The Referee's comments below are in italics, our answer in plain font in blue

This paper from Dalaiden and co-authors addresses the question of the relationship between surface air temperature (SAT) and surface mass balance (SMB) in Antarctica, from the past 1000 years to the last decades, in view of using the SMB information for reconstructing past SAT. Given the short and sparse observational coverage in Antarctica, reconstruction of the Antarctic climate further than the last decades rely on the interpretation of proxies. The isotopic composition of the snow (in particularδ18O in ice cores) is the most widely used proxy of SAT in Antarctica. First the authors show that the strong link between SMB and SAT, already acknowledged in the literature (e.g. Frieler et al 2015), remain valid in GCMs during the past1000 years and the past 200 years. They also show that the relationship does not stand when considering the last two reconstructions of surface air temperature (based on ice cores δ18O, Stenni et al., 2017) and surface mass balance (based on ice cores accumulation, Thomas et al., 2017), but does exist when using an Antarctic SAT reconstruction based on weather stations (Nicolas and Bromwich 2014, NB14) instead of the SAT reconstruction based on ice cores δ18O. Then the authors use isotope-enabled global climate models to perform an offline data assimilation of δ18O and SMB over the past 200 years. They obtain more consistent results with NB14 SAT over West and East Antarctic ice sheets when combining the assimilation of δ18O and SMB. I think using both SMB and δ18O for reconstructing SAT with an assimilation method is novel and relevant for the cryosphere and climate community. The overall presentation is clear and figures are nicely shaped. Conclusions seem robust and interesting. However I have some concerns about some of the interpretations, and I also have comments on the methodology. Therefore I recommend this article to be published after addressing the following issues.

We would like to thank the Referee for the careful evaluation and for all the suggestions that helped to improve the manuscript.

Major

1) I think the GCM evaluation is of interest, in particular the plots comparing SMB by elevation bins, but I disagree with the conclusion that GCM are doing a good job in Antarctica. I think this is not a critical point for this study, so the authors should minimize or remove the section about GCM evaluation (Section 4.1, one or two sentences and citing supplementary would be enough) and extend the analysis on the SMB/SAT relationship (Section 4.2). Fig. 2 is not necessary, Fig. 3 and Fig. 4 could be moved to the SAM/SAT section, Fig. 4 could be extended with a scatterplot comparing SMB/SAT sensitivity factors (% K-1) of West vs East. This way the result section would follow the plan detailed in the introduction: i) SMB/SAT in GCMS over the past millennia and centuries ii) data assimilation for the past centuries.

As suggested by the reviewer, we have trimmed the GCM evaluation. The evaluation over the recent past (1979-2005; i.e. the comparison to RACMO outputs) has been moved to Supplementary Materials. However, we have kept the section on the comparison between the simulated and reconstructed (i.e. Thomas et al., 2017) SMB changes during the last two centuries. Therefore, we have adapted the title section: "Reconstructed and simulated SMB changes over the last centuries".

The Fig. 4 has been extended with a scatter plot comparing the SMB/SAT sensitivity factors:

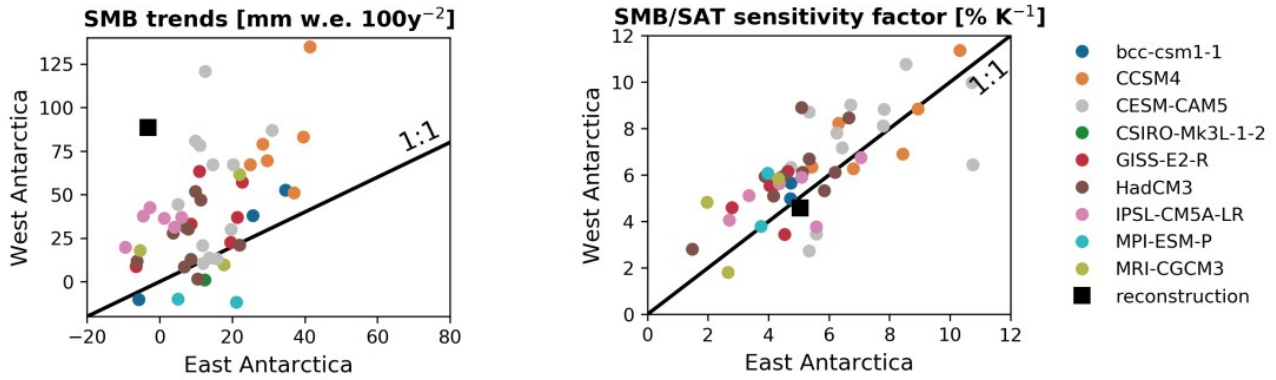

Figure 4. (left) Comparison between the reconstructed and the simulated SMB trends (mm w.e./100y^{-2}) over the period 1950–2000 CE in West Antarctica (y axis) and East Antarctica (x axis). (right) As on the left but for SMB/SAT sensitivity factors (% K^{-1}). For the reconstruction, data from Thomas et al. (2017) and Nicolas and Bromwich (2014) are used.

