

**Final author response to referees 1 and 2 for
"The regional scale surface mass balance of Pine Island Glacier, West Antarctica over the period 2005–2014, derived from airborne radar soundings and neutron probe measurements" by S.Kowalewski et al.**

The Cryosphere Discuss.

<https://doi.org/10.5194/tc-2020-102>

We thank Dr. Brooke Medley and the second anonymous referee for their thoughtful comments and suggestions to improve the manuscript. We respond to each of the referee comments below. Line numbers given in our response refer to the discussion version of the manuscript prior to the revision and our response being in plain text.

Stefan Kowalewski

Response to referee 1, Brooke Medley

The authors present a very thorough and well conducted analysis of airborne radar data collected over the Pine Island Glacier catchment to recover estimates of surface mass balance. By tracking single horizon (circa 2005), dated through unique neutron probe depth–age and depth-density information, they successfully mapped SMB over a large portion of the Pine Island catchment area. The radar technique presented is robust, and the authors apply an improved kriging technique to spatially interpolate the tracks to a more complete grid. The results are largely in agreement with a prior work, suggesting little change in snow accumulation over the region and also highlighting the robustness of the approach.

The paper details a significant amount of work that is praiseworthy. A few clarifications are required to improve transparency in a few locations (see Major Comments). The paper is generally well-written and outlined but could use substantial editing for improved flow and clarity. It is great to read such interesting studies; I have made a few major and minor comments below for the authors to consider to help improve the presented work.

Major Comments

(MC1) - *On line 49 there is mention of "adjust"ing layers at crossovers. How was this accomplished? Will this impact the results? In the description later on line 132–134, there should be discussion regarding the magnitude of the adjustments. Some discussion of what this means for the total uncertainty would be welcome. Are only the layer traces impacted or are the layer ages adjusted to the adjusted layer depth?*

We agree that "adjusting layers" can be misleading in this context. We changed line 49 from

"In addition, the ASIRAS flight track contains several self-intersections so that we can compare and adjust traced layers at those points." to

"In addition, the ASIRAS flight track contains several crossovers which we used to validate the same isochronal reflector from different directions."

How was this accomplished?

Previous version: "We compared the traced layer at 34 intersections and 8 nearby flight track segments. In case of layer mismatches, the layer tracing was iteratively adjusted to find a consistent layer for the entire flight track."

We clarified this point in the new version as follows: "We checked the traced layer for possible mismatches at 34 crossover points and 8 nearby flight track segments, which may have resulted from systematic errors in the initial manual layer tracing.

Such mismatches were particularly observed along challenging profile sections and fixed by retracing the reflection layers which yield the best match at the crossover points."

Will this impact the results? In the description later on line 132-134, there should be discussion regarding the magnitude of the adjustments. Some discussion of what this means for the total uncertainty would be welcome. Are only the layer traces impacted or are the layer ages adjusted to the adjusted layer depth?

We think that our previous terminology, i.e. "adjusting layers", led to some confusion at this point. As outlined in the revised version of Line 132, our corrections refer to potential errors in the initial layer tracing, which we fixed after the inspection of crossover points. There is no remaining "adjustment" to the final layer trace, which would impact our SMB results.

(MC2) - *Regarding the analysis of the NP density profiles: please clarify in the Fig. 2 caption and in the paragraph beginning Line 88 that only $n = 22$ profiles are used in the calculation of the depth-density profile? If the entire 43 individual profiles are used, it would act to minimize the standard deviations of the profile. Are the "43 NP profiles" in line 89 all from traverse T2?*

As discussed on line 89, we are merging the shallow NP profile with the nearby deeper profile to generate a single density depth profile at the considered iSTAR site. However, we agree that counting both profiles as single entities can be confusing. Therefore, we removed the counting number "43" from the text and consider the merged density depth profiles as single entities in the revised text.

(MC3) - *The authors comment that the picked layer depths (Lines 142-143) do not necessarily associate with a peak in the density profile yet argue that density-driven dielectric contrasts form the radar reflection horizons. Do the authors have any insight into why that might be? Some further discussion of this would strengthen the paper.*

We agree that this point must be clarified in the text, as it is important to differentiate between the annual density modulation in the NP profiles and formation of hoar and ice layers, which generate a pronounced dielectric contrast due to their density contrast to the surrounding firn matrix. First, we changed Line 116 from

"We assume that the annual peaks in density lead to reflection layers because of the associated dielectric constant (Eisen et al., 2003)." to

"We assume that internal reflection layers are generated by the dielectric contrast at embedded thin ice and hoar layers (Arcone et al., 2004, 2005) and that these layers are formed at regional scales around summer/autumn (Medley et al., 2013). These layers may coincide with the annual density modulation, which we observe with the NP measurements."

