Response to Reviewers Round 2:

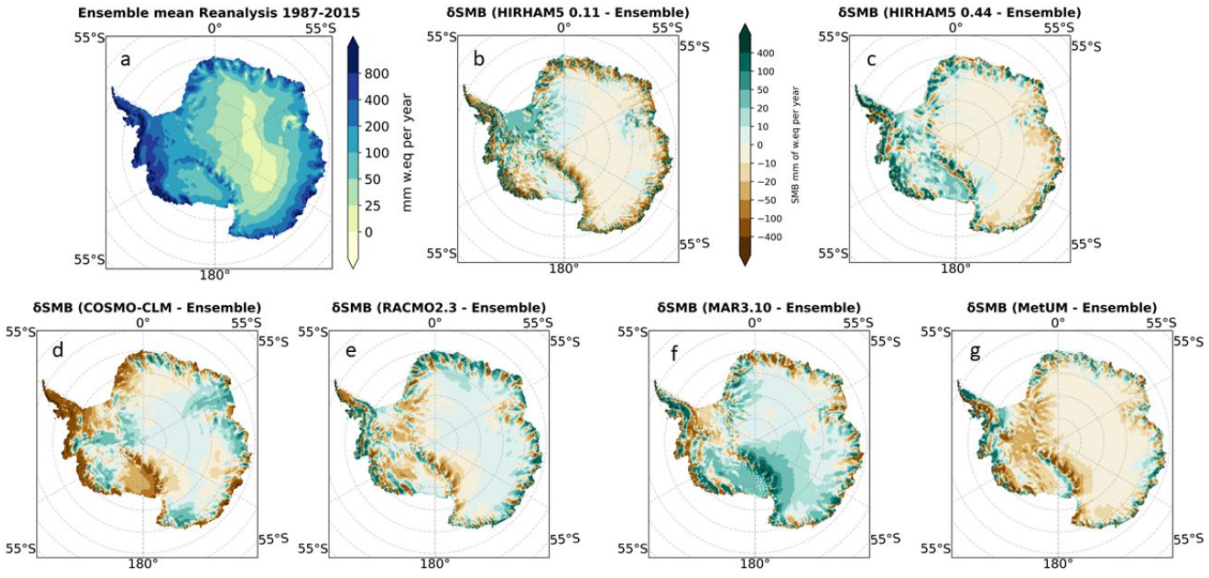
We thank the reviewers and editor again for their very helpful feedback and have endeavoured to implement their suggestions where appropriate, including adding a new figure and substantially revising others to improve readability. Please see below for detailed responses to the comments in blue text..

Editor:
Although still quantitative words are missing as pointed out by the referees,

We have conducted a thorough proofread and inserted more quantitative description of our findings throughout to give a more quantitative understanding. .

figures are providing substantial amount of new information in this work.
Figure 4 compares model SMB and observed SMB. It shows both the mean and the bias with the color bar reflecting the wide range of SMB of 10s kg m-2 yr-1 in the broad inland (above 3000m altitude) to nearly 1000 kg m-2 yr-1 by using unequally spaced color scale, nicely. On the other hand, Figure 6 summarizes the ensemble mean of SMB reanalysis with yet another color bar, without using unequally space color scale and even the other SMB unit (mm yr -1 water equivalent, although the number is the same). Figure 6 will provide more information by using the same unequally spaced color scales as in Figure 4 both for mean and bias. Not only the coast region but also the inland information is very useful for the ice sheet modellers once this work is published.
I am looking forward to seeing the nearly finalized improved version reflecting the comments by the referees.

Thank you very much for this suggestion. We have implemented the suggested changes to Figure 6 to bring it into line with Figure 4 and we feel that it really improves the information provided in this figure. Please see below:



Reviewer 1:

While the authors have added some additional discussion that addresses our initial comments (and that of the other reviewers), we would encourage them to be more specific and quantitative. For example, the discussion about the relevance to Antarctic mass balance (input-output method) is useful but currently very vague, qualitative, and not supported by any visual evidence (a figure would be useful).