In detail:
** Abstract "Here, we show that Global Climate Models (GCMs) can reproduce the present-day (1979–2005) AIS SMB and the temporal variations over the last two centuries."*

We have removed this sentence to stay focused in the abstract on the SMB-SAT relationship and on our reconstructions.

** P17 "The GCMs are able to simulate relativity well the current AIS SMB"*
-> Should be rephrased or removed (see hereafter).

We have removed the SMB evaluation in the discussion/conclusions section.

** P8 "Overall, the AIS SMB simulated by GCMs is in good agreement with the SMB simulated by the regional climate model RACMO2 over the last decades (1979–2005,R2 = 0.53; Fig. 2 and S1 for the SMB of each model)."*
-> I see huge differences, spatially and integrated over the ice sheet (Fig. S1 and S2). How is computed this correlation coefficient? What is the bias?

We have made a correlation plot (new figure: see below, Fig. S2) of the SMB climatology as simulated by the average of the GCMs as a function of the climatology of RACMO over the 1979-2005 period. The correlation is computed between the model mean spatial distribution (averaged over 1979-2005) and the spatial distribution of RACMO over the same period. The model mean has been interpolated on the RACMO grid to compute the correlations. The bias is the average of the difference between the GCM mean and RACMO (in mm w.e. year^{-1}).

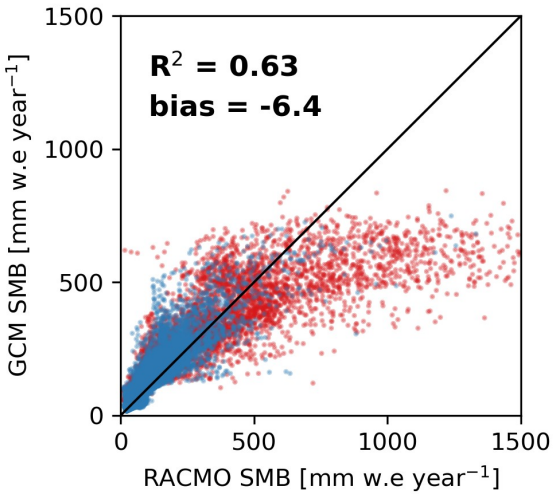

Figure S2. Correlation plot of SMB climatology from GCM mean (average over all the GCMs including isotope-enabled models) as a function of SMB RACMO over the 1979–2005 period at the same location. R^2 is the determination coefficient and the estimation of the bias is the average of the difference between GCM mean and RACMO (in mm w.e. year^{-1}). Red (blue) dots are for places where the altitude is lower (higher) than 1500m. See Fig. S4 for the equivalent for each model.

Because we have added the isotope-enable models in the evaluation, we have updated the following sentence:

"The mean of the SMB over the entire AIS simulated by the selected CMIP5 models is 87 Gt year -1 higher than the SMB simulated by RACMO2 (relative bias: -3.7%; see Fig. S2 for the integrated SMB over the entire AIS for each model)."

by:

"The mean of the SMB over the entire AIS simulated by the selected models (including isotope-enable models) is 6.4 mm w.e. year -1 lower than the SMB simulated by RACMO2 over the 1979-2005 period (relative bias: -3.4%; see Fig. S4 for the correlation plots for each model and Fig. S5 for the integrated SMB over the entire AIS for each model)."

** P8 "Both display high values of SMB along the coast (>300 mm w.e. year-1) – especially for West Antarctica and the Antarctic Peninsula – and lower values at high elevations (e.g. the Plateau: <100 mm w.e. year-1)."*
-> This is really the minimum feature a model can do, because of the general circulation and the ice sheet topography.

Yes, we totally agree with your remark, but we think that it is important to notice the main Antarctic SMB pattern. Therefore, we have added "As expected" at the beginning of the sentence to show that is not something surprising.

2) I found interpretations in contradiction with the figures.
** P9 "Nevertheless, when analyzing the individual simulations of the ensemble performed with CESM1-CAM5, the contrast between East Antarctica and West Antarctica is as large as in recent observations (Fig. 4). This indicates that 1) the observed SMB trends between the two regions are within the range of the simulated values; 2) internal variability has an important role in the current Antarctic SMB changes."*
-> Reconstruction is a clear outlier of the GCM's scatterplots, so reformulate the conclusion in agreement with your figure.

We have changed the paragraph following the suggestion to be in better agreement with the figure:

“When analyzing the ensemble of simulations performed with CESM1-CAM5, the ensemble mean also shows a relatively homogeneous increase, but some simulations display a contrast between East Antarctica and West Antarctica close to the one observed in the reconstruction (Fig. 3). This suggests that internal variability has a dominant contribution in the current Antarctic SMB changes and might explain why the observed contrast between East and West Antarctica is only present in a few simulations.”