We removed the following Lines 118-122 and shifted the discussion of intra-annual layers to the following Measurement Error section. On line 136 we discuss our assumptions made on stratigraphic characteristics (see our response to the first major comment of Referee 2).

(MC4) - *Uncertainty Analysis: The authors clearly put in a significant amount of time designing the uncertainty analysis, which is quite commendable and appreciated. Some additional details would help clarify the exact plan since the authors produce two estimates (Fig 4 a and c), and it become unclear which is being used in what way.*

This should be clear from the revised caption of Figure 4 (see our response to mc3) and we added the following sentence to Line 187:

"We consider the combined measurement error based on Fig. 4 (c-d) for our SMB estimates."

"It is not entirely clear why there is a "picking" uncertainty in the age uncertainty as well as an evaluation of the stan-

standard deviation of the ages in Table 2. Wouldn't any uncertainty in picking layers manifest in the numbers in Table 2? There is also uncertainty in the NP dating; perhaps this is what the +/- 1 year uncertainty is meant for?"

We agree that the reasoning behind the "picking" uncertainty requires clarification in the text. According to our revised "Measurement error estimation" section, the +/- 1 year uncertainty is meant for the potential removal of an annual layer or addition of an intra-annual layer (see our response to MC1 of Referee 2).

What does the small-scale variability in snow accumulation look like? Can comparisons be made from over 2 km away?

We assume that this question refers to the lateral variability in the stratigraphy (Line 159) at the range of displacement between the iSTAR location and ASIRAS point of closest approach. Typically the stratigraphy is not flat at the iSTAR locations and may contain pronounced gradients which are visible in the radar imagery. We agree that the question is justified as to whether a layer dating based on measurements separated by 2 km in one case is reliable. We included this extreme case in the analysis as the stratigraphy was rather flat according to the radar imagery. We suppose that the reliability of such comparisons strongly depends on the degree of folding in the stratigraphy, which is quite variable along the flight track. Even though we are limited to the flight direction (Line 149), we consider the standard deviation σ_x (Line 148) as a benchmark for the uncertainty which results from the lateral folding in the stratigraphy between the closest ASIRAS measurements and iSTAR locations. As proposed in our reply to mC5 of Referee 2, we plan to make the radar data available at PANGAEA to allow for a further inspection of small scale variability.

Perhaps a good alternative would be a simple weighted mean of the age by distance (as well as standard deviation).

The suggested alternative sounds reasonable, but we do not expect a noticeable impact on the results due to the already small σ_x values according to Table 2. As we are limited in our assessment of layer folding towards the flight direction, we think that it is reasonable to make a more conservative assumption by giving all age estimates along the flight track the same weight.

Please also clarify that in Eqn 5, the "spatial" component refers to the density profile.

We may already refer to Line 187:

"It is evident from Eq. (4) that the spatial uncertainty of the density profile affects both the integration depth and incremental mass."

which explains that the spatial term in Eq. 5 relates to the variability in density. In addition, we relate the spatial error to the variability in density in the new caption of Fig. 4 (see our response to mc3).

Finally, is it justifiable to assume that the spatial errors in density are uncorrelated with depth? While firn depth-density is noisy with depth, there is no reason to expect that under different accumulation/temperature regimes, there will be large biases between measurements (that will accumulate with depth, rather than cancel out).

We assume for Eq.(5) that errors are uncorrelated (Line 171). This also includes spatial errors with depth as questioned above. However, we replace the root-sum-squares (RSS) of spatial error terms with their absolute values to estimate the measurement uncertainty for our results. This implies that different errors cannot cancel each other out (Line 182), i.e. we are considering the most conservative scenario by this assumption. Any error correlation with depth will yield to an uncertainty equal or below this uncertainty threshold value. In this connection, we are over- rather than underestimating the measurement error by neglecting the covariance terms in Eq.(5) that account for any spatial error correlation with depth. Therefore, we think that this question should already be answered on Lines (181-185).