We have added more detail in this section of the discussion and a new figure showing the comparison with IMBIE derived mass budget estimate for Antarctic see below:

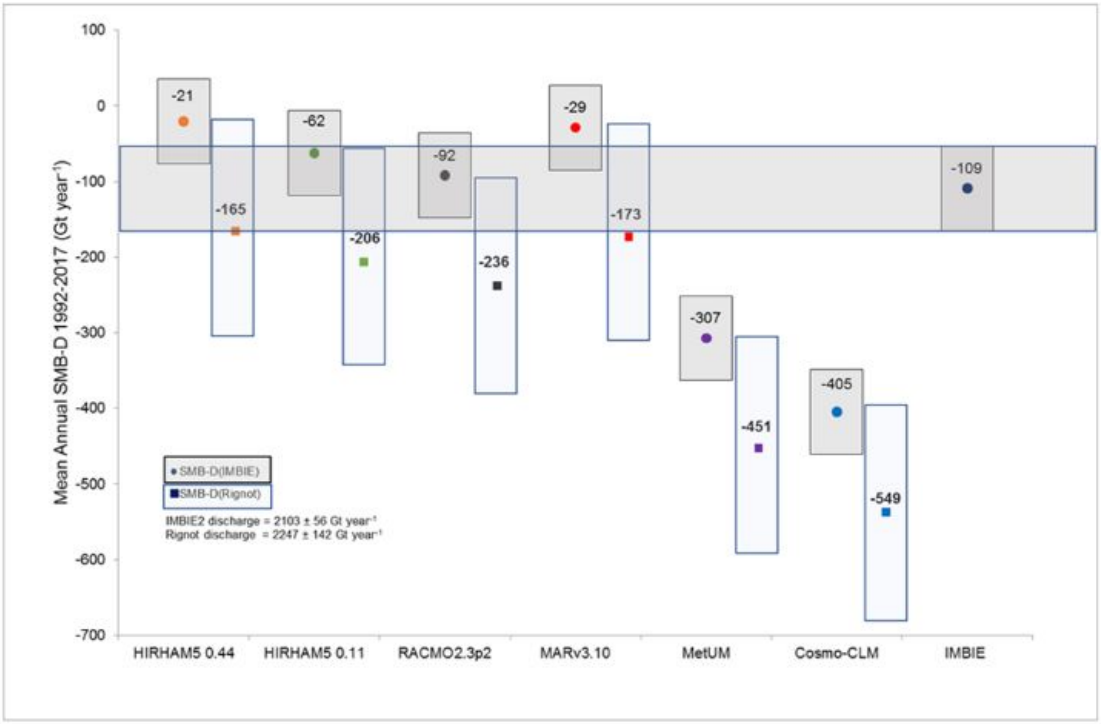


Figure 10. Modelled SMB minus discharge calculated from IMBIE2 results (Shepherd et al., 2018b) (filled circles indicate mean, light grey box indicates IMBIE2 uncertainty range of ±56 Gt year^{-1}) and Rignot et al. (2019) (mean showed in filled square, uncertainty range of 142 Gt year^{-1} shown by narrow blue shaded box). The uncertainty for is taken from table 1 in Rignot et al. (2019), assuming the same uncertainty range for the period 2009 to 2017 is applicable over the longer 1992 to 2017 period.The total mass budget estimated by IMBIE2 is also shown by the horizontal dark grey shaded box for ease of comparison. Numbers are mean annual SMB-D for the 1992 to 2017 IMBIE period for each model.

We can compare our results for the total mass budget of Antarctica with those produced by the IMBIE2 study (Shepherdet al., 2018b). In figure 10 we show the SMB-Discharge for two different datasets, where the IMBIE2 (Shepherd et al., 2018b) reconciled estimate of mean annual discharge is 2103±56Gt year−1and the Rignot et al. (2019) estimated discharge of 2247±140Gt year−1 for the same period is subtracted from SMB calculated from each model. We use the simple SMB calculation in equation 1 for the period 1992 to 2017 over the grounded ice sheet only. The Rignot et al. (2019) dataset has a wider uncertainty range than the Shepherd et al. (2018b) estimate and a larger discharge that gives a lower total mass budget overall, but in all cases the two overlap within the uncertainty ranges. Note that the RACMO2.3p2 model was used to produce both the IMBIE2 and Rignot et al. (2019) estimates and it is thus not a truly independent comparison. The earlier MARv3.6 model was