** P12 "For most regions, the link between surface temperature and SMB (r=0.70 on average over the seven subregions for the 1850–2000 period) is higher than that between surface temperatures andδ18O (r=0.55 on average over the seven subregions for the 1850–2000 period)." (...) "The results with the outputs of ECHAM5-wiso and ECHAM5/MPI-OM are similar (Figs. S6 and S7)."*
-> It does not appear to be true when looking at Fig. S6 and S7: blue dots (SAT/δ18O) are often higher than green dots (SMB/SAT). I regret this over-interpretation and the fact that the authors focused on the iHadCM3 in the main text without specifying it and explaining this choice.

We mostly focused on iHadCM3 outputs and not on the other isotope-enable models in the main the text because, in contrast to the other isotope-enabled models (ECHAM5-wiso and ECHAM5/MPI-OM), iHadCM3 offers an ensemble of simulations which is a significant advantage for data assimilation. Indeed, dealing with an ensemble of simulations allows increasing the probability to find a good match between the assimilated records and model results during the assimilation process.

Regarding the ECHAM5-wiso and ECHAM5/MPI-OM models, we have modified the figures S6 and S7 to replace them by the Figure S9:

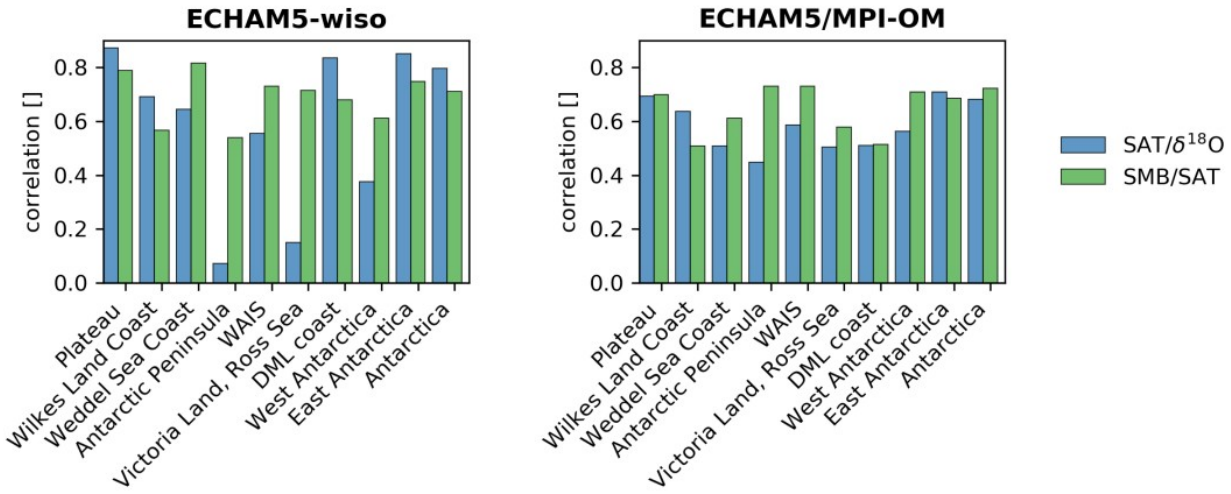

Figure S9. 5-year mean correlations between surface temperature and $\delta^{18}O$ (blue) and, SMB and surface temperature (green) for the seven Antarctic regions for the entire period simulation (1871–2010 for ECHAM5-wiso and 801–2000 for ECHAM5/MPI-OM).

This new figure allows for an easier comparison between the potential of SMB and $\delta^{18}O$ in reconstructing regional surface temperatures. As the reviewer mentioned, the results of ECHAM models are a little different than those of iHadCM3. We thus propose to discuss in more details those results of the ECHAM in the main text:

“The results of ECHAM5-wiso and ECHAM5/MPI-OM confirm this strong link between SMB and temperature but, in contrast to iHadCM3, the correlations are not systematically higher than between $\delta^{18}O$ and temperature (Fig. S9). When analyzing the long ECHAM5/MPI-OM simulation (800–2000), the relationship between SMB and surface temperature is generally higher than between $\delta^{18}O$ and surface temperature but the difference is small. For some regions, the SMB-surface temperature link is much higher than the $\delta^{18}O$-surface temperature link but it is weaker for other regions. In contrast to the $\delta^{18}O$-surface temperature link, the SMB-surface temperature is less spatially variable (minimum regional correlation is 0.54 against 0.07 for the $\delta^{18}O$-surface temperature link).”

P18 "On the one hand, models show a strong correlation between δ18O and SMB for all the Antarctic regions"-
> It’s not true: red dots in Fig 7, S6 and S7. Is there a typo here? But even SAT-SMB relationship is not strong for all regions (Fig S5).

Indeed, we made a mistake here (it is the SAT-SMB relationship and not the $\delta^{18}O$-SMB relationship that shows a strong correlation for all regions). Thank you for that.

We propose to replace “for all the Antarctic regions” by “many Antarctic regions”.

