sheet. To satisfy complaints from both reviewers we will expand our analysis in section 3.4.1 to include a comparison of ice shelf SMB from CESM2 with CESM1, ERA5, RACMO2.3, and MAR by adding the following paragraph:
>
> "Over ice shelves, CESM2 has an average SMB of 559 +/- 27 Gt yr^{-1} between 1979 and 2015, significantly greater (p~<~0.05) than the average SMB over ice shelves from CESM1 (520 +/- 26 Gt yr^{-1}), ERA5 (506 +/- 26 Gt yr^{-1}), RACMO2.3 (523 +/- 24 Gt yr^{-1}), and MAR (459 +/- 23 Gt yr^{-1}). The MT2019 reconstruction only covers the grounded ice sheet and thus ice shelf SMB cannot be calculated from this product."

Are you well using a common mask? if yes, specify it to help the readers interpretation, if not I strongly recommend to do it to compare something similar. (Also true for CMIP5 and CMIP6 values in Fig6.).

We are using the Zwally mask which has been regridded for all of the modeling products used in this study. We will specify this and note the AIS grounded and ice shelf areas for each products in a new methods section 2.6 AIS Model masks:

"For area-integrated quantities we use the Zwally et al. (2012) AIS mask which has been re-gridded for all of the modeling products used in this study. The resulting grounded AIS areas from these models are as follows: 12043565 km^2 for CESM1 and CESM2, 12059084 km^2 for ERA5, 12063497 km^2 for RACMO2.3, 12154338 km^2 for MAR, and 12028208 km^2 for the MT2019 reconstruction. The resulting ice shelf areas from these models are: 1738581 km^2 for CESM1 and CESM2, 1755916 km^2 for ERA5, 1734991 km^2 for RACMO2.3, and 1749205 km^2 for MAR. Ice shelves are not included in the MT2019 reconstruction."

The annual AIS SMB values from all CMIP5 and CMIP6 models are from Gorte et al. 2020. The mask used for this analysis was the Ice Sheet Mass Balance Inter-comparison Exercise Team's (IMBIE Team) ice sheet mask.

Trends: Why do you use normalized trends? I understand that it better highlights the importance of small changes over areas with low values (eg., SMB over the high plateau) but at the same time it masks the real changes. Importance of small changes that are significant can still be highlighted by dots or crosses as you did. For Figures 3b and 7c indicate in the caption what crosses represent.

We agree that showing the normalized trends in Figure 7 is not necessary and will remove panel (c) from this figure. We will add cross-hatched areas to panel (b) to show areas where this trend is significant and add the following to Figures 3b and 7b: "Cross-hatched areas represent regions where this trend is significant ($p < 0.05$)."

Most CESM2 trends are compared against other products (which is really interesting), but not the melt trend. I'd suggest to perform a similar comparison or at least cite a study (eg. Kuipers Munneke et al., 2012) that presents melt trends.

We will cite the Kuipers Munneke et al., 2012 study to compare CESM2 historical melt trends with observations in the following sentence on line 193: "Historical (1979-2015) surface melt in CESM2 has increased across much of the AIS, a trend that is absent from observations (Kuipers Munneke et al., 2012)."

P14 L246 and after: SSP scenarios are only mentioned using their first category (SSP5 instead of SSP5-8.5). Since there are several under scenarios in each category, keep mentioning the full name to remain clear.

We will update all references of SSP5 to SSP5-8.5, SSP3 to SSP3-7.0, and SSP1 to SSP1-2.6.

Specific comments and stylistic suggestions

P1 L4 : maybe « climate models » in general is enough than only ESM.

We will keep ESM here because "climate models" could also include regional climate models or reanalysis climate models which we are not referring to in this line.

P1 L14 : I suggest to replace « a coupled Antarctic Ice Sheet » by an « coupled ice sheet model » as it's not the real AIS that will be integrated into CESM3. This is only a suggestion which the authors can obviously accept or refuse.

We will change "coupled Antarctic Ice Sheet" to "coupled ice sheet model" on line 14.

P1 L24 : I agree about the stronger regional warming over these regions but the references are not adequate. The mass losses in West Antarctica are mainly due to ocean warming and not to the atmosphere that the references refer to. Increasing air-temperatures are more likely to contribute to hydrofracturing over the AP and subsequent glacier speed-up, but this is still a small contribution against the total mass loss over these two regions. Please reformulate/change your references.