Minor Comments

(mc1) - Some description of the range of "N" in Section 2.4 would help clarify the robustness of the methodology. Maybe the authors could add the N to Table 2.

We added N values to Table 2

Also, be consistent with significant digits. Why do some have 2 decimal places and others only one? Based on the remainder of the text, it looks like the standard is just 1 decimal place.

The number of significant digits was selected based on the precision of lateral uncertainty and therefore we stick to the least significant digit, which varies.

Also, please explain the "comment"s in the Table 2 caption.

Old caption, " Dating (Year) with associated uncertainty of closest ASIRAS reflection layer to ith iSTAR site. "Track number" refers to the ASIRAS flight track naming convention and " ΔD " is the closest distance between the ASIRAS track and the iSTAR site. Years in brackets are discarded from the regional layer age estimation for the reasons given in the Comment column."

New caption, "Dated reflection layer year at nearby iSTAR sites. "Track number" refers to the ASIRAS flight track naming convention (year of measurement season [4 digits], measurement type [1 digit], profile segment number [3 digits]), " ΔD " is the closest distance between the ASIRAS track and iSTAR site, and N is the number of picking samples considered for layer dating. Years in brackets are discarded from the regional layer age estimation as follows: Significant departure between traced layer and depth–age scale at site 2 (see main text), layer gaps due to high noise levels in the radargram, layer is significantly exceeding the dated NP profile depth (values in brackets indicate extrapolated depth–age values).

(mc2) - *Is the uncertainty in the depth of the tracked layer in determination of the age uncertainty accounted for? From Line 148, it sounds like only a standard deviation of points is used. Additional uncertainty is imposed from the translation of the uncertainty in depth to an uncertainty in age based on the slope of the depth–age profile.*

Indeed, we are considering the standard deviation from N layer age estimates around each site for the given $2\Delta D$ interval. However, we are also considering the local density profile for the TWT to depth conversion (Line 140) and as such, the scattering follows the slope of the associated depth–age scale.

We noticed that we may determine the layer age at each site by averaging its N layer age estimates instead of calculating a mean depth for all points and relate this value to an age value. We tested and implemented the former method in the revised version but it didn't have an impact on the final layer age of $a = 10.1 \pm 1.4$.

We revised the Lines 139-151. Old version: "At N points along the closest approach of the ASIRAS track to each iSTAR site we estimated the depth of the reflection layer from its TWT, using the fitted density profile from that site. We also tested the use of the measured density profile for the TWT conversion instead, but we find that the impact on our results is negligible similar to Morris et al. (2017). The estimated depths do not necessarily coincide with the depth of a peak in the density profile i.e with the start of a mass balance year. We assign an estimated date to the mean value of N points by interpolation using the depth–age scale from the iSTAR site. These points are centred around the point of closest approach to the nearby iSTAR GPS location. We chose the associated interval length to be twice the distance ΔD between the iSTAR GPS location and point of closest approach of the flight track. Finally, we added six months to the estimated dates according to the mass balance year definition. This yields the estimated Year values for the reflection layer at each iSTAR site in Tab. 2. Following Konrad et al. (2019) we express the dating uncertainty as the standard deviation σ_x in the N measurements. In this sense, our error estimate is more conservative than the standard error of the mean. Furthermore, we assume that the uncertainty due to local variation in

the stratigraphy is isotropic, which does not generally need to be true. However, according to Tab. 2 in most cases the overall impact of this effect is one order of magnitude smaller than the variability of dated years among all iSTAR sites."

New version: "Based on annual density markers, we can relate the snow and firn depth at each iSTAR site to its associated age and determine the reflection layer age from its depth at each cross section. Here, we use an exponential fit of the local density–depth profile for the TWT to depth conversion. The lateral displacement ΔD between the point of closest approach of the flight track and iSTAR site adds to the reflection layer dating uncertainty. We therefore consider all N points which lie within a $2\Delta D$ interval along the flight track and which is centred at the point of closest approach for the layer dating. Based on the local depth–age scale, we relate the estimated depths of N points to their ages and assign the final layer date to their mean value. To account for the mass balance year definition above, we add six months to the mean layer date, which is listed for all iSTAR positions in Tab. 2. In addition, we estimate the dating uncertainty from the N lateral estimates by their standard deviation σ_x . In this sense, our error estimate is more conservative than the standard error of the mean. Furthermore, we assume that the uncertainty due to local variation in the stratigraphy is isotropic, which does not generally need to be true. However, according to Tab. 2 the overall impact of this effect is one order of magnitude smaller than the variability of layer age values among all iSTAR sites in most cases."