"we showed that the relationship between SMB and surface temperature is often higher than the one between surface temperature andδ18O. This is true both on the continental and regional scale."
-> That’s not true when considering ECHAMwiso and ECHAM/MPI-OM

Even though the ECHAM models do not always display stronger regional correlations between SMB and surface temperature than between $\delta^{18}O$ and surface temperature, on average over all the isotope-enable models, the SMB-surface temperature link is stronger (90% of the time for iHadCM3, 80% for ECHAM/MPI-OM and 50% for ECHAM5-wiso) and more stable than the $\delta^{18}O$ -surface temperature link. We propose to modify slightly this sentence:

“By analyzing isotope-enabled climate models, we show that the relationship between SMB and surface temperature is often higher than the one between surface temperature and δ 18 O.”

by:

“By analyzing isotope-enabled climate models, we showed that on average over the models, the relationship between SMB and surface temperature is often higher (or at least equivalent) and more stable than the one between surface temperature and $\delta^{18}O$.”

3) Methodology
Data assimilation (DA) must be evaluated with independent datasets. It is the case for SAT (NB14 is not assimilated) but not for SMB. The authors assimilate SMB from Thomas et al. (2017) and evaluate their results with Thomas et al. (2017). I suggest to use independent and annually resolved datasets, such as the radar transects resolved annually in West Antarctica (Medley et al. 2014 https://doi.org/10.5194/tc-8-1375-2014) and stake line transects (JARE, CHINARE).
** P19 "Considering our good results regarding surface temperatures and SMB reconstructions,"*
-> This sentence is not fair if you evaluate your result with the data you assimilate.

We totally agree with the reviewer. Our goal is to propose a new reconstruction method for surface temperature. It is thus needed to evaluate this new reconstruction with an independent dataset. Unfortunately, we did not find any suitable dataset to evaluate our data assimilation-based

reconstruction. The radar transects that you suggest (Medley et al., 2014) cover a small part of the West Antarctic Ice Sheet over the 1985-2009 period. It is thus not possible to make an evaluation at the scale of Antarctica. Furthermore, because we applied a 5-year smoothing on our SMB and surface temperature reconstruction to remove the non-climatic noise, any validation would be based on a too small sample (applying a 5-year smoothing on the NB2014 dataset which covers the 1958-2012 period reduces the time series to 12 points which is already low for making correlations).

This absence of independent datasets forbids us to evaluate the skill of the new reconstruction. The comparison of our data assimilating-based SMB reconstruction to Thomas et al. (2017) is thus only done to check if the reconstruction is consistent with all the input information or if major incompatibilities are present. If model results (used as prior) and data are too different or if the uncertainty is not well estimated, the particle filter may degenerate. The resulting reconstruction can also be far away from the assimilated records if there is no model result that fits with the signal recorded in those data. Our comparison to Thomas et al. (2017) is not independent but at least shows that our reconstruction is consistent with Thomas et al. (2017). This is indeed expected but good to verify.

We specified in the experimental design (section 3.2) that we are not able to independently evaluate our SMB reconstruction:

“SMB estimates are also available for the last decades (e.g. Medley et al. 2014), but they cover a too short period or have a too small spatial coverage to provide an independent validation of our reconstruction. It is thus not possible to estimate if the assimilation of SMB and $\delta^{18}O$ measurements provides an improvement for this field.”

We have also specified in the discussion/conclusions section that we cannot independently simulate our SMB reconstruction:

“Although it is not possible to independently evaluate our SMB reconstruction, our good results regarding surface temperatures and SMB reconstructions suggest that the strong simulated correlation between surface temperatures and SMB in GCMs is not a model artefact.”

** P19 "our data assimilation-based reconstructions suggest that the strong simulated correlation between surface temperatures and SMB in GCMs is not a model artefact"*
-> DA is a weighted average, so if the SMB-SAT relationship exists in the models, isn’tit conserved in the reconstruction by construction?

Yes, this link should be preserved as the reconstruction is based on the covariance between those two variables as displayed in models. However, if the models were overestimating this link, the particle filter would give more weight to the model results that display the weakest correlation. Furthermore, the increased skill of the surface temperature reconstruction when including SMB data also indicates that the model covariance is bringing additional information. This is not a formal proof. This is the reason why in the corresponding sentence, we propose to use ‘suggest’ (see the new proposed sentence just above), but it remains consistent with the fact that the strong correlation between SMB and surface temperature is not a model artefact.

4) A remark
Results of data assimilation seem less variable than the other reconstructions (Fig 8 and Fig 9). Is it due to the assimilation method? What is the confidence on the DA temporal variability?

The mean reconstruction provided by data assimilation may underestimate the variability if the data is too uncertain or if there is not enough data. In the extreme case when you have no data (or with

data displaying a very large uncertainty), the particle filter will just give a reconstruction that is the model ensemble mean which consists here, because of the experiment design, in a value of zero for the whole period. However, in that case, the uncertainty of the ensemble would be very large, and this of course must be taken into account when discussing the temporal variability of the reconstruction. More specifically, with only a few uncertain data, it is expected that the reconstruction based on our data assimilation method may show less variance than reconstructions provided by some other methods (as observed previously; e.g. Goosse et al. 2010). Nevertheless, we did not discuss much this point in the manuscript as it critically depends on the uncertainty of the input data, that is itself not well known.

