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. See below:

p2, line 25 “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 precipitation and ablation on grounded parts of the ice sheet is. As the models used in this study by and large do not distinguish between grounded and floating ice in their ice masksm in this paper, when we refer to SMB over an area we include ice shelves unless otherwise specifically noted."

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.

P27, L17 "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 and ice dynamic 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 (see first comment 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:

P12, L4 “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.

P13 L2 “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 to “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

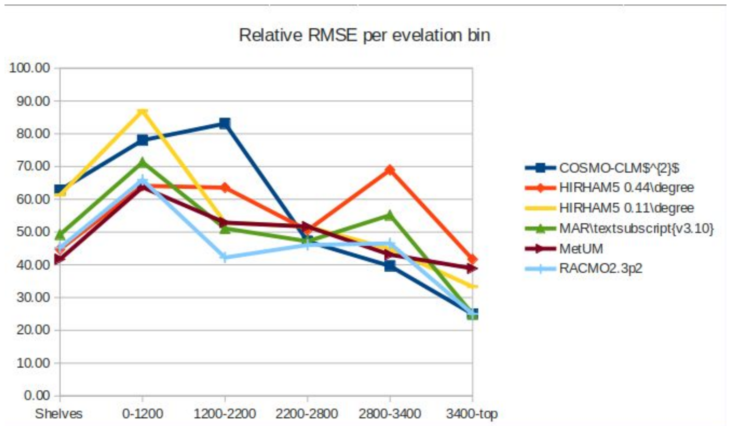
P15 L9 “The blowing snow module included in RACMO2.3p2 may explain the lower bias in results between 0 and 1200 and especially 1200 and 2200 m (Table \ref{tab:samba_smb} b and c), compared to the other models.”

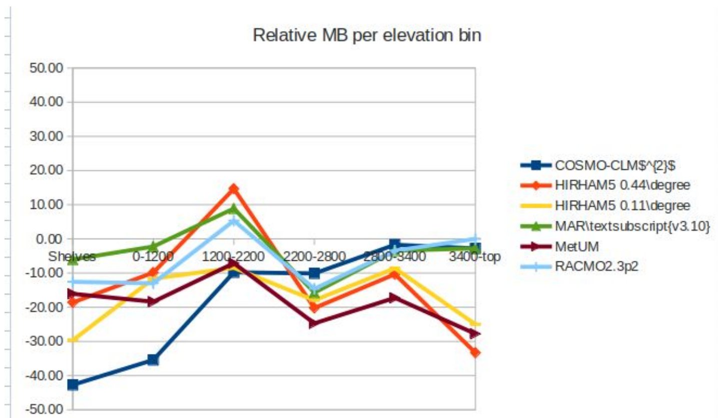
P16L2: “...we here show...” → “...here we show...”

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.

"Table 3.Integrated mean annual SMB for the six models used in this study, for the period 1980 to 2010 except for COSMO-CLM2 where the period was 1987 to 2010. Three older model versions, ensemble mean and standard deviation as shown in Figure 5. All calculations done on the original grid of the individual models using a common set of drainage basins and ice mask defined by IMBIE2 Shepherd et al. (2018b). The ensemble mean was calculated by transforming all models to the RACMO2.3p2 grid. GIS denotes grounded ice sheet, IS denotes ice shelves and ToTIS denotes the full Antarctic ice sheet including ice shelves"

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:

P27 L9 "The HIRHAM5 0.11◦ and MARv3.10 numbers 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 138 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.

P28 L4 "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 in the Results section as below

"Figure 1: Taylor diagrams showing model performance compared to daily observations of surface pressure (a), near-surface temperature (b), and observed wind speeds (c).The horizontal and vertical axes represent the standard deviation, the dashed line in bold shows the standard deviation of the observations . The Taylor plot also shows the correlation which is measured by the angle with the x-axis. Finally, the CRMSE is represented by the curved lines in light grey. A perfect model should therefore 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)."

P11L6 "In figure \ref{fig:taylor} we show Taylor diagrams for pressure, temperature and wind velocities. Taylor diagrams offer an efficient way to assess model skill by comparing the Pearson correlation coefficient, the centred root mean square error (RMSE) and the standard deviation of the modelled output with the observed values. 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 should be in the same place as the observations (black star, correlation of 1, same standard deviation, and zero CRMSE). The further away a model is from the observations, the worse it matches the observed weather. Mean biases and observation mean are also indicated. In this case modelled values closest to the dashed line have a more correct representation of the standard deviation and the closer to the black reference star the closer the model correlates to the observations values. We also list the bias below the diagrams."

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 2 is not a central figure for understanding the paper and we have included Table 2 with substantial statistics relating to the SMB.

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 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.

Reviewer 2

General Comments:

This paper presents results from a first intercomparison of polar regional climate models (RCM) applied in the Antarctic Ice Sheet (AIS). The model performances were compared and assessed in terms of surface pressure, near-surface air temperature, near-surface wind speed, surface temperature, and surface mass balance (SMB) of the AIS. The models that participate in this intercomparison project are COSMO-CLM2, HIRHAM5, MetUM, MAR, and RACMO. For some models, results from different versions are provided additionally.

We thank the reviewer for their thoughtful comments and have addressed in detail the points they raise below.

My first honest impression after reading through this manuscript is that the current title "What is the Surface Mass Balance of Antarctica?" is a bit misleading, because meltwater runoff is not considered in the most participating models except for MAR and RACMO. It is true that a contribution by runoff to the changes in the present- day AIS SMB is relatively small than contributions from precipitation and sublimation/evaporation. But, runoff in the present-day AIS already cannot be neglected as presented by several studies cited in this manuscript. In the future in a warming world, the contribution by runoff to the changes in AIS SMB will become much higher almost certainly as pointed out by the authors (P. 2, L. 18 ~ 19). Therefore, this reviewer expected that all the models calculated runoff in the present study, and as a result, I was a bit disappointed when I found the relevant description in Sect. 2.2.1.

Related to the point indicated above, the intercomparison procedure for SMB sounds a bit inadequate to me, because the authors employ different definitions of SMB (Sect. 2.2.1). If the authors focus intercomparisons only for precipitation and sublimation/evaporation (in addition to the three surface meteorological properties as well as the surface temperature), it makes sense and highlights key differences in model physics employed by these participating models more clearly. This reviewer recommends the authors to reconsider the title of this manuscript: maybe something like “intercomparison of Antarctic ice sheet surface meteorological conditions simulated by five different regional climate models” would be appropriate. However, the intercomparison of RCMs performed in the AIS itself is a considerable new challenge, so provides the latest comprehensive information related to these RCMs, which is very informative for readers certainly, so deserved to be published

In this paper we focus on the Surface Mass Budget of Antarctica and the uncertainties introduced by using different regional climate models to calculate it, even when those models are forced by the same global model. Overall, precipitation dominates SMB to such an extent at the present day that even subtracting all runoff from the models that calculate melt and refreezing leaves the overall SMB virtually unchanged. The difference is negligible even on a basin scale. We agree that melt is likely to become more important in the future, but at the present day melt and runoff are only observed at a few very specific locations.

However, while all of the models simulate melt, they have varying degrees of complexity to calculate refreezing so purely to simplify the comparison here we focus on precipitation, evaporation and sublimation terms. Two of the models include sublimation from blowing snow subroutines, and as the physical parameterisations have been developed with these processes in the model, we have also used the drifting snow schemes in the results.

We absolutely agree that there are important questions around melt extent that have important implications for future SMB projections and we plan to extend this study with a detailed look at how the models simulate melt and runoff in a paper currently in preparation. For now, we have however added in the introduction and discussion sections more detail on how SMB is computed (see response to reviewer 1) and the processes that can be and are included and address the issue the reviewer raises in more details.

We note also that we do compare modelled SMB with measured SMB from stakes and from other studies such as the IMBIE study (see reply to reviewer 1 here also) and we think therefore it is justified to keep the title as it is.

Specific comments (major)

P. 2, L. 7 ~ 8: What kind of measurements do the authors think here (observational campaigns)? Maybe it is not necessary to indicate explicitly here; however, please suggest something at least in the discussion and/or conclusion sections.

Our results suggest that in particular stake measurements of SMB are crucial. These need to cover locations where there are very few recent observations and where there are large disagreements between the models. We propose adding a new figure (see below) to the supplementary materials that show the locations of SMB observations as well as locations of weather stations in order to demonstrate the significant data gaps. We have made this clearer in the conclusions.

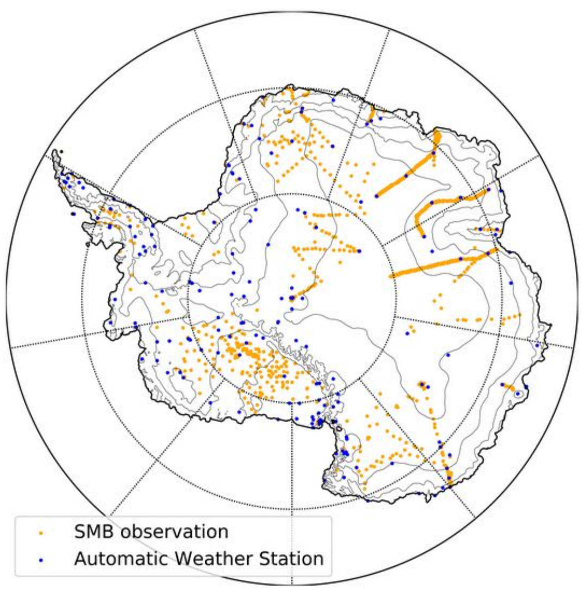


Figure A8. Location of automatic weather stations and SMB observations in Antarctica and used in this study

P. 12, L. 2 ~ 3: What kind of physical mechanisms do the authors think here? Please detail more.

In locations with melt we expect that the lack of refreezing will affect the latent heat release in the snowpack which will in turn affect observed 10m depth temperatures. Furthermore, the diffusion of and conductivity of the surface snow layers is affected by the presence or absence of ice layers and by density which is dependent on densification schemes that this version of HIRHAM and COSMO-CLM2 do not have. We have clarified this to read:

P13, L2 “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 van Wessem et al. (2014). In their study, significant improvements in the RACMO2.3p2 model were obtained by adjustments to the cloud microphysics. Furthermore, the lack of detailed subsurface snow pack schemes including processes such as refreezing (and subsequent latent heat release) and densification also likely has an impact on the temperature bias in HIRHAM5 and MetUM (see also figure 2)”

P. 12, L. 5: How large do the authors think the uncertainties are?

It is very difficult to quantify uncertainties on these observations, particularly wind, as they are made at mostly automatic unstaffed stations and are subject to different biases depending on location from effects such as burial by snow, changes in orientation due to wind and breakdown of sensors among others.

Modified to:
P13, L8. ”This is likely in part due to large uncertainties in the observations especially at unattended stations where burial by snow, changes in orientation and sensor breakdown are more likely. However, the effects of different resolution and differences in turbulent schemes between the models may also be important. In particular the extremely stable boundary layer over most of Antarctica is hard to represent in models particularly at lower resolutions”

P. 12, L. 5 ~ 6: Readers cannot know the difference in turbulent heat schemes, because they are not described in this manuscript.

A detailed description of all of the models turbulent energy schemes is beyond the scope of this paper, we have however added relevant references to these for each model to support our interpretations of model biases in an expanded section 3.1.

P. 24, L. 13 ~ 15: It is interesting to see the model-simulated precipitation integrated over the common ice sheet mask by COSMO-CLM2 tends to be lower than that by the parent data ERA. It is because precipitation in a dynamically downscaled data is higher than precipitation in its parent data in general. Please discuss.

Our analysis suggests that the COSMO-CLM2 model used in this simulation has indeed a dry bias compared to the other models and the reasons behind this are the subject of ongoing work. The bias was first identified by Souverijns et al., 2019 and seems in part to be a consequence of a particularly dry bias on the coast, especially in the peninsular and west Antarctica but there is conversely an overestimate of precipitation in the interior. We have added these details to the paper.

P30, L7. “The driest model COSMO-CLM2 underestimates SMB close to the coast, a region very relevant for total ice sheet mass balance. This is due to an overestimated sublimation amplified by an underestimated snowfall rate close to the coast. High values for the sublimation originate from an underestimated albedo due to aging of the snow that occurs too fast in the model (Souverijns et al., 2019). The low values for the snowfall rate is likely related to cloud microphysics, namely a too slow conversion of ice to snow or a too slow deposition of water vapor on the solid hydrometeors. Currently efforts are ongoing to improve the coastal SMB performance in COSMO-CLM2.”

P. 29, L. 27: Please suggest what kind of measurements do the authors think necessary in the “observational campaigns”?

Given the importance of precipitation and snow processes to the SMB in Antarctica, stake and radar measurements, supplemented with firn cores are a clear priority. New observations should focus where possible in regions where there is a lack of measurements, but also where there is strong disagreement between models, as identified in the Results section. We have added these points explicitly to the conclusions.

P32, L6 “In particular, we argue that given the importance of precipitation for SMB, new observational programmes are needed that focus on accumulation and snow processes, e.g. stakes, firn cores and radar. Furthermore, focusing new observations in regions (see for example,A8) where there is both a lack of current data and strong disagreement between models will be valuable for understanding climate in Antarctica”

Specific comments (minor)

P. 4, L. 13 ~ 18: Is it OK to understand MAR 3.6 is older than MAR v3.10? If yes, it is a bit confusing isn’t it?

MAR v.3.10 is the more recent version of MAR

P. 8, L. 6: What do the authors mean by “cloud physics”? To me, it is difficult to understand why “cloud physics” is resolved better in nudged models.

As nudged models better represent cyclones when compared to observations, the presence or absence of clouds is more likely to be closer to observed, however we agree saying “cloud physics” is not quite technically correct so we have modified to “the presence of clouds”

P. 12, L. 24 ~ 25: It is not clear why the authors think so. Please explain more.

We have added the reference to van Wessem et al., 2014 and updated the text to:

P13, L2 “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 van Wessem et al. (2014). In their study, significant improvements in the RACMO2.3p2 model were obtained by adjustments to the cloud microphysics. Furthermore, the lack of detailed subsurface snow pack schemes including processes such as refreezing (and subsequent latent heat release) and densification also likely has an impact on the temperature bias in HIRHAM5 and MetUM (see also figure 2)”

P. 12, L. 31: “For the warmer coastal regions”: From which data can we see this argument?

As this sentence is a bity unclear we have changed to

P.14, L3 “For the lower elevation, mostly coastal regions most models have a cold bias.”

Figure 2: This figure is a bit difficult to see. Please provide a table indicating ME, RMSE, and R2.

We have included table 2.

P. 23, L. 4 ~ 7: What is an interesting point here? I don’t think the lower panel of Fig. 8 is necessary; however, if the authors think it is necessary, please discuss more about the figure. Maybe, inter-annual variations of these model results should be discussed more.

We have added more discussion in the section following Figure 8 to make explicit the finding that while all the models have the same anomaly when compared to their own mean, the sign of the anomaly compared to the ERA-Interim value can be different. Since the most highly constrained models show the lowest anomaly compare to ERA-Interim, we suggest that most of the variation is related to internal variability (weather) within the domain.

P25, Line 1 “Figure 8 emphasises the large variability in SMB on an annual to decadal scale by plotting the variation from the mean for each model and the variation from ERA-Interim for each model. This shows that while all the models have more or less the same anomaly when compared to their own mean, the sign of the anomaly compared to the ERA-Interim value can be different. Since the most highly constrained models show the lowest anomaly compared to ERA-Interim, we suggest that most of the variation is related to internal variability (weather) within the domain.Both HIRHAM5 0.11\degree and 0.44\degree shows the highest values of variability, probably due to the unconstrained nature of the runs, but in different years different models show higher variability than the others. The lower panel in Figure \ref{fig_SMB_trend} shows that MetUM is by far the closest to the driving model with much less variability than the others, HIRHAM5 again shows the highest difference compared to the driving model but from year to year the model showing maximum difference varies and there appears to be no systematic pattern as to whether or not modelled SMB is higher or lower than the ERA-Interim reanalysis when quantified on the common mask and over the whole of Antarctica. The implication is that while the driving model controls broad scale pattern of SMB, the downscaling model adds its own weather variability to the broad scale pattern. The variability, or weather noise is unsurprisingly, largest in un-nudged models. The effect of this noise on ice sheet dynamics may be small overall but as for example, Mikkelsen et al. (2018) show, small variations in SMB can have a non-negligible impact on ice sheet dynamics.”

Technical corrections

P. 2, L. 1: “compar”: typo

P. 2, L. 11: “a potentially important potential contributor” -> “an important potential contributor”?

Fixed

P. 2, L. 22: surface mass balance -> SMB; Note this term is already defined.

Fixed

P. 3, L. 22 ~ 25: This sentence especially after “and better understand drive sea level rise . . .” is a bit difficult to understand. Please reformulate it.

Fixed

P. 5, L. 11: “regional mesoscale model e.g.” -> “regional mesoscale model as presented by e.g.”?

Fixed

P. 7, L. 14: Define SU_{ds} and ER_{ds}.

Fixed

P.9, L. 2: “GrIS”: typo, right?

Fixed

C4P. 10, L. 27: “assessing” -> “assess”

Fixed

P. 12, L. 7: Indicate the publication year for Zentek and Heinemann.

Fixed

P. 12, L. 25: “downwelling longwave and surface albedo” -> “downwelling longwave radiation flux and surface albedo”

Fixed

P. 14, L. 18: “HIRHAM5.011”: typo

Fixed

Review of Mottram et al., What is the Surface Mass Balance of Antarctica? An Inter-comparison of Regional Climate Model EstimatesSummaryThe authors present an intercomparison exercise of five different regional climate model surface mass balance estimates, as well as the near surface climate, over Antarctica. The authors find a large spread in total SMB (1961 to 2519 Gt year-1), which largely stems from differences in West Antarctica and the Antarctic Peninsula.Variability is quite consistent between models, which is unsurprisingly since they are all forced by ERA-Interim, but the trends differ in sign and magnitude and are quite sensitive to the time period selected.
Also, not surprisingly, the nudged models simulate the near-surface climate better as they are not allowed to deviate as substantially fromERA-Interim as the un-nudged models. Finally, the authors discover that the biases are typically consistent between models. The paper presents a significant amount of work but still requires improvements.

We thank the reviewer for their careful reading and thoughtful comments, which have led to significant improvements in the manuscript.. We address the specific points in the text below.

First, the manuscript has numerous mistakes throughout and needs refinement of the language in several places (see Minor Comments).

This point has also been addressed by the other reviewers and the whole manuscript has been thoroughly proof-read and made simpler to read and more consistent in language and structure.

More importantly, there are several major issues with the analysis that need to be addressed to improve the scientific rigor of the paper.

Major Comments

1. Throughout the manuscript, it is not clear what time periods are being used. There Is the common model interval, climatological mean, etc. The authors need to be very clear throughout the paper because it often seems that different intervals are being confused in nomenclature. It's not clear to me why the common reference intervals are not the common period between all models: 1987 – 2015. Throughout the paper sometimes its 1980-2010, 1987-2015, and 1987-2018. I recommend using the same interval through to avoid confusion. If the authors have a reason to use different intervals, then please make it clear what interval is being used. It is additionally unclear why 1980-2010 is representative of the climatological period, please explain.

We realise that it is confusing that the models were run for slightly different periods and this also makes comparisons between them more complicated. We have added a paragraph explaining the different periods to the Methods section and clarified all the way through the paper which periods are being used during the results and discussion when relevant.

P4. L13 "Unfortunately, as we are constrained to using existing simulations, the models cover slightly differing periods (see \ref{tab:model_overview} for details). We have therefore defined a common 30 year climatological period of 1980 to 2010 for all models to simplify the integrated mass budget comparison, except for COSMO-CLM^{2} where the period covers 1987 to 2010. Figures that show time series of data show the full period relevant for each model."

2. Similarly, there is no discussion of significance for the statistics presented. There are claims within the text that certain models perform better than, but without significance levels, these claims lack strength and are more speculative.

Trends are discussed at both long (1987-2015) and short (decadal) time intervals, but the significance is never discussed. I would caution the authors' descriptions of trends, especially at short time scales, since it is very hard to observe a significant trend in SMB since its highly variable year to year.

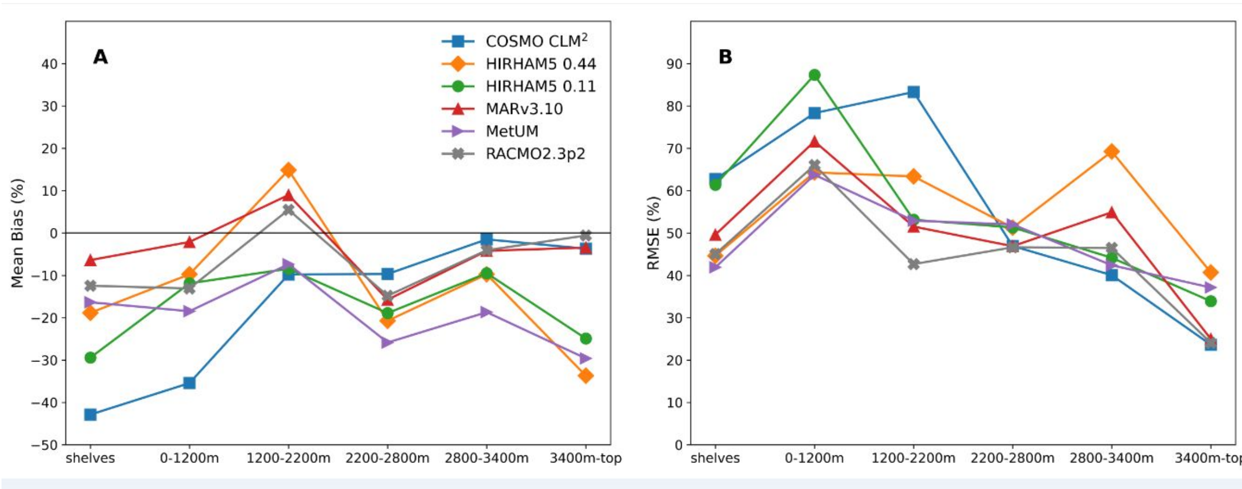
We absolutely agree that detecting a significant trend in SMB is almost impossible and in fact that was one of the points of Figures 7-9. However, given all reviewers comments we have clearly not described this well enough. We have added a paragraph making this point explicitly and setting the SMB in context.

P28, L4 “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 on first look 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.”

Furthermore, because this is an intercomparison paper, it’s important for the authors to be very clear concerning the metrics of how they conclude one model outperforms the other. Is it RMSE? R2? Bias? And what is the threshold? Is an RMSE of 93 better than 97? What if one model performs differently at different elevation bands?
I did not find the argument compelling that the models tuned to specific Antarctic conditions outperformed the others because there was not a clear frame-work for comparison. The authors need to make clear the evaluation metrics and how they evaluate model performance, which will require more detailed statistical analysis throughout.

Model means are compared, but its not clear if the paper considers even a simple statistic of the standard error of the mean. The Student’s t-test can be used to evaluate whether the means are different. Please be transparent with the limitations of the analysis and provide meaningful significance tests on all of the comparisons,otherwise the conclusions are speculative rather than significant.

This is a very important point and in part one of the drivers for this paper. We do not attempt to rank the models because it is clear from our results that on different measures, (bias, RMSE etc) the different models perform quite differently for different variables (both SMB and the meteorological variables). There is also a spatial component as the reviewer points out with different biases apparent at different elevation bands and in different locations. This means that most likely different models should be used for different purposes. It is also an aim of this paper however, to give clarity on exactly how the models compare, for which reason we give extensive statistics in figure 1 and table 2, which we have also expanded to include the mean value of SMB observations for the elevation bands as requested by reviewer 1. As the paper is already very long, we have added 2 new figures and associated statistics in the supplementary section showing how the relative RMSE and mean bias compares between models for the different elevation bands. These for instance show that at high elevation COSMO_CLM, MAR and RACMO better represent SMB but HIRHAM, MetUM and COSMO-CLM have a lower mean bias in the middle elevations and MAR, HIRHAM and RACMO have a lower mean bias in the lowest elevations. In addition we have added some extra discussion comparing the different statistical methods and their use in evaluating the models in the discussion of table 2.



3. All of the RCMs presented are forced by the ERA-Interim reanalysis product. I find it concerning that there is no discussion of the role of using a single reanalysis to force all of the RCMs. Thus, this is not a definitive evaluation of the full range of possibilities in SMB, but rather a range due to RCM differences alone. There should be more discussion about how there would be additional spread due to varying choice of forcing; specifically, what is the impact of comparing models that are all forced by the same reanalysis. I think the paper needs to tone done the claims about the work providing the "likely range of SMB" in the first sentence of section 4.1, as it is more the likely range of RCMs forced by ERA-Interim. Basically, this explores the range in RCM space, but not reanalysis forcing space.