also included in the Shepherd et al. (2018b) study. When taking into account the published uncertainties on the observational mass budget estimates of discharge, only the COSMO-CLM2 and MetUM estimates are completely outside the range defined by the IMBIE study (109±56Gt year−1) for the total mass budget of Antarctica. However, both models perform well compared to the weather station observation, particularly MetUM, and both have higher correlations and lower biases than the two HIRHAM simulations (see figure 1) for pressure and temperature. Comparison with the SMB observations shows that COSMO-CLM2 has a large dry bias (of 40%) over ice shelves and at lower elevations, this bias is larger than the other models over the ice shelves and up to 1200m in elevation, but at higher elevations the mean bias is close to zero for the COSMO-CLM2 model and in fact much lower than the other models in the 2800 to 3400m elevation range (see figure 4). MetUM on the other hand has a middle of the range mean bias at low elevations compared to other models but a much higher (-25 to -30%) mean bias as shown in figure 4 at the upper elevations. The combination of these results, bearing in mind also the undersampling in the dataset, thus indicate either that some of the components of SMB are poorly captured by the models or that there are compensating errors in the modelledSMB components and/or their spatial variability and most likely a combination of factors. This means that there are large uncertainties in both observations as well as the biases in models that we discuss in this paper, that complicate assessing the contribution to sea level rise from Antarctica from SMB processes.

More generally, the comparisons between the models in the text remain rather qualitative in many instances,sometimes subjective,and or not supported by numbers and their associated uncertainties.There are numerous examples of this throughout the text, e.g. ‘overestimate/underestimate’ (by how much?); ‘lower bias’; ‘best comparison’; ‘rather small’; etc. etc.

We have conducted a thorough proofread and revised text where necessary to provide more quantitative description of our findings with statistics to support qualitative statements.

While much better than the previous iteration of the manuscript, we find that the authors still switch from passive to active voice (and back), sporadically. For instance:

Thankyou for the very thorough comments on our text. We have again proofread the final draft and we have hopefully managed to remove all instances of passive voice now. See below for specific edits.

“The model was used to run a series of consecutive twice-daily 24-hour forecasts at 00 and 12 UTC 25 from the beginning of 1980 to the end of 2018.” (P5L24-25)

Fixed to “For this study we ran a series of consecutive twice-daily 24-hour forecasts..”

“... in these simulations we used a simplified single-layer scheme with for example, no refreezing (Cox et al., 1999).” (P5L27-28).

“SMB was calculated based on output precipitation and sublimation and evaporation.” (P5L28)

“We therefore calculate SMB based on output precipitation and sublimation and evaporation.”

The figure quality needs additional improvements. Several figures are still hard to read(see ‘Specific Comments’ for more details). The font size for some figures, particularly figures with maps, could be increased a bit to help with readability.

Figure 3 is especially problematic, since the individual dots are not readable, the lines are difficult to separate, and most importantly, the comparison of observed and modeled SMB is plotted on a double-logarithmic axis, which hides a lot of the scatter between observations and models. A linear fit that is shown on a double logarithmic axis(as is the case in Figure 3), does not imply a linear relation between the observations and models–it might just show that their exponential behavior is linearly related. We would highly encourage the authors to revise Figure 3 substantially to provide a fair and visually engaging and understandable.

We have revised all figures in line with the specific comments (below) to make them more readable. It is unfortunately difficult to include all of the information we would like on all the plots we would like when including so many models. However, we agree it is difficult to read some of the text and hopefully the revisions have substantially improved.

Figure 3 is now as shown below with larger labels and point markers and is therefore we hope much clearer. As before, we have presented the individual model comparisons in the appendix to avoid filling the paper up with too many figures.We prefer to keep the double log axes because the distribution of both the observed and simulated SMB is not Gaussian. This means a linear regression will be strongly influenced by the extreme values, especially the r^2, which is skewed by the errors for the largest values but only weakly influenced by the errors on the smallest values. This means that in the absolute numbers, the error is greater for large (observed and simulated) SMB values than for small values, while in relative terms, it may be exactly the same error. Using log/log axes makes the distributions Gaussian and enables us to calculate a linear regression with unbiased coefficients. Using linear axes would mean that the comparison is biased by the errors for the largest values and we can now see how the models reproduce the low values. The associated statistics show both r and rlog so the linear relationship is also given.

