Response to Review of Mottram et al: “What is the Surface Mass Balance of Antarctica? An Intercomparison of Regional Climate Model Estimates”

Review 1:

By Tessa Gorte and Jan Lenaerts, University of Colorado Boulder.

Mottram and co-authors present an intercomparison of different regional climate models regarding their performance in simulating Antarctic Ice Sheet surface mass balance. They show that the RCMs, all forced by ERA-Interim at their boundaries, show overall satisfying (but to a varying degree) correspondence with available weather and SMB observations, and that many remaining biases are common between the different models. The integrated ice sheet SMB varies widely from model to model, but interannual variability is very similar. Overall, we think that this is an interesting paper, containing relevant and important results for the climate, SMB, and ice sheet modeling communities, and very fitting for potential publication in The Cryosphere.

However, this paper lacks a bit of context and broader impacts in its current form, and it suffers from some internal inconsistencies, ambiguities, and poor figure and language quality in places. We would invite the authors to consider our general and more specific comments, highlighted below.

We thank the reviewers for their very considerate and thoughtful review, we agree with many of their comments and in the process of addressing these we feel that the paper has been considerably improved.

General comments

First of all, in many places it is not clear to what the ice sheet integrated SMB numbers refer to, i.e. grounded ice sheet or full ice sheet (including ice shelves)? That’s an important issue to improve, not only to enhance clarity, but also since the former is directly translatable to sea level equivalent, the latter is not. An obvious place to start is the abstract (e.g. page 1, line 7 and 11). Using appropriate labels and explanations, and clearly separating grounded and full ice sheet throughout the paper would be essential.

This is a very good point and we have added a paragraph clarifying the difference between grounded SMB and SMB on ice shelves in the introduction section. The abstract has been completely revised and rewritten to reflect this and several other points raised by reviewers. Where SMB is discussed throughout the paper we clarify if we refer to the whole continent including ice shelves or only the grounded part.

“It is important to distinguish between the continental grounded ice sheet and ice shelves when considering values for SMB integrated over a wider area whether regional or continent wide. Snowfall and melt on ice shelves is not directly relevant to sea level rise contributions as they are already floating but that on grounded parts of the ice sheet is. In this paper when we refer to SMB over an area we include ice shelves, unless otherwise specified as the models used in this study by and large do not distinguish between grounded and floating ice in their ice masks.”

Second, although we understand that the authors want to refrain from ‘ranking’ the models, we would argue that, based on the input-output method of determining mass balance (in e.g. the IMBIE assessments), one could qualify the new RACMO2 and MARv3 models more realistic than other models. Using other models would draw a completely different picture of AIS mass balance; based on Table 3, using e.g. COSMO-CLM would more than double current AIS mass loss, and HIRHAM would suggest AIS mass gain, both of which cannot be reconciled with other methods that determine AIS mass balance (GRACE, altimetry, etc.). A discussion on this would strengthen the impact of this paper beyond a straightforward intercomparison, and inform the community on strengths and weaknesses of the different models.

The reviewers are correct that the aim of this study is not to rank the models. Our analysis shows that the different models tend to have different strengths both spatially and in terms of different processes in reproducing climate and weather in Antarctica. However, it is also an important point that the modelled SMB should be consistent with observational constraints from the input-output method and we have therefore explored this further. We have added a new short section in the discussion where we analyse the model output on the same ice sheet mask and over the same time period as that used in the IMBIE (Shepherd et al., 2019) study of Antarctic mass budget and discuss the implications. Our analysis shows that, given the published uncertainties on the observational estimates from the input-output method, the COSMO-CLM and MetUM estimates are outside the range defined by the IMBIE study based on altimetry and GRACE data. However, as these models, particularly MetUM, perform well in comparison to meteorological observations, the source of the mismatch is less clear and indicates that some of the components of SMB are being poorly captured by the models and/or that there are compensating errors in the modelled SMB. This is an important point and we have therefore also included it in the discussion as below and in the summary conclusions.