Reference:

Goosse, H., E. Crespin, A. de Montety, M. E. Mann, H. Renssen, and A. Timmermann (2010), Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation, J. Geophys. Res.,115, D09108, doi:10.1029/2009JD012737.

Minor
Abstract
"with a linear correlation coefficient with the observed surface temperatures (1958–2010 CE) of 0.73"
I don't think this number is meaningful, I suggest to remove it.

It has been removed.

P2
"(Rignot et al., 2011)"
Update with Rignot et al. (2019) https://www.pnas.org/content/116/4/1095"

Thank you for the updated reference. It has been updated in the new version of the manuscript.

(Wouters et al., 2013).
"Idem, update the reference.

It is done: we have replaced the old reference by the new one: Martín-Español, A., et al. (2016), *Spatial and temporal Antarctic Ice Sheet mass trends, glacio-isostatic adjustment, and surface processes from a joint inversion of satellite altimeter, gravity, and GPS data*, J. Geophys. Res. Earth Surf.,121, 182–200, doi:10.1002/2015JF003550.

"from stable isotope ratios of oxygen"
From water stable isotopes, and in particular δ18O

Thank you for the specification. We have added it in the text.

P3
"According to Monaghan et al. (2008), the observed sensitivity of Antarctic snowfall accumulation to surface temperature was about 5% K-1 during the 1960–1999 period."
Why Monaghan and not a most recent and complete reference? (e.g. Frieler 2015)

We have replaced Monaghan et al. by Frieler et al. as suggested.

"These results suggest that in some regions, especially along the AIS coasts, the variability of thermodynamic processes (such as the Clausius-Clapeyron effect) on SMB is dominated by the large-scale atmospheric circulation, limiting the correlation with δ18O."
Do you mean: SMB variability is dominated by large-scale atmospheric circulation rather than by thermodynamic processes?

Yes, as mentioned by Philippe et al. (2016), we think that the SMB variability along the coasts is more related to large-scale atmospheric circulation than the thermodynamic processes.

We have changed the sentence to make it clearer:

"These results suggest that in some regions, especially along the AIS coasts, the variability of thermodynamic processes (such as the Clausius-Clapeyron effect) on SMB is dominated by the large-scale atmospheric circulation, limiting the correlation with δ18O."

by this:

"These results suggest that in some regions, especially along the AIS coasts, the SMB variability is dominated by large-scale atmospheric circulation rather than by thermodynamic processes (such as the Clausius-Clapeyron relation), limiting the correlation with $\delta^{18}O$."

"While the statistical methods classically used to infer past surface temperature (see for instance Stenni et al., 2017) rely on the length of the calibration period, on the quality of the record during this period, and on the stationarity of the link between the proxy and the variable of interest, which can be strong assumptions in the case of the δ18O-temperature relationship (Klein et al., 2019), data assimilation does not.
"Doesn't data assimilation rely on the quality of the assimilated record too? One step further, a short sentence about the limits of the assimilation method is missing, to be fair. E.g. changes in the number and quality of assimilated data?

We agree that all the reconstruction methods, including data assimilation, rely on the quality of the input data. The point here is that statistical methods are based on strong assumptions such as the stationarity of the link between the proxy and the climate variable. As this relationship is estimated over the instrumental period (i.e. calibration period), statistical methods highly depend on the data quality during this period. Because data assimilation methods do not require any calibration period, these methods are not dependent on the quality of assimilated records over the calibration period used in the statistical periods. Therefore, we propose to keep this sentence in the text, but we have added a general sentence to state that all methods depend on the quality of the input records to be fair:

"All reconstruction methods depend on the number and quality of the input data."

P4
"The simulation of ECHAM5-wiso, which only includes an atmospheric component, was performed by Steiger et al. (2017) and covers the period 1871–2011 CE at 1◦resolution. The model is driven by the sea surface temperature and sea ice from the Rayner et al. (2003) dataset."
You have to mention that the Rayner et al. (2003) dataset is not relevant before 1973: "2.1.3. Antarctic Atlas Climatologies Before the advent of satellite area based imagery in 1973, sea ice concentration data for the Antarctic are not available, and sea ice extent data are not readily available for individual months, seasons or years, although some visible and infrared data do exist for 1966–1972 [Zwally et al., 1983] and some undigitized charts reside in national archives (e.g.,

V. Smolyanitsky, personal communication,2002). Readily available information was limited to two historical climatologies of sea ice extent. Therefore our sea ice concentration analysis before 1973 is derived indirectly, and does not include any interannual variability, though there are some trends resulting from the differences between climatologies for different periods."

Thank you for the specification. We have added this information in the text:

"Due to a lack of Antarctic sea ice data before 1973, this dataset is based on historical climatologies of sea ice concentration for the period 1871-1973 CE, with no interannual variability."

