Reply to comments by Christoph Kittel: The influence of present-day regional surface mass balance uncertainties on the future evolution of the Antarctic Ice Sheet (egusphere-2023-2233)

Summary of Changes

We are grateful to C. Kittel for evaluating our work, and the valuable and constructive comments that help improve the manuscript. In response, we now

- Relax our restriction on stability for 100 kyrs under RACMO forcing for our parameter ensemble, to account for parameter configurations which might not work with RACMO but with other RCM forcings.
- Perform additional pre-industrial control runs and improve the spinup of our centennial simulations to ensure minimal model drift.

Below, we respond to C. Kittel's individual comments in detail and describe the actions we took to address them.

Detailed response

(Original report cited in italics)

This study focuses on the impact of anthropogenic global warming on rising sea levels, specifically examining the Antarctic Ice Sheet (AIS). It underscores the crucial role of selecting appropriate regional climate model (RCM) references for predicting future sea level rise contributions from ice sheets. By using the Parallel Ice Sheet Model (PISM), the researchers find that the choice of RCM reference forcing introduces uncertainties in sea level rise predictions. Additionally, the study highlights how the choice of RCM reference influences grounding line retreat in West Antarctica.

We thank, C. Kittel for evaluating our manuscript and his constructive and helpful comments. We address the raised points below and propose actions to clarify the points to further improve the manuscript.

Overall the manuscript is clear but some sections could be improved (notably the Methods). The topic is interesting as the influence of the SMB baseline has not yet been assessed. Differences of less than 100Gt/yr ie lower than the annual variability (for instance between MAR and RACMO whose results are really close to the observations following Mottram et al., 2021) seem to lead to large mass differences.

In order to address the raised points by all reviewers, the applied methodology (e.g. model initialization) will be revised. Reflecting this we will update and refine the Methods section of our manuscript.

Major comments

The paper is worth publishing, but I'm particularly concerned about the initialisation method and wonder to what extent the results are influenced by it. Like most models, PISM was originally calibrated to RACMO over Antarctica, and its development was based on this forcing. We can already assume that part of the model's behaviour is linked to RACMO (or at least a similar SMB field). The idea of redoing calibrations with other parameters seems to me to be an interesting way of overcoming this problem, but I have the impression that the results are still essentially influenced by RACMO and PISM's intrinsic behaviour, and therefore favour differences as soon as another forcing is applied. Only "good" calibrations for RACMO (or giving correct results and stability under the RACMO forcing) are conserved while other combinations could work for other models but not for the RACMO forcing. My concern is that the authors want to analyse the influence of the SMB between different models, but that they rely heavily on one of the models in question.

To address the fact that we have chosen the parameter combinations for our ensemble from calibration against the RACMO model, we relax the restriction on long term (100kyrs) stability under RACMO forcing. By doing so, we now account for parameter combinations which won't work for RACMO but might work for other RCM forcings.

To further, decrease the dependency of the simulation outcome on the thermal initialization performed with the RACMO forcing, we aim to perform individual thermal spinups for every RCM forcing as we described in our reply to Reviewer 1.

Furthermore, as the reviewer stated, PISM is often used in combination with RACMO forcing. Therefore, we can assume that some of the parameters we didn't touch in our simulation might have a bias towards the RACMO model, in the case RACMO was used in the calibration process. However, as PISM is a complex model it is unfeasible to individually recalibrate all parameters for this study. In conclusion we had to limit ourselves to parameters governing or affecting ice flow, grounding line behavior, ice shelf mass balance etc.. This is reflected in the choice of the "flow parameters" (sia_e, pQ), basal friction and the ice-ocean heat exchange coefficient (gamma), which have the strongest influence on the evolution of the ice sheet. It is important to note, that many default model parameters might represent Greenland conditions or are simply initial guesses. The selection of the baseline parameter combinations derives from a longer history of PISM studies and mostly deviates from the default parameter settings when running PISM as a black box.