(mc3) - *For Figure 4, please indicate that the darker shades of grey refer to a higher density of layers. Also, the caption should have more information. It's difficult to attribute what calculations are made in Figure 4 c,d. Perhaps clarify that the "Spatial" error is in reference to the variability in density.*

We changed the old caption, "Spatial, temporal, digitization, and combined relative errors (left panels) and error partitioning (right panels). Grey background shades indicate the depth distribution of the traced 2005 reflection layer. (a,b) Based on error propagation according to Eq.(5). (c,d) Excluded error cancellation for the spatial error terms (see text)."

as follows: "Spatial, temporal, digitization, and combined SMB measurement errors, which relate to the variability in density, dating uncertainty, and ASIRAS sampling accuracy respectively: Relative errors (left panels) and error partitioning (right panels). Grey background shades indicate the depth distribution of the traced 2005 reflection layer with higher number concentrations towards darker shaded. (a,b) Based on error propagation according to Eq.(5). (c,d) Excluded spatial error cancellation in Eq.(5) [see main text] and considered for the final error estimation of this study."

(mc4) - *Several studies have shown that RACMO is biased (often too low accumulation in the interior), so is it sufficient to take the RACMO values at face value when adding them to the regions that you cannot resolve. Perhaps a correction could be made to the RACMO data first using overlap between the radar estimates and the model over the region of overlap.*

Despite the significant correlation, which we found between the elevation and SMB bias, the correlation remains weak and as such, one may question as to whether a correction based on a weak correlation is justified. The inclusion of MAR results to the revised manuscript should at least provide an alternative estimate, which helps to quantify the impact of an elevation dependent bias to our estimates.

(mc5) - *The uncertainties provided are found to be larger than in ME14, which is expected as the authors note. From the text, the temporal error from this work is $1.4 \text{ y} / 10.1 \text{ y} = 13.9\%$, which indicates that the smallest error for any radar-derived SMB estimate is effectively 14%. In ME14, the temporal uncertainty is $1 \text{ y} / 25 \text{ yr} = 4\%$, suggesting that the temporal uncertainties are much lower in ME14. That reduction is likely due to the robust dating techniques used on the firn cores used in that study. Based on this alone, it is not clear how the shorter time window and larger dating uncertainties in this analysis do not at least account for a substantial amount of the increased basin-wide uncertainty values found in this work.*

We agree that the smaller layer age uncertainty and the larger temporal averaging interval reduce the temporal uncertainty in M14 noticeably. Even if we assumed the same layer age uncertainty of 1.4 y for M14, the resulting temporal uncertainty of $1.4/25 = 5.6\%$ would still remain noticeably smaller compared to this study. It is difficult though to give an expectation about its

impact on the measurement error grid based on the percentages mentioned above. The displayed error grids in M14 are limited to the combined error, hence, we lack information about the spatial distribution of the partitioning between the interpolation and measurement errors in M14. Furthermore, both studies use different Krige methodologies which use different expressions for the interpolation error (Line 234-240) and may differ in their response to added noise to the SMB input values.

Locally, it does appear that uncertainties from the kriging technique are much larger in this work than in ME14; however, they appear smaller than some of the uncertainties (say, in the southernmost reaches of the catchment) than in ME14. The total ME14 basin-wide uncertainty is 9% (6.8/78.3), whereas this work is 24% (19.2/79.9). Adding an additional 10 to the ME14 estimates puts values at 19%. Therefore, the likely impact of the kriging technique is on the order of 5%. This is a very simplistic take but should be robust from an order of magnitude perspective.

Indeed, a local cluster of high interpolation errors North of the flight track between iSTAR 1 to 6 (Line 375) is significantly higher compared to the combined errors in M14. Care must be taken though when comparing the combined error grid maps due to the considered larger bin range of this study [0%,100%) and M14 [0%,30%]. For instance, the second and third colour shades (i.e.[10%,20%] and [20%,30%]) of this study already reach the maximum colour bin of [25%,30%] of M14 to the southernmost reaches of the catchment.