“It is interesting to compare our results with those used in the IMBIE study of Antarctic mass budget \citep{shepherd2018}. When taking into account the published uncertainties on the observational mass budget estimates from the input-output method, only the COSMO-CLM and MetUM estimates are outside the range defined by the IMBIE study based only on altimetry and GRACE data. However, as these two models, particularly MetUM, perform well in comparison to meteorological observations, the source of the mismatch is unclear and an area that requires significant future work. It may also indicate either that some of the components of SMB are poorly captured by the models or that there are compensating errors in the modelled SMB components and/or their spatial variability.”

Thirdly, many of the figures are very difficult to read, and colors showing different models are difficult to separate. Moreover, the figures could use a bit more explanation in the text as well as in the caption. A lot is left to the reader to decipher these figures (which potentially convey very interesting information).

We agree that more explanation of the figures is necessary and as well as revising them to make them clearer we have added additional explanatory text for each throughout the paper.

Lastly, language needs to be improved throughout. A few places are consistently lacking commas: after/around thus, therefore, moreover, etc. Several sentences were a bit long and could be broken up to make them easier to read. The authors switch between active and passive voice quite often throughout the text (i.e. “parameterizations are included” instead of “the models include parameterizations), suggesting that various authors have contributed to the writing and the end result is somewhat heterogeneous. We have pointed out a few locations below, but there are many more in the paper. Try to avoid phrases like ‘clearly’ throughout the paper. This is a subjective statement, and findings may not be so clear to the reader as it is to the authors.

A multi-author paper of this type is indeed vulnerable to inconsistent language and we have therefore proof-read and thoroughly revised all text and reverted all passive voice to active to make the paper more readable. We have also removed more subjective language and (also with reference to Reviewer 3’s comments) tightened up the statistical basis of statements where necessary.

Specific Comments

P1: Why are SMB and Gt given as abbreviations but not AIS which is abbreviated later?

We have added the AIS acronym and made use consistent throughout.

P1L1-2: Technically, Antarctica loses mass through enhanced ice discharge across the grounding line into ice shelves (not compensated by an increase in SMB), and ice shelves lose mass by enhanced calving and basal melt (not compensated by an increase in ice shelf SMB and/or solid ice influx). Separating these various processes can help to separate the grounded and full ice sheet (see General Comment 1).

We have revised the abstract considerably to make it shorter and easier to read, the separation of the mass budget components, including this point has now been included in the introduction section (as above).

P1L3-4: “... of crucial importance...” → “crucially important”

Removed - see previous comment

P2L1: “compar” → “compare”

Fixed

P2L12: “... a significant part of the climate system” is a bit vague and could be expanded upon

Adjusted to: “The Antarctic Ice Sheet (AIS) is the largest body of freshwater on the planet and thus a potentially important contributor to global sea level rise as well as a significant part of the climate system contributing freshwater to the ocean and with it's high relief significantly affecting atmospheric circulation.”

P2L15: Is “submarine melting” a common phrase for basal melting?

We use submarine melting here to distinguish from basal melting at the bed of the ice sheet generated by e.g. geothermal flux, friction processes etc.

P2L20-21: Scambos and Shuman maybe shouldn't be in all caps.

Fixed

P2L27-28: "In the future... climate change" → this sentence requires a change in punctuation for readability for me. For instance, consider changing to "In the future, a "greenlandification" of the ice sheet climate (increased melt and refreezing within the snowpack) is projected due ..."

Changed to: In the future, a "greenlandification" of the ice sheet climate is projected due to anthropogenically induced climate change \citep{trusel2018nonlinear}. This will lead to more melt with more refreezing in the snowpack as well as increasing runoff.

P3L12-16: "Souverijns et al... peer review literature" → this is quite a long sentence. Perhaps consider breaking it up for readability.