The point of this study is to determine the RCM uncertainty space rather than different boundary conditions. We specifically excluded models that ran different reanalyses as we would like to determine how different models compare with each other. However, having said that, analysis by Agosta et al., 2019 used different reanalyses to force the same model and found that the results were quite similar. We have added two extra sentences and this reference stating this in the methods section.

P3, L31. "All models were forced on the lateral boundaries with the ERA-Interim climate reanalysis (Dee et al., 2011) but downscaling used different grids, over slightly different domains and at different resolutions with slightly different ice masks used in the different model versions (see A1 in the appendix). Simulations with MAR forced by different reanalyses (e.g. Agosta et al., 2019) found that results were rather similar to ERA-Interim. However, in order to exclude additional variability potentially introduced by using different boundary forcings, we chose to use a single common reanalysis only”

4. It’s obviously quite a challenge to compare these models, which have differing levels of complexity. But it seems that the comparison would be better suited by comparing all the variables consistent between models (Precip-Evap-Subl). Otherwise, an inter-comparison doesn’t shed much light on direct model to model differences. In fact, it appears that the authors could investigate whether these Antarctic specific physics actually provide improvement, which would be of great interest to the community. Therefore, the paper should do an ideal comparison of all 5 models with common variables (P – E – S) and evaluate performance. Then evaluate the models with extra physics (RACMO/MAR) to see if and how much model performance improves. Otherwise, it is difficult to untangle whether those additional processes provide any more information.

We have also addressed this point in our response to the second reviewer. Melt is likely to become more important in the future, but at the present day melt and runoff are only observed at a few very specific locations. However, while all of the models simulate melt, they have varying degrees of complexity to calculate refreezing so purely to simplify the comparison here we focus only on the precipitation, evaporation and sublimation terms. Then, RACMO and MAR include sublimation from blowing snow subroutines, while it would be ideal to separate these out, the physical parameterisations have been developed and tuned with these processes in the model, so it is difficult to remove them without negatively and unfairly affecting the results- we have therefore also used the sublimation from snow schemes in calculating SMB. We note that both RACMO and MAR groups have published articles demonstrating the improvement from the enhanced snow schemes (Van Wessem et al and Agosta et al., 2019). We have added extra text to make this point in the description of the SMB in the methods section.
P4, L26 “As the RACMO and MAR models have been developed to include the wind blown snow sublimation terms, they cannot easily be removed without retuning the models, and for this reason we have opted to include these within the SMB calculation for these two models.”

5. The manuscript needs to justify the use of SMB observations starting in 1950.There are regions of strong trends in snow accumulation that might end up biasingthe comparison. If the issue relates mainly to reducing the number, the authors could present a comparison of only coincident SMB observations with the data, but then also provide the more liberal comparison as it currently exists in the text.

Unfortunately there are relatively few observations in Antarctica and including only those that were taken during the period of the simulations would make the model - observation comparison less robust. Including observations taken from 1950 increases the number of observations available to 923 comparisons from 469. More importantly, while the total number still sounds substantial, the main benefit is in fact in spatial representativeness. The 1987-2015 observations cover only a very small part of Antarctica. We discuss this problem in section 2.3.3 but we have expanded the discussion and we propose to include a new figure showing the spread of observations by date in the supplementary material. As we point out in the discussion and conclusions, the difficulty is also that the places where the models disagree most are also the areas with the sparsest observations of SMB.
We have mitigated the problem of unrepresentativeness as much as possible by for example excluding observations before 1987 that cover too short a period (less than 5 years) in order to keep only observations representative of a mean climate. However, we are not immune to introducing biases because these observations include biases arising from trends that the models cannot represent, nonetheless the comparison is more robust if it represents a larger area. We have added these points in the expanded methods section.

P10, L14 “To evaluate the models, we selected observations of SMB belonging to the common ice mask and for which the measurement period began after 1950 to 2018. These conditions reduced the total number of observations used in the comparison to 3671. We used the observations between 1950 and 1987, or 2015 and 2018 that are not fully included in the common modelling period of 1987 to 2015, for evaluation only if they covered more than 5 years. These 1849 SMB observations are compared to modelled values averaged over the common modelling period in order to compute a climatological mean while we averaged modelled SMB values over the exactly same period for the observations between 1987 to 2015 (1822 observations). Since the models have different resolutions and grids, we do not directly compare the modelled SMB values to the observations.”

6. Finally, the paper needs to discuss the impacts of its findings. With the present day mass loss from Antarctica on the order of 100 Gt per year, this is quite concerning finding the differences in SMB from RCM choice alone are several hundred Gt per year.Please contextualize the findings in regard to how we can measure the mass balance of the ice sheets.

We have added a paragraph in response to reviewer 1’s comments along these lines, where we relate the modelled SMB to the latest analysis of Antarctic mass budget from altimetry and GRACE observations:

P27, L17 “It is interesting to compare our results with those used in the IMBIE study of Antarctic mass budget (Shepherd et al., 2018). 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. Nevertheless it is therefore also important to consider the wide uncertainties in both observations and the likely biases in models discussed in this paper, in assessing the contribution to sea level rise from Antarctica"

Minor Comments

Several model names and versions are discussed before they are described, which makes it quite hard to follow. Please reorder the sections to ease.

For instance, section2.1 and the end of Section 1 mention several models and different version, but there is no description, so it's hard for the reader to follow. It would also be appropriate to cite the papers that refer to these model versions.

We have reordered and expanded this whole section to make it easier to read and to follow which models are under discussion and how they relate to each other and to give further details on the different schemes and parameterisations.

P1, Line 7: Is this for grounded ice only? Does it include islands and ice shelves?

This was for the whole We have added a section discussing SMB and discharge and the differences between grounded ice sheet and ice shelves to clarify our ice mask definitions.

P2, L25 "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 precipitation 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 mask"

P1, Line 7-8: What do the values after the ± represent? The standard deviation of all the models?

The values represent the standard deviation of all models and this has been clarified in the paper.

P1, Line 10-11: Why is 1980-2010 chosen as the climatological period? Later in section 2.3.3, it appears that 1987-2015 is the common modeling period that is used to"compute a climatological mean" (P10, Line 12). Please rectify.

We realise that it is confusing that the models were run for slightly different periods and this also makes comparisons between them more complicated. We have added a paragraph explaining the different periods to the Methods section and clarified all the way through the paper which periods are being used during the results and discussion (see reply to comment 1)

"Unfortunately, as we are constrained to using existing simulations, the models cover slightly differing periods (see \ref{tab:model_overview} for details). We have therefore defined a common 30 year climatological period of 1980 to 2010 for all models to simplify the integrated mass budget comparison, except for COSMO-CLM^{2} where the period covers 1987 to 2010. Figures that show time series of data show the full period relevant for each model."

P2, Line 1: change "compar" to "compare"
Fixed

P2, Line 11: remove either "potentially" or "potential" since its repetitive
Fixed

P2, Line 13: add "and" after "2002,"
Fixed

P3, Line 16: remove the comma after "published"
Fixed

P3, Line 23: remove "drive"
Fixed

P4, Line 1: please describe what a "reinitialized hindcast" is
A reinitialised hindcast is a model run in weather forecast mode that is reinitialised by observations every 48 hours. We have added a line to explain this.

P4, Lines 5-7: While this is true, it might have a limit. See Lenaerts et al., 2018, which shows that often the snow is not dumped in the proper place when moving from 27 km to 5.5 km. Please add a sentence clarifying this.

This is a good point and in fact one of the reasons we have undertaken this comparison. We have added a line mentioning this and also included this point in the discussion section P4, L9 "Lucas-Picher et al. (2012); Lenaerts et al. (2012b); Franco et al. (2012); van Wessem et al. (2018) among others have found that a higher spatial model resolution gives more physically plausible results, especially with respect to precipitation processes in areas with steep terrain. However, there is also evidence that moving to high resolution (~5.5km) can lead to precipitation falling in the wrong place due to e.g. upslope effects (e.g. Lenaerts et al., 2018; Schmidt et al., 2017)."

P4, Line 8: add "is used" after "ensemble mean"
Fixed

P5, Line 6: change "developed in" to "developed for"
Fixed

P5, Line 18: the end of this sentence needs to be reworded
Edited for clarity
P5, L26"Although the mesoscale version includes a multi-layer snowscheme (Walters et al., 2019), in these simulations we used a simplified single-layer scheme with for example, no refreezing (Cox et al., 1999). SMB was calculated based on output precipitation and sublimation and evaporation."

P6, Line 4: Do you mean "processes" not "process"?
Fixed
P6, Line 22: change "includes no" to "does not include"
Fixed

Table 1: What does SMB scheme mean?
SMB scheme refers to whether or not the regional climate model has been modified to take into account atmosphere - ice sheet interactions, or if it is run in a standard mode without explicitly calculating SMB.

P7, Line 10: change "schemes" to "scheme"
Fixed

P10, Line 10: add "that are" after "2015 and 2018"
Fixed

P10, Lines 14-20 need clarification
P10, L21 "As in Kittel et al. (2018) and Agosta et al. (2019), we compute modelled and observed SMB values in 2 steps. Firstly, the original resolution modelled SMB values were interpolated, as for AWS observations, to the observation location using a four-nearest inverse-distance-weighted method. Secondly, all the interpolated SMB values contained in the same grid cell from the common ice mask were averaged as well as the observations for finally creating 923 comparison pairs. This leads to a fair comparison for each model that takes into account the benefit of using a higher resolution for a specific model and removing the very high spatial variability of the observations that cannot be reproduced by the models."

P12, Lines 4-7: This sentence is very long and needs to be split in two.

Fixed

P12, Line13-14: Please reword the sentence as its confusing.
Fixed

P13, Line 6: Remove "In"
Fixed

Figure 3. Please add the statistics to these plots (RMSE, etc.). Also, its very difficult to distinguish the colors here. Maybe large dots would help.
The statistics in these plots are given in Table 2 for clarity, we also present the models separately in the supplement and (in response to a request by reviewer 1) we have expanded this analysis to include a relative RMSE plot that we include in the supplementary material.

Figure 4. This figure should be much bigger. It's very hard to see the colors. Also, in the caption there are "a", "b", etc., but they do not exist on the plots.
Figure 4 has been revised and enlarged to make it easier to read, we have removed the superfluous letters from the caption.

P18, Lines 9-11: are these values consistent with what is listed in the abstract?
We have revised the text of the manuscript to be more clear about when values are consistent with a given dataset. The abstract has been completely rewritten to simplify it and summarise the conclusions further

Figure 5. This needs to be in landscape orientation. The numbers are much too small to read.
We have edited this figure to make it larger and the labels clearer, and we have kept it in landscape orientation as it makes it easier to read and interpret the figure.

P20, Lines 5-6: What does "much clearer mean," please clarify
The topography in the regions noted in the text have a substantial influence on the modelled SMB, this allows physical features such as the Transantarctic mountains to be picked out in the SMB maps. We have updated the sentence to reflect this.

P22. L 4 "The figure shows quite substantial agreement between models over large areas of Antarctica but also some considerable local variability. Features such as the Transantarctic mountains and the rugged coastal topography in West Antarctica both of which substantially influence local weather patterns are picked out in the spatial pattern of the SMB. These features are more clearly delineated in the higher resolution runs."

Figure 6. Again, these plots are too small, and the numbers are nearly impossible to read.
Figure 6 has been made bigger and restructured for ease of reading with larger font on the labels.

P22, Line 2: remove "below" There should be significance values associated with the trends. It looks like none would be statistically significant and thus are effectively no different than zero.

We fully agree about the lack of significance of the trends and have added a paragraph discussing this point.

P24, Line 6: remove "very"

Fixed

P24, Line 9: change "bring" to "brings"

Fixed

P25, Line 7: Should the interval be 1987-2015?

Yes, good catch! Here we show results for the common model period 1987-2015

P27, Line 29-30: Do your results actually support "Models that have not undergone specific adjustments for Antarctica clearly represent the SMB in Antarctica more poorly".Look at the RMSE value in Table 2, it looks like sometimes they perform better. Please Clarify.

As discussed above, assessing how well the models perform is complex. The new figures discussed above help to clarify this somewhat but we have modified the text here to take into account the spatial and process variability.

P30, L4. "Models that have not undergone specific adjustments for Antarctica clearly represent the SMB in Antarctica more poorly than those that have been adjusted in some regions . However table 2 shows this is not unambiguous as in some elevation bands the unmodified models have lower bias and RMSE (see section 3.3)."

P 28, Line 11-12: please give values in Gt of these processes to show that they are effectively negligible

Fixed

P 28, Line 20: add "fig." before "7"

Fixed

P28, Lines 20-21: the sentence needs to be improved.

Modified to:

P30, L30 "The higher resolution version adds value with higher spatial variability that should better capture local topography and associated weather phenomena. This is especially important in areas of high relief such as in the coastal areas and around the Transantarctic Mountains. These are also the areas where models vary from each other and the ensemble mean the most. While there are very few observations to confirm the better performance on a local scale, the pattern of SMB suggests that the high relief rugged topography is better captured in HIRHAM5 0.11\degree than

0.44\degree. However, the higher resolution model is not only more computationally expensive, in a simulation where there is no nudging, like here, the larger number of grid points gives increased degrees of freedom for the model to evolve freely and thus introduces more internal variability.

P29, Line 2: replace "mod-latitudes" with "mid-latitudes"
Fixed

What is the Surface Mass Balance of Antarctica? An Intercomparison of Regional Climate Model Estimates

The Polar CORDEX Community[1], Ruth Mottram[1], Nicolaj Hansen[1,2], Christoph Kittel[3], Melchior van Wessem[4], Cécile Agosta[5], Charles Amory[3], Fredrik Boberg[1], Willem Jan van de Berg[5], Xavier Fettweis[3], Alexandra Gossart[6], Nicole P.M. van Lipzig[6], Erik van Meijgaard[7], Andrew Orr[8], Tony Phillips[8], Stuart Webster[9], Sebastian B. Simonsen[2], and Niels Souverijns[6,10]

[1]DMI, Lyngbyvej 100, Copenhagen, 2100, Denmark
[2]DTU-Space, Kongens Lyngby, Denmark
[3]Laboratory of Climatology, Department of Geography, SPHERES, University of Liège, Liège, Belgium
[4]Institute for Marine and Atmospheric Research Utrecht, Utrecht University, Utrecht, the Netherlands
[5]Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[6]Department of Earth and Environmental Sciences, KU Leuven, Belgium
[7]Royal Netherlands Meteorological Institute, De Bilt, the Netherlands
[8]British Antarctic Survey, High Cross, Madingley Road, Cambridge, UK
[9]UK Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, UK
[10]Unit Remote Sensing and Earth Observation Processes, Flemish Institute for Technological Research (VITO), Mol, Belgium

Correspondence: Ruth Mottram (rum@dmi.dk)

Abstract. ~~Antarctic ice sheet mass loss is currently equivalent to around 1 mm year^{-1} of global mean sea level rise. Most mass is lost due to sub-ice shelf melting and calving of icebergs. Ice sheet models of the Antarctic ice sheet have thus largely concentrated on parameterising sub-shelf and calving processes. However, surface mass balance (SMB) is also of crucial importance in controlling the stability and evolution of the vast Antarctic ice sheet. In this paper we~~ We compare the performance of five different regional climate models (COSMO-CLM^2, HIRHAM5, MAR3.10, MetUM and RACMO2.3p2) forced by ERA-Interim reanalysis, in simulating the near surface climate and SMB of Antarctica. ~~Our results show that , when regional climate models (RCMs) are forced by the ERA-Interim reanalysis, the integrated Antarctic ice sheet~~ Evaluation of the models shows that they simulate Antarctic climate well when compared with daily observed temperature and pressure though nudged models perform slightly better than un-nudged models. The ensemble mean annual SMB over the AIS including ice shelves is 2329 ± 94 Gigatonnes (Gt) year^{-1} over the common 1987 to 2015 period covered by all models. However, mean annual SMB is sensitive to the chosen period with large interannual variability. Over a defined 30 year climatological mean period of 1980 to 2010, the ensemble mean is 2486 Gt year^{-1}. However, individual model estimates vary from an annual mean of 1961 ± 70 to 2519 ± 118 Gt year^{-1}. The ~~large differences are mostly explained by different SMB estimates~~ largest spatial differences between model SMB estimates are in West Antarctica and the ~~peninsula~~ Antarctic Peninsula as well as around the Transantarctic mountains. ~~The calculated annual average SMB is very sensitive to the period chosen but over the climatological mean period of 1980 to 2010 the ensemble mean is 2486 Gt year^{-1}. The interannual variability in SMB~~ Interannual variability is consistent between ~~the~~ models and dominated by variability in the driving ERA-Interim reanalysis. ~~The declining trend in~~

~~Antarctic SMB reported in other studies is also very sensitive to period chosen and models disagree on the sign and magnitude of the~~ We find no significant trend in Antarctic SMB~~over the ERA-Interim period~~.

~~Evaluation of models shows that they simulate Antarctic climate well when compared with daily observed temperature (Pearson correlation of 0.85 and higher) and pressure (bias ranges from -0.39 hPa in HIRHAM5 to -6.01 hPa in MAR with a mean of -3.49 hPa over all models) and nudged models, constrained within the domain as well as at lateral boundaries, perform better than un-nudged models.~~ We compare modelled surface mass balance with a large dataset of observations which, though biased by undersampling in some regions, indicates that many of the biases in modelled SMB are common between models. ~~The inclusion of drifting snow schemes improves~~ Drifting-snow schemes improve modelled SMB on ice sheet slopes between 1000 and 2000 m where strong katabatic winds form but other regions where precipitation rates are high lack observations needed for ~~the evaluationof different SMB estimates~~evaluation. Different ice masks have a substantial impact on the integrated total SMB and along with model resolution is ~~therefore~~ factored into our analysis. ~~The majority of the different values for continental SMBare due to differences in modelled precipitation at relatively few grid points in coastal areas. Our analysis suggests that targeting~~ Antarctic ice sheet (AIS) mass loss is currently equivalent to around half a millimetre year^{-1} of global mean sea level rise (Shepherd et al., 2018b) and our results indicate some substantial uncertainty in the surface mass balance (SMB) contribution based on regional climate models. Targeting coastal areas for ~~observational campaigns will be~~ observations is key to improving ~~and refining~~ estimates of the ~~total~~ surface mass balance of Antarctica.

1 Introduction

The Antarctic Ice Sheet (AIS) is the largest body of freshwater on the planet and thus a potentially important ~~potential~~ contributor to global sea level rise as well as a significant part of the climate system contributing freshwater to the ocean and with its high relief significantly affecting atmospheric circulation. Studies by Rignot et al. (2011) and Shepherd et al. (2018a) showed the AIS to have had a net loss since at least ~~2002, current~~ 2002. Current estimates suggest that around 10 % of observed sea level rise since 1993 is from Antarctica, however that rate ~~is~~ of contribution is also increasing (IPCC SROCC Chap. 4 Oppenheimer et al. (2019)). Most ice loss in Antarctica occurs as a result of submarine melting and calving from ice shelves and recent ice dynamics studies (DeConto and Pollard, 2003; Edwards et al., 2019; Sutter et al., 2016; Shepherd et al., 2018a) have shown that there is potential for rapid ice sheet loss owing to ice sheet dynamics that are currently poorly understood, especially in West Antarctica. Ice sheet models of the AIS have thus largely concentrated on parameterising sub-shelf and calving processes. However, surface mass balance (SMB) is also of crucial importance in controlling the stability and evolution of the vast ice sheet. Changes in precipitation and increases in surface melt and runoff will change the mass budget and therefore both ice dynamics and the sea level rise contribution from Antarctica in the future. Moreover there has been disagreement between studies focused on the SMB contribution to the total mass budget of Antarctica and therefore the contribution to sea level rise (Scambos and Shuman, 2016; Zwally et al., 2015), that makes it essential to understand potential biases and uncertainties.

Surface mass budget (also known as surface mass balance or climate mass balance (Cogley et al., 2011)) is the difference between accumulation and ablation at the surface of a glacier. In Antarctica, accumulation is derived primarily from solid precipitation, but on local or regional scales wind-driven processes can have a significant effect on accumulation rates. Surface ablation in Antarctica is primarily a result of ~~evaporation~~ erosion and sublimation due to the high winds and generally dry atmosphere (Scambos et al., 2012; Das et al., 2013; Agosta et al., 2019), although increasing melt rates are documented in some areas (Stokes et al., 2019). In the future, a "greenlandification" of the ice sheet climate ~~with increasing melt and refreezing within the snowpack~~ is projected due to ~~anthropogenically~~ anthropologically induced climate change (Trusel et al., 2018). This will lead to more melt with more refreezing in the snowpack as well as increasing runoff. 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 precipitation and ablation on grounded parts of the ice sheet is. As the models used in this study by and large do not distinguish between grounded and floating ice in their ice masks, in this paper when we refer to SMB over an area we include ice shelves unless otherwise specifically noted.

Currently runoff is a relatively minor contribution (Lenaerts et al., 2019) to mass loss in Antarctica and an increase in snowfall associated with higher saturated vapour pressure is expected to dominate future changes in SMB, compensating for the projected increase in surface runoff (Krinner et al., 2008; Lenaerts et al., 2016) but the balance between these processes is still a matter of debate. This makes it even more important to evaluate the effectiveness of modelled precipitation and sublimation across the continent to be able to estimate SMB at present. Accurate SMB estimates are required to both drive ice sheet dynamical models and to accurately partition sea level rise contributions determined from observations. SMB from regional climate models (RCMs) is also used to correct altimetry measurements by accounting for firn compaction processes for remote sensing applications.

The most common way to observe SMB is by geodetic mass balance stakes (Lenaerts et al., 2019) but this is challenging due to the size and environmental conditions in Antarctica and the most practical alternative is to use output from (high-resolution) RCMs to make continent-wide estimates. There are now an increasing number of RCMs downscaling Antarctic climate simulations available via the CORDEX (CoOrdinated Regional climate Downscaling EXperiments) database. CORDEX is a project of the World Climate Research Programme that aims to produce representative ensembles of regional climate models for different regions of the world. The purpose is to better understand regional climate change, assess regional impacts and improve adaptation to future climate conditions (http://www.cordex.org/).

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. Souverijns et al. (2019) made a 30 years hindcast with COSMO-CLM^2, and Agosta et al. (2019) estimated the SMB using MAR, while various versions of RACMO2 have been used to estimate the SMB of the AIS (Van Wessem et al., 2014; van Wessem et al., 2018)~~but while both~~ . 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 (Hansen, 2019). Here, we use the framework of the Polar CORDEX project to assess climate model performance in Antarctica for the period 1979-2018 derived from an ensemble of six simulations from five different RCMs. The RCMs cover a range of resolutions, physical and dynamical schemes in the atmosphere and types of surface and snow/ice schemes. This allows us to determine the relative importance of individual model components needed to accurately model the climate by comparing the modelled

SMB against the sparse observational data-sets available in Antarctica. We also investigate some of the uncertainties within the individual models and between the ensemble members.

In this paper we seek to quantify present-day Antarctic SMB and understand the sources of variation as a baseline to assess mass budget changes and better understand ~~drive~~ sea level rise observations and projections both directly in terms of the amount of meltwater added to oceans and indirectly as surface forcing for ice sheet dynamical models (Robel et al., 2019; Nowicki et al., 2016).

2 Methods

We compare six climate simulations made with five different RCMs (COSMO-CLM^2, HIRHAM5, MAR, MetUM, RACMO) in the newest available version of the given RCM. However, to provide backwards continuity we also briefly compare three older versions that have been widely used in earlier studies, to examine how results have varied (or not) as RCMs have been developed. We assess the climate of Antarctica in the models and derive estimates for SMB. All models were forced on the lateral boundaries with the ERA-Interim climate reanalysis (Dee et al., 2011) but downscaling used different grids, over slightly different domains and at different resolutions with slightly different ice masks used in the different model versions (see A1~~in the appendix). Some of the models (~~). Simulations with MAR forced by different reanalyses (Agosta et al., 2019) found that results were rather similar to ERA-Interim but to exclude additional variability potentially introduced by using different boundary forcings, we chose to use a single common reanalysis only. The MAR, RACMO, COSMO-CLM^2) models were nudged within the domain using upper air relaxation and ~~one model (MetUM)~~ MetUM was run as a 12 hour reinitialised hindcast~~, while~~ . With this technique the model is run in weather forecast mode and restarted with new boundary conditions every 12 hours. The two versions (~~one high resolution, one lower~~ high and low resolution) of the HIRHAM5 model were allowed to run freely within the domain and forced only on the boundaries.