Minor Edits

Line 4: consider replacing "allowing" since "allow" is used in the previous sentence We replaced "Ground based neutron probe measurements of snow density at 22 locations allow us to derive SMB from the annual internal radar reflection layers. The 2005 layer was traced for a total distance of 2367 km allowing us to determine annual mean SMB for the period 2005–2014."

with "Ground based neutron probe measurements provide information of snow and firn density with depth at 22 locations and were used to date internal annual reflection layers. The 2005 layer was traced for a total distance of 2367 km to determine annual mean SMB for the period 2005–2014."

Line 17: remove "the" in front of "upwelling" Line 18: consider replacing "stimulating" with "initiating" Line 28: replace "in particularly" with "in particular". Line 30: replace "in the following" with "hereinafter" Lines 32-33: remove "the" before "logistical. Line 33: change to "dielectric properties of snow and firn. Line 37: replace "In the following," with "Hereinafter,". Line 44: replace "trace" with "track" Line 48: replace "self-intersections" with "crossovers" Line 50: change "measurement" to "measurements. Figure 3: add a righthand y-axis with depth equivalent.

We applied all suggested corrections.

Line 29: clarify what is meant regarding the sentence starting with "Basin wide mass..." It is unclear what it means in this context and might require a citation.

We replaced "mass balance estimates" with "total mass input"

Note, there are several minor fixes needed beyond section 1, which will require further refinement by the authors.

Several minor fixes, e.g. restructuring of phrases, correcting grammar errors, etc. were applied beyond section 1.

This paper presents recent surface mass balance (SMB) estimates derived from airborne radar observations, and ground based neutron probe measurements of snow density along the iSTAR traverse (2013,2014) at Pine Island Glacier (PIG). This paper interestingly focuses on data uncertainties resulting from methodology and assumptions made on the interpolation error, demonstration and compares estimates with those given in previous publication from Medley et al., 2014, and with RACMO2.3p2 simulations. This paper is well written, presents an interesting new dataset for model validation and deserves to be published.

However, before publication, I have a few questions and suggestions. I hope these could improve the final paper.

Major Comments

(MC1) - *My main comment concerns the dating of reflection layers. Here, the dating relies only on big assumptions made on stratigraphy characteristics. This technique is based on differences in winter and summer snow due to changes in atmospheric conditions and radiative fluxes. I suppose that dating of reflection layers and NP data is accurate but there is no comparison with stake networks, or with a clear “absolute” dating based on anthropogenic radionuclides or volcanic horizons at several cores.*

There were no stake networks available for comparison. We added to Line 85:

"According to internal communication with the iSTAR drilling team, no absolute volcanic reference horizon was accessible for the processed iSTAR cores and therefore limits their annual markers to the H2O2 and density profiles"

Since layers sometimes mismatch at several intersections, or are excluded (around iSTAR sites 2 and 19 for instance), the final dating may be not fully robust.

This directly relates to the concerns of referee 1 in her first major comment. As outlined in our response, we clarified in the new version that initial layer mismatches were caused by systematic errors in the manual layer tracing, which were corrected during the tracing process.

As requested by referee 1 in her first minor comment, we added explanations for the excluded iSTAR sites in the text.

I understand that the authors define a layer age uncertainty of ± 1.4 year and assess the associated uncertainty in the surface accumulation, but is it possible that the layer dating uncertainty exceeds 1 year? In particular, melt or rain is expected to occur mainly in summer; but is it possible that significant surface melt (or rain) occurrences occurred at the beginning and end of summer but were separated by an “extreme” solid precipitation event (Turner et al., 2019), leading to a 2 maxima in density and in other snow characteristics used for the layer counting? Is there any snow erosion, which could remove the surface layer at locations in the area? Are there any stake farms or well constrained ice cores (with an absolute age of one or various layers) in the area, on which the authors could validate their estimates in snow accumulation?