We will change the sentence in L24 to read: "Ice shelves in the Amundsen and Bellingshausen sea regions are thinning in large part due to increased basal melting (Pritchard et al., 2012), a process that reduces the buttressing effect of ice shelves and leads to increased ice discharge (Rignot et al., 2019, Milillo et al., 2022)"

P2 L30 : Consider to remove « Studies have shown that »

We will remove the words "Studies have shown that" in line 30 such that this sentence now begins: "Historical increases in AIS SMB…"

P2 L32 : Barthel et al. 2020 do not discuss the SMB uncertainty.

We will remove the "Barthel et al. 2020" citation in line 32.

P2 L54 : Add a reference (Gorthe et al., 2020?)

We will add the "Gorte et al. (2020)" citation to the sentence "CMIP6 modeled annual SMB values… mean of 2127 Gt yr^{-1}." in line 54.

P3 and P4 (Section 2.1 and 2.2): Do you use a specific member for the comparison or also the average of the 11 members?

We will clarify that we use the average of the 11 ensemble members by stating in section 2.1.1: “For comparing the CESM2 mean and uncertainty of these output variables to other products we calculated the 11-member ensemble average mean and standard deviation.”

P4 L106-108 : I suggest to specify that the SMB of the RCMs (and CESM2?) also includes the runoff.

Runoff is included in SMB calculations from all model simulations except for ERA5. We will update the ERA5 SMB calculation to include runoff and will clarify this by changing line 106 to “(approximated by precipitation - evaporation/sublimation - runoff)”

P4 L111-112 : « The » reconstruction is perhaps a little over-emphasized given that the other products (MAR and RACMO) also give reconstructions. (Again a suggestion, feel free to take into account or not). I’d suggest to refer to something like “the SMB reconstruction of Medley and Thomas (2019)” (or any abbreviation like MT2019 reconstruction).

We will change instances of “the reconstruction” to “the MT2019 reconstruction”, which we will define in Methods section 2.5.

P6 L132 and 136 : Consider to replace « affect » by « effect ».

We will make these changes to lines 132 and 136.

P7 L139-149 : Are the temperature trends in ERA5 reliable? If I’m not mistaken, most evaluations (eg., Gossart et al., 2019) only assessed the mean climate and not the trends. I would like more discussion on the potential reasons for these differences. Perhaps just mentioning that CESM2 is not constrained would be enough. Do you have a simulation where CESM2 is constrained that you could also compare to ERA5 (or AWS if ERA5 is not reliable)? (see also the minor comment about trends above)

ERA5 temperature trends are generally deemed reliable in those regions where we have long-term, high quality temperature observations (Zhu et al., 2021). As the reviewer indicates, CESM2 is fully coupled so is not constrained by any temperature observations. This will be clearly indicated in the Methods section of the text.

P11 L205-204: It’s confusing that CESM2 SMB is significantly greater than RACMO SMB (1997 Gt/yr) but not significantly greater than “the reconstruction” (1953 Gt/yr). I guess this come from the large variability in the reconstruction. Do you know why this variability is so large? Is the variability computed on the same period (as all the other products have almost the same variability)?

We have checked our significance tests for the SMB difference between CESM2 and the various other products and found that CESM2 SMB is significantly greater than all products (including MAR, for which we recalculated SMB using the Zwally mask used for all other products, and the reconstruction). For the reconstruction, the "variability" is so large because here we use the reconstruction error as provided by Medley and Thomas, 2019. We will clarify this at the end of section 2.5 by adding the following sentence:

"We use the SMB error provided by Medley and Thomas (2019) as the variability for this dataset."

P13 L236-240: Consider to divide the sentence in several ones to make it clearer.

We will split this sentence into 2 separate sentences so that it reads: "Differences in historical precipitation trend between ERA5 and CESM2 exist across much of the AIS, but particularly in Wilkes Land and Princess Elizabeth Land (~75 °E – 136 °E), with precipitation largely decreasing in ERA5 but increasing in CESM. Additionally, over the eastern AP (~63 °W) in DJF, precipitation decreases strongly in ERA5 but remains roughly constant in CESM2."

P14 L246: Specify if you're presenting temperatures over the (grounded or full) ice sheet or over the regions.

We will specify that we are presenting temperature over the full ice sheet in line 246.