Changed to 2 sentences "In the polar regions, CORDEX simulations can also be used to assess the mass budget of the large polar ice sheets, but have not yet been evaluated together for Antarctica. \citet{Souverijns2019} made a 30 years hindcast with COSMO-CLM^{2}, and \citet{Agosta2019} estimated the SMB using MAR, while various versions of RACMO2 have been used to estimate the SMB of the AIS \citep{van2014improved, VanWessem2018}. Both MetUM and HIRHAM5 have been run for the Antarctic domain but evaluation of the SMB results have not yet been published in peer review literature \citep{hansen2019}"

P3L27: It might be good to list all 5 RCMs at the beginning of the Methods section

Added in brackets on first line.

P7L4: "Parameterizations are included..." → "The models include parameterizations..."

Fixed

P8L2: "... nudging whether spectral or with simpler techniques keeps..." → "nudging, whether spectral or with simpler techniques, keeps..."

Fixed

P9L6: "Weather observations are used..." → change to active voice

Fixed

P9L22-27: Change paragraph to active voice

Fixed to : "As the different models have different ice masks and topographies we only retain stations on the common mask where the difference in elevation is lower than 500 m for each model, this gives a total of 184 AWS (See the supplementary material for locations of AWS used in this study). We compute the modelled surface pressure, near-surface temperature

and wind speed, as well as the model elevation, using a four-nearest inverse-distance-weighted method. Finally, since the measurement height is not known for every station, we use the vertical level closest to the surface (10 m or 2 m) of the models for all comparisons with the observations. “

P10L10-11: “Observations between... 5 years” → consider rephrasing for readability

Fixed to “Observations between 1950 and 1987, or 2015 and 2018 that are not fully included in the common modelling period of 1987 to 2015, were used for evaluation only if they covered more than 5 years.”

P10L15-20: Authors say SMB was computed in 3 steps but only two seem to be explicitly mentioned.

Typo, FIXED

P11L10: So you’re saying that the higher resolution the model, the poorer skill it will show due to increased internal variability? Please clarify, since this is essentially contradicting many other studies that are suggesting enhanced performance with resolution.

The main issue here is that the Antarctic domain is very large, without nudging or relaxation the higher resolution models have many more degrees of freedom to evolve, we have clarified this here:

“Without nudging, the large domain size in Antarctica means that synoptic scale systems have more degrees of freedom to evolve away from the observed quantities. This is likely to be a particular problem for higher resolution models where there are more grid points between the boundary and a given station, compared to a lower resolution model with fewer grid points. Our results show that the high resolution (0.11\degree) version of HIRHAM5 that has many more grid cells than the low resolution (0.44\degree) version has a higher divergence due to internal variability. MetUM is not nudged by surface relaxation but is run in daily reinitialisation mode and while this probably also helps to keep surface pressure close to observed it is also likely that the large number of atmospheric levels in MetUM also improves modelled surface pressures.”

P11L17: What causes you to suspect “...biases in cloud cover and long-wave radiation...” are the leading factors in divergence from observation? How would you expect a model bias that overestimates cloud cover to effect observations, for example?

The analysis of Van Wessem et al 2014 shows that significant improvement of the RACMO2.3 model was derived from improved cloud microphysics parameterisations. We have clarified this further.

“However, biases in cloud cover and long-wave radiation reaching the surface are likely the main explanation for divergence from observations and should be investigated for all RCMs run for Antarctica as shown by \citet{vanWessem2014}. IN theri study, significant improvements in the RACMO2.3p2 model were obtained by adjustments to the cloud microphysics.”

P12L7-8: “The models can be divided into two groups...” → how are you dividing these groups? Not sure we understand the origin or the purpose of having two groups here.

Here we were referring to a visual contrast in the placement of the models on the Taylor diagrams. We have clarified it “The models appear to fall in two groups on the Taylor Diagram”

P12L22: Extra parenthesis.

Fixed

P12L23: “...in the colder, and therefore higher elevation locations, while...” → is this supposed to be “...in the colder, and therefore higher elevation, locations while...”? Also, perhaps consider changing the order to “in the higher elevation, and therefore colder, ...” such that it seems like temperature is a function of elevation and not the other way around.