We added the following text on line 136 to discuss the mentioned concerns about systematic errors in the layer dating:

"So far, we assumed that internal reflection layers form on an annual basis during summer/autumn but the potential formation of intra-annual reflection layers may challenge this assumption. For instance, Nicolas et al. (2017) found evidence of surface melt episodes over large parts of WAIS in response to warm air intrusion events. Scott et al. (2010) observed a strong reflection layer, which coincides with an exceptional melt layer at 22 m depth at one PIG ice core location. While such findings appear to be of sporadic nature at the basin scale and may be related to the coupling between atmospheric modes (Nicolas et

al., 2017), we find evidence of intra-annual reflection layers at coastward iSTAR sites where the snow accumulation is high. Extreme solid precipitation events may also impact the density modulation with depth (Turner et al., 2019), which is considered for the depth–age scale based on the NP measurements. Snow erosion may remove annual markers where accumulation rates are low. In addition to annual layer counting errors, the timing between the reflection layer formation and snow densification may be offset during summer/autumn. All these factors challenge the tracing and dating of the 2005 reflection layer but combining the stratigraphic information from the ASIRAS and iSTAR observations helps reducing the risk of systematic errors from erroneous layer counting. To account for the remaining risk in terms of isochronal accuracy, we assign an annual layer tracing uncertainty of $\overline{\delta t} = \pm 1$ years."

(MC2) - *Neutron probe is a really interesting way to retrieve snow density and the authors clearly took profit from this technique in the past and in the present paper. However, data rely on a few calibration steps. I had a look to previous papers from Konrad et al., 2019 and Morris et al., 2017, and I did not really understand how density data were validated before being used in the present paper. My concern is because in Figure 2 we observe that snow density is 550 kg m⁻³ at 7m below the ground level, whereas it is 600 kg m⁻³ at the same depth in Morris et al., 2017. Snow density in firn cores from Konrad et al., 2019 are hard compare here because their Figure 3 is developed until 50 m.*

We are afraid that different units for depth, i.e. water equivalent depth (Fig. 2, Morris et al., 2017) and geometric depth (Fig. 2, this study) led to some confusion. Conversion of geometric to water equivalent depth yields an approximate density of 600 kgm⁻³ at site 21 and 7 m level for this study. Even though, some of the referenced earlier studies use water equivalent scales, we prefer to keep SI units in this study to avoid potential pitfalls from the conversion to a specific reference scale as these are unambiguous. Nonetheless, the difference is well spotted by the referee and highlights the required caution to be taken when comparing different scales.

Could the author describe whether they calibrated the NP snow density data with snow pit data in the present paper or not? If not, is there a difference in the density/depth relationships between (Konrad et al., 2019), (Morris et al., 2017) and present paper. If snow density from NP measurements is biased, how will this impact the final SMB values?

Calibration of NP firn and snow density was done independently of the additional core data and based on theoretical considerations. We added a citation of Morris (2008) on line 75, which discusses the calibration equation of NP measurements. In addition, we added the following Line:

"A comparison with gravimetric density measurements at existing core profiles did not indicate a systematic bias between both measurement methods."

(MC3) - *Concerning the kriging method, is it worth using northing, easting, and elevation as explanatory variables? Would it be more relevant to use the distance from the coastline as explanatory variable? Or even the distance from Amundsen sea coast?*

Our initial motivation was to follow the methodology by M14 as close as possible for our data comparison. It is an interesting idea, if changes to the parametrisation of the regression model improved the statistical characteristics of residual SMB to the regression surface in our case. However, limiting the explanatory variables to the coastal distance would also limit the degrees of freedom in the regression model. We can only speculate at this point, but orographic effects in response to atmospheric circulation patterns may not be adequately captured by such a parametrisation. We clarified on Line 202 that we do not consider any regression model for our results:

Previous version: "An alternative approach, which is also mentioned in ME14, is a logarithmic transformation [...]"

New version: "We therefore searched for an alternative approach to generate kriged estimates from the SMB sample population of this study without the use of a regression model. Such alternative, which is also mentioned in ME14, is a logarithmic transformation [...]"

(MC4) - Comparison of ASIRAS data with RACMO2.3p2 simulations are really interesting but differences are not fully justified/explained in the text. Since differences are model dependant, it would be interesting to see potential differences with another model used in Antarctica (e.g., COSMO-CLM, HIRHAM5, MAR3.10, MetUM , see Mottram et al., 2020). Since Agosta et al. (2019) proposed potential justifications of the differences existing between RACMO2.3p2 and at least the MAR model, I believe that a comparison with the MAR model would make sense here. Indeed, in Agosta et al., 2019, large differences between RACMO and MAR are observed in regions where the RACMO2.3p2 model presents the largest differences with the ASIRAS data. A quick comparison could be relevant to discriminate whether the precipitation formation, advection of hydrometeors, and sublimation of precipitating hydrometeors (Agosta et al., 2019) are important or not in the PIG area. Data from Agosta et al. 2019 are available here: <https://zenodo.org/record/2548848#.X0St8TXgphE> If the authors are interested in higher resolution simulations, data from a more recent paper from Donat-Magnin et al., 2020, focusing in the Amundsen region are also available at: <https://doi.org/10.5281/zenodo.2815907>.