We first give a brief overview of each of the participating models, summarised in Table 1. The CORDEX protocol (Christensen et al. (2014)) prescribes a simulation domain for Antarctica with a minimum common analysis extent and a resolution of 0.44°. Lucas-Picher et al. (2012); Lenaerts et al. (2012b); Franco et al. (2012); van Wessem et al. (2018) among others have found that a higher spatial model resolution gives more physically plausible results especially with respect to precipitation processes in areas with steep terrain. Hence, several participating groups have chosen to run their RCMs at higher spatial resolution. ~~Outputs from the different models are compared with each other and the ensemble mean to~~ To quantify both the absolute and relative integrated and basin scale SMB for the continent~~. The models are also compared to~~ , we compare outputs from the different models with each other and the ensemble mean. We also evaluate the models with SMB observations (including ice cores and stakes) and near-surface climate observations (surface pressure, temperature and wind speed) measured across the continent. Unfortunately, as we are constrained to using existing simulations, the models cover slightly differing periods (see 1 for details). We have therefore defined a common 30 year climatological period of 1980 to 2010 for all models to simplify the integrated mass budget comparison, except for COSMO-CLM^2 where the period covers 1987 to 2010. Figures that show time series of data show the full period relevant for each model.

2.1 Models

The model versions included in this paper were selected as fulfilling the two requirements of being the most up-to-date model version as well as being forced on the boundaries with ERA-Interim reanalysis. We include the earlier RACMO version 2.1 and MAR 3.6 as part of the initial SMB comparison as these models have been widely used and are still available for scientific use online; for example, results from RACMO2.1P were used in compiling the IPCC AR5 climate atlas. However, they are no longer considered up to date and have been replaced by RACMO2.3p2 and $MAR_{v3.10}$ respectively therefore we do not consider them in the detailed results analysis in this paper. The models also have snow schemes of differing complexity so the comparison of SMB necessarily includes slightly different terms for different models. For example, the RACMO and MAR models have been developed to include the wind blown snow sublimation terms in SMB. As these terms cannot easily be removed without retuning the models, we have opted to include these within the SMB calculation for these two models.The individual model descriptions give further details of each models outputs.

2.1.1 COSMO-CLM^2

COSMO-CLM^2 is a non-hydrostatic RCM developed at the German Weather service together with an extensive scientific community (Rockel et al., 2008). The model is applied over the Antarctic at a spatial resolution of ~25 km and 40 vertical levels in the atmosphere. The model is forced every 6 hours at the boundaries by ERA-Interim. Additionally this model is coupled to the Community Land Model (version 4.5; Oleson and Lawrence, 2013), with adjustments in the perennial snow proposed by van Kampenhout et al. (2017) to better represent the SMB of ice sheets (COSMO-CLM^2). Apart from this, several model parameters were adjusted for polar regions, particularly those related to the turbulent kinetic energy scheme and the cloud scheme. A full description of the setup over Antarctica including an evaluation of its performance in simulating the Antarctic climate and SMB is available in Souverijns et al. (2019). In this paper, precipitation minus sublimation is taken as a proxy for the SMB.

2.1.2 HIRHAM5

HIRHAM5 is an RCM developed at the Danish Meteorological Institute and run in this study at both low (∼50 km) and high (∼12 km) resolution, with all other model elements being kept identical. The model combines the atmospheric dynamics of the HIRLAM7 numerical weather prediction model (Eerola, 2006), and the physics of the ECHAM5 global climate model (Roeckner et al., 2003). There are 31 vertical levels in the atmosphere and the model is forced at 6 hourly intervals on the lateral boundaries with temperature, pressure, relative humidity and the wind vectors. Sea surface temperatures (SST) and sea ice concentration (SIC) are forced on the lower boundary at daily intervals. The set-up for Antarctica is similar to that of Lucas-Picher et al. (2012) in Greenland, that is with only a very simple surface physics scheme over glacier ice. A subsurface scheme developed ~~in~~ for Greenland by Langen et al. (2017) is currently undergoing optimisation for Antarctic SMB processes but was not available for use in these simulations. ~~The~~ We used the model outputs of precipitation, evaporation and sublimation ~~are therefore used to make~~ to compute a simple SMB~~calculation~~.

2.1.3 MetUM

The UK Met Office Unified Model ~~(MetUM)~~ **(MetUM)** is a numerical modelling system based on non-hydrostatic dynamics (Walters et al., 2017), which can be run either as a global model or a regional mesoscale model, as presented by e.g Orr et al. (2015). Here we run version 11.1 of the mesoscale model over the standard Antarctic CORDEX domain at a spatial resolution of 50 km and 70 vertical levels (reaching up to 80 km). The mesoscale model is nested within a global version of the MetUM with a horizontal resolution of N320 (i.e. 640 × 480 longitude-latitude grid implying a nominal 40 km horizontal mesh), which was initialised by ERA-Interim. The model was used to run a series of consecutive twice-daily 24-hour forecasts at 00 and 12 UTC from the beginning of 1980 to the end of 2018. The first 12 hours of each forecast were discarded as spin-up, with the remaining output concatenated together to form a continuous time-series. ~~Note that although~~ Although the mesoscale model includes a multi-layer snow scheme (Walters et al., 2019), ~~these simulations~~ in these simulations we used a simplified single-layer scheme with for example, no refreezing (Cox et al., 1999)~~and therefore the simplified~~ . SMB was calculated based on ~~ouput~~ output precipitation and sublimation and evaporation ~~.~~

2.1.4 MAR$_{v3.10}$

The «Modèle Atmosphérique Régional»~~(MAR)~~ **(MAR)** (Gallée and Schayes, 1994) is a hydrostatic RCM specifically designed for polar areas (e.g., Fettweis et al., 2017; Kittel et al., 2018; Agosta et al., 2019). The model has 24 atmospheric vertical levels and an horizontal resolution of 35 km. MAR is coupled to the 1-D multilayer surface scheme SISVAT (Soil Ice Snow Vegetation Atmosphere Transfer; De Ridder and Gallée, 1998), which simulates mass and energy fluxes between the atmosphere and the surface. The snow-ice module, based on the CROCUS model (Brun et al., 1992), represents the evolution of the snowpack for 30 snow layers through subroutines of snow metamorphism, surface albedo, meltwater runoff, percolation, retention and refreezing. MAR is forced with ERA-Interim every 6 hours over 1979 – 2018 at its atmospheric lateral and upper boundaries (pressure, wind, specific humidity and temperature at each vertical level) and over the ocean surface (SST and SIC). Furthermore, an upper-air relaxation is used to constrain the MAR general atmospheric circulation (van de Berg and Medley, 2016). Relative to previous studies over the ~~Antarctic ice sheet~~ AIS (Kittel et al., 2018; Agosta et al., 2019), the version used in this study (MAR$_{v3.10}$) only improves the cloud lifetime, the model stability and its computational efficiency enhancing a larger independence of MAR to its timesteps. Furthermore, the definition of the ~~Antarctic ice sheet~~ AIS mask has also been improved by taking into account rock outcrops. An extensive description of the adaptation of MAR to the ~~Antarctic ice sheet~~ AIS can be found in Agosta et al. (2019).

2.1.5 RACMO2.3p2

The Regional Atmospheric Climate Model ~~RACMO2.3p2~~ **RACMO2.3p2** combines the dynamical ~~process~~ processes of the High Resolution Limited Area Model (HIRLAM) (Undén et al., 2002) ~~with~~ and the physics package CY33r1 of the European Centre for Medium-range Weather Forecasts (ECMWF) Integrated Forecast System (IFS). RACMO2.3p1 was built by porting the polar physics components that were part of RACMO2.1P into the standard climate model RACMO2.3 developed at the Royal Netherlands Meteorology Institute (KNMI). RACMO2.3p2 is the follow-up of RACMO2.3p1 and has been applied to the polar ice sheets of Greenland and Antarctica by the Institute for Marine and Atmospheric research Utrecht (IMAU). RACMO2.3p2 includes a multilayer snow model that calculates melt, percolation, refreezing and runoff of liquid water (Ettema et al., 2010). RACMO2.3p2 also uses a prognostic scheme for snow grain size used to calculate surface albedo (Kuipers Munneke et al., 2011); and a drifting snow routine that simulates the interaction of drifting snow with the surface and the lower atmosphere (Lenaerts et al., 2012a). For this study, the model operates at a horizontal resolution of $\sim$27 km, with 40 vertical atmospheric levels. Surface topography is based on Cook et al. (2012) and Bamber and Gomez-Dans (2009). At the lateral and the upper atmospheric boundaries the model is forced by ERA-Interim reanalysis data every 6 hours, and at the ocean boundaries by prescribed ocean temperatures and sea ice cover. The model atmosphere is initialised Jan, 1st, 1979 with the ERA-Interim reanalysis data, and the snow/firn layer with data generated by the IMAU Firn Densification Model (IMAU-FDM) (Ligtenberg et al., 2011). The precursor version, RACMO2.3p1 includes an older ice mask and surface topography, no upper air nudging, a more severe drifting snow formulation eroding more snow and changes in the formulations of surface melting and precipitation. Further details can be found in van Wessem et al. (2018) that intercompares versions p1 and p2 more fully.

2.1.6 RACMO2.1P

RACMO2.1P is an earlier version of RACMO2 using the ECMWF-IFS physics package CY23r4 ~~which includes no~~ that does not include ice cloud super-saturation and utilizes earlier parameterizations for short wave radiation and boundary-layer turbulence as described in Van Wessem et al. (2014). This version of RACMO2.1 includes the polar multi-layer snow routines, as well as the ~~the~~ schemes for drifting snow and albedo as described for RACMO2.3p2 above. In essence, its polar physics components are identical to those in RACMO2.3p1. Simulations with RACMO2.1P have been performed on a modelling domain matching the CORDEX ANT-44 domain in the interior plus a 16-point extension on each domain side for boundary relaxation of ERA-Interim fields. There is also no nudging within the domain in this version.

Model	Period	Resolution [km] (degrees)	Nudging	SMB scheme	Topography	Atmos. Levels
COSMO-CLM^2	1987-2016	25 (0.22)	Yes	Yes	GLOBE	40
HIRHAM5	1979-2017	50 (0.44); 12.5 (0.11)	No	No	GTOPO	31
MetUM	1979-2018	50 (0.44)	Reinit.	No	GLOBE	70
MAR$_{v3.6}$	1979-2018	35	Yes	Yes	Bedmap2	23
MAR$_{v3.10}$	1979-2018	35	Yes	Yes	Bedmap2	24
RACMO2.1P$_{v1}$	1979-2012	50 (0.44)	No	Yes	RAMPv2	40
RACMO2.3p2	1979-2018	27 (0.25)	Yes	Yes	Cook, Bamber	40

Table 1. Summary of differences and similarities between the RCMs. Horizontal resolution is given in degrees and (kilometres), while the number of atmospheric levels refers to the vertical resolution. Nudging refers to the level of forcing within the domain, refer to the individual model descriptions for more details.

2.2 Model Set-up and Outputs

2.2.1 Surface Mass Balance Calculations in RCMs

Two of the models (RACMO and MAR) have subsurface schemes optimised over snow and ice for Antarctica (see references under the model descriptions). ~~Parameterisations are included that~~ The models include parameterisations to account for retention and refreezing of meltwater and also in the case of RACMO2.3p2 wind-driven processes such as erosion at the surface and sublimation of blowing snow. Thus, the definition of the calculation of the SMB changes depending on the complexity of the model. Three models (HIRHAM5, METUM, COSMO-CLM^2,) have only simple surface snow physics over ice surfaces in these experiments. The basic SMB we calculate for them in this study is:

$$\text{SMB} = \text{precipitation} - \text{evaporation} - \text{sublimation} \tag{1}$$

For MAR with an optimised subsurface schemes the SMB is calculated from Eq.2:

$$\text{SMB} = \text{precipitation} - \text{evaporation} - \text{sublimation} - \text{runoff} \tag{2}$$

This differs slightly in RACMO2.3p2/RACMO2.1P as sublimation and erosion of drifting snow (SU$_{ds}$ and ER$_{ds}$ respectively) are also included as a mass loss term as in Equation 3:

$$\text{SMB} = \text{precipitation} - \text{evaporation} - \text{sublimation} - \text{runoff} - \text{SU}_{ds} - \text{ER}_{ds} \tag{3}$$

Both models account for refreezing and retention and thus use runoff rather than melt. Due to the low temperatures in Antarctica, most meltwater refreezes and runoff is ~~largely~~ negligible in the current climate (van Wessem et al., 2018; Agosta et al., 2019) so for the remaining models without the multi-layer subsurface schemes SMB is calculated without the runoff component.

2.2.2 Nudging and upper atmosphere relaxation

As von Storch et al. (2000) pointed out, nudging, whether spectral or with simpler techniques, keeps a regional model closer to the driving large-scale fields (GCM or reanalysis) and is thus a valuable technique where a close match to observations or to a driving GCM is required. Within Polar CORDEX, upper-air relaxation and other forms of nudging have been included as a standard where observational campaigns in large domains require close matches between modelled and observed weather. For example, Arctic cyclone systems and ~~cloud physics~~ cthe presence of clouds in particular appear to be better resolved in models that include nudging (Akperov et al., 2018) and (Sedlar et al., 2011). Similarly, nudging of RCMs run over Antarctica ties their synoptic evolution to these of the driving reanalysis, improving the representation of the interannual variability in SMB to similar levels as in the reanalysis as shown in van de Berg and Medley (2016).

In the experiments presented here, COSMO-CLM^2, MAR$_{v3.10}$, and RACMO2.3p2 are nudged by adjusting temperature and wind fields to the global fields with a minimum relaxation time scale of 6 hours. Strongest relaxation is applied at the top of the atmosphere and relaxation decreases gradually for lower levels. Below typically 4 km (ocean) to 6.2 km (4 km land topography) no relaxation is applied. In the case of MAR$_{v3.10}$, the relaxation of the temperature is weaker than the relaxation of the wind between the highest cloud level and the lowest nudging level. This prevents inconsistency between the temperature inherited from the reanalyses and the humidity and clouds conditioned by the MAR microphysics scheme. Moisture fields are not adjusted by nudging as this would introduce artificial uphill moisture transport. HIRHAM5 and MetUM are not nudged but MetUM is run in a 12-hourly reinitialisation hindcast that keeps the model evolution close to the driving reanalysis.

2.2.3 Grids and land-sea-ice masks

All models have been run for a domain covering the entire Antarctic continent but not all of the domains are the same. HIRHAM5 0.44°and MetUM use the standard CORDEX domain and grid. However, COSMO-CLM^2 extends this slightly to cover more ocean around Dronning Maud Land while the HIRHAM5 0.11°simulations and MAR$_{v3.10}$ were run over slightly smaller domains than the CORDEX domain to reduce computational time, though only after running experiments to determine that e.g. precipitation was not affected. RACMO2.3p2 and RACMO2.1 are run for a domain slightly larger than CORDEX but are trimmed back to remove the relaxation zone such that final results are presented on the CORDEX domain. As the model resolutions are different and each model had its own land-sea mask, the area of Antarctica is not the same in all models, which complicates the SMB results when integrated over the continent. To correct for this areal difference, all the data have been bilinearly regridded to the HIRHAM5 0.11° grid, with the unglaciated land of MAR$_{v3.10}$ included, with a threshold for the ice mask of 50%. This was used to generate a common ice mask for the models in order to calculate the integrated SMB over the ice sheet. In the appendix, figure A1 shows all masks compared to the common mask. Most models had very few grid points different from the common mask but these were also areas with high precipitation rates and this therefore accounts for some fairly large differences in annual SMB on the native masks.

Modelled SMB is integrated over drainage basins defined as in Shepherd et al. (2018b). The horizontal resolution of the models is not altered and the drainage basin masks are defined by selecting all model grid points that fall within the drainage basin outlines. In addition to the drainage basins, that are by definition grounded ice, outlines of the ice shelves that the basins drain into are also used. This allows us to partition SMB over grounded ice (~~GrIS~~GRIS) and ice shelves (IS) as well as over the ice sheet as a whole including ice shelves (TotIS).

2.3 Observations

2.3.1 AWS observations

~~Weather observations are used~~ We use weather observations to assess how well RCMs reproduce the meteorological conditions over the ~~Antarctic ice sheet~~AIS. Although a detailed evaluation of the near-surface model climates is not the purpose of this study, this comparison helps to explain model biases in simulating SMB and especially the coherence between the modelled SMB and the near-surface climate. The original dataset is a compilation of surface pressure, near-surface temperature and wind speed from 307 AWS over the ice sheet used in the MET-READER database (Turner et al., 2004), but also collected by the BAS, IMAU (van Wessem et al., 2014), and the IGE/IPEV (Amory, 2019). The original data were available at several sampling time steps (sub-hourly, hourly, 3-hourly) and were averaged to obtain daily values. Only daily averages computed from more than 75% of the original data are considered as representative of the entire measurement (UTC) day and are used for comparison. Several stations displayed suspicious measurements (sudden discontinuity in pressure and temperature, temperature values capped to the lower bound of the measurement range during the whole winter season, etc) and these were removed from the dataset. Stations occasionally exhibited wind speeds of 0 m/s for day-long periods, probably as a result of sensor riming. For these cases the daily averages were considered as no data (See Kittel et al. in preparation for details on the full list of AWS and data selection protocol). Although we use a homogenised and quality-controlled dataset for the comparison, observations may still be biased in ways that are hard to quantify due to e.g. burial of stations by snow, battery failures, tilt due to strong winds and other instrument failures that remained undetected, reflecting the difficulties involved in collecting data in the harsh and remote Antarctic environment.

~~In order to take into account the different~~ As the different models have different ice masks and topographies~~used by the models, only the stations belonging to~~ , we only retain stations on the common mask ~~and having a~~ where the difference in elevation is lower than 500 m for each model~~are retained, leading to the use~~ , this gives a total of 184 AWS (See the supplementary material for locations of AWS used in this study). ~~The~~ We compute the modelled surface pressure, near-surface temperature and wind speed, as well as the model elevation, ~~are computed~~ 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 ~~is used~~ for all comparisons with the observations~~, since the measurement height of the observations is not known for every stations~~.

2.3.2 Comparison with 10 m snow temperature observations

Deep snow temperatures in Antarctica are indicative of the annual long-term mean surface air temperature. Here, 64 observations of 10 m snow temperature are used ~~that are~~ and collected from a broad range of climatic regions of Antarctica, representing a spatially complete picture of climatological surface temperature (Van Wessem et al., 2014).

2.3.3 Observed SMB

Observations of SMB are sparse over the wide Antarctic continent, and have been obtained from diverse measurement techniques such as stake measurements, ice core, or radar stratigraphy. For the purpose of our evaluation, we use the SAMBA dataset from Favier et al. (2013), that has been completed with Wang et al. (2016) and yearly values of the ice cores from Thomas et al. (2017) to obtain an original dataset of around 7136 observations for various time periods and for a wide range of locations scattered across the ~~Antarctic ice sheet. The~~ AIS. We did not used the radar measurements published by Medley et al. (2014) ~~are not used~~ in this study as the spatial variability is very high and difficult to smooth appropriately for all model grids.

To evaluate the models, we selected observations of SMB belonging to the common ice mask and for which the measurement period began after 1950 to 2018. These conditions reduced the total number of observations used in the comparison to 3671. ~~Observations~~ We used the observations between 1950 and 1987, or 2015 and 2018 that are not fully included in the common modelling period ~~(ie.,~~ of 1987 to 2015~~), were used~~ , for evaluation only if they covered more than 5 years. These 1849 SMB observations are compared to modelled values averaged over the common modelling period in order to compute a climatological mean while we averaged modelled SMB values over the exactly same period for the observations between 1987 to 2015 (1822 observations).

Since the models have different resolutions and grids, we do not directly compare the modelled SMB values to the observations. As in Kittel et al. (2018) and Agosta et al. (2019), we compute modelled and observed SMB values in ~~3~~ 2 steps. Firstly, the original resolution modelled SMB values were interpolated, as for AWS observations, to the observation location using a four-nearest inverse-distance-weighted method. Secondly, all the interpolated SMB values contained in the same grid cell from the common ice mask were averaged as well as the observations for finally creating 923 comparison pairs. This leads to a fair comparison for each model that takes into account the benefit of using a higher resolution for a specific model and removing the very high spatial variability of the observations that cannot be reproduced by the models.

Like the meteorological data, SMB observations are subject to measurement biases notably due to post-depositional redistribution of snow and the related formation of sastrugi that can considerably complicate the interpretation of measurements at the very local scale (Andersen et al., 2006). SMB observations should therefore be considered as a best estimate of accumulation rather than an absolute value. As SMB observations are not evenly distributed over the ice sheet, the comparison statistics may be artificially influenced by over- and/or under-sampled regions.

3 Results

We first focus on how the RCMs characterise the surface climate over the ice sheet before turning to assessing the SMB and taking note of the differences in precipitation distribution.

3.1 Temperature, Surface Pressure and Wind Speed from Models and Observations

Weather observations in Antarctica extend further back in time and there is generally better coverage than for direct SMB measurements. In figure 1 we show Taylor diagrams for pressure, temperature and wind velocities. Taylor diagrams offer an efficient way to assess model skill by comparing the Pearson correlation coefficient, the centred root mean square error (~~RMSE~~CRMSE) and the standard deviation of the modelled output with the observed values. CRMSE is equivalent to the Root Mean Square Error but systematic biases are removed by subtracting the mean observation and mean modelled values from each value. A perfect model should be in the same place as the observations (black star, correlation of 1, same standard deviation, and zero CRMSE). The further away a model is from the observations, the worse it matches the observed weather. Mean biases and observation mean are also indicated. In this case modelled values closest to the dashed line have a more correct representation of the standard deviation and the closer to the black reference star the closer the model correlates to the observations values. We also list the bias below the diagrams.

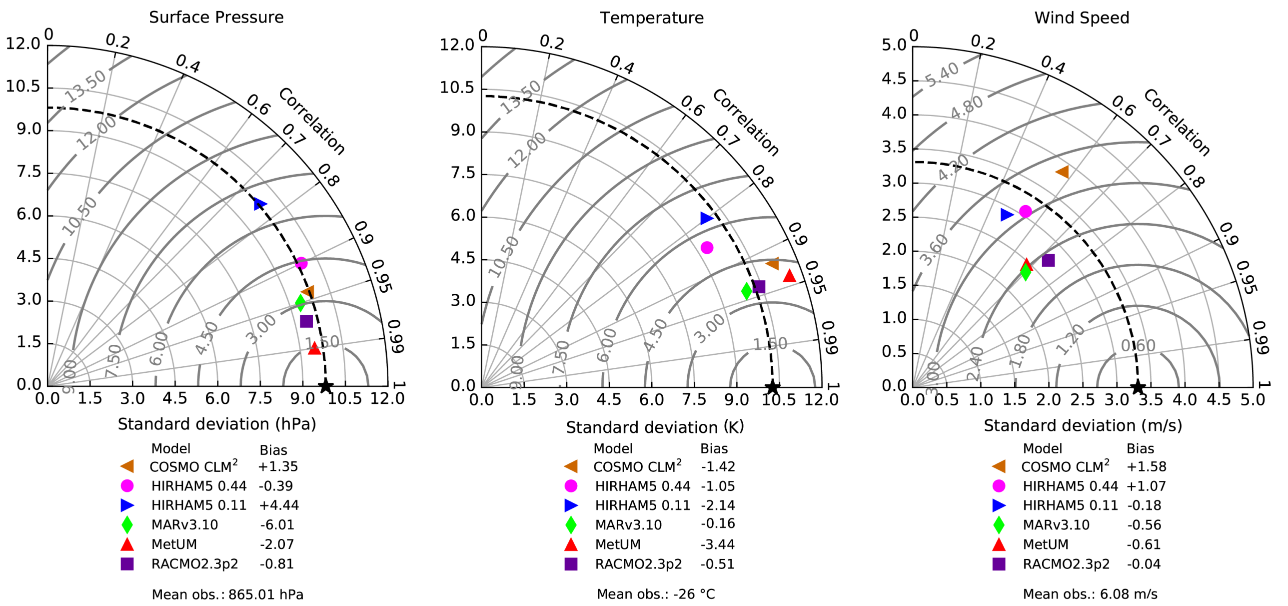


Figure 1. Taylor diagrams showing model performance compared to daily observations of surface pressure (a), near-surface temperature (b), and observed wind speeds ~~as well as~~ (c).The horizontal and vertical axes represent the standard deviation, the dashed line in bold shows the standard deviation of the observations. The Taylor plot also shows the correlation which is measured by the angle with the x-axis. Finally, the CRMSE is represented by the curved lines in light grey. The units of standard deviation, CRMSE, mean bias ~~statistics~~ and mean of the observations are the same (hPa for ~~each model~~surface pressure, K for near-surface temperature, and m/s for wind speed).