"Comparisons of the results of these three isotope-enabled models with modern δ18O observations indicate that they all reproduce the main characteristics of the spatial distribution of the isotopic composition of precipitation over Antarctica (see reference for each model)."
Add a word about their known biases.

We have added a few sentences in the text regarding the modelled biases:

"According to Tindall et al. (2009) and Sime et al. (2008), the small biases in $\delta^{18}O$ (for example, an underestimation of the spatial $\delta^{18}O$ variability in rugged areas) in the iHadCM3 simulation mainly come from the coarse horizontal resolution of the model and not from the isotopic model itself. ECHAM5-wiso and ECHAM5/MPI-OM display an overall underestimation of $\delta^{18}O$ in Antarctica but reproduce well the general Antarctic $\delta^{18}O$ pattern (Goursaud et al., 2018; Klein et al., 2019, see reference of each model for more details)."

P5
"(4) the output of RACMO2 for the AIS SMB agrees very well with available measurements (correlation coefficient with observations of 0.9; van Wessem et al., 2018)."
A high correlation coefficient alone is not a proof of good performance. Correlation can be equal to one with a very large bias.

Thank you for your remark. We have removed the part with the correlation and modified the previous sentence:

"(4) the output of RACMO2 for the AIS SMB agrees very well with available measurements (correlation coefficient with observations of 0.9; van Wessem et al., 2018)."

by this:

"(4) RACMO2 has been extensively evaluated against available measurements and displays a very good agreement (e.g. van Wessem et al., 2018; Lenaerts et al., 2012)."

References:

Lenaerts, J. T. M., M. R. van den Broeke, W. J. van de Berg, E. van Meijgaard, and P. Kuipers Munneke (2012), A new, high-resolution surface mass balance map of Antarctica (1979–2010) based on regional atmospheric climate modeling, Geophys. Res. Lett.,39, L04501, doi:10.1029/2011GL050713.

P6
"This temporal averaging reduces uncertainties in dating linked to the noise induced by non-climatic processes (e.g. Laepple et al., 2018; Fan et al., 2014)."

The temporal averaging is not described before, and I understood latter in the paragraph that you were talking about the 5-year and 10-year average. The whole paragraph is strangely shaped, please rephrase.

Thank you for your remark. We made a mistake here. This sentence is not at the right location. We have moved it at the end of the third paragraph of the experimental design section (3.2).

P7
"each ensemble member, called particle, is compared to the proxy-based reconstruction by computing its likelihood, taking into account data uncertainties."
Give a description of this likelihood function. How do you compute it?

In our data assimilation method, the weights given to each particle are computed using a Gaussian likelihood. All the details can be found in Dubinkina et al. (2011). It is now specified in the new version of the manuscript:

"At each time step of the data assimilation procedure (yearly, see Sec. 3.2), each ensemble member, called particle, is compared to the proxy-based reconstruction by computing its likelihood, assumed here to be Gaussian, taking into account data uncertainties (see Dubinkina et al. (2011) for details)."

Reference:

Dubinkina, S., Goosse, H., Sallaz-Damaz, Y., Crespin, E., and Crucifix, M.: Testing a Particle Filter To Reconstruct Climate Changes Over the Past Centuries (2011), International Journal of Bifurcation and Chaos, 21, 3611–3618, https://doi.org/10.1142/S0218127411030763.

P8
"The median of the SMB over the entire AIS simulated by CMIP5 models is 1.16"
A median computed from 12 values is not robust. This number is hiding large discrepancies between the models.

We have replaced the median by the mean in the text (absolute and relative biases):

"The SMB integrated over the entire AIS is 87 Gt year $^{-1}$ higher for the mean of the selected CMIP5 models than in RACMO2 (relative bias: -3.7%; see Fig. S2 for the integrated SMB over the entire AIS for each model)."

As mentioned in the comment, there are large discrepancies between the models. Especially the MRI-CGCM3 model largely overestimates the AIS SMB compared to RACMO2 (+1320 Gt year $^{-1}$, see the figure below).

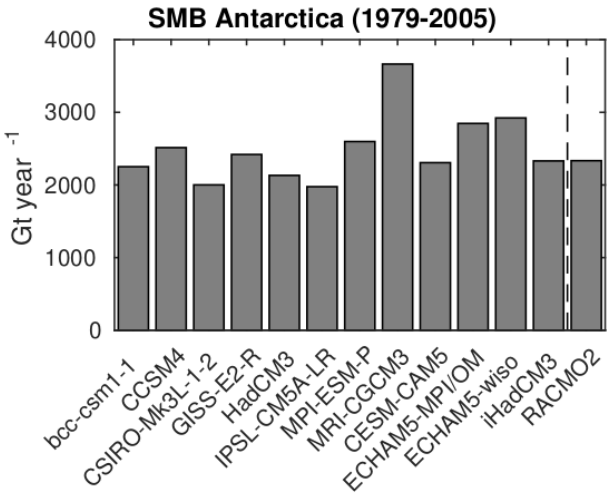

Figure S5. Mean Antarctic Ice Sheet surface mass balance (Gt year^{-1}) simulated by all the models used in this study.