We included simulated SMBs from the high resolution MAR runs by Donat-Magnin, as suggested. We mainly focussed on the model impact on the hybrid total mass input estimates, which is mainly limited to the high elevation interior.

(MC5) - The paper largely describes differences with Medley et al. (2014) paper, but the authors never include any figure presenting the differences. I propose to include a map presenting ME14 route, SMB results and differences with the ASIRAS data.

We do not have access to the required data, while adding additional maps may affect the conciseness of the manuscript. Our main focus is on the estimates of quantitative SMB properties, which we compare with the study of M14 in the provided tables.

(MC6) - The paper is sometimes quite hard to follow for a non-expert of this area. Different datasets are used here, and the difference between ISTAR/ASIRAS and ME14 is not always clear. The authors use a many acronyms. I suggest that the authors include a table where they clearly describe the difference between ASIRAS/iSTAR data used here, and more particularly the difference with the ME14. For instance, Table 1 presents different radars (ASIRAS, pulseEKKO PRO GPR , CReSIS AccuR) part of the information is given in the caption, but perhaps the authors could also precise in the table if field campaigns were deployed on the ground or by plane,

We indicate now which radar measurements were performed airborne or groundbased in Table 1.

over what distance?

We haven't received flight track coordinates from M14 yet, hence, we cannot make a precise statement in Table 1. Mentioned distances in M14 refer to the combined flight tracks above the Pine Island and Thwaites catchment areas.

which were the reflection layers used for SMB estimates? which density measurements were considered (NP? Firn cores?)? how was performed the dating of the reflection layers?

We included the considered SMB averaging periods in Table 1.

which density measurements were considered (NP? Firn cores?)

We added the new row "considered density profiles" to Table 1.

how was performed the dating of the reflection layers? We added a row to Table 1, which lists the considered annual markers for each study.

(MC7) - *Figures could display the route followed by ME14, and where GPR from Konrad et al., 2019 was carried out.*

We still don't have access to flight lines from M14.

and where GPR from Konrad et al., 2019 was carried out.

We added the following sentence on Line 65: "The route of these observations closely follows the ASIRAS flight track and are both available at <http://gis.istar.ac.uk/> ."

Minor comments: Abstract

(mc1) - *"Thus there is no evidence of a secular trend in mass input to the PIG basin." => please be more accurate because secular may be misinterpreted here. I suggest to replace secular (here and elsewhere) by decadal.*

We now use "secular trend at decadal scales" in the abstract. "Decadal trend" alone would be more limiting. It is our intention to relate the steady acceleration in ice loss to a potential change in total mass input, i.e does a non-periodic change in total mass input amplify or compensate the mass imbalance of PIG at the observational time scale. Later passages which contain "secular" relate this term to the considered decadal time scales, which should clarify its meaning.

(mc2) - *"in particularly"*

Changed to "in particular".

(mc3) - *Lines 38-41: firn cores are only used to retrieve the depth of dated snow layer? Are they used to calibrate the neutron probe density data?*

We added on Line 75: " Further details on the calibration procedure, which is based on theoretical considerations, can be found in Morris (2008). A comparison with gravimetric density measurements at existing core profiles did not indicate a systematic bias between both measurement methods."

(mc4) - *Figure 1. ASIRAS-iSTAR survey : please also include the location of ground GPR observations from Konrad et al., 2019?*

See our response to MC6. We added the following sentence on Line 65:

"The route of these observations closely follows the ASIRAS flight track and are both available at <http://gis.istar.ac.uk/> ."