Figure 1 analysis shows that depending on variable the models perform reasonably well with some variation. With respect to surface pressure, the majority of models are similarly skillful with the exception of HIRHAM5 0.11°, though the model is still close to the pattern of the standard deviation. The other models have quite a high degree of nudging including upper atmosphere pressure fields within the domain so it is not so surprising to see the good performance here ~~. The~~ as the nudging forces the models to be closer to the observed pressure. 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°) version of HIRHAM5 that has many more grid cells than the low resolution (0.44°) version ~~, which in effect increases the degrees of freedom associated with the high-resolution run and allows~~ 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. The near-surface temperatures in figure 1 show that although overall the models perform well (Pearson correlation of 0.85 and higher) on average all the models are too cold and only $MAR_{v3.10}$ and RACMO2.3p2 have a bias of less than 1 degree (respectively -0.16 and

-0.51 K), with MetUM having the highest bias (-3.44 K). As with the surface pressure analysis, the HIRHAM5 high resolution simulations have a lower correlation coefficient and this may well be again the consequence of the unnudged simulations. However, biases in cloud cover and long-wave radiation reaching the surface are ~~probably~~ likely the main explanation for divergence from observations and should be investigated for all RCMs run for Antarctica as shown by van Wessem et al. (2014) . In their study, significant improvements in the RACMO2.3p2 model were obtained by adjustments to the cloud microphysics. Furthermore, the lack of detailed subsurface snow pack schemes including processes such as refreezing (and subsequent latent heat release) and densification also likely has an impact on the temperature bias in HIRHAM5 and MetUM (see also figure 2).

It is clear in figure 1 that all of the models perform less well for wind speeds than for temperature or pressure. This is likely in part due to large uncertainties in the observations ~~and~~ especially at unattended stations where burial by snow, changes in orientation and sensor breakdown are more likely. However, the effects of ~~resolution but may also relate to~~ different resolution and differences in turbulent schemes between the models ~~and in~~ may also be important. In particular the extremely stable boundary layer over most of Antarctica ~~that~~ is hard to represent in models particularly at lower resolutions (Zentek and Heinemann, 2019). The models ~~can be roughly divided into two groups~~ appear to fall in two groups on the Taylor Diagram: $\mathrm{MAR}_{v3.10}$, MetUM and RACMO2.3p2 on the one hand, and the two HIRHAM5 runs and COSMO-CLM^2 on the other hand. In the case of COSMO-CLM^2 wind speeds are output at 20 m and then interpolated to 10 m using Monin-Obukhov theory (Souverijns et al., 2019), which may not be sufficient to properly represent near-surface winds and associated interactions. The HIRHAM5 results may again be biased due to the lack of nudging within the domain. However it is worth pointing out that HIRHAM5 correctly represents the mean spatial variability (both runs are the closest to the dashed line indicating the standard deviation) and in the case of the high resolution run the mean observed wind speed, but not the daily evolution of the wind, resulting from the biases in the atmospheric circulation.

3.2 Comparison with 10 m snow temperature observations

Figure 2 shows the modelled surface temperature of the RCMs as a function of 64 measurements of temperature at 10 m depth as also used by Van Wessem et al. (2014). The majority of the AIS has negligible snow melt and in these regions the 10 m snow temperature is representative of the annual long-term average surface temperature. This comparison therefore is a robust assessment of the climatological surface signal calculated by the models, also because the observations are evenly scattered across the continent and represent most climatic regions. All models capture the wide range of surface temperatures from $\approx$ 218 K to 260 K. HIRHAM5 0.44°consistently underestimates temperature for most locations, a bias that closely resembles RACMO2.1 in Van Wessem et al. (2014) ~~)~~ and which the authors concluded was predominantly related to biases in the down-welling long-wave radiation. The other models overestimate temperature in the ~~colder, and therefore higher elevation~~higher elevation, colder locations, while underestimating temperature ~~for~~ at lower elevations in the coastal regions. For the colder regions below $\approx$ 240 K, these biases are ~~mostly~~ most likely related to discrepancies in cloud cover ~~and most~~ likely snowfall, affecting downwelling longwave radiation and surface albedo. Some of the Antarctic models have been tuned to improve the dry and cold biases in the interior that were persistent in earlier model versions (see RACMO2.1P; Van Wessem et al. (2014); van Wessem et al. (2018)), but now overestimate temperature slightly instead. While subsequent model updates have led to significant improvements in simulated SMB, this has come at the expense of surface temperature due to excessive increases in downwelling radiative fluxes that accompany increases in snowfall.

For the warmer lower elevation, mostly coastal regions most models have a cold bias. This bias is likely related to the effects of surface meltwater percolating into the firn and refreezing within, raising deeper snow temperature, implying modelled surface temperature is not a good metric for observed 10 m snow temperatures in the percolation zone. A more accurate comparison would therefore be to directly compare 10 m snow temperatures from the models with the observations. However, not all models calculate snow temperatures, and given the scope of this manuscript, we only intercompare the surface temperature. Here, Figure 2 illustrates a consistent intermodel scatter, with mainly the models that do not include a sophisticated snow model outside of this range. This clearly points to a significant potential source of improvements for modelled SMB in the future.

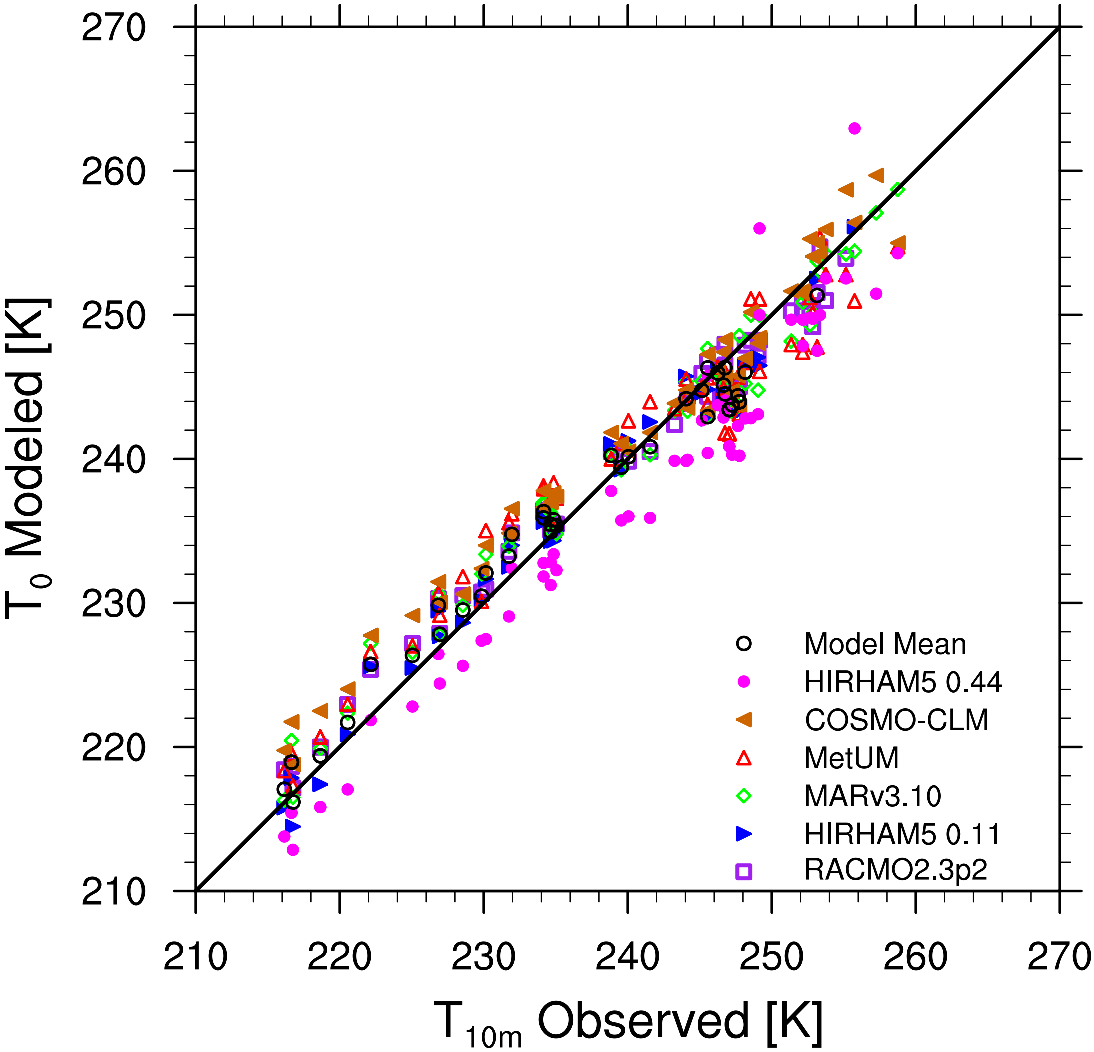


Figure 2. Modelled surface temperature as a function of observed 10 m-snow temperature (Van Wessem et al., 2014). Observations ~~that are~~ not fully located on the model ice-mask are excluded.

3.3 Comparison with Observed SMB

Evaluating SMB is hindered by poor observations across the cryosphere, particularly in Antarctica where remoteness and extreme weather conditions add to the challenge of observing SMB. In our analysis, in spite of using a large dataset of observations, shows that there are large areas ~~that are~~ significantly undersampled (See for example, Figure 4). We therefore separate the comparison of modelled and observed SMB into elevation bins in Figure 3 in order to make the results clearer. It is important to mention that for the scatter plots by elevation class, if an observation or one of the models had a negative value, the observation and modelled values were removed from the analysis using logarithmic values (hereafter, rlog is the correlation computed on the logarithm of SMB values), but are retained in the analysis using the original populations. Statistics for the SMB comparison are given in Table 2 but to show the large scatter in the observations and the models clearly we show all models plotted against observations in Figure 4 and plotted against each model individually in the appendix (Figures A2 , A3, A4 A5, A6,A7).

Apart from COSMO-CLM^2 and HIRHAM5 0.11°the RCMs show similar Root Mean Square Error (RMSE) and r^2 values when compared over the full dataset but broken down by elevation class or locally by regions shows a more complex story. In general all models underestimate SMB over the ice shelves and at the low elevation coastal regions of Antarctica (see also statistics in Table 2 a. and b. and Figure 3). This is particularly true for COSMO-CLM^2 over both locations and HIRHAM5 0.11°over the ice shelves leading to higher RMSE compared to the others RCMs, but while all the RCMs underestimate SMB over the Ross ice shelf, MAR$_{v3.10}$ overestimates it, probably related to a poorer representation of the surface climate by the model over this ice shelf. The blowing snow module included in RACMO2.3p2 may explain the ~~good~~ lower bias in results between 0 and 1200 and especially 1200 and 2200 m (Table 2 b and c), ~~where it outperforms~~ compared to the other models. A previous comparison shows higher sublimation in RACMO2.3p2 than in MAR$_{v3.10}$ (Agosta et al., 2019) notably at the elevations where katabatic winds are strong due to the slope of the ice sheet and where the atmosphere is not too cold enabling large amounts of sublimation from blowing snow particles. COSMO-CLM^2 and HIRHAM5 0.44°have the highest RMSE while HIRHAM5 ~~.011~~0.11, MAR$_{v3.10}$ and MetUM have similar statistics at this elevation. For the highest elevations (above 2200 m), all the model RMSE scores are relatively low and similar to each other except HIRHAM5 0.44°(and to a lesser extent MAR$_{v3.10}$) between 2800 m and 3400 m (Table 2 e). However, the less extensively evaluated models (HIRHAM5 at both resolutions and MetUM) are both too dry over the high plateau of the ~~Antarctic ice sheet~~AIS. If we look at all the elevation ranges, no model is systematically in the top 3 for every range but, RACMO2.3p2 has the best comparison with all the observations, closely followed by MetUM with MAR$_{v3.10}$ and HIRHAM5 0.44°performing almost equally. It is worth emphasising though that as Fig. 4 shows, the observations in this elevation class are also very noisy and the poor relative performance of the models may result as much from unrepresentative and sparse repeat observations as it does from missing or poorly resolved processes in models. Analysis of these results not only indicates areas where models need to be improved but also areas where more observations to test models are desirable, notably between 1200 and 2200 where the mean biases of the models used in this study display large discrepancies (Table 2 c). It is also likely that there are compensating errors within each model that hide the true performance. For example, the mean bias between the two different HIRHAM runs has opposite signs in the 1200 - 2800 m range, likely reflecting the difference in model resolution. Orographic precipitation is very sensitive to slope effects and the presence of steep topography is very different between the two resolutions, affecting where precipitation falls across the continent. The wide scatter in modelled SMB in the 2200-2800 m elevation range is therefore also likely to reflect in part the resolution of the different models and how well they capture orography and the consequent precipitation. Studies by for example Hermann et al. (2018) and Schmidt et al. (2017) show that hydrostatic models like HIRHAM5 and RACMO2.3 typically overestimate precipitation on the upslope and have a dry bias downwind of initial steep topography, this pattern seems to some extent to be repeated here in Figure 3 and 4. Comparing the observations used in this analysis with the RCM ensemble modelled SMB in Figure 6 also highlights that the largest differences between models and compared with the ensemble mean are mostly in regions with very few or no observations,. These are also regions where precipitation is typically high, making it difficult to assess the ability of models to truly simulate the SMB of Antarctica. Our analysis therefore also helps to identify areas where increased observations will be most useful to help assess and improve model processes.

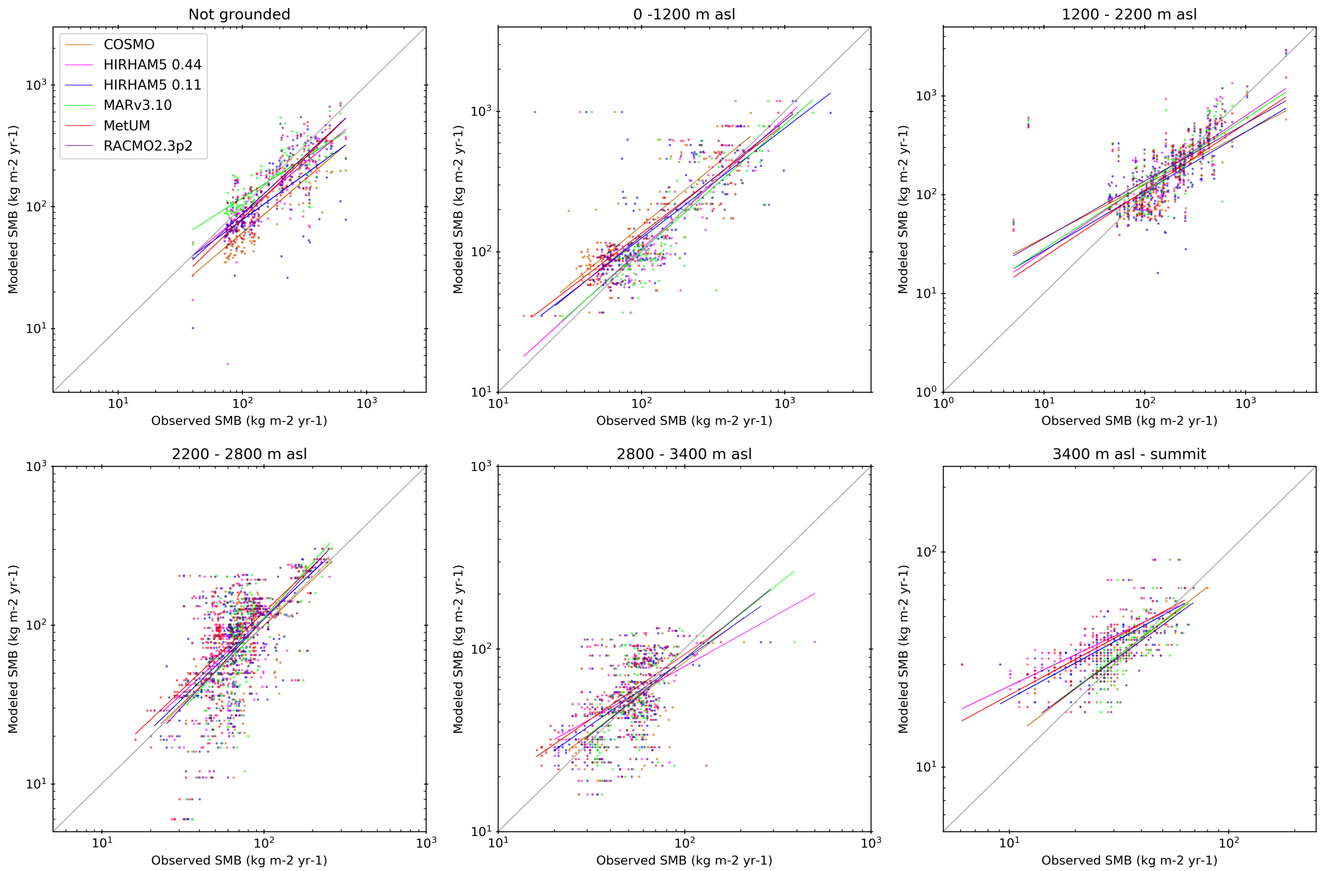


Figure 3. Comparison between modelled SMB and observed SMB in a gridded dataset. Trend lines and points are plotted for each model in a different colour, note different x and y axes for different elevation bins.

Mean bias and RMSE for each model by elevation bin is summed up in the supplementary materials in Figure A8. However as Figure 4 also shows, this is not a straightforward comparison either due to the large areas with only few observations.

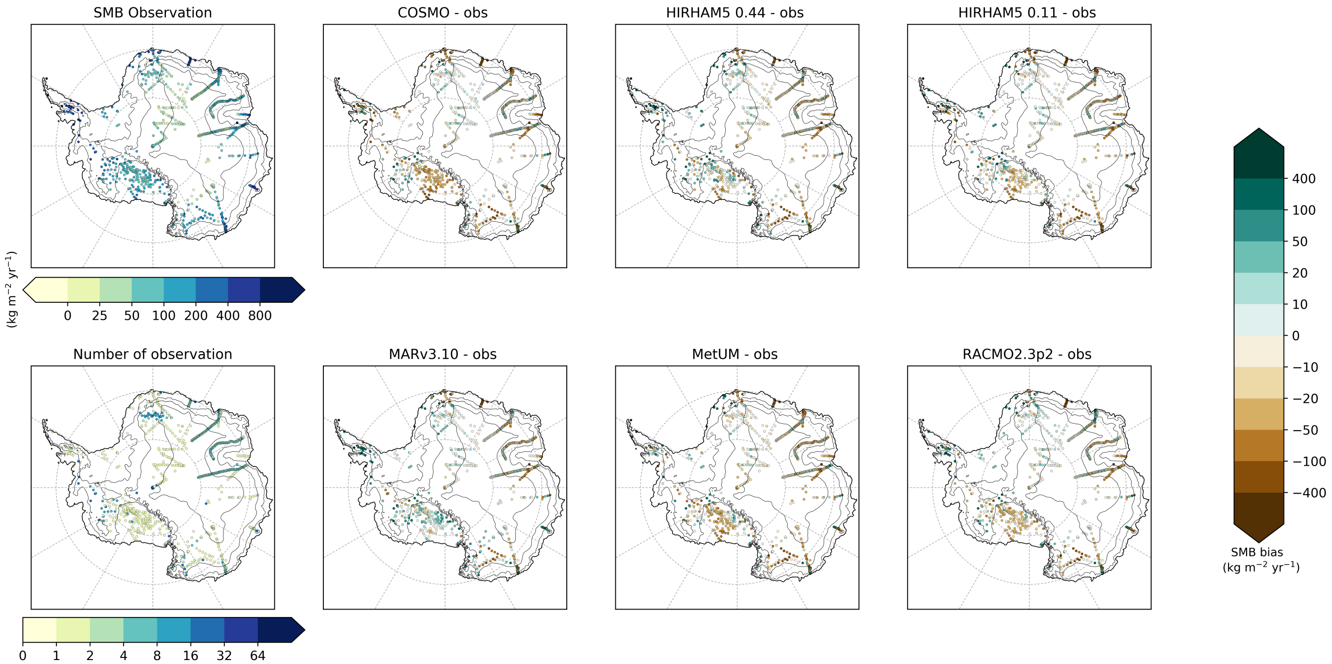


Figure 4. ~~a)~~ Observed SMB values ~~, b) Number~~ and the number of observations ~~per pixel~~at each point. ~~Difference~~ The model plots show the difference between observed and modelled SMB ~~from~~ for COSMO-CLM^2~~(c)~~, HIRHAM5 0.44°~~(d)~~, HIRHAM5 0.11°~~(e)~~, MAR$_{v3.10}$~~(f)~~, MetUM~~(g)~~, RACMO2.3p2~~(h)~~.

3.4 Assessing the Surface Mass Balance of Antarctica

Bearing in mind the results presented in the preceding section evaluating the RCMs, we ~~here show~~ show here the range of best estimates for Antarctic SMB based on RCMs. Figure 5 shows the modelled Specific Surface Mass Balance (SSMB) and integrated modelled SMB for the AIS of the nine climate simulations for 19 drainage basins as defined in Shepherd et al. (2018b). The SSMB is shown by the colour shading and is defined as the integrated SMB (in mm per year), divided by the area of the basin. The SMB is the total integrated SMB of the basin in units of Gigatonnes (shown by the numbers in a box in each basin). Figure 5 illustrates that all models simulate a comparable SSMB for EAIS, with values between 100 and 400 mm per year. Due to the moist coastal climates over the ice shelves, SSMB values here reach values as high as 1000 mm per year. The main intermodel differences are found over the WAIS and the Antarctic ~~peninsula~~ Peninsula (AP) and are most likely related to differences in horizontal resolution and therefore orographic precipitation. The higher resolution models (RACMO2.3p2, HIRHAM5 0.11°and MAR$_{v3.10}$), generate the highest SSMB values over the AP and WAIS basins, up to 2000 mm per year. The other models have considerably lower SSMB, especially over the adjacent ice shelves. The exception is COSMO-CLM^2 which is drier than all other models in all basins with the exception of the Queen Mary Land basin in the EAIS, where HIRHAM5 0.11°is slightly drier, and the interior of the EAIS where MAR$_{v3.10}$ is slightly drier. The two areas with the largest ensemble

	a. Shelves (N=112, L=112) Mean obs: 199±132				b. 0 - 1200 m (N=130, L=128) Mean obs: 223±224			
	MB	RMSE	r	rlog	MB	RMSE	r	rlog
COSMO-CLM^2	-85	125	0.75	0.84	-79	174	0.73	0.81
HIRHAM5 0.44°	-37	89	0.79	0.75	-22	143	0.77	0.82
HIRHAM5 0.11°	-59	122	0.60	0.67	-26	194	0.68	0.76
$MAR_{v3.10}$	-12	98	0.69	0.79	-5	159	0.74	0.79
MetUM	-32	83	0.82	0.82	-41	142	0.79	0.84
RACMO2.3p2	-25	90	0.78	0.78	-29	147	0.78	0.87

	c. 1200 - 2200 m (N=158, L=154) Mean obs: 225±240				d. 2200 - 2800 m (N=259, L=258) Mean obs: 89±55			
	MB	RMSE	r	rlog	MB	RMSE	r	rlog
COSMO-CLM^2	-22	187	0.63	0.75	-9	42	0.67	0.61
HIRHAM5 0.44°	33	143	0.89	0.78	-18	45	0.65	0.59
HIRHAM5 0.11°	-19	119	0.89	0.68	-16	46	0.64	0.56
$MAR_{v3.10}$	20	115	0.90	0.79	-14	42	0.70	0.63
MetUM	-16	119	0.87	0.80	-22	46	0.68	0.63
RACMO2.3p2	12	95	0.94	0.77	-13	41	0.68	0.66

	e. 2800 - 3400 m (N=161, L=161) Mean obs: 58±27				f. 3400 m - top (N=103, L=103) Mean obs: 36±12			
	MB	RMSE	r	rlog	MB	RMSE	r	rlog
COSMO-CLM^2	-1	23	0.59	0.61	-1	9	0.70	0.72
HIRHAM5 0.44°	-6	40	0.35	0.53	-12	15	0.72	0.72
HIRHAM5 0.11°	-5	26	0.55	0.62	-9	12	0.72	0.72
$MAR_{v3.10}$	-2	32	0.41	0.54	-1	9	0.67	0.69
MetUM	-10	25	0.59	0.61	-10	14	0.73	0.73
RACMO2.3p2	-2	27	0.46	0.56	0	9	0.70	0.72

	g. All (N=923, L=916) Mean obs: 133±160			
	MB	RMSE	r	rlog
COSMO-CLM^2	-28	113	0.74	0.79
HIRHAM5 0.44°	-9	91	0.85	0.82
HIRHAM5 0.11°	-20	101	0.81	0.79
$MAR_{v3.10}$	-3	88	0.85	0.83
MetUM	-22	82	0.87	0.84
RACMO2.3p2	-9	79	0.88	0.85

Table 2. Comparison of the modelled SMB to the SMB observations over the ice shelves (A), by elevation bins (B-F) and over the whole ~~Antarctic ice sheet~~ AIS (G). Unit of Mean Biases (MB), Root Mean Square error (RMSE), and Mean of the observation is kg/m²yr. N denotes the number of comparison used for each bin while L represents the number of comparison used the log distribution (See the supplementary materials for more details)

mean deviation are the western peninsula basin but also the interior of the EAIS bordering the Transantarctic mountains and including the South Pole. In this region the $MAR_{v3.10}$ model has the highest SMB (196 Gt) but MetUM has the lowest (77 Gt). Figure 5 also shows some of the striking features in the pattern of SMB present in all the models where the magnitude differs, for example, all models have a steep gradient in the SMB over the Antarctic ~~peninsula~~Peninsula, but this is much more pronounced in HIRHAM5 0.11°than in HIRHAM5 0.44°, demonstrating the importance of resolution in this region. MetUM and COSMO-CLM^2 also show the same pattern but with considerably lower absolute values particularly on the western side, than the other models. These differences in modelled SMB on the basin scale may have considerable impact on dynamic ice sheet models used to determine the evolution of the ~~Antarctic ice sheet~~ AIS and are consequently important to take into account when selecting SMB to force ice dynamics models. Looking at the total surface mass budget including ice shelves for the period 1980 to 2010 (numbers in the caption and summarised in Table 3) generated by the models, the HIRHAM5 0.44°simulation is the wettest model (2752 Gt per year; 2328 Gt excluding ice shelves), while COSMO-CLM^2 is ~~clearly~~ the driest (2031 Gt per year; 1751 Gt excluding ice shelves). The other simulations are all closer to each other and are within an SMB range of $\pm$ 200 GT per year, while the two dedicated polar models (RACMO2.3p2 and $MAR_{v3.10}$) have only a small difference of 83 Gt per year on average, corresponding to around 3% of the total budget. These two models have been evaluated and optimised for Antarctica the most intensely of all the models (van Wessem et al., 2018; Agosta et al., 2019). We also include MAR3.6 and RACMO2.1 in this figure to give context to earlier studies. The two closest models overall are in fact HIRHAM5 0.11°and $MAR_{v3.10}$ which differ by only 26 Gt overall with much of the difference accounted for by the SMB of the ice shelves.