**Figure 2:* You show the average while above you give the number for the median.*

We have now replaced the median by the average.

"who have shown that due to the lower spatial resolution of GCMs in comparison to the regional model, SMB is underestimated at the coasts while an overestimation occurs in the interior of the ice sheet."
Resolution might play a role but model's physics also plays a major role. E.g. Fig S1 shows that MRI-CGCM3 and ECHAM-wiso have much large SMB at the margins than RACMO2, whereas they have a lower resolution.

Thank for your remark. We have added this sentence in the text:

However, models with similar resolutions may also have very different results, in particular in coastal regions (relative SMB biases of +47% and +100% for CCSM4 and MRI-CGCM3 respectively compared to RACMO for DML coast over the 1979-2005 period), suggesting a critical role of model physics in some of the GCM biases.

**Fig. S3:* Add the isotope-enabled models*

As suggested, the isotope-enabled models have been added on the figure.

"confirming that the spatial resolution has a crucial impact on the simulated SMB."
This is not convincing and not the dominant factor in my point of view.

We have added a new sentence on the role of the model's physics in the new version of the manuscript (see the previous answer on the same topic).

P11
"According to these reconstructions, this sensitivity has increased a lot for the recent period (1950–2005; 15.52 Do you think it is realistic? I don't find such an increase in sensitivity in Frieler et al. (2015)?

We totally agree that this large increase in the SMB sensitivity to surface temperature using these reconstructions is quite surprising. Actually, as you mentioned, Frieler et al. (2015) do not obtain

such an increase. This could suggest that the reconstructions used in this study suffer from issues. We have added a sentence accordingly to this result:

“However, Frieler et al. (2015) do not obtain such an increase in SMB sensitivity (only ∼+40%).”

**Figure 6:* I don’t understand why for WAIS and AP, ‘reconstructions’ (black line) is lower than model mean, while for the combination of both (West Antarctica), ‘reconstructions’ is larger than the model mean? + typos in the legend.*

The sensitivity factor for West Antarctica is not the average of the sensitivity factors of AP and WAIS. For the three aggregated regions (i.e. West Antarctica, East Antarctica, and Antarctica), our resulting sensitivity factors are based on SMB and SAT averaged over the regions. Because of some compensations between regions, what is observed for AP and WAIS can be different from what is observed for West Antarctica. The same behavior is noticed for Antarctica as a whole. Sensitivity factors deduced from the reconstruction for all sub-Antarctic regions are lower than the model mean, while for the continent as a whole, the value for the reconstruction is very close to the model mean.

P12
"The analysis of isotope-enabled model results reinforces this hypothesis (Fig. 7): the iHadCM3 outputs show high correlations between these two variables."
In the sub-section 4.3, you only focus on the iHadCM3 outputs without explicitly announcing it and explaining why you did this choice.

Throughout the text we mainly focused on the iHadCM3 model because, in contrast to the other isotope-enabled models (ECHAM5-wiso and ECHAM5/MPI-OM), iHadCM3 offers an ensemble of simulations, which is a significant advantage for data assimilation.

We added a few words on the reason of our choice at the end of the section 3.1.:

“Because iHadCM3 offers an ensemble of seven simulations, while the other isotope-enable models have only a single realization, we mainly focus on the iHadCM3 outputs in the manuscript. Dealing with an ensemble instead of a single simulation increases the probability of finding model results close to the assimilated records during the data assimilation process.”

P16
"(estimated by the weighted variance of the particles with non-zero weight)"
Define this weight/metric in the method section. What is the threshold?

After each particle has received a weight depending on its likelihood, all the weights are multiplied by the total particle number. Then, the weights are rounded to the nearest integer toward negative infinity. Therefore, the maximum value of the weight is the number of particles and the minimum value is zero. We have specified in the new version of the manuscript how the weights are computed:

“Depending on its likelihood, each particle receives a weight. Then, all the weights are multiplied by the number of particles and rounded to the nearest integer toward negative infinity by ensuring that the sum of the weights equals the number of particles throughout the data assimilation process (see Dubinkina et al., 2011 for details).”

"When assimilating both δ18O and SMB, the SMB reconstruction is in good agreement with the reconstruction of Thomas et al. (2017)."

As expected as Thomas is assimilated.

Indeed, this is expected. However, we assimilate both $\delta^{18}O$ and SMB and not only SMB. Therefore, we constrain the model with two types of information. This can lead to a SMB reconstruction different from the reconstruction of Thomas et al. (2017) and indeed the reconstruction is different than the one assimilating only SMB (Figure S8). Additionally, if model outputs and assimilated records are too different, the resulting data assimilation-based reconstruction can highly differ from the data assimilated. If the resulting data assimilation-based reconstruction is close to the assimilated records, it means that no inconsistency is found between model results and the assimilate records.

Nevertheless, as this is not a surprising result, we have added "as expected" at the end of the sentence.

P18
"who suggest an increase of the SMB sensitivity to surface temperature for the future in Antarctica,"
Can you give a number?