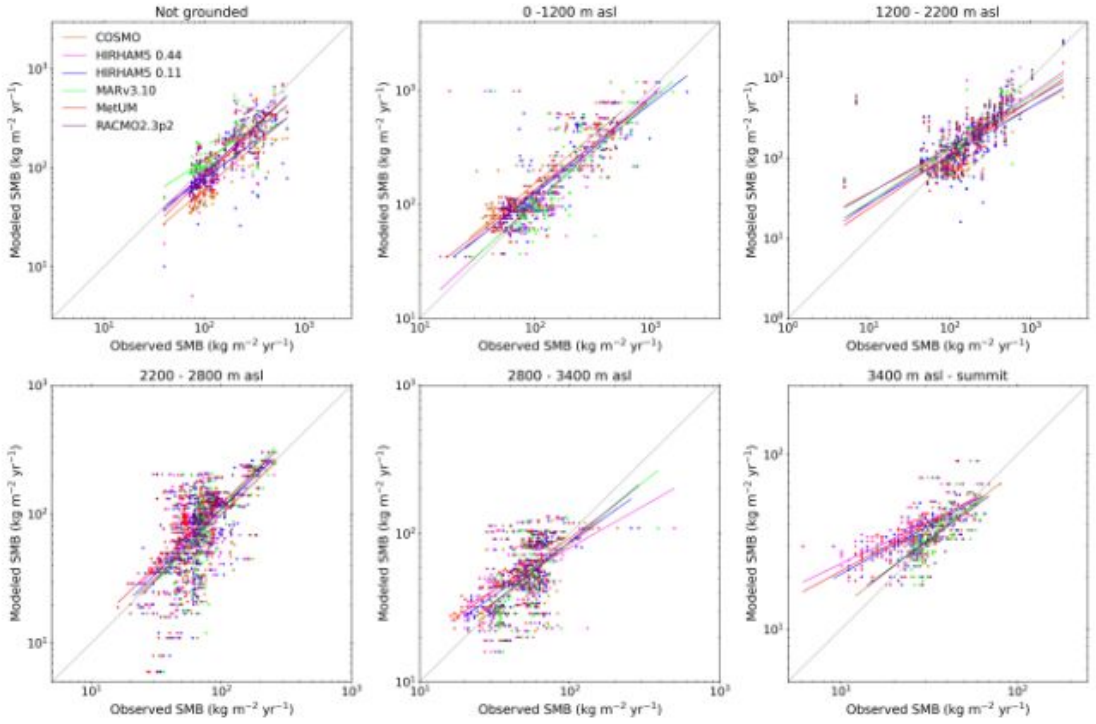


Figure 3. Comparison between modelled SMB and observed SMB in a gridded dataset. Trend lines and points are plotted for each model in a different colour, note different x and y axes for different elevation bins. The figures are plotted on logarithmic axes because the datasets are not gaussian distributed and this better represents the relative error in both high and low SMB regions than linear axes.

We have clarified this in the figure caption (see above) and inserted the following text :

"Note that Figure 3 is plotted on logarithmic axes because the distribution of both the observed and simulated SMB is not Gaussian. As linear regression is strongly influenced by the extreme values, which skew r2 errors in both modelled and observed SMB for the largest values but is only weakly influenced by the errors on the smallest absolute values, a logarithmic plot better displays how well models reproduce SMB in both high and low SMB regions."

Lastly, the third reviewer brought up a very important point, and we feel that hasn't yet been adequately dealt with. If a true intercomparison of models is presented, the common P-E-S budget should be evaluated. So that means that, while blowing snow sublimation can be included, runoff (for MAR and RACMO2) and ER_{ds}(for RACMO2) should not. The author's argument that 'these cannot be easily removed without retuning the models', is only relevant for blowing snow sublimation, not for runoff erosion.