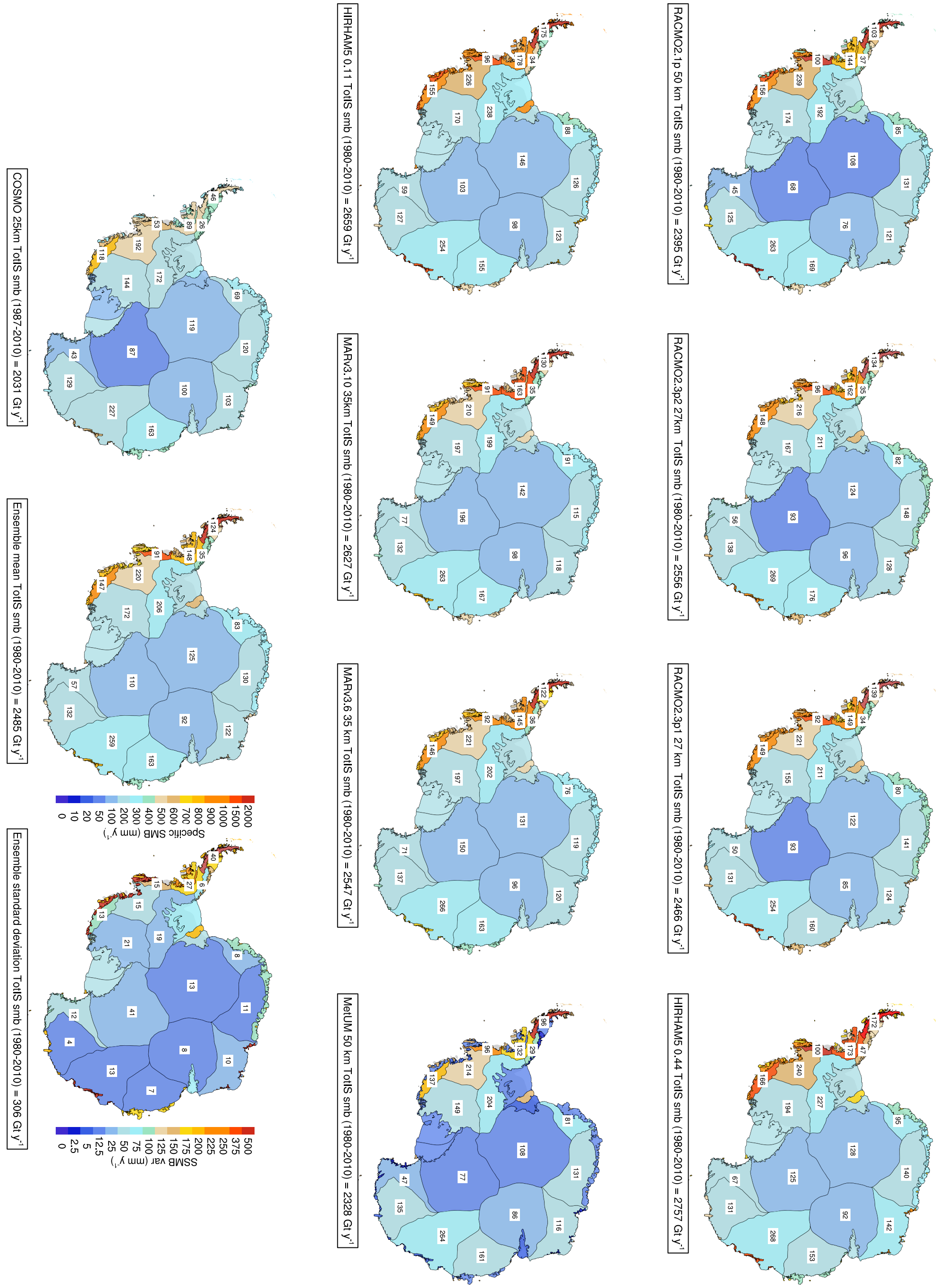


Figure 5. Integrated SMB and specific SMB (SSMB) for the 9 models included in this study: RACMO2.1, RACMO2.3p2, RACMO2.3p1, HIRHAM5 50 km, HIRHAM5 12.5 km, $MAR_{v3.10}$, $MAR_{v3.6}$, MetUM and COSMO-CLM^2, as well as the ensemble mean and standard deviation shown. Colours denote the SSMB in mm w.e. per year for all grounded ice sheet basins as well as the ice shelves these drain into, defined in Shepherd et al. (2019). The numbers included in the basins denote the basin integrated SMB in Gt $year^{-1}$ for the grounded ice sheet for the period 1980 to 2010 with the exception of COSMO-CLM^2 where the time series starts in 1987. Finally, the total integrated number for the grounded ice sheet including ice shelves is shown in the figure label.

Model	GIS (Gt y^{-1})	IS (Gt y^{-1})	ToTIS (Gt y^{-1})	Area (km^{-2})
RACMO2.1p	1933	471	2395	13.85
RACMO2.3p2	2133	430	2556	13.85
RACMO2.3p1	2035	438	2466	13.85
MAR$_{v3.10}$	2227	413	2633	13.92
MAR$_{v3.10}$	2158	396	2547	13.92
HIRHAM-5 0.44°	2328	438	2757	13.87
HIRHAM-5 0.11°	2235	434	2659	13.83
MetUM	1884	452	2328	13.82
COSMO-CLM^2	1751	288	2031	13.84
Model Mean	2076	418	2486	13.86
Model $/sigma$	306	77	266	0.085

Table 3. Integrated mean annual SMB for the six models used in this study, for the period 1980 to 2010 except for COSMO-CLM^2 where the period was 1987 to 2010. Three older model versions, ensemble mean and standard deviation as shown in Figure 5. All calculations done on the original grid of the individual models using a common set of drainage basins and ice mask defined by IMBIE2 Shepherd et al. (2018b). The ensemble mean was calculated by transforming all models to the RACMO2.3p2 grid. GIS denotes grounded ice sheet, IS denotes ice shelves and ToTIS denotes the full Antarctic ice sheet AIS including ice shelves.

As the basin scale SMB values differ quite substantially between models, in Figure 6 we plot the mean annual SMB from the ensemble mean and the anomaly to that for each of the different models. The ensemble mean is calculated on a common grid but the model anomalies are calculated from it on their own gridswhich shows better , which more clearly shows the effects of the different resolutions on the SMB. The figure shows quite substantial agreement between models over large areas of Antarctica but also some considerable local variability. Features such as the Transantarctic mountains and the rugged coastal topography in west West Antarctica both of which substantially influence local weather patterns are for example much clearer picked out in the spatial pattern of the SMB. These features are more clearly delineated in the higher resolution runs. However, the ensemble mean can also hide large disagreements between the models. For example there is an interesting asymmetry in the model results for the region of the Queen Maud Mountains and Queen Elizabeth ranges of the Transantarctic mountains.The MAR model and to a lesser extent the HIRHAM5 0.44°model show rather different patterns in SMB compared to the other models, with higher SMB south of the Range and lower than ensemble mean values north of the range. The other models show the reverse with lower than the mean south of the range and higher to the north. A similar, but less clear pattern is also seen along the Ross and Amundsen Sea coastal sectors. The coastal margin of the whole continent in general shows a blotchy pattern in the SMB anomaly plots that reflects rugged topography. In these regions the resolution of the model determines the location of orographic precipitation. Analysis of similar SMB simulations in Greenland with the HIRHAM, MAR and RACMO models (Hermann et al., 2018; Schmidt et al., 2017) suggests that in these types of locations HIRHAM and RACMO overestimate precipitation at lower elevations in steep terrain, whereas MAR tends to have a wet bias at a slightly higher elevation where the other two models are drier. Agosta et al. (2019) related this different pattern of biases in MAR to the advection of precipitation in the models prognostic precipitation scheme. Understanding these biases is crucial to understanding and interpreting modelled SMB and comparing Figure 6 with Figure 4 it is clear that the locations where there is the highest disagreement between models are also the regions with the poorest systematic observational coverage of SMB, especially in coastal regions and in ~~west~~ West Antarctica.

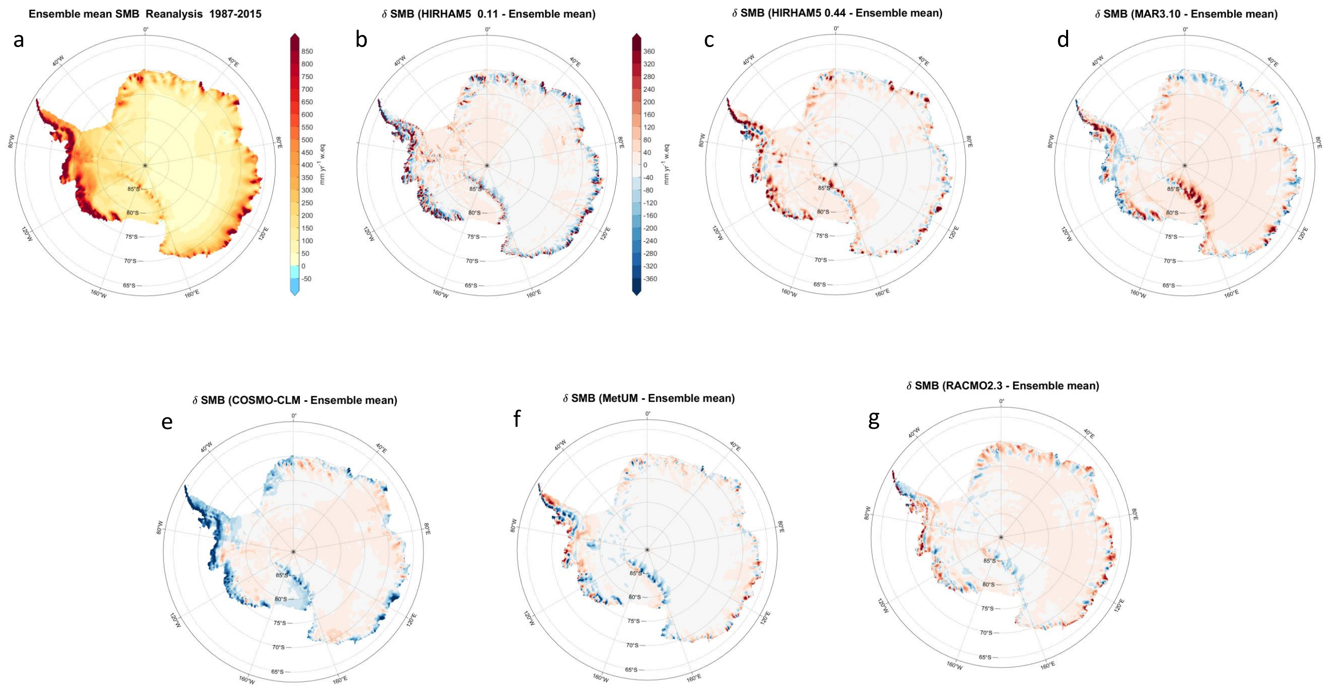


Figure 6. Sub-figure **a** shows the SMB ensemble mean for the common period, on the common mask. Sub-figure **b-g** show the difference between each model and the ensemble mean.

SMB varies not only spatially but also temporally and average annual SMB values hide very large interannual variability in SMB as depicted in Figure 7. The spread in the range of estimates of SMB is however, consistent from year to year. The integrated continental SMB calculated over the common mask has a spread of more that 550 GT between the highest and lowest estimate on average (see also Table 4) but all the models show similar annual and decadal scale variability.This implies that the driving model, in this case ERA-Interim, is the most important source of SMB variability but that the individual models are important when considering both the absolute number and the local spatial variability.

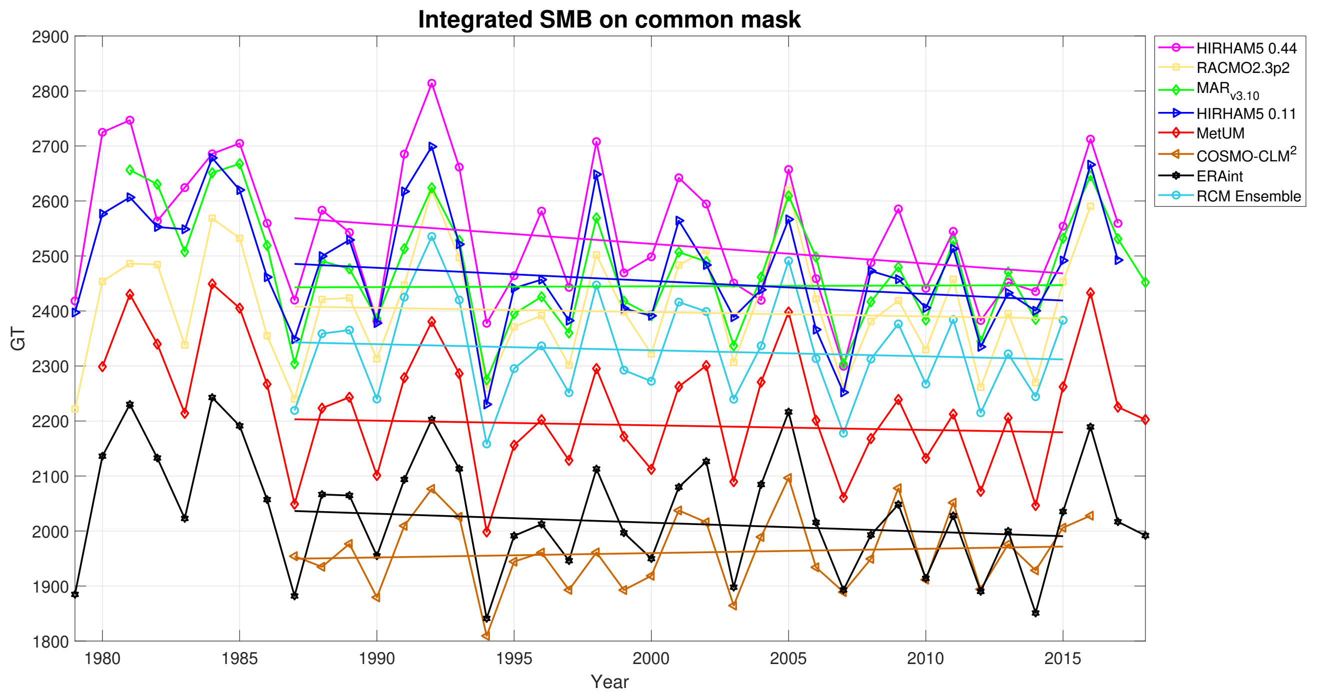


Figure 7. Annually resolved SMB integrated over the common ice mask for the different RCMs, in the period 1979-2018. All RCMs are driven by ERA-Interim. The ensemble is a mean calculated from all 6 RCM's in the period 1987-2015 where there is data from all the models. All the trend lines are calculated for the period 1987-2015.

~~Over the period 1987 to 2015 for which data are available for all the models,~~ We calculated the mean annual SMB and components ~~are calculated~~ across the continent including ice shelves, as given in Table 4~~below, note~~, over the period 1987 to 2015 for which outputs are available for all the models. Note that this is calculated over a common ice mask and a common simulation period and results are therefore slightly different to those already published for different models or shown in Figure 5 or Table 3. In this time series HIRHAM5 0.11°and $MAR_{v3.10}$ are again the closest two models to each other. RACMO2.3p2 is closest to the ensemble mean but COSMO-CLM^2 is closest to the driving ERA-Interim modelled values. The trend lines are very sensitive to starting and ending years and in some cases change sign if a longer period is chosen, but as we have only a short common period we have chosen to calculate the trend over the common period. For this chosen period COSMO-CLM^2, and $MAR_{v3.10}$ show a slightly increasing trend in SMB, whereas the rest show a slightly declining trend in SMB although the trend in RACMO2.3 and MetUM are almost flat. The ERA-Interim trend over the period declines slightly more than the MetUM trend, which is otherwise extremely close. The different trends from the models and in particular the sensitivity to different start and end points does not give us confidence to ascribe a statistically significant trend to Antarctic SMB over the whole continent. We note though that all models show a declining trend in the 1990s and early 2000s but with a recent increase in SMB since 2014. The early part of the record appears to have higher variability but this may be related to changes in data assimilation in the driving reanalysis (Dee et al., 2011).

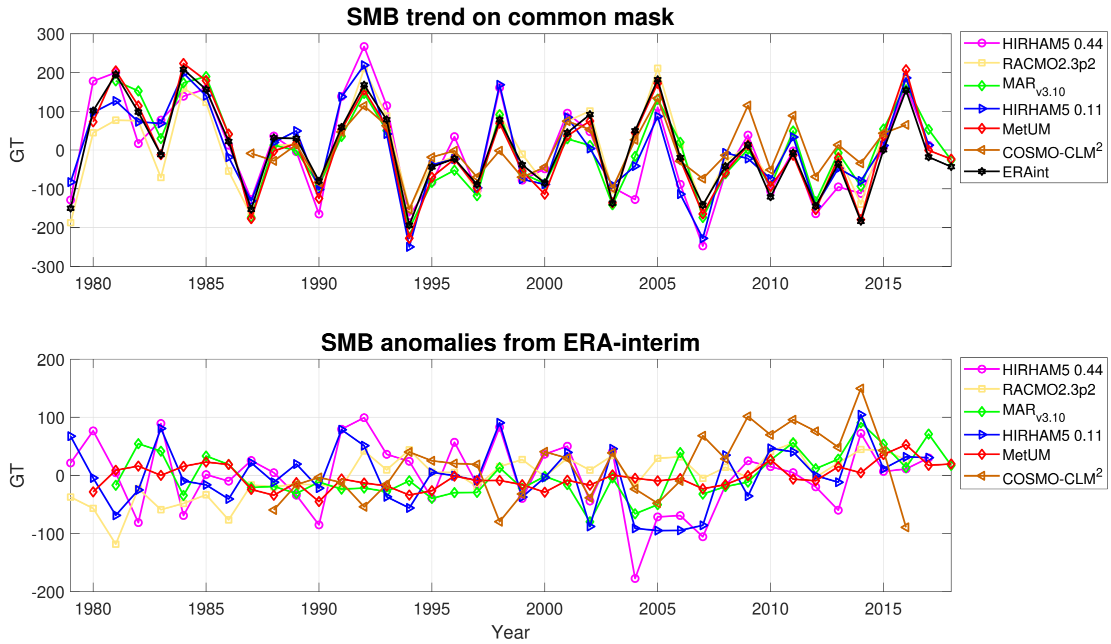


Figure 8. The upper figure shows the variability in surface mass balance over the common mask for each of the different RCMs, in the period 1979-2018. Calculated by subtracting the respective model mean from each RCM's SMB time series. The bottom figure displays how modelled SMB from each RCM deviates from the ERA-interim SMB.

Figure 8 emphasises the large variability in SMB on an annual to decadal scale by plotting the variation from the mean for each model and the variation from ERA-Interim for each model. This shows that while all the models have more or less the same anomaly when compared to their own mean, the sign of the anomaly compared to the ERA-Interim value can be different. Since the most highly constrained models show the lowest anomaly compare to ERA-Interim, we suggest that most of the variation is related to internal variability (weather) within the domain.Both HIRHAM5 0.11°and 0.44°shows the highest values of variability, probably due to the unconstrained nature of the runs, but in different years different models show higher variability than the others. The lower panel in Figure 8 shows that MetUM is by far the closest to the driving model with much less variability than the others (likely due to its frequent reinitialisation), HIRHAM5 again shows the highest difference compared to the driving model but from year to year the model showing maximum difference varies and there appears to be no systematic pattern as to whether or not modelled SMB is higher or lower than the ERA-Interim reanalysis when ~~quanitifed~~ quantified on the common mask and over the whole of Antarctica. The implication is that while the driving model controls broad scale pattern of SMB, the downscaling model adds its own weather variability to the broad scale pattern. The variability, or weather noise is unsurprisingly, largest in un-nudged models. The effect of this noise on ice sheet dynamics may be small overall but as for example, Mikkelsen et al. (2018) show, small variations in SMB can have a non-negligible impact on ice sheet dynamics.

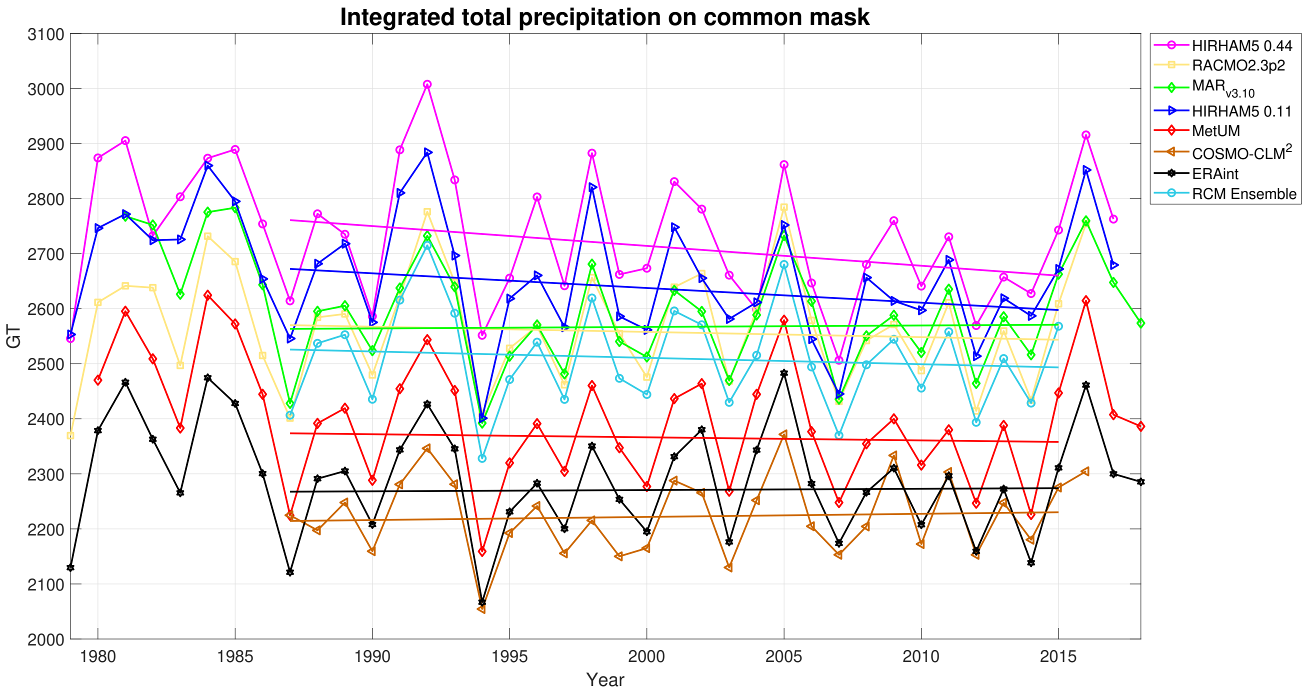


Figure 9. Annually resolved precipitation integrated over the common mask for the different RCMs, in the period 1979-2018. All RCMs use ERA-Interim. The ensemble is a mean calculated from all 5 RCMs in the period 1987-2015 where all models have data.