Fixed to: “the other models overestimate temperature in the higher elevation, colder locations, while underestimating temperature at lower elevations in the coastal regions”

P14L14: What do you mean by “good results” exactly?

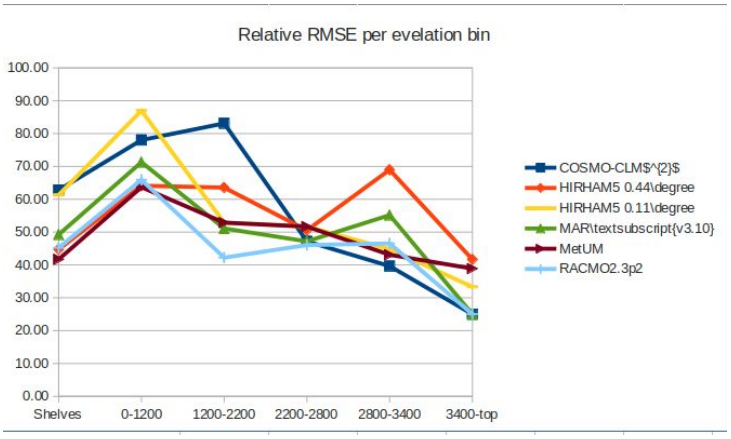
In this case we mean that compared to the other models, RACMO2.3p2 has a lower bias in the SMB and a higher correlation, however as the word good is unclear we have changed the sentence to read “The blowing snow module included in RACMO2.3p2 may explain the lower bias in results between 0 and 1200 and especially 1200 and 2200 m (Table \ref{tab:samba_smb} b and c), compared to the other models.”

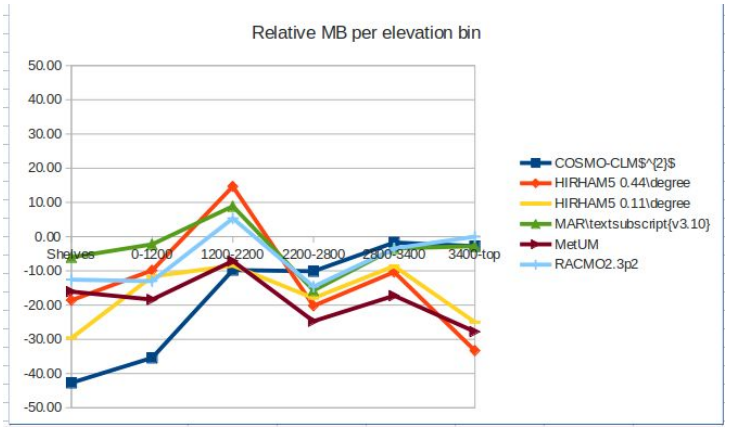
P16L2: “...we here show...” → “...here we show...”

Fixed

P17, Table 2: Arguable showing an RMSE with absolute SMB numbers that decrease rapidly from the coast to the interior is not justified, since the RMSE will tend to decrease along with the SMB itself. Adding relative RMSE (i.e. as a ratio to the mean) is required to compare apples to apples across the elevation bins.

This is a fair comment, we have updated table 2 to include the relative RMSE for each model by elevation bin, however it does not alter our results substantially. We propose to add the following new plot showing this relative RMSE for each model in the supplementary information as it shows visually how the percentage RMSE varies for each model according to different elevation bins.





P20: When looking at the ensemble mean, have you considered how your results may change if you calculate the mean on different grids? What grid did you use for this (i.e. how does this common grid resolution compare to that of any of the given models)?

We computed the ensemble mean SMB of the 9 models using each model's own grid. First we calculate the basin averaged SMB for all the models on their own grids and then we take the common grid points that fall within the defined basins. This means that numbers are independent of the model grids, and can be averaged into an ensemble mean.