We apologise for causing some confusion here. It is true that the versions of the MAR and RACMO models presented here have fully optimised SMB schemes including runoff, and in the case of RACMO the wind blown erosion of snow,compared to the other models. The SMB from using these full schemes is presented in Table 3. However, in table 4 we also present SMB as calculated using only the simple SMB as described in equation 1 along with the components precipitation and sublimation for all models and the ERA-Interim reanalysis in order to enable exactly this kind of comparison. However we had not noted this very

clearly in the text. This has now been clarified with the following text in the discussion and updated in the table caption.

$$\text{SMB} = \text{precipitation} - \text{evaporation} - \text{sublimation} \quad (1)$$

“We calculate the mean annual SMB and components across the continent including ice shelves, as given in Table 4, 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 using the simple SMB calculation given in equation 1 and results are therefore slightly different to those already published for different models or shown in Figure 5 or Table 3.”

Specific Comments

P1L2: SMB is introduced without spelling out the acronym. Later in the abstract, the phrase is entirely spelled out, though.
Fixed

P1L5: Similar to SMB, AIS should be spelled out the first time.
Fixed

P1L11: Consider adding over what period you refer to finding no trend (i.e. 1987-2015 or 1980-2010). It may be over both, in which case you can add “over either period.”
Fixed

P1L14: “between 1000 and 2000 m” what? I’m not sure what that scale is referring to. Is it the length of the slope?
It refers to the surface elevation of the ice sheet. This has been clarified
“Drifting-snow schemes improve modelled SMB on ice sheet surface slopes with an elevation between 1000 and 2000 m where strong katabatic winds form”

P2L9: fine to use submarine melting here so as not to confuse it with basal melting but still think this term should be defined either within this sentence or in a following sentence.
Fixed
“Most ice loss in Antarctica occurs as a result of submarine melting, that is melt at the water-ice interface underneath ice shelves, or by the calving of icebergs from ice shelves.”

P2L18: We are a bit confused why the authors introduce the term “surface mass budget” here. If it is synonymous with surface mass balance, surely just using SMB throughout the paper is sufficient. If the authors want to note this term and its equivalence to surface mass balance or climate mass balance, perhaps it would be best to do so when SMB is first introduced.
This was a relic of an earlier draft. We have moved the phrase “ also known as surface mass balance or climate mass balance” to where the abbreviation SMB is described for the first time for the sake of completeness.

P2L30: “Currently runoff...” —> “Currently, runoff...”
Fixed

P2L30-34: This is one of the few remaining run-on sentences that I think could be broken up.
Adjusted to: Currently, runoff is a relatively minor contribution (Lenaerts et al., 2019) to mass loss in Antarctica. Increasing 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.

P3L23: “In this paper we seek...” —> “In this paper, we seek...”
Fixed

P3L29: “...backwards continuity we also...” —> “...backwards continuity, we also...”
Fixed

P4L9: There should be an and at the end of this list and commas around “among others.”
Fixed

P4L28: Missing space after the end of the sentence starting with “As these terms cannot easily...”
Fixed
P4L33: “Additionally this model...” —> “Additionally, this model...”
Fixed

P5L21: “Here we run...” —> “Here, we run...”

Fixed

P11L2: “...taking note of the differences...” —> “...take note of the differences...”
Fixed
P11L6: “In figure 1 we show...” —> “In figure 1, we show...”
Fixed

P12L1: “Figure 1 analysis shows that depending on the variable the models...” —> “Figure 1 analysis shows that, depending on the variable, the models...”
Fixed

P12L12-13: “...show that although the models perform well (...) on average...” —> “...show that, although the models perform well (...), on average...”
Fixed

P13L8: Avoid “It is clear.” This is a complex figure that, while you describe what’s going on well, is still fairly high-level.
Fixed to: “Figure 1 shows that all of the models perform less well for wind speeds than for temperature or pressure observations. The wind speed plot shows all models have higher CRMSE, higher standard deviation and lower correlation values when compared with observations.”

P13L23: “This comparison therefore is...” —> “This comparison, therefore, is...”
Fixed

P15L8-P16L2: This sentence 1) requires commas around “to show... models clearly” and 2) switches from passive voice in the first half to active voice in the second (“are given in Table 2”, “we show all models”).
Fixed: “We show detailed statistics for the SMB comparison in Table 2. In order to show the large scatter in the observations and the models clearly, we also plot all modelled SMB values against observed SMB values in Figure”

P16L3: “Apart from COSMO-CLM2 and HIRHAM5 0.11◦the...” —> “Apart from COSMO-CLM2 and HIRHAM5 0.11◦, the...”
Fixed

P16L5: “In general all models...” —> “In general, all models...”