Since SMB is made up of accumulation and ablation components, and in Antarctica precipitation is the dominant term, Figure 9 shows the precipitation component only over the common mask for the different models and ERA-Interim. There is a very similar pattern to that in Figure 7 but compensating effects from melt and sublimation explain the bigger offset between HIRHAM5 0.11°and MAR$_{v3.10}$, which in turn is closer to RACMO2.3 in terms of precipitation. The mean values for the SMB components of precipitation, evaporation and sublimation as well as SMB for the common period 1987-2015 over the common ice mask are also displayed in Table 4. These values confirm that the very much higher precipitation in both HIRHAM5 runs compared to the other models is to some extent compensated for by higher values of sublimation. The higher sublimation rates in the RACMO2.3 model results from the drifting snow scheme, as the RACMO2.3 model includes sublimation from ventilated snow, while MAR$_{v3.10}$ only includes wind erosion of surface snow. This also bring the SMB further away from that of MAR$_{v3.10}$, even though the total precipitation is rather similar in the two models, even if differently distributed. MetUM, which performs similarly to RACMO2.3 when compared with SMB observations has lower precipitation and higher sublimation rates than RACMO2.3 however, suggesting that ventilation of drifting snow alone does not explain the higher sublimation rates. MAR$_{v3.10}$ has the lowest of all sublimation rates and COSMO-CLM^2 the highest. In fact our results suggest that the dry bias in COSMO-CLM^2 is a result in part of the lower precipitation values, which are very close to those of the driving ERA-Interim model, but also a consequence of the much higher sublimation values. This dry bias is mostly confined to the coastal regions and peninsula and is identified and discussed in Souverijns et al. (2019). The RACMO2.3 model is still closest to the ensemble mean annual precipitation but the $MAR_{v3.10}$ model mean values are only 10 Gt different to RACMO2.3 and in some years shown in fig. 9 it is actually even closer.

Model	SMB (Gt yr^{-1})	Precipitation (Gt yr^{-1})	Sublimation (Gt yr^{-1})
HIRHAM5(0.44°, 0.11°)	2519 ± 118, 2452 ± 107	2715 ± 117, 2635 ± 107	192 ± 12, 183 ± 10
$MAR_{v3.10}$	2445 ± 91	2567 ±87	122 ± 11
RACMO2.3p2	2397 ± 101	2557 ± 100	158 ± 7
MetUM	2191 ± 101	2366 ± 100	175 ± 9
COSMO-CLM^2	1961 ± 70	2222 ± 72	262 ± 10
ERA-Interim	2016 ± 99	2271 ± 95	255 ± 18
Ensemble mean	2329 ± 94	2498 ± 93	194 ± 9

Table 4. Mean annual SMB and components on common mask for each model averaged over the 1987-2015 period where all the models overlap, standard deviations are also show.

4 Discussion

4.1 The Surface Mass Budget of Antarctica

The range of models in this intercomparison study allows us to estimate not only the likely range of SMB over Antarctica, but also to identify sources of disagreement and bias within and between models. Accounting for differences in ice mask, the ensemble mean annual SMB integrated over the whole of Antarctica between 1987 and ~~2015~~ 2018 is 2329 ± 94 Gt per year. The RACMO2.3p2 model has a value closest to the ensemble mean with the high resolution HIRHAM5 model 190 GT over this number and the COSMO-CLM^2 model 368 Gt below. The HIRHAM5 0.11°and $MAR_{v3.10}$ numbers are almost exactly the same ~~however~~ at 2452 and 2445 respectively around 150 Gt above the mean~~but~~ . MetUM, like COSMO-CLM^2, is much lower at about 138 Gt and 368 Gt below the mean respectively. Given that the models perform fairly similarly when evaluated against SMB observations we here give all models equal weight, although we suspect that there is a dry bias in COSMO-CLM^2 and a wet bias in HIRHAM5 0.44°. With an identical forcing from ERA-Interim, the present day estimate of the surface mass budget of Antarctica ranges from 2519 Gt to 1961 Gt per year, a 558 Gigatonne range that alone is equivalent to around 1.5 mm of global mean sea level rise. ~~Clearly narrowing~~ Narrowing this range for the purposes of estimate sea level change at present and in the future is an important task and for this reason we have evaluated the models against observations in Antarctica (see below).

~~There is~~ It is interesting to compare our results with those used in the IMBIE study of Antarctic mass budget (Shepherd et al., 2019) . 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. Nevertheless it is therefore also important to consider the wide uncertainties in both observations and the likely biases in models discussed in this paper, in assessing the contribution to sea level rise from Antarctica

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. However, shorter Shorter periods within the time series can 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, suggesting that distinguishing 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 emphasising this also emphasises the importance of long time series of observations. SMB variability is a result of low- and mid-latitude weather variability, interannual variability is particularly large at the beginning of the ERA-Interim period up to 1990 and we hypothesise this is related to improved data assimilation in the southern hemisphere in the period between 1979 and 1989 (Dee et al., 2011). The models disagree on both the magnitude and the sign of the overall trend in the 1987 to 2018 common period of all models. It is clear from figure 8 that the external forcing model, in this case ERA-Interim, is extremely important in determining both the total SMB and the year-to-year variability in the SMB trend, even though the absolute values are somewhat dependent on the individual RCM. This is not an unexpected result given that these are all limited area models forced at the boundaries but it has important implications for estimates of future projections of SMB in Antarctica. Decadal and multidecadal scale climate variability expressed in global climate models will have a strong influence on Antarctica mass budget (including the dynamical components via ocean forcing) that may suppress or enhance the anthropogenic forcing in ways that are difficult to predict given the large internal variability in the system. Long climate simulations with large ensembles will be necessary to define the likely range of internal climate variability and this poses challenges of computing resources when regional downscaling is required to represent the spatial patterns of SMB over the ice sheet at high resolution.

Even between models with similar values for the integrated SMB there is substantial spatial variability in the pattern of SMB, as shown by the basin level breakdown in Figure 5 and the variation from the ensemble mean in Figure 6. These together show a nuanced picture. Over most of Antarctica, particularly in the east, the variation between models is rather small, the biggest deviations are largely around the coast. These small areas have a disproportionate influence on the continental integrated SMB values due to high accumulation rates. Basins in west West Antarctica, and particularly on the Antarctica peninsula have very large differences where for example, HIRHAM5 0.11°shows an average annual SMB of 176 Gigatonnes but COSMO-CLM^2 has the lowest estimate of 46 Gigatonnes in the same basin. The MAR model which shows an integrated SMB value similar to HIRHAM5 over the whole continent gives 130 Gt in the same basin, closer to the RACMO2.3p2 value of 134 Gt while MetUM is again lower at 96 Gigatonnes. Clearly, averaging Averaging SMB over the whole continent smooths out a good deal of the spatial variability which in turn is also important for driving ice dynamics. Equally, as some basins especially in west West Antarctica have very high precipitation rates, differences between models in a relatively small areas here can make a large contribution to the difference in the integrated numbers over the whole continent.

Similarly, relatively small differences in ice masks that are primarily in coastal regions with high accumulation rates can lead to relative large differences in SMB estimates (see Figure A1) as Vernon et al. (2013) have also shown in Greenland. Figure A1 in the Appendix compares the ice masks of all the models. We found that although the variation looks quite small, the grid points affected include some of the highest precipitation points within the domain and thus small differences can have large effects. This is one of the main differences between the earlier RACMO2.1, with one of the smallest ice masks, and RACMO2.3 for example. Almost all the other models were larger around the entire coastline. The total SMB integrated over the continent is therefore highly sensitive to the size of the common mask. For example, the SMB for HIRHAM5 0.11°is computed on its native mask gives an integrated SMB on average 9.95% higher compared to the common mask result, even though the native mask is only 2.93% larger than the common mask. These differences suggest that the CORDEX community should agree a common protocol to calculate the ice mask to reduce uncertainties in Antarctic ~~ice sheet~~ SMB. The deviation from ensemble mean SMB shown in Figure 6 suggests that while over the high plateau of east Antarctica there is little deviation in general, much bigger differences occur between model SMB estimates around the Transantarctic mountains where the effect of higher resolution becomes obvious in resolving the topography but model physics also likely plays a role. We see a similar effect in the high relief topography of ~~west~~ West Antarctica. Finally, our results show that between 14% (COSMO-CLM^2) and 19% (MetUM) of the SMB is accounted for by the ice shelves around Antarctica.

A comparison of the high and low resolution HIRHAM5 simulations is interesting here as the models are identical other than resolution. There is a substantial difference in the location of the maximum upslope precipitation as well as the downslope precipitation shadow. We attribute these differences to resolution that allows high resolution simulations to better represent steep topography. A similar but less marked impact is seen between the earlier RACMO2.1P and newer RACMO2.3p2 though in this case changes in model physics may also be responsible.

4.2 Model Evaluation with Observations

~~Clearly, evaluating~~ Evaluating the models against observations is very important for assessing where there are significant biases, but evaluation of model performance is significantly hampered by the lack of observations in key regions. Nonetheless we are able to show that the models have some skill in simulating surface climate, particularly temperature and pressure, as well as SMB. The skill in simulating climate does not however translate perfectly to simulating SMB, partly due to the difficulties of modelling and evaluating precipitation, as our analysis shows, where e.g. COSMO-CLM^2 better simulated surface climate than HIRHAM5 but has an lower skill in SMB. The difference can be explained as variables such as temperature and pressure are more easily measured and models have been optimised to give good performance. Antarctic SMB is dominated by the precipitation term that is much harder to measure accurately and also has much higher uncertainty in models.

SMB observations themselves, are not always very reliable and sub-grid scale surface snow processes, such as the build up of sastrugi can give substantially different results over short spatial scales (Andersen et al., 2006). It is therefore important to break down the data into different regions and elevation classes to see where models have better or weaker performance. We note the scatter in both models and observations within the different elevation bins and that the two polar optimised models (MAR and RACMO) perform, broadly speaking better than the others (see also Figures A2 to A7 in the appendix), though the differences are rather small in some of the elevation bins and not always very significant. It is clear that more work needs to be done to understand exactly how SMB varies spatially over the continent in order to better optimise parameterisations. The use of nudging in models does however seem to make it easier to replicate both observed climate and SMB in RCMs, we discuss further below the use of nudging in regional climate simulations.

4.3 Processes Important for Ice Sheet SMB Processes

Evaluation against observations helps to identify missing and mischaracterised processes within RCMs. Models that have not undergone specific adjustments for Antarctica clearly represent the SMB in Antarctica more poorly than those that have bee adjusted been adjusted in some regions . However table 2 shows this is not unambiguous as in some elevation bands the unmodified models have lower bias and RMSE as also shown in the supplementary material (Figure A8). Other biases are also clear in this analysis. The driest model COSMO-CLM^2, underestimates SMB close to the coast, a region very relevant for total ice sheet mass balance. This is due to an overestimated sublimation amplified by an underestimated snowfall rate close to the coast. High values for the sublimation originate from an underestimated albedo due to aging of the snow that occurs too fast in the model (Souverijns et al., 2019). The low values for the snowfall rate is likely related to cloud microphysics, namely a too slow conversion of ice to snow or a too slow deposition of water vapor on the solid hydrometeors. Currently efforts are ongoing to improve the coastal SMB performance in COSMO-CLM^2. The HIRHAM5 climate simulations both appear to have a wet bias, likely again related to the cloud microphysics and precipitation schemes but also probably a result of a diagnostic precipitation scheme commonly used in hydrostatic models. The models typically have a wet bias on the upslope of steep topography and a dry bias on the downslope. The RACMO2.3 model shows a similar, though less pronounced effect that derives also from the IFS physical schemes (Hermann et al., 2018; Schmidt et al., 2017). New prognostic precipitation schemes have been developed in for numerical weather prediction models to solve this problem (Forbes, 2011) and implementation of a similar prognostic scheme in MAR probably explains the different pattern of SMB in areas with steep topography (Agosta et al., 2019). As RACMO and MAR are the only two models that have a specific subsurface scheme for ice sheets in this model comparison we have excluded discussion of melt and runoff and this will likely be the subject of future work. Given the high amount of precipitation over Antarctica this runoff is still very small in absolute and relative senses but as a warming future climate is expected to bring increasing amounts of melt a more sophisticated treatment that includes refreezing within the snowpack will become increasingly important. More importantly, with respect to the radiative schemes within the models, adding an ice sheet specific snowpack to the surface module in MAR and RACMO does improve the surface temperature (and 10 m snow temperature) and therefore the air temperature. This is clear in fig. 2 and may also be a factor in some of the biases shown in fig. 1. Improving these surface schemes is therefore important not just for future projections of SMB but also to improve the near-surface climate.

4.4 Model Topography and Resolution

The inclusion of two simulations with the HIRHAM5 model, varying only the resolution, allows us to assess the impact that higher resolution has on the results as shown in Fig. 7 and Table 4. The added value from a higher resolution is higher resolution version adds value with higher spatial variability that should better capture local topography and associated weather phenomena. This is especially important in areas of high relief such as in the coastal areas and around the Transantarctic Mountains. These are also the areas where models vary from each other and the ensemble mean the most. While there are very few observations to confirm the better performance on a local scale, the pattern of SMB suggests that the high relief rugged topography is better captured in HIRHAM5 0.11°than 0.44°. However, there is a cost of a high resolution model also. Not only is the higher resolution model is not only more computationally expensive, in a set-up like the one described here simulation where there is no nudging, like here, the larger number of grid points gives increased degrees of freedom for the model to evolve freely and thus introduces more internal variability. While this is not necessarily a problem for climate simulations in the future, the enhanced internal variability is inevitably punished when compared with observations and models that have been internally nudged.

Nudged models (MAR, RACMO, COSMO-CLM^2) show a generally lower variance from the ERA-Interim mean SMB compared to the unnudged models (HIRHAM5, MetUM), though MetUM, run as a hindcast, shows the closest values to ERA-Interim overall. They also show a closer match to observed climate than the unnudged model runs. The advantages of nudged runs are thoroughly explored in van de Berg and Medley (2016) who run two versions of RACMO2 for Antarctica one nudged and one not nudged. They find that RACMO2 nudged gives SMB results that better represent the temporal variability of the observations, because the top of the atmosphere is constrained, thus avoiding the model deviating too far from large scale systems in the mod-latitudesmid-latitudes. The nudging as applied in RACMO is not spectral nudging but relaxation of temperature, pressure and wind fields and this leads to some systematic mid-tropospheric warming, and hence to slightly lower SMB in the interior of Antarctica also. Other studies (Alexandru et al., 2009; Berg et al., 2013) show that spectral nudging can also lead to lower precipitation extremes and reduced vorticity while Akperov et al. (2018) shows better representation of Arctic cyclones in nudged models. The daily reinitialisation and close forcing by ERA-Interim also explains why the MetUM modelled SMB is closest to the ERA-Interim values when integrated over the common mask.The MetUM simulation is a hindcast series where the full prognostic model state is replaced daily or twice-daily. The series is technically made continuous by construction, but it is in fact likely to be discontinuous in terms of energy, momentum and moisture budgets and like all nudged models, they are in general not energy, moisture or momentum conserving. Berg et al. (2013), argue for caution in applying nudging during climate simulations as while it compensates for the RCM's deficiencies in meso and large scale circulation, the assumption is that the driving model represents the large-scale circulation well. In the ERA-Interim re-analysis dataset, this is a minor problem, but for free-running GCMs, large-scale circulation may well be more poorly simulated. As the external forcing controls what s is delivered on the boundaries, future projections of Antarctic climate and ice sheet change will be highly controlled by the quality of the forcing on the RCM boundaries. Models nudged internally within the domain will be further constrained in estimates of SMB by the driving models, implying rigorous assessment of global climate models should be performed before downscaling GCMs for future projections to determine which biases will be introduced (Agosta et al., 2015; Barthel et al., 2019).

5 Conclusions

The Polar CORDEX regional climate simulations for Antarctica are a valuable and freely available dataset for climate researchers. In this paper we have compared the models against each other and against observational datasets. Much more analysis is possible and will be followed up by this group. We hope also to encourage other scientists to make use of the CORDEX dynamically downscaled models. Analysis and model intercomparison is a useful technique to evaluate models and to show directions for model improvements. Our results can be summarised as showing that the RCMs in this analysis produced skillful climate simulations over the Antarctic continent though with more uncertainty surrounding estimates of SMB. There is a high annual and decadal and variability in SMB across Antarctica and no clear long-term trend. Model resolution and model dynamics interact in interesting ways in areas with high relief and complex topography that make it important to focus on observational campaigns in these regions. In particular, we argue that given the importance of precipitation for SMB, new observational programmes are needed that focus on accumulation and snow processes, e.g. stakes, firn cores and radar. Furthermore, focusing new observations in regions (see for example, A9) where there is both a lack of current data and strong disagreement between models will be valuable for understanding climate in Antarctica.

There is closer model agreement on SMB for the interior of the Antarctic ice sheet than there is in the margins and on the Antarctic ~~peninsula~~Peninsula. The largest areas of disagreement between models are primarily in ~~west~~ West Antarctica. In this paper we focus mostly on precipitation and sublimation/evaporation, but reliable subsurface snow and firm schemes will become increasingly important, particularly when making projections of SMB in the future. Models that have been optimised for the Antarctic climate and which incorporate nudging, typically demonstrate more model skill than those which do not.

Data availability. Model outputs used in this paper are available to download from the CORDEX archive, see https://www.cordex.org/data-access/how-to-access-the-data/ for instructions. In addition The COSMO-CLM^2 monthly output of key variables is open-access available (https://doi.org/10.5281/zenodo.2539147). Output for key variables from the high resolution HIRHAM5 simulations are available here: http://ensemblesrt3.dmi.dk/data/prudence/temp/RUM/HIRHAM/ANTARCTICA/ERAI/, further data is available on request. MAR3.10 monthly outputs are available here: ftp://ftp.climato.be/climato/ckittel/MARv3.10/ while all other variables are available on request.

Appendix A: Appendix A: Ice Masks

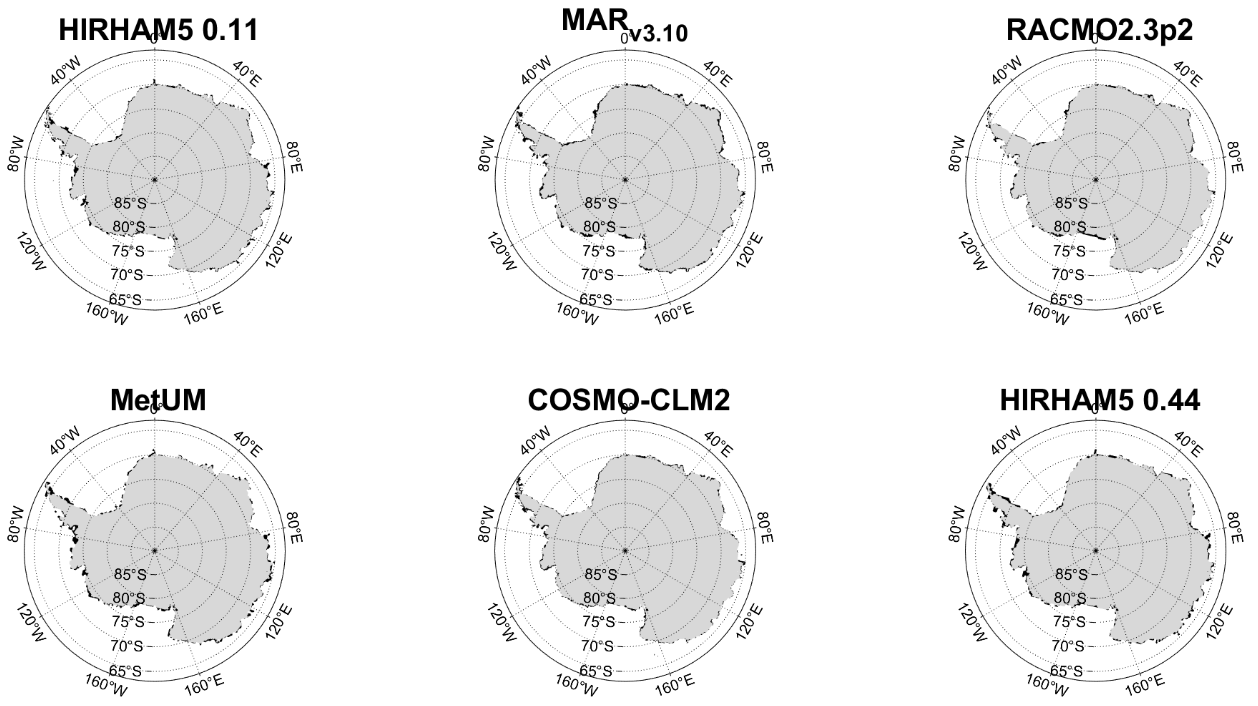


Figure A1. Each sub-figure shows where the common mask and the individual model masks are identical, grey is ice sheet or land in both masks. Black shows variation between model and common masks.

All title masks are larger than the common mask, HIRHAM5 0.11°is 2.43 % larger, MAR$_{v3.10}$ is 2.89 % larger, RACMO2.3P2 is 1.85 % larger, MetUM is 2.49 % larger, COSMO-CLM^2 is 1.94 % larger, and HIRHAM5 0.44°is 2.49 % larger. Some of the differences is due to inclusion of nunataks and mountain ranges within the continent. The common mask also includes nunataks.

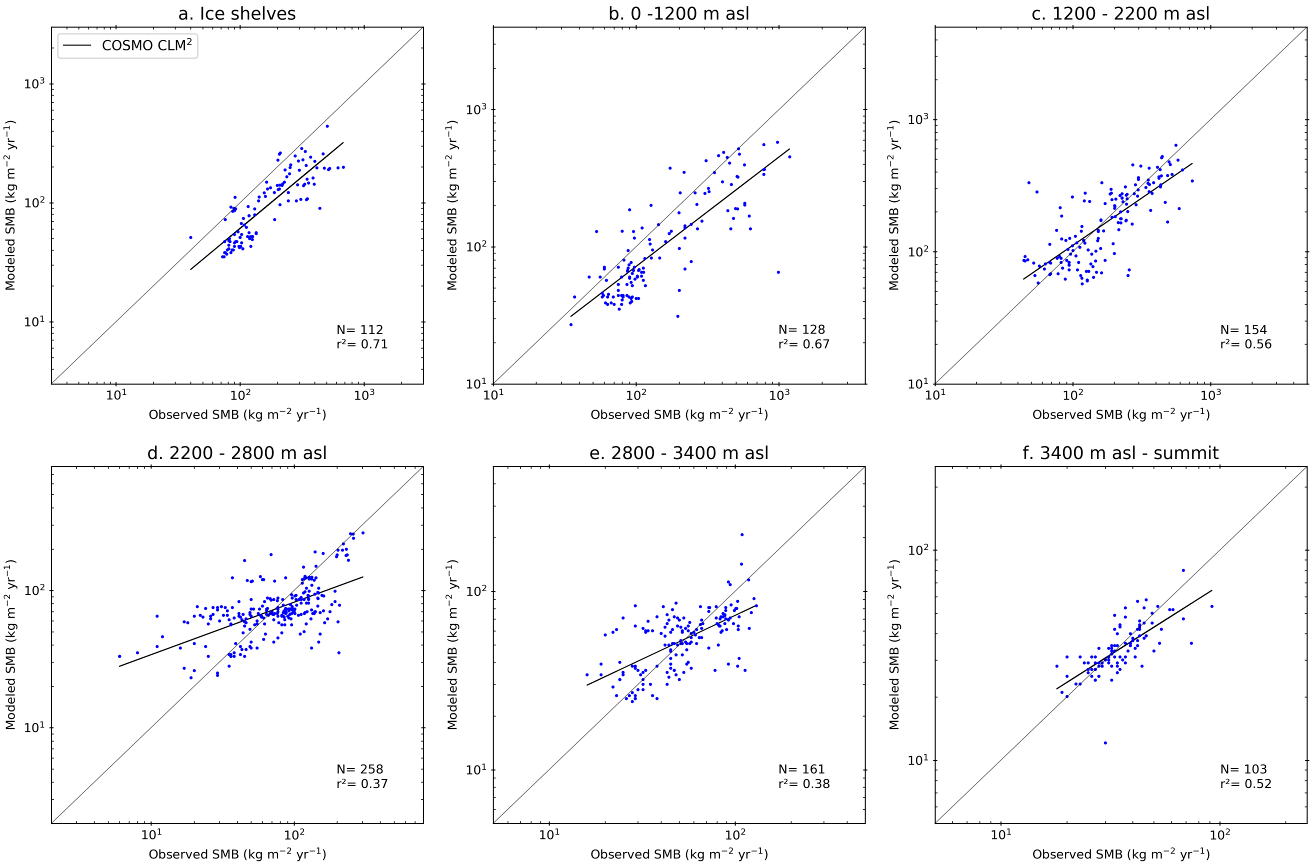


Figure A2. Comparison between COSMO-CLM^2 and observed SMB (units= kg m^{-2} yr^{-1}) over the ice shelves (a) and by elevation classes (b-f). Due to the use of logarithmic axes, only positive values for the observed and modelled SMB from all the RCMs in this study are used (number for each bin N). Finally, the regression coefficient of each regression line is also shown (r^2).