According to Frierler et al. (2015), this increase is about 40% (Table 1). It has been added in the new version of the manuscript.

"The GCMs may have biases in the simulated temperature changes or in their response to anthropogenic forcing."
This is very general, what are the known biases in GCMs?

We agree that this sentence in the discussion/conclusions section is very general. We have added a couple of sentences regarding the GCM biases:

"The GCMs may have biases in the simulated temperature changes. For example, as shown by Klein et al. (2019), GCMs display on average a homogeneous warming over Antarctica during the last decades while observations mainly show a warming for West Antarctica with no significant change for East Antarctica. Additionally, climate model simulations generally display a warming starting in the 19th century in Antarctica while it begins much later in proxy-based reconstructions (Abram et al., 2016)."

"This may contribute to an overestimation of the contribution of the simple thermodynamic link between temperature and precipitation and thus snow accumulation while it underestimates the role of changes in atmospheric circulation variability.
"Any reference on this point?

We have added three papers supporting this point.

1. Abram, N. J., McGregor, H. V., Tierney, J. E., Evans, M. N., McKay, N. P., Kaufman, D. S., Thirumalai, K., Martrat, B., Goosse, H., Phipps,S. J., Steig, E. J., Kilbourne, K. H., Saenger, C. P., Zinke, J., Leduc, G., Addison, J. A., Mortyn, P. G., Seidenkrantz, M. S., Sicre, M. A.,Selvaraj, K., Filipsson, H. L., Neukom, R., Gergis, J., Curran, M. A., and Von Gunten, L. (2016): Early onset of industrial-era warming across the oceans and continents, Nature, 536, 411–418, https://doi.org/10.1038/nature19082.

2. Klein, F., Abram, N. J., Curran, M. A. J., Goosse, H., Goursaud, S., Masson-Delmotte, V., Moy, A., Neukom, R., Orsi, A., Sjolte, J., Steiger, N., Stenni, B., and Werner, M. (2019): Assessing the

robustness of Antarctic temperature reconstructions over the past 2 millennia using pseudoproxy and data assimilation experiments, Clim. Past, 15, 661–684, https://doi.org/10.5194/cp-15-661-2019.

3. PAGES 2k-PMIP3 group: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium (2015), Clim. Past, 11, 1673–1699, https://doi.org/10.5194/cp-11-1673-2015.

The first paper shows that GCMs may imperfectly simulate the main mode of atmospheric variability over the last millennium. The other papers suggest that the model response to anthropogenic forcing (radiative forcing) is too important relatively to changes in general atmospheric circulation.

"According to Neukom et al. (2018), uncertainties in the reconstructions (the noise in proxy data and the deficiencies in the reconstruction methods) and the data sampling could be an explanation of the observed discrepancy between models and reconstructions."
Give some key details on how it is proven.

We have added the method used by Neukom et al. (2018) in the new version of the manuscript:

“To understand the potential origin of the disagreements between model results and reconstructions over the last millennium, Neukom et al. (2018) used pseudoproxy experiments. They found that uncertainties in the reconstructions (the noise in proxy data and the properties of the reconstruction methods) and the data sampling could be an explanation for many observed discrepancies between models and reconstructions."

"surface temperature over the period 1958–2010"
Add the reference (Nicolas and Bromwich, 2014)

Done.

P19
"Regarding changes in SMB over the last two centuries, our reconstruction shows large regional differences in SMB trends, both in magnitude and in sign, in accordance with Medley and Thomas (2019; Fig. S12)."
A word on the fact that DA assimilate Thomas 2017, which use the same ice core dataset as in Medley and Thomas 2019? So it is not surprising that patterns are similar?

As the method used by Medley and Thomas (2019) is different than ours, we could have had different results (even if the ice core dataset is the same). Unlike their method, we do not make any assumption on the stationarity of the link between the reanalysis (that they use) and the ice core dataset. Getting similar results thus shows that by using different methods, we obtain similar results, which gives more robustness to these results. However, we have added something in the corresponding sentence accordingly:

“Regarding changes in SMB over the last two centuries, our reconstruction shows large regional differences in SMB trends, both in magnitude and in sign, in accordance with Medley and Thomas (2019; Fig. S12) who used the same ice core dataset but a different method.”

"This is supported by a strong link between these two variables in observations, in particular for East Antarctica (r=0.82, statistically significant)."
Specify that is between Thomas et al 2017 and NB14, and does not work with Stenni2017

The specification has been added in the text:

“This is supported by a strong link between these two variables in observations when using snow accumulation data from Thomas et al. (2017) and surface temperatures from Nicolas and Bromwich (2014), in particular for East Antarctica (r=0.82, statistically significant).”

"By using data assimilation, no assumption such as stationarity or long calibration periods is required to estimate the link between variables"
Please also include the limitations of the data assimilation method

We propose to add this sentence:

“However, to get a skillful data assimilation-based reconstruction, it is essential that the selected climate models have an adequate representation of climate variability and that good uncertainty estimates are available for the chosen datasets.”