Fixed

P16L6-9: This is another long sentence that could benefit from being broken into two.
Fixed: “In general, all models underestimate SMB over the ice shelves and at the low elevation coastal regions of Antarctica (see also statistics in Table \ref{tab:samba_smb} a. and b. and Figure \ref{smb_obs_comp}). The highest mean bias, lowest RMSE and lowest r values in particular are given in the COSMO-CLM^{2} and HIRHAM5 0.11\degree models at the lowest elevations.”

P18L3: Consider introducing the concept of SSMB as well as the acronym to help the reader.
We have slightly restructured the paragraph to bring the explanation of SSMB forward.

P18L10: “...and therefore orographic precipitation.” —> “...and, therefore, orographic precipitation.”
Fixed

P22L8: “For example there is an...” —> “For example, there is an...”
Fixed

P25L5: Missing spaces after the degree symbols.
Fixed

P27L5: “...allows us to estimate not only the likely range of SMB over Antarctica, but also to identify...” —> “...allows us not only to estimate the likely range of SMB over Antarctica, but also to identify...”
Fixed

P28L1: “Nevertheless it is therefore also important...” —>“Nevertheless, it is also important...”
Fixed

P28L5: “...have quite strong trends, for example a steady...” —> “...have quite strong trends: for example, a steady...” or “...have quite strong trends (for example, a steady...).”

Fixed to: “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 reverse after 2014.”

P28L7: I believe the comma after “very difficult” is meant to be a period.
Yes it was. Fixed

P28L12: Again, try to avoid “it is clear” as it may not be clear to all readers.
Fixed to “Figure 8 demonstrates..”

P28L26: “Basins in West Antarctica, and particularly on the Antarctic peninsula have very large...” —> “Basins in West Antarctica, and particularly on the Antarctic peninsula, have very large...”
Fixed

P29L1: “We found that although the variation...” —> “We find that, although the variation...”
Fixed

P29L21: “Nonetheless we are able...” —> “Nonetheless, we are able...”
Fixed - see below

P29L22-25: This is a long and confusing sentence that could be reworked.
The whole paragraph has been reworked:
“Evaluating the models against observations is very important for assessing where there are important biases, but evaluation of model performance is significantly hampered by the lack of observations in key regions. Nonetheless, Figure 1 shows that the models do have skill in simulating surface climate, particularly temperature and pressure. The skill in simulating surface climate does not however translate perfectly to simulating SMB, partly due to the difficulties of modelling and evaluating precipitation. Our analysis shows that for example, COSMO-CLM2 better simulated surface climate compared to observations than HIRHAM5 but it has a lower skill in SMB. Variables such as temperature and pressure are more easily measured and are assimilated into the reanalysis used to drive the models. RCMs have also been optimised to give good performance compared to these kinds of observations. However, Antarctic SMB is dominated by the precipitation term that is much harder to measure accurately and also has much higher uncertainty in models

P29L29: “It is therefore important...” —> “Therefore, it is important...”
Fixed

P29L32: “...perform, broadly speaking better than...” —> “...perform, broadly speaking, better than...”

Fixed

P30L6: “However table 2 shows...” —> “However, table 2 shows...”
Fixed

P30L7: Consider changing “clear” to “evident” (or something similar).
Fixed

P30L12: “Currently efforts are...” —> “Currently, efforts are...”P31L2: Missing space after the degree symbol.P21L18: Missing space after the period.
Fixed

P30L21: “Berg et al. (2013), argue...” —> “Berg et al. (2013) argue...”
Fixed

P31L31: “In this paper we have compared...” —> “In this paper, we have compared...”
Fixed

Figure 1 caption: Toward the end of the first sentence, there is an extra space in the (c) notation and a missing space after the period.The curved, centric lines are not explained (do they show CRMSE? What is CRMSE?).

We have remade the figure to enlarge it and remove the typo errors. As we explain in the caption: "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. 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)."
We also explain CRMSE in more detail just above Figure 1 and have added the equation to show more fully how it is calculated. '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.'