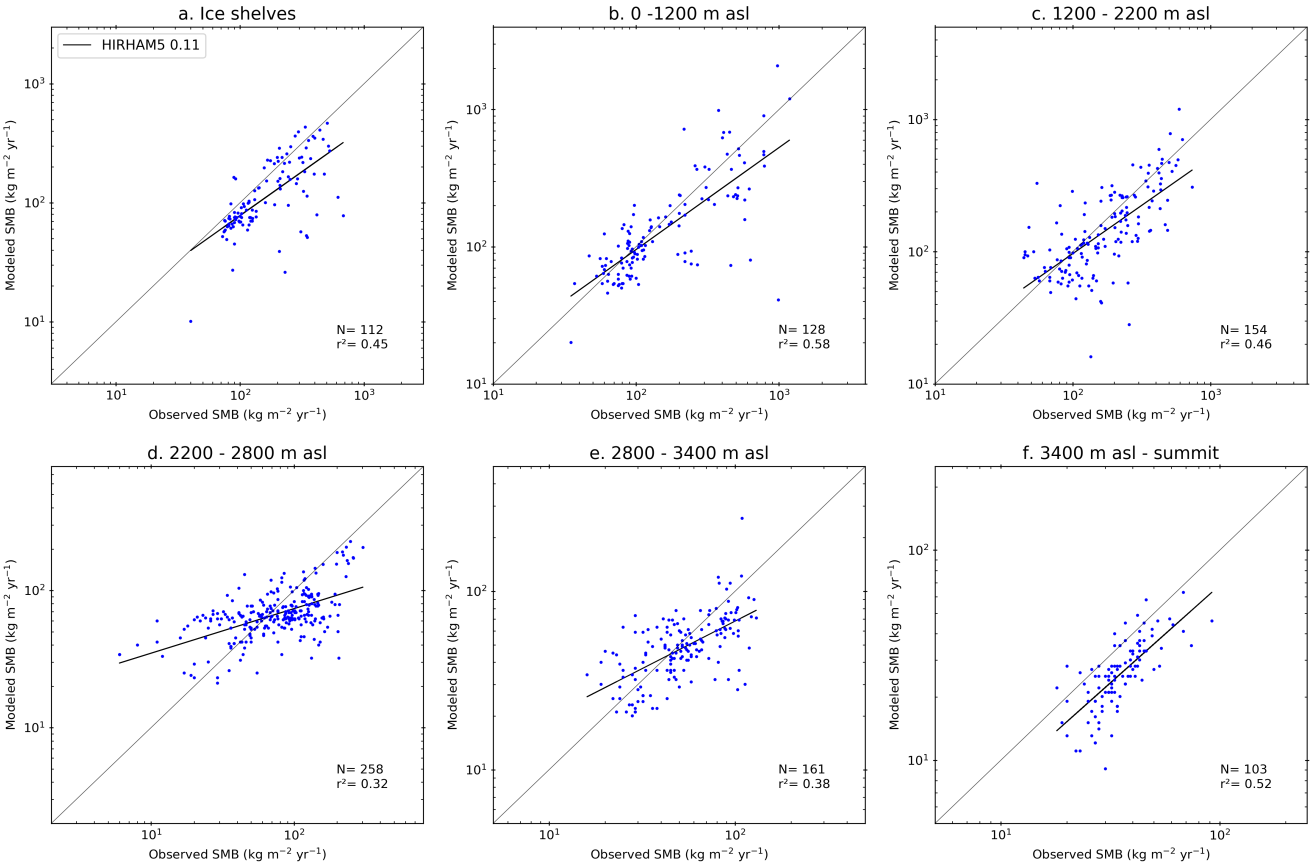


Figure A3. Comparison between HIRHAM5 0.11°and observed SMB (units= kg m^{-2} yr^{-1}) over the ice shelves (a) and by elevation classes (b-f). Due to the use of logarithmic axes, only positive values for the observed and modelled SMB from all the RCMs in this study are used (number for each bin N). Finally, the regression coefficient of each regression line is also shown (r^2).

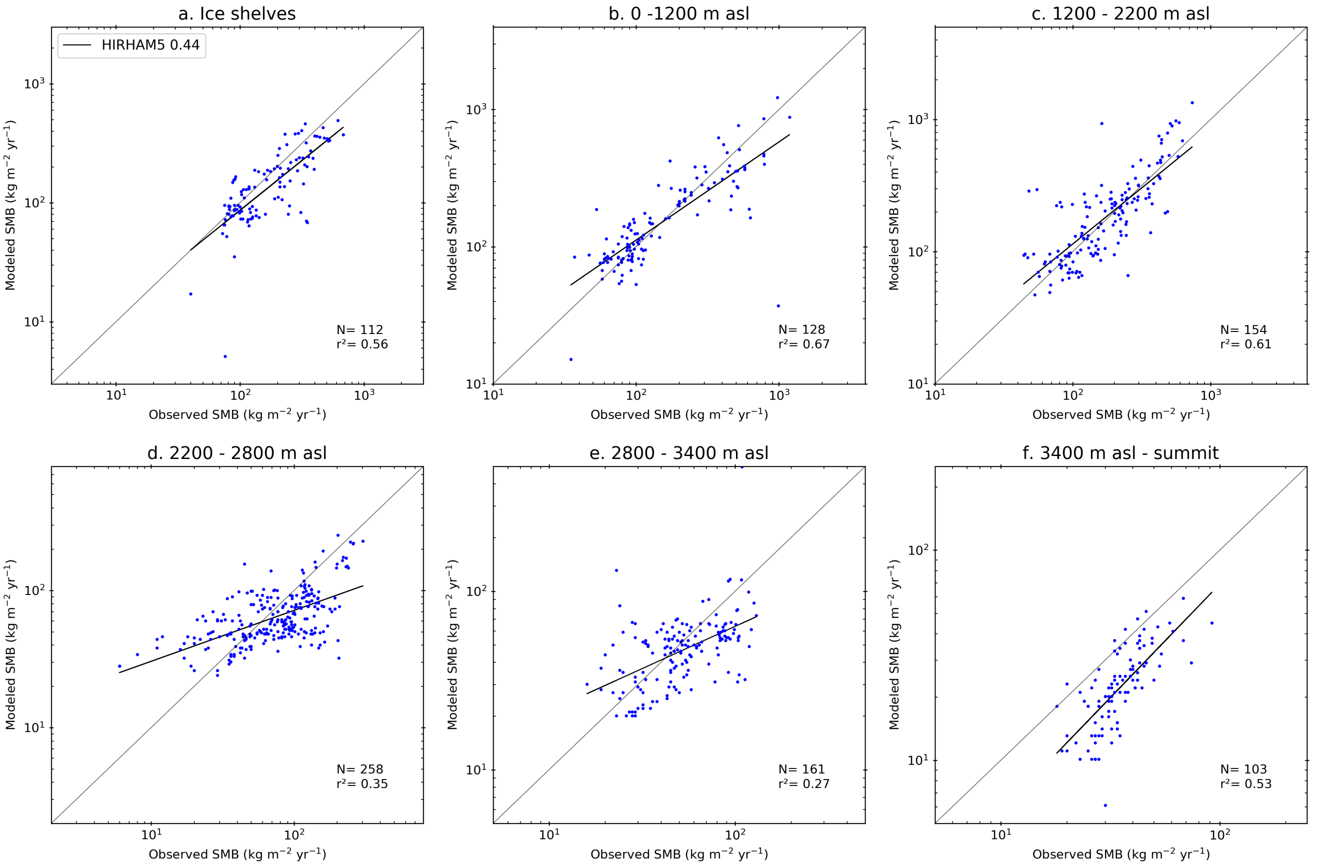


Figure A4. Comparison between HIRHAM5 0.44°and observed SMB (units= kg m^{-2} yr^{-1}) over the ice shelves (a) and by elevation classes (b-f). Due to the use of logarithmic axes, only positive values for the observed and modelled SMB from all the RCMs in this study are used (number for each bin N). Finally, the regression coefficient of each regression line is also shown (r^2).

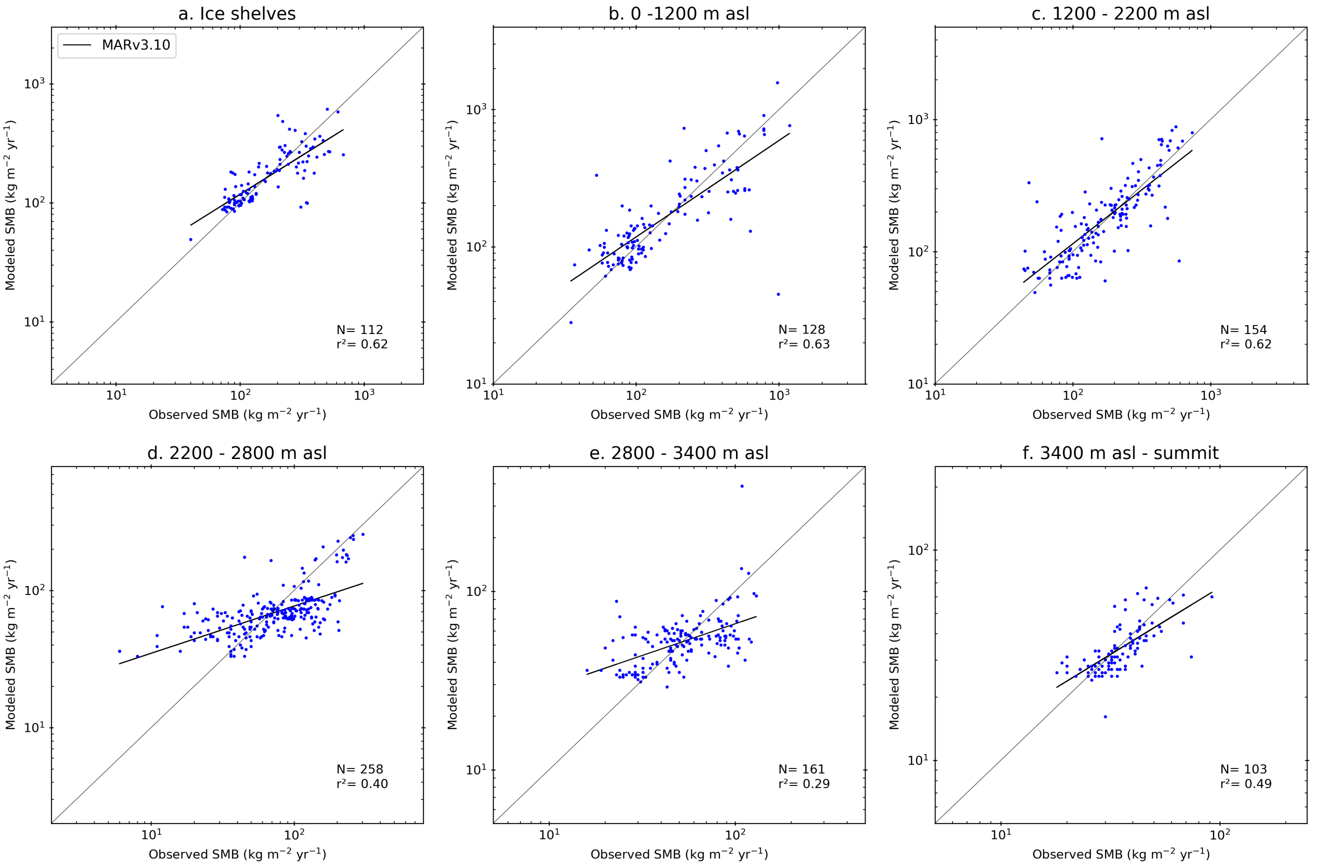


Figure A5. Comparison between $MAR_{v3.10}$ and observed SMB (units= kg m^{-2} yr^{-1}) over the ice shelves (a) and by elevation classes (b-f). Due to the use of logarithmic axes, only positive values for the observed and modelled SMB from all the RCMs in this study are used (number for each bin N). Finally, the regression coefficient of each regression line is also shown (r^2).

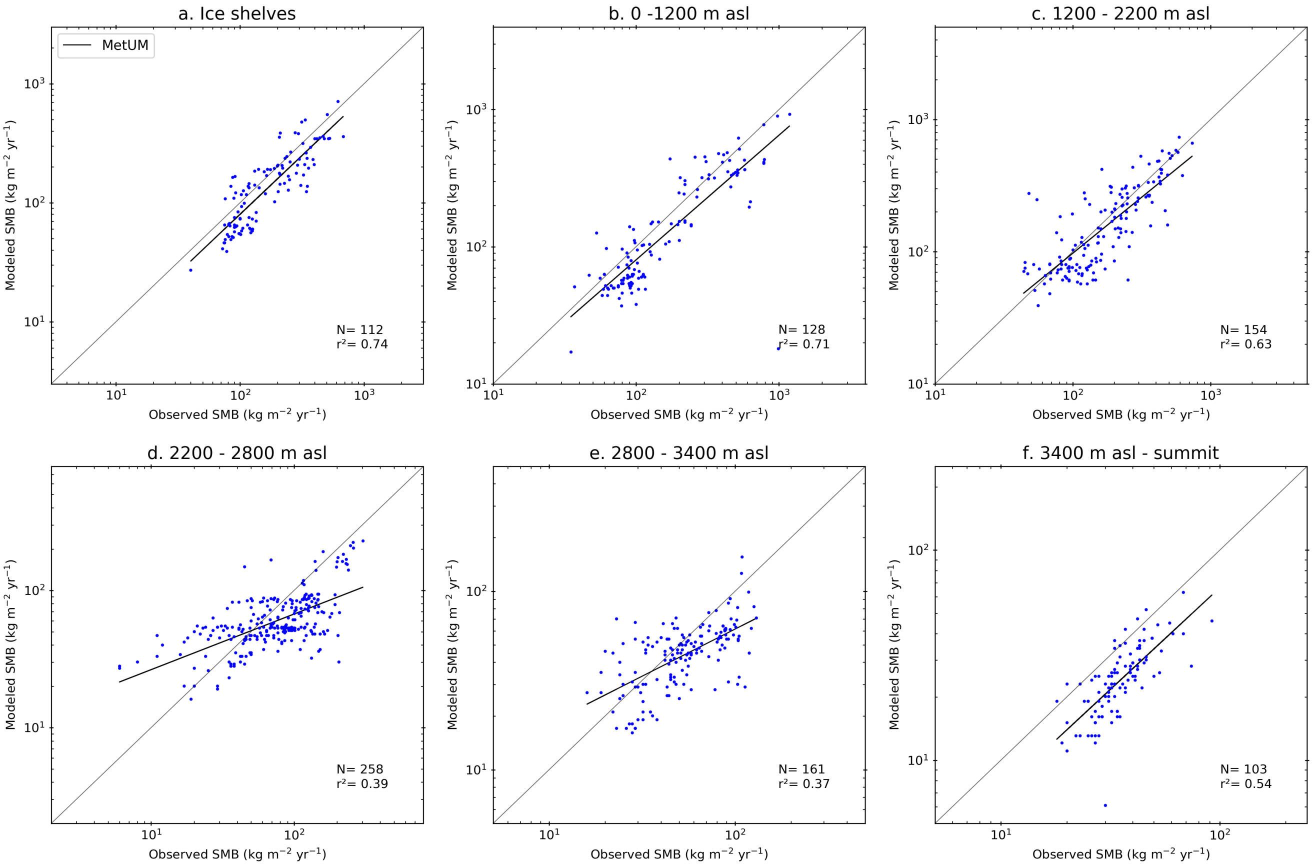


Figure A6. Comparison between MetUM and observed SMB (units= kg m^{-2} yr^{-1}) over the ice shelves (a) and by elevation classes (b-f). Due to the use of logarithmic axes, only positive values for the observed and modelled SMB from all the RCMs in this study are used (number for each bin N). Finally, the regression coefficient of each regression line is also shown (r^2).

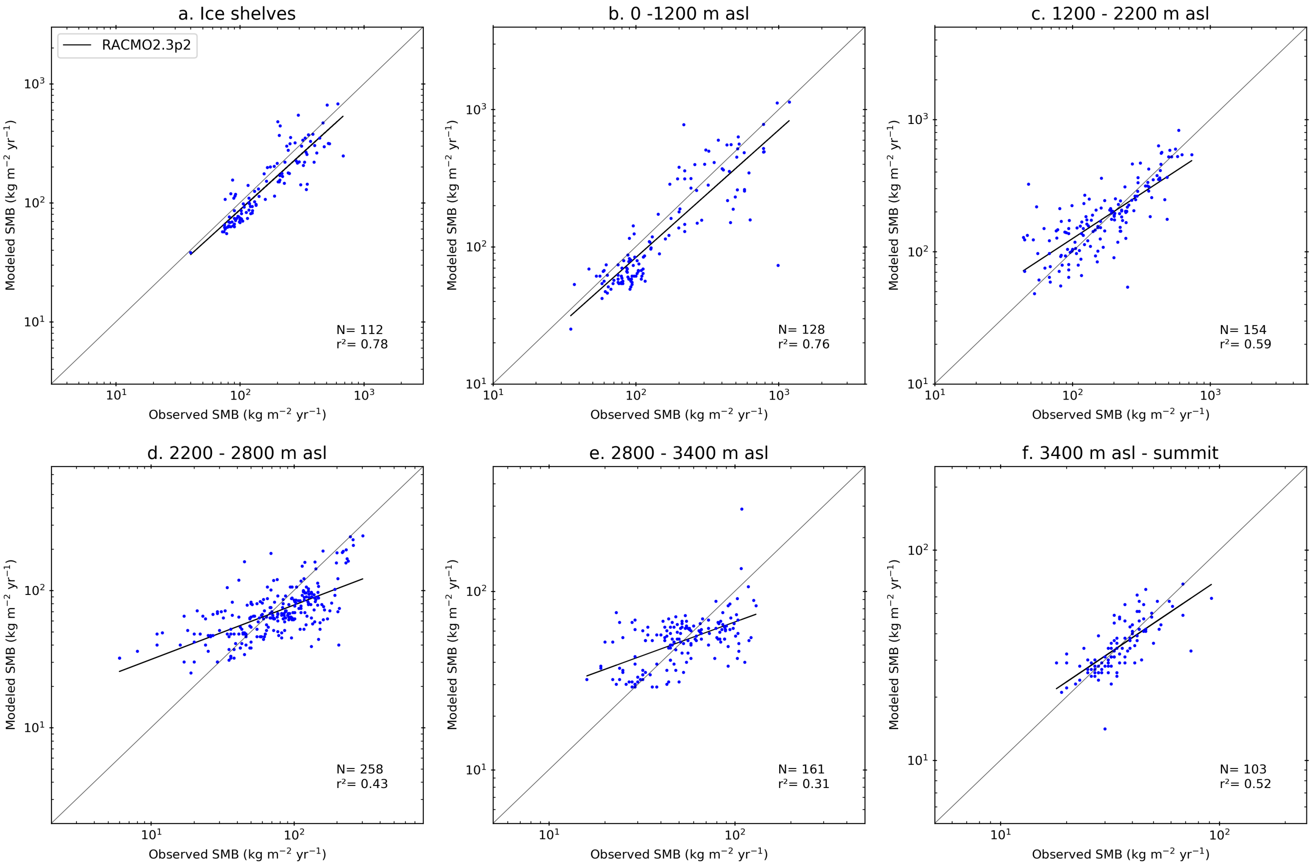


Figure A7. Comparison between RACMO2.3p2 and observed SMB (units= kg m^{-2} yr^{-1}) over the ice shelves (a) and by elevation classes (b-f). Due to the use of logarithmic axes, only positive values for the observed and modelled SMB from all the RCMs in this study are used (number for each bin N). Finally, the regression coefficient of each regression line is also shown (r^2).

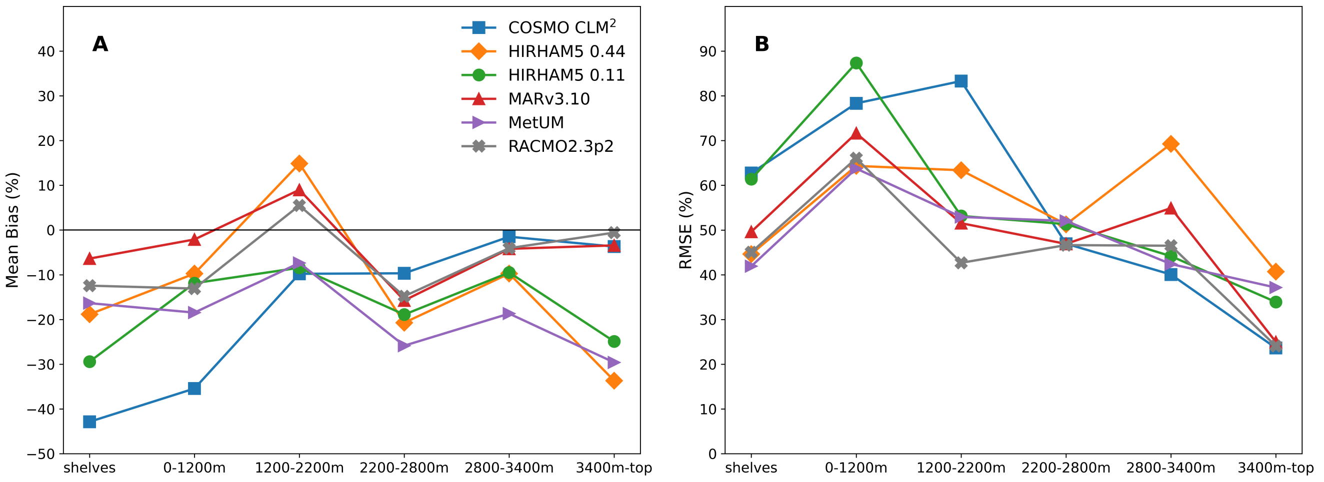


Figure A8. Mean bias and RMSE by elevations bin for each RCM compared to SMB observations as shown in Figures 3 and 4.