P14 L247: "the first ten years of the future scenario (2015-2025) to the final ten years of the scenario (1990-2100)" Is there a mistake for the second period? (Shouldn't be 2090-2100?).

Correct, it should be 2090-2100. We will fix this mistake in line 247 of the manuscript.

Note that changes are more often compared to a selected period over the historical period than over a "future" period (I mean by "future", after 2014 where the scenario is no more the "historical" concentrations). The choice of the period should be consistent with P14 L268.

For the calculation of the $\frac{\partial SMB}{\partial T}$ values presented in section 3.5 we will change the period to be consistent with P14 L268 such that we are comparing the final 10 years of the historical simulation (2005-2015) with the final 10 years of the future simulation (2090-2100).

Furthermore, are 10 years representative of the climates of both the "historical/start of the future period" and the end of the century?

The objective of this analysis and of Figure 9 is to capture the change in SMB over the future period. We are not comparing representative climates because the climate is changing so rapidly throughout the future period, even within 10 years. We chose to use 10 years because it is a long enough period to account for inter-annual variability and allows us to evaluate the change in SMB over the future period in different scenarios.

P14 L250: Could you explain these differences? Are they due to the inertia of the system?

We believe that the diverging SMB trend on ice shelves contributes to the differences in $\frac{\partial SMB}{\partial T}$ between the different SSP scenarios. These rate change numbers are for the full ice sheet, something we have clarified on line 246. A strongly decreasing ice shelf SMB in SSP5-8.5 contributes to a lower rate of SMB change with increasing temperatures. This is stated on line 253-254: "A diverging future SMB trend on ice shelves and the grounded ice sheet, of which CESM2 agrees with previous studies (Kittel et al., 2021), is responsible for the varying $\frac{\partial SMB}{\partial T}$ between different emission scenarios."

P14 L270: This is a really interesting analysis. The negative SMB in summer for all the scenario suggests high runoff values and in general strong melt and melt ponds. Since runoff indicates remaining liquid water at the surface (sometimes considered to be a proxy of potential hydrofracturing– Donat-Magnin et al., 2021; Gilbert and Kittel, 2021), this might suggest that even for the low-emission scenario, surface melt could lead to severe damages over the ice shelves and strongly contribute to their disintegration with large consequences for the ice sheet stability. Maybe you could discuss/mention this in your manuscript.

Given the limitations in CESM2's ability to simulate meltwater processes, we would like to refrain from making substantial claims on ice shelf vulnerability to hydrofracture, and stick with the analysis of SMB and components. Surface melt and processes related to surface melt (ie refreezing, liquid water storage at the surface) are still not well captured by CESM2. Additionally, there are lots of other factors that contribute to meltwater ponding and the potential for hydrofracture (i.e. the availability of near-surface firn, the formation of impermeable ice lenses) that are likely not captured by CESM2. We propose that atmospheric forcing from CESM2 would need to be used as input for a more sophisticated firn model to analyze the impacts on ice shelf surface hydrology.

Appendix: Change the order of the figures to match their order of appearance in the manuscript.

We will update the order of the appendix figures to match their reference in the manuscript.

Figures (clear and adapted. I particularly appreciated Fig6.) For Figures 3b and 7c indicate in the caption what crosses represent.

> Thank you, we will indicate in the captions of Figures 3 and 7 what cross-hatching represents.

References

Gorte, T., Lenaerts, J. T., and Medley, B.: Scoring Antarctic surface mass balance in climate models to refine future projections, Cryosphere, 14, 4719–4733, https://doi.org/10.5194/tc-14-4719-2020, 2020.

Kuipers Munneke, P., G. Picard, M. R. van den Broeke, J. T. M. Lenaerts, and E. van Meijgaard: Insignificant change in Antarctic snowmelt volume since 1979, Geophysical Research Letters., 39, L01501, doi:10.1029/2011GL050207, 2012.

Milillo, P., Rignot, E., Rizzoli, P., et al: Rapid glacier retreat rates observed in West Antarctica, Nature Geoscience, 15, 48-53, 2022.

Pritchard, H.D., Ligtenberg, S. R. M., Fricker, H. A., et al: Antarctic ice-sheet loss driven by basal melting of ice shelves, Nature, 484, 502-505, 2012.

Zhu, J., Xie, A., Qin, X., et al.: An assessment of ERA5 reanalysis for antarctic near-surface air temperature, Atmosphere, 12, 217, 2021.