10 value as shown in equation 4

$$CRMSE = \sqrt{\frac{\sum_{i=0}^{n}(m_i - o_i)^2}{n} - (\overline{m} - \overline{o})^2} \quad (4)$$

Where n is the number of observations, m_i is the modelled value, o_i is the observed value and $\overline{m}$ and $\overline{o}$ are the average of the modelled and observed values respectively.

Figure 2: I realize this is in contrast from our previous review so at the risk of sounding contradictory and incredibly nitpick-y, would it be possible to make this figure a touch smaller such that it fits onto the previous page? The font size increase will help significantly if the figure size is reduced and (I think) the figure will still be readable. Also, the y-axis reads “Modeled” which is fine American English but the authors use “modelled” (which is fine British English) throughout the rest of the paper.

We have resized the image and the font to make it fit better on the page. The typo has been fixed back to British English to align with the rest of the manuscript.

Figure 4: This seems like a potentially really great figure but the images are a bit too small to glean a lot of high detail information. Perhaps consider rearranging to make the individual subplots larger.

We have reshaped this figure to make it clearer as below:

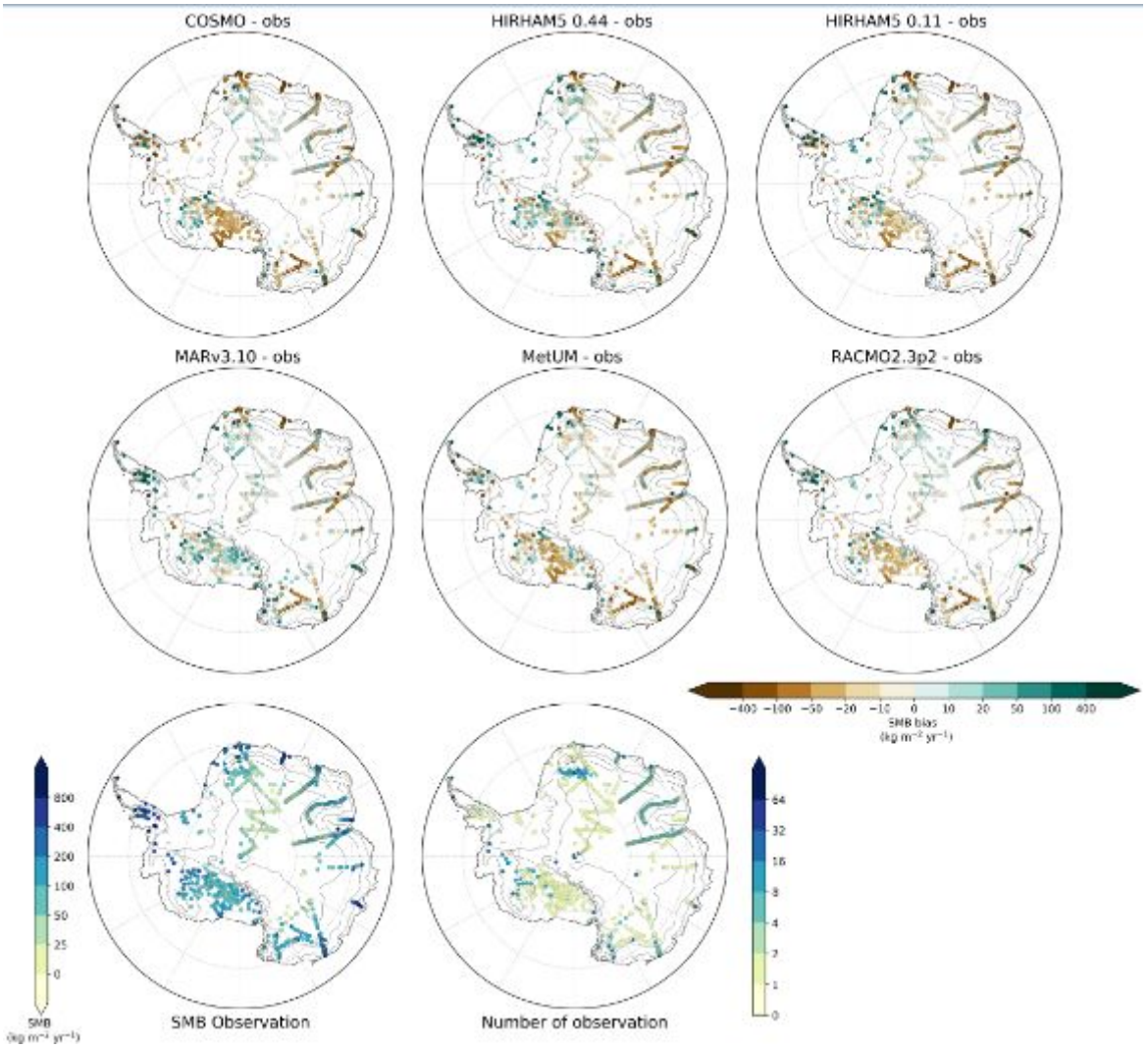


Table 3: Is the last column meant to have sigma spelled out?
Yes, a typo was introduced in the latex code. Now fixed.

Figure 6: The text on this figure (both the title and the numbers on the scale) are still quite small and hard to read.

We have substantially revised Figure 6 with a different colour scheme consistent with that in Figure 4 and larger text.

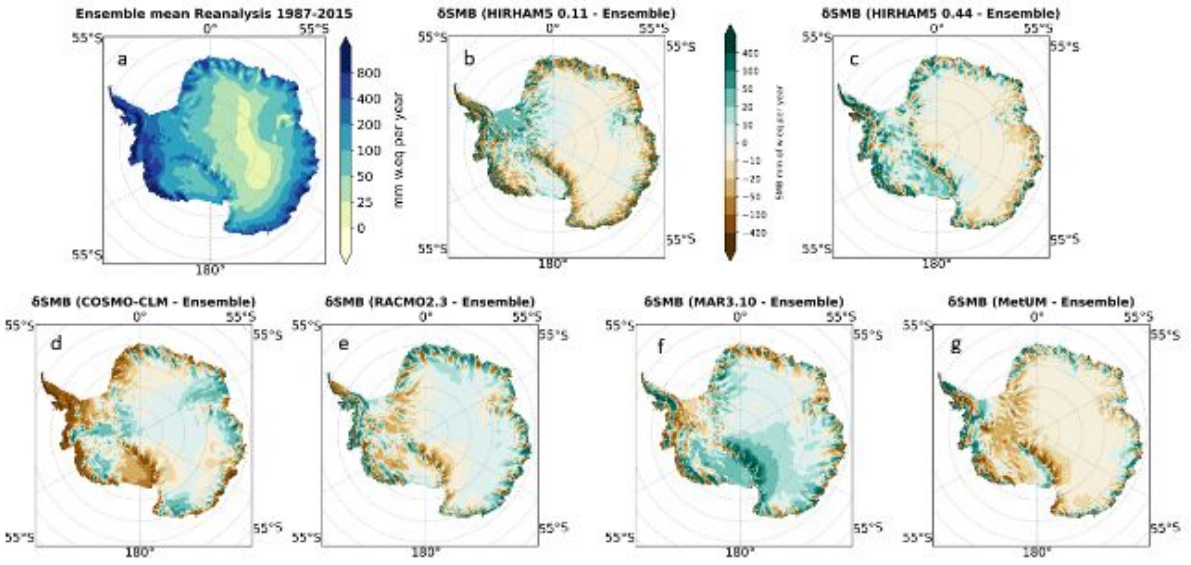


Figure 6. Sub-figure **a** shows the SMB ensemble mean for the common period, on the common mask. Sub-figure **b-g** show the difference between each model and the ensemble mean.

Figures 7, 8, & 9: Shouldn't the units for SMB (and precipitation) be Gt/yr and SMB trend Gt/yr2, respectively?

We recognise that the figure titles are perhaps a bit confusing here so we have relabelled these graphs to reduce the confusion around what they present. As they show the year on the x-axis the y-axis unit should be Gigatonnes. Figure 7 shows the annual variability in SMB for each model by subtracting the mean value so in fact the quantities are also Gigatonnes here.

Table 4 caption: "...standard deviations are also show." —> "...standard deviations are also shown

Fixed