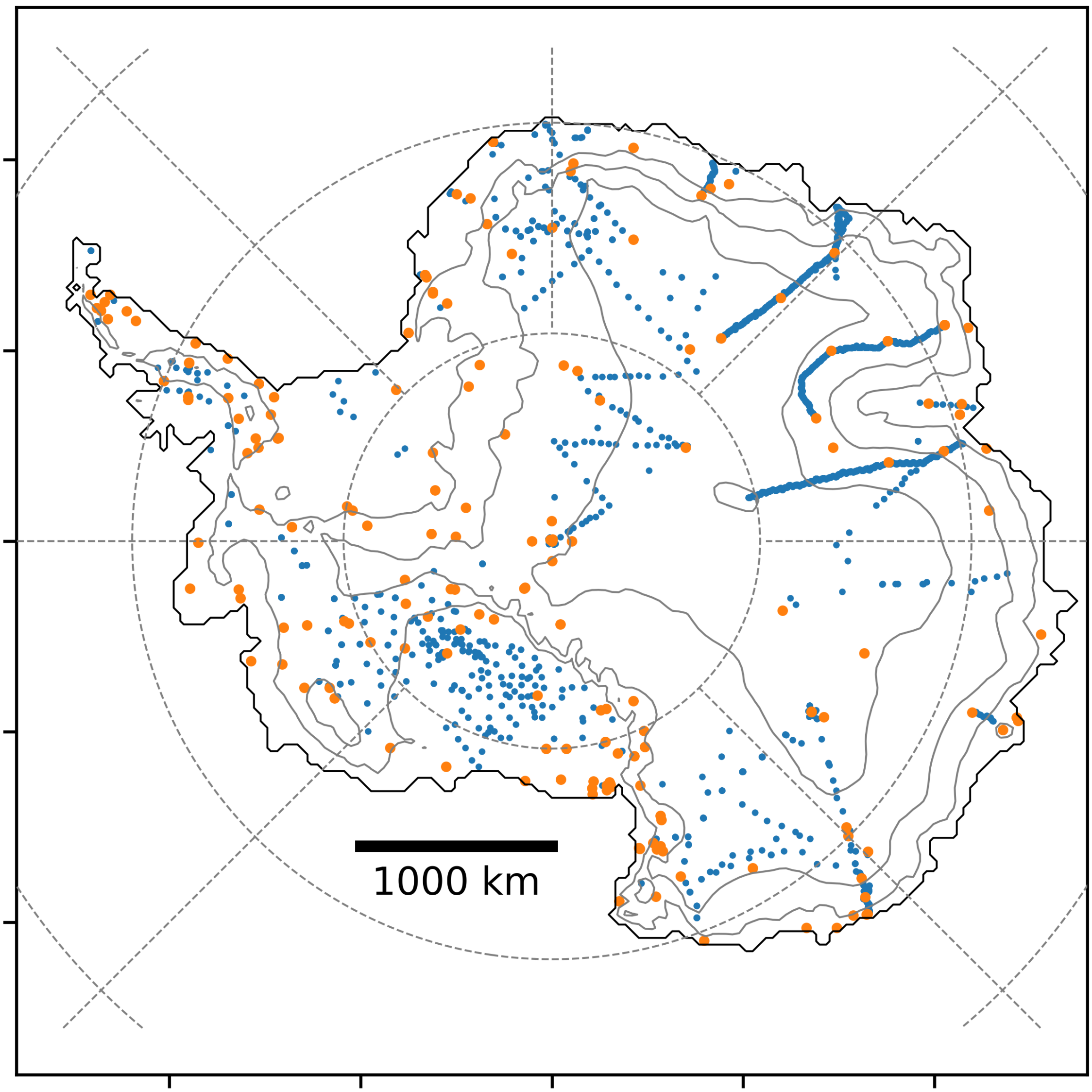


Figure A9. Location of automatic weather stations and SMB observations in Antarctica and used in this study

A1

Author contributions. RM and SBS conceived the study, analysis of simulations was carried out by CK, JMVW, NH and RM. Model simulations were carried out by FB, CK, NS, AG, AO, SW, TP, JWVM, and EVM. All authors contributed to the manuscript.

Competing interests. We declare no competing interests.

Disclaimer. TEXT

Acknowledgements. Acquisition of meteorological data in Adélie Land has been made with the financial and logistical support of the French Polar Institute IPEV (programme CALVA-1013). The authors appreciate the support of the University of Wisconsin-Madison Automatic Weather Station Program for the data set, data display, and information, NSF grant number 1924730.

The COSMO-CLM^2 integrations were supported by the Belgian Science Policy Office (BELSPO; grant number 747 BR/143/A2/AEROCLOUD) and the Research Foundation Flanders (FWO; grant number 748 G0C2215N and GOF5318N (EOS ID: 30454083)). Computational resources and services were provided by the Flemish Supercomputer Center, funded FWO and the Flemish Government - department EWI. Matthias Demuzere, Jan Lenaerts, Irina Gorodetskaya and Sam Vanden Broucke are gratefully acknowledged for supporting the COSMO-CLM^2 integrations. COSMO-CLM^2 is the community model of the German regional climate research jointly further developed by the CLM-Community. Computational resources for MAR simulations have been provided by the Consortiumdes Équipements de Calcul Intensif (CÉCI), funded by the Fondsde la Recherche Scientifique de Belgique (F.R.S. – FNRS) undergrant no. 2.5020.11 and the Tier-1 supercomputer (Zenobe) of theFédération Wallonie Bruxelles infrastructure funded by the WalloonRegion under grant agreement no. 1117545. C. Kittel's work was suppported by the Fonds de la Recherche Scientifique – FNRS undergrant no. T.0002.16

R Mottram, N. Hansen and S.B. Simonsen acknowledge the ESA Climate change initiative for the Greenland ice sheet funded via ESA-ESRIN contract number 4000104815/11/I-NB and the Sea Level Budget Closure CCI Project funded via ESA-ESRIN contract number 4000119910/17/I-NB. HIRHAM5 regional climate model simulations were carried out by R.~~M.~~ Mottram and F.~~B.~~ Boberg as part of the ice2ice project, a European Research Council project under the European Community's Seventh Framework Programme (FP7/ 2007-2013)/ ERC grant agreement 610055. Data analysis was supported by the Danish State through the National Centre for Climate Research (NCKF).

References

Agosta, C., Fettweis, X., and Datta, R.: Evaluation of the CMIP5 models in the aim of regional modelling of the Antarctic surface mass balance, The Cryosphere, 9, 2311–2321, 2015.

Agosta, C., Amory, C., Kittel, C., Orsi, A., Favier, V., Gallée, H., van den Broeke, M. R., Lenaerts, J. T. M., van Wessem, J. M., van de Berg, W. J., and Fettweis, X.: Estimation of the Antarctic surface mass balance using the regional climate model MAR (1979-2015) and identification of dominant processes, The Cryosphere, 13, 281–296, https://doi.org/10.5194/tc-13-281-2019, https://www.the-cryosphere.net/13/281/2019/, 2019.

Akperov, M., Rinke, A., Mokhov, I. I., Matthes, H., Semenov, V. A., Adakudlu, M., Cassano, J., Christensen, J. H., Dembitskaya, M. A., Dethloff, K., et al.: Cyclone activity in the Arctic from an ensemble of regional climate models (Arctic CORDEX), Journal of Geophysical Research: Atmospheres, 123, 2537–2554, 2018.

Alexandru, A., de Elia, R., Laprise, R., Separovic, L., and Biner, S.: Sensitivity Study of Regional Climate Model Simulations to Large-Scale Nudging Parameters, Monthly Weather Review, 137, 1666–1686, https://doi.org/10.1175/2008MWR2620.1, http://journals.ametsoc.org/doi/10.1175/2008MWR2620.1, 2009.

Amory, C.: Drifting snow statistics from multiple-year autonomous measurements in Adelie Land, eastern Antarctica, The Cryosphere Discussions, 2019, 1–15, https://doi.org/10.5194/tc-2019-164, https://www.the-cryosphere-discuss.net/tc-2019-164/, 2019.

Andersen, K. K., Ditlevsen, P. D., Rasmussen, S. O., Clausen, H. B., Vinther, B. M., Johnsen, S. J., and Steffensen, J. P.: Retrieving a common accumulation record from Greenland ice cores for the past 1800 years, Journal of Geophysical Research: Atmospheres, 111, 2006.

Bamber, J. and Gomez-Dans, J.: A new 1 km digital elevation model of the Antarctic derived from combined satellite radar and laser data – Part 1: Data and methods, The Cryosphere, 3, 101–111, http://content.imamu.edu.sa/Scholars/it/net/tc-3-101-2009.pdf, 2009.

Barthel, A., Agosta, C., Little, C. M., Hattermann, T., Jourdain, N. C., Goelzer, H., Nowicki, S., Seroussi10, H., Straneo11, F., and Bracegirdle12, T. J.: CMIP5 model selection for ISMIP6 ice sheet model forcing: Greenland and Antarctica, The Cryosphere Discussions, pp. 1–34, 2019.

Berg, P., DÃ¶scher, R., and Koenigk, T.: Impacts of using spectral nudging on regional climate model RCA4 simulations of the Arctic, Geoscientific Model Development, 6, 849–859, https://doi.org/10.5194/gmd-6-849-2013, https://www.geosci-model-dev.net/6/849/2013/, 2013.

Brun, E., David, P., Subul, M., and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forescating, Journal of Glaciology, 38, 13–22, 1992.

Christensen, O., Gutowski, W., Nikulin, G., and Legutke, S.: CORDEX Archive design, Danish Meteorological Institute, 2014.

Cogley, J. G., Hock, R., Rasmussen, L., Arendt, A., Bauder, A., Braithwaite, R., Jansson, P., Kaser, G., Möller, M., Nicholson, L., et al.: Glossary of glacier mass balance and related terms, IHP-VII technical documents in hydrology, 86, 965, 2011.

Cook, A. J., Murray, T., Luckman, A., Vaughan, D. G., and Barrand, N. E.: A new 100-m Digital Elevation Model of the Antarctic Peninsula derived from ASTER Global DEM: methods and accuracy assessment, Earth System Science Data, 4, 129–142, https://doi.org/10.5194/essd-4-129-2012, http://www.earth-syst-sci-data.net/4/129/2012/, 2012.

Cox, P., Betts, R., Bunton, C., Essery, R., Rowntree, P., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Climate Dynamics, 15, 183–203, 1999.

Das, I., Bell, R. E., Scambos, T. A., Wolovick, M., Creyts, T. T., Studinger, M., Frearson, N., Nicolas, J. P., Lenaerts, J. T., and Van Den Broeke, M. R.: Influence of persistent wind scour on the surface mass balance of Antarctica, Nature Geoscience, 6, 367, 2013.

De Ridder, K. and Gallée, H.: Land surface-induce regional climate change in Southern Israel, Journal of Applied Meteorology, 37, 1470–1485, 1998.

DeConto, R. M. and Pollard, D.: Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO 2, Nature, 421, 245, 2003.

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, 2011.

Edwards, T. L., Brandon, M. A., Durand, G., Edwards, N. R., Golledge, N. R., Holden, P. B., Nias, I. J., Payne, A. J., Ritz, C., and Wernecke, A.: Revisiting Antarctic ice loss due to marine ice-cliff instability, Nature, 566, 58, 2019.

Eerola, K.: About the performance of HIRLAM version 7.0, Hirlam Newsletter, 51, 93–102, 2006.

Ettema, J., Van den Broeke, M., Van Meijgaard, E., and Van de Berg, W.: Climate of the Greenland ice sheet using a high-resolution climate model-Part 2: Near-surface climate and energy balance, The Cryosphere, 4, 529–544, 2010.

Favier, V., Agosta, C., Parouty, S., Durand, G., Delaygue, G., Gallée, H., Drouet, A.-S., Trouvilliez, A., and Krinner, G.: An updated and quality controlled surface mass balance dataset for Antarctica, The Cryosphere, 2013.

Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., and Gallée, H.: Reconstructions of the 1900-2015 Greenland ice sheet surface mass balance using the regional climate MAR model, The Cryosphere, 11, 1015–1033, https://doi.org/10.5194/tc-2016-268, 2017.

Forbes, R., T. A. U. A.: A new prognostic bulk microphysics scheme for the IFS, Technical Memorandum 649, ECMWF, http://www.ecmwf.int/sites/default/files/elibrary/2011/9441-new-prognostic-bulk-microphysics-scheme-ifs.pdf, 2011.

Franco, B., Fettweis, X., Lang, C., and Erpicum, M.: Impact of spatial resolution on the modelling of the Greenland ice sheet surface mass balance between 1990–2010, using the regional climate model MAR, Cryosphere, 6, 695–711, 2012.

Gallée, H. and Schayes, G.: Development of a Three-Dimensional Meso-γ Primitive Equation Model: Katabatic Winds Simulation in the Area of Terra Nova Bay, Antarctica, Monthly Weather Review, 122, 671–685, https://doi.org/10.1175/1520-0493(1994)122<0671:DOATDM>2.0.CO;2, 1994.

Hansen, N.: Modelling the present day and future climate of Antarctica, Master's thesis, The Technical University of Denmark, Anker Engelunds Vej 1 Bygning 101A 2800 Kgs. Lyngby, Denmark, 2019.

Hermann, M., Box, J. E., Fausto, R. S., Colgan, W. T., Langen, P. L., Mottram, R., Wuite, J., Noël, B., van den Broeke, M. R., and van As, D.: Application of PROMICE Q-Transect in Situ Accumulation and Ablation Measurements (2000–2017) to Constrain Mass Balance at the Southern Tip of the Greenland Ice Sheet, Journal of Geophysical Research: Earth Surface, 123, 1235–1256, https://doi.org/10.1029/2017JF004408, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JF004408, 2018.

Kittel, C., Amory, C., Agosta, C., Delhasse, A., Doutreloup, S., Huot, P.-V., Wyard, C., Fichefet, T., and Fettweis, X.: Sensitivity of the current Antarctic surface mass balance to sea surface conditions using MAR, The Cryosphere, 12, 3827–3839, https://doi.org/10.5194/tc-12-3827-2018, https://www.the-cryosphere.net/12/3827/2018/, 2018.

Krinner, G., Guicherd, B., Ox, K., Genthon, C., and Magand, O.: Influence of oceanic boundary conditions in simulations of Antarctic climate and surface mass balance change during the coming century, Journal of Climate, 21, 938–962, 2008.

Kuipers Munneke, P., Van den Broeke, M., Lenaerts, J., Flanner, M., Gardner, A., and Van de Berg, W.: A new albedo parameterization for use in climate models over the Antarctic ice sheet, Journal of Geophysical Research: Atmospheres, 116, 2011.

Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H., and Box, J. E.: Liquid water flow and retention on the Greenland ice sheet in the regional climate model HIRHAM5: Local and large-scale impacts, Frontiers in Earth Science, 4, 110, 2017.

Lenaerts, J. T., Van den Broeke, M., Déry, S., Van Meijgaard, E., Van de Berg, W., Palm, S. P., and Sanz Rodrigo, J.: Modeling drifting snow in Antarctica with a regional climate model: 1. Methods and model evaluation, Journal of Geophysical Research: Atmospheres, 117, 2012a.

Lenaerts, J. T., Van Den Broeke, M. R., Scarchilli, C., and Agosta, C.: Impact of model resolution on simulated wind, drifting snow and surface mass balance in Terre Adélie, East Antarctica, Journal of Glaciology, 58, 821–829, 2012b.

Lenaerts, J. T., Vizcaino, M., Fyke, J., Van Kampenhout, L., and van den Broeke, M. R.: Present-day and future Antarctic ice sheet climate and surface mass balance in the Community Earth System Model, Climate Dynamics, 47, 1367–1381, 2016.

Lenaerts, J. T., Medley, B., van den Broeke, M. R., and Wouters, B.: Observing and Modeling Ice-Sheet Surface Mass Balance, Reviews of Geophysics, 2019.

Ligtenberg, S., Helsen, M., and Van den Broeke, M.: An improved semi-empirical model for the densification of Antarctic firn, The Cryosphere, 5, 809–819, 2011.

Lucas-Picher, P., Wulff-Nielsen, M., Christensen, J. H., Aðalgeirsdóttir, G., Mottram, R., and Simonsen, S. B.: Very high resolution regional climate model simulations over Greenland: Identifying added value, Journal of Geophysical Research: Atmospheres, 117, 2012.

Medley, B., Joughin, I., Smith, B., Das, S. B., Steig, E. J., Conway, H., Gogineni, S., Lewis, C., Criscitiello, A. S., McConnell, J. R., et al.: Constraining the recent mass balance of Pine Island and Thwaites glaciers, West Antarctica, with airborne observations of snow accumulation, The Cryosphere, 8, 1375–1392, 2014.

Mikkelsen, T. B., Grinsted, A., and Ditlevsen, P.: Influence of temperature fluctuations on equilibrium ice sheet volume, The Cryosphere, 12, 39–47, https://doi.org/10.5194/tc-12-39-2018, https://tc.copernicus.org/articles/12/39/2018/, 2018.

Nowicki, S. M., Payne, T., Larour, E., Seroussi, H., Goelzer, H., Lipscomb, W., Gregory, J., Abe-Ouchi, A., and Shepherd, A.: Ice sheet model intercomparison project (ISMIP6) contribution to CMIP6, Geoscientific model development, 9, 4521, 2016.

Oleson, K. W. and Lawrence, D. M.: Technical description of version 4.5 of the Community Land Model (CLM), Tech. Rep. July, NCAR, Boulder, Colorado, https://doi.org/10.5065/D6RR1W7M, 2013.

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A. K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., Deconto, R. M., Ghosh, T. J. H., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvar, Z.: Sea level rise and implications for low lying Islands, coasts and communities, 2019.

Orr, A., Hosking, J., Hoffmann, L., Keeble, J., Dean, S., Roscoe, H., Abraham, N., Vosper, S., and Braesicke, P.: Inclusion of mountain wave-induced cooling for the formation of PSCs over the Antarctic Peninsula in a chemistry-climate model, Atmospheric chemistry and physics, 15, 1071–1086, 2015.

Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A., and Lenaerts, J. T.: Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise, Geophysical Research Letters, 38, 2011.

Robel, A. A., Seroussi, H., and Roe, G. H.: Marine ice sheet instability amplifies and skews uncertainty in projections of future sea-level rise, Proceedings of the National Academy of Sciences, 116, 14 887–14 892, 2019.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), Meteorologische Zeitschrift, 17, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, 2008.

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., et al.: The atmospheric general circulation model ECHAM 5. PART I: Model description, 2003.

Scambos, T. and Shuman, C.: Comment on 'Mass gains of the Antarctic ice sheet exceed losses' by H. J. Zwally and others, Journal of Glaciology, 62, 599–603, https://doi.org/10.1017/jog.2016.59, 2016.

Scambos, T. A., Frezzotti, M., Haran, T., Bohlander, J., Lenaerts, J., Van Den Broeke, M., Jezek, K., Long, D., Urbini, S., Farness, K., et al.: Extent of low-accumulation'wind glaze'areas on the East Antarctic plateau: implications for continental ice mass balance, Journal of glaciology, 58, 633–647, 2012.

Schmidt, L. S., Adalgeirsdóttir, G., GuÃ°mundsson, S., Langen, P. L., Pálsson, F., Mottram, R., Gascoin, S., and Björnsson, H.: The importance of accurate glacier albedo for estimates of surface mass balance on Vatnajökull: evaluating the surface energy budget in a regional climate model with automatic weather station observations, The Cryosphere, 11, 1665–1684, https://doi.org/10.5194/tc-11-1665-2017, https://www.the-cryosphere.net/11/1665/2017/, 2017.

Sedlar, J., Tjernström, M., Mauritsen, T., Shupe, M. D., Brooks, I. M., Persson, P. O. G., Birch, C. E., Leck, C., Sirevaag, A., and Nicolaus, M.: A transitioning Arctic surface energy budget: the impacts of solar zenith angle, surface albedo and cloud radiative forcing, Climate dynamics, 37, 1643–1660, 2011.

Shepherd, A., Fricker, H. A., and Farrell, S. L.: Trends and connections across the Antarctic cryosphere, Nature, 558, 223, 2018a.

Shepherd, A., Ivins, E., Rignot, E., Smith, B., Van Den Broeke, M., Velicogna, I., Whitehouse, P., Briggs, K., Joughin, I., Krinner, G., et al.: Mass balance of the Antarctic Ice Sheet from 1992 to 2017, Nature, 558, 219–222, 2018b.

Shepherd, A., Gilbert, L., Muir, A. S., Konrad, H., McMillan, M., Slater, T., Briggs, K. H., Sundal, A. V., Hogg, A. E., and Engdahl, M.: Trends in Antarctic Ice Sheet Elevation and Mass, Geophysical Research Letters, 2019.

Souverijns, N., Gossart, A., Demuzere, M., Lenaerts, J. T. M., Medley, B., Gorodetskaya, I. V., Vanden Broucke, S., and van Lipzig, N. P. M.: A New Regional Climate Model for POLAR-CORDEX: Evaluation of a 30-Year Hindcast with COSMO-CLM2 Over Antarctica, Journal of Geophysical Research: Atmospheres, 124, 1405–1427, https://doi.org/10.1029/2018JD028862, 2019.

Stokes, C. R., Sanderson, J. E., Miles, B. W., Jamieson, S. S., and Leeson, A. A.: Widespread distribution of supraglacial lakes around the margin of the East Antarctic Ice Sheet, Scientific reports, 9, 1–14, 2019.

Sutter, J., Gierz, P., Grosfeld, K., Thoma, M., and Lohmann, G.: Ocean temperature thresholds for Last Interglacial West Antarctic Ice Sheet collapse, Geophysical Research Letters, 43, 2675–2682, 2016.

Thomas, E. R., Melchior Van Wessem, J., Roberts, J., Isaksson, E., Schlosser, E., Fudge, T. J., Vallelonga, P., Medley, B., Lenaerts, J., Bertler, N., et al.: Regional Antarctic snow accumulation over the past 1000 years, Climate of the Past, 13, 1491–1513, 2017.

Trusel, L. D., Das, S. B., Osman, M. B., Evans, M. J., Smith, B. E., Fettweis, X., McConnell, J. R., Noël, B. P., and van den Broeke, M. R.: Nonlinear rise in Greenland runoff in response to post-industrial Arctic warming, Nature, 564, 104, 2018.

Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A., and Iagovkina, S.: The SCAR READER project: toward a high-quality database of mean Antarctic meteorological observations, Journal of Climate, 17, 2890–2898, 2004.

Undén, P., Rontu, L., Jarvinen, H., Lynch, P., Calvo Sánchez, F. J., Cats, G., Cuxart, J., Eerola, K., Fortelius, C., García-Moya, J. A., et al.: HIRLAM-5 scientific documentation, 2002.

van de Berg, W. J. and Medley, B.: Brief Communication: Upper-air relaxation in RACMO2 significantly improves modelled interannual surface mass balance variability in Antarctica, The Cryosphere, 10, 459–463, 2016.

van Kampenhout, L., Lenaerts, J. T. M., Lipscomb, W. H., Sacks, W. J., Lawrence, D. M., Slater, A. G., and van den Broeke, M. R.: Improving the Representation of Polar Snow and Firn in the Community Earth System Model, J. Adv. Model. Earth Syst., 9, 2583—-2600, https://doi.org/10.1002/2017MS000988, 2017.

Van Wessem, J., Reijmer, C., Morlighem, M., Mouginot, J., Rignot, E., Medley, B., Joughin, I., Wouters, B., Depoorter, M., Bamber, J., et al.: Improved representation of East Antarctic surface mass balance in a regional atmospheric climate model, Journal of Glaciology, 2014.

van Wessem, J. M., Reijmer, C. H., Lenaerts, J. T. M., van de Berg, W. J., van den Broeke, M. R., and van Meijgaard, E.: Updated cloud physics in a regional atmospheric climate model improves the modelled surface energy balance of Antarctica, The Cryosphere, 8, 125–135, https://doi.org/10.5194/tc-8-125-2014, https://www.the-cryosphere.net/8/125/2014/, 2014.

van Wessem, J. M., van de Berg, W. J., Noël, B. P. Y., van Meijgaard, E., Birnbaum, G., Jakobs, C. L., Krüger, K., Lenaerts, J. T. M., Lhermitte, S., Ligtenberg, S. R. M., Medley, B., Reijmer, C. H., van Tricht, K., Trusel, L. D., van Ulft, L. H., Wouters, B., Wuite, J., Amory, C., and van den Broeke, M. R.: Modelling the climate and surface mass balance of polar ice sheets using RACMO2, part 2: Antarctica (1979–2016), The Cryosphere, 2018, 1–35, https://doi.org/10.5194/tc-2017-202, 2018.

Vernon, C. L., Bamber, J. L., Box, J. E., van den Broeke, M. R., Fettweis, X., Hanna, E., and Huybrechts, P.: Surface mass balance model intercomparison for the Greenland ice sheet, The Cryosphere, 7, 599–614, https://doi.org/10.5194/tc-7-599-2013, http://www.the-cryosphere.net/7/599/2013/, 2013.

von Storch, H., Langenberg, H., and Feser, F.: A Spectral Nudging Technique for Dynamical Downscaling Purposes, Monthly Weather Review, 128, 3664–3673, https://doi.org/10.1175/1520-0493(2000)128<3664:ASNTFD>2.0.CO;2, 2000.

Walters, D., Brooks, M., Boutle, I., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., et al.: The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations, Geoscientific Model Development, 10, 1487–1520, 2017.

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., et al.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geoscientific Model Development, 12, 1909–1963, 2019.

Wang, Y., Ding, M., Van Wessem, J., Schlosser, E., Altnau, S., van den Broeke, M. R., Lenaerts, J. T., Thomas, E. R., Isaksson, E., Wang, J., et al.: A comparison of Antarctic Ice Sheet surface mass balance from atmospheric climate models and in situ observations, Journal of Climate, 29, 5317–5337, 2016.

Zentek, R. and Heinemann, G.: Verification of the regional atmospheric model CCLM v5. 0 with conventional data and Lidar measurements in Antarctica, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-141, 2019.

Zwally, H. J., Li, J., Robbins, J. W., Saba, J. L., Yi, D., and Brenner, A. C.: Mass gains of the Antarctic ice sheet exceed losses, Journal of Glaciology, 61, 1019–1036, https://doi.org/10.3189/2015JoG15J071, 2015.