(mc5) - *Line 68, the authors write: " Due to the reported consistency between the GPR and airborne SMB measurements in Konrad et al. (2019), we limit the comparison of our results to the basin wide estimates by ME14". I feel that a figure (perhaps in the supplementary material) showing the different radargrams could help the reader. A quick data comparison, before interpolation, could also be done to see how snow density and radar data uncertainty impact the final SMB value.*

We agree that a comparison of radargrams from different platforms can be useful for an ongoing discussion on the error assumptions made by different studies. However, we are cautious about the expectation that this may be done by means of a quick data comparison as suggested herein. The radar studies of M14 and Konrad et al. (2019) use a different reflection layer for their analysis compared to this study, which already impacts the uncertainty of their SMB estimates. Local noise in the stratigraphy varies along the flight track and impacts related uncertainties in addition. Our error discussion builds upon the developed methodologies in M14 and Konrad et al. (2019), and already considers impacts from the uncertainty in the assumed regional density profile and dating precision for the selected reflection horizons. Figure 9 (b) already gives a benchmark for

other studies to the estimated measurement uncertainty of our results. However, we think that it is of help for potential future studies to add the ASIRAS radargrams, interpolated, and non-interpolated reflection layer picks in units of annual SMB and TWT to the PANGAEA repository.

(mc6,mc7) - Line 85 “*Morris et al. (2017) applied an automatic annual layer identification routine to their snow density profiles and used the annual H₂O₂ peak depths as an additional guidance for the annual layer dating.*” => *is it possible to observe the removal of on year of snow in the PIG area (due to erosion) or multiple summer maxima?*

Line 88: Here I understand that the authors did not consider the density obtained from the firn cores to compute the final SMB. What is the difference in the final SMB if the authors use the density from firn cores to calibrate the snow density profiles?

See our response to MC1

(mc8) - Line 89: *43 profiles => I suppose this means that profiles were done twice at the 22 sites?*

Yes, that’s correct, but we are now only counting the merged NP profiles as single entities at each site to avoid any confusion about the number of iSTAR sites and density profiles. (See also our response to the second major comment of referee #1).

(mc9) - Line 92: *is there any relationship between snow density profile and Accumulation/Temperature as suggested by Herron & Langway 1980 equation?*

we replaced Line 84, “ Finally we fit an exponential function to the regional mean profile (red dashed line), which we apply to the TWT to depth conversion.”

with “Morris et al. (2017) observed a two-stage Herron and Langway (1980) type densification at PIG, with the stages separated by an additional transition zone. We achieve a good fit to our regional mean profile with a simple exponential function (red dashed line, Fig. 2), which we apply to the TWT to depth conversion.”

(mc10) - Line 120: *how does hoar produce thin ice layers? Is it possible to have short rain our melt events in spring or late summer? Could wind erosion or sublimation create wind crusts in this region?*

According to our cited reference to Arcone et al. (2004) on Line 120, i.e. “[...] a hoar layer, which frequently occurs beneath thin ice layers because it is the source of vapor that creates the ice layer”, therefore, the answer to this question should already be provided by the cited reference.

Is it possible to have short rain our melt events in spring or late summer? Could wind erosion or sublimation create wind crusts in this region?

See our response to MC1.

(mc11) - Line 144: “*We assign an estimated date to the mean value of N points*” => *do you mean “the mean value of depth”?*

Yes, we mean “the mean value of depth”. To make our methodology more clear, we rephrased Lines 139 to 147 as follows:

“Based on annual density markers, we can relate the snow and firn depth at each iSTAR site to its associated age and determine the reflection layer age from its depth at each cross section. Here, we use an exponential fit of the local density–depth profile for the TWT to depth conversion. The lateral displacement ΔD between the point of closest approach of the flight track and iSTAR site adds to the reflection layer dating uncertainty. We therefore consider all N points which lie within a $2\Delta D$ interval along

the flight track and which is centred at the point of closest approach for the layer dating. Based on the local depth–age scale, we relate the estimated depths of N points to their ages and assign the final layer date to their mean value. To account for the mass balance year definition above, we add six months to the mean layer date, which is listed for all iSTAR positions in Tab. 2".

(mc12) - Line 198: "may by due"

Corrected typo: "may be due"

(mc13) - Line 275: *please discuss this sentence according to Agosta et al. (2019) results. Indeed, according to this publication, sublimation of precipitating hydrometeors are missed at low elevation in RACMO2.3p2. This could justify that RACMO 2.3p2 presents a positive bias at low elevation. Conversely, they suggest that MAR snowfall