The ensemble set could be enlarged by keeping the combinations that also work for another forcing, and a comparison of the best combinations for each forcing would also allow us to *see how a less good combination influences the results. It would also be possible to take the ideal calibration for one forcing and apply it to the others, to see how PISM responds in this case. What about PISM bias? Despite the different calibrations, isn't there a PISM component in the results? If PISM has a tendency to discharge the ice too slowly or too quickly (poor discharge due to poor basal or dynamic) obviously a "better adapted" SMB will always work better, especially when only the parameters that fit a model are kept.*

As already described above we aim to enlarge the ensemble by keeping parameter combinations which might not work for long term stability under RACMO forcing, but potentially work under different forcings.

Concerning the PIMS bias, we agree that PISM, as any other ice sheet model, tends in some regions to discharge ice to slow or fast. Nevertheless, since we perform all simulations with PISM, we assume that all results contain the same ice sheet model specific bias. Since we then calculate the difference between the individual simulations, we can assume that the ice sheet model bias will vanish in first order. There, might still be a higher order bias one could tackle by repeating our study with different ice sheet models. However, this is far beyond the scope of this study.

*Similarly, what is the impact of model drift? From figure d1 (a,b,c) (*which should be in the main text), only one simulation seems to have no drift. What happens to these differences if we remove the drift from PISM? For the scenarios with little warming, apart from MAR it looks like most of the differences could be caused by drift alone. Is it the drift of the model itself or also the result of an ice sheet that was out of balance at the start of the simulation because of another shape? This drift or imbalance would then be less significant in the simulation with a stronger anthropogenic forcing (rcp8.5).*

Figure D1 (a,b,c,) illustrates the ice mass change (in meters sea level equivalent) for the historical as well as the RCP scenarios. The dashed lines indicate control runs with the individual RCM forcings and constant 2005 HadGem2-ES anomalies. It is correct is that a significant portion of the absolute ice mass change, especially in the RCP2.6 and RCP4.5 scenario, occurs as well in the control run. Since we don’t expect the AIS to be necessary stable under 2005 climate conditions, we would also not assume the control run to be constant. In addition, both the control run as well as the RCP scenario simulations might be affected by a model drift. To assess this drift, we now perform an additional control run initialized at the start of the simulation with constant preindustrial climate conditions, as we have also described in the reply to Reviewer 1 and 3.

However, the aim of this study is to investigate how different RCM forcings affect the evolution of the ice sheet. To do so we calculate the maximum sea level contribution difference (Δslr_max) given in equation 4. In first order approximation Δslr_max should be invariant against the model specific drift, since we only look at differences of individual simulations and not at absolute ice mass change.

Specific and minor comments

P3L66 *: Please refer to Mottram et al., 2021 where MAR is described and not the dataset on Zenodo. I also encourage the authors to respect the data usage notice concerning MAR outputs that are available on Zenodo.*

Our apologies for failing to acknowledge the MAR team and will add this to the revised version of the manuscript. We will also update the reference to the publication and not the stand-alone dataset.

*The authors use models with the same forcing (ERA-Interim), which is a good point. They refer to Mottram et al., 2021 (**P4 L75- L76**) for the comparison between these models. However, RACMO2.3p2 is a more recent version of RACMO than the one use in Mottram et al., 2021. I won't say that the conclusions remain valid.*

Thanks for spotting this. We will clarify this in the manuscript.

Figure 1f*: This is not SMB ERAint vs the Ensemble Mean, but rather the SMB from ERAinterim. Please check your caption as some of them are not clear.*

We will fix this in the manuscript.

Generally speaking, I recognise that a lot of work has been done to free ourselves from the problems of initialisation and calibration depending on a single model, which is already a good thing, but I'm not convinced that we're free enough. I hope that the authors can improve this aspect of their study because I really think that this article is interesting and highlights the importance of multi-model studies.

Best regards,

C. Kittel