We opted to use the RACMO2.3p2 grid to present the ensemble mean as it is an intermediate resolution for all the models and we compare it with the Shepherd et al 2019 study that also used this grid. We have updated the caption to reflect this.

P24L9: “bring” → “brings”

Fixed

P25L9-11: “The HIRHam5 ... below the mean respectively” → This sentence is long and difficult to read due to the lack of commas.

Fixed to read:

“The HIRHAM5 0.11◦and MARv3.10numbers are almost exactly the same at 2452 Gt and 2445 Gt respectively around 150 Gt above the mean. MetUM, like COSMO-CLM2, is much lower at about 10138 Gt and 368 Gt below the mean respectively.”

P25L17-20: The authors address the period of the “1990s and 2000s” for SMB trend, but since SMB is so highly variable, can you really say that this is significant/important?

This is actually one of our main points, that it’s almost meaningless to suggest significant trends over short periods given the large variability which Figures 7-9 clearly show. As all reviewers have a similar comment here we have added a section in the discussion where we explicitly state this.

“Unlike previous studies, we detect no obvious strong trend in the modelled SMB in any of the models or in the driving ERA-Interim model. Shorter periods within the time series appear to have quite strong trends, for example a steady declining trend is apparent through the 1990s and 2000s but appears to have reversed since 2014. Our results suggest that strong interannual and decadal variability makes the identification of meaningful trends over short periods very difficult, Distinguishing noise from signal will be challenging in coming decades and this also emphasises the importance of long time series of observations.”

P26L16: “west Antarctica” and “Antarctic peninsula” → “West Antarctica” and “Antarctic Peninsula”

Fixed

P27L29: “bee” → “been”

Fixed

Figure Comments:

Figure 1: These Taylor diagrams are a very interesting way to convey information, but many readers will have never seen something like this before. It will be important to better clarify the metrics conveyed by the figure. For instance, we are unsure what the curved lines (i.e. ranging 1.60 to 13.50 in the left panel) are supposed to represent.

We have expanded the explanation of the Taylor diagrams in the caption and the analysis of the results shown here both and in the Results section as below

From:
“Taylor diagrams showing model performance compared to daily observations of surface pressure, near-surface temperature and observed wind speeds as well as the bias statistics for each model”

To:

“We use Taylor diagrams to show model performance compared to daily observations of a) surface pressure, b) near-surface temperature, and c) observed wind speeds. The horizontal and vertical axes represent the standard deviation, the

dashed line in bold shows the standard deviation of the observations. The Taylor plots also show the correlation which is measured by the angle with the x-axis. Finally, the centered root mean square error (CRMSE) is represented by the curved lines in light grey. CRMSE is equivalent to the RMSE but systematic biases are removed by subtracting the mean observation and mean modelled values from each value. A perfect model would be in the same place as the observations (black star, correlation of 1, same standard deviation, and zero CRMSE). Similarly, the further away a model is from the observations, the worse it is. Mean biases and observation mean are also indicated. The units of standard deviation, CRMSE, mean bias and mean of the observations are the same (hPa for surface pressure, K for near-surface temperature, and m/s for wind speed)."

NB: $cRMSE = \sqrt{1/n\Sigma}\,[(m_i - \overline{m}) - (o_i - \overline{o})]$

Figure 2: Could you perhaps also include a table of correlation and/or bias for each model?

As the paper is already very long and we have also added substantial new material in response to the reviewers comments we don't want to add further tables or figures unless absolutely necessary.

Figure 4: The same comment as figure 3, but with the color bars

There appears to be a comment about Figure 3 missing that makes this comment difficult to answer - we have however revised Figure 4 to make it easier to read along the lines suggested by the other reviewers

Figure 5: Is this meant to be rotated? Also, increase the font size again.

This figure has been deliberately rotated to make it easier to fit on the page. We have also increased the font size.

Figure 6: Increase label size

The labels have been increased and the plot has been restructured and enlarged to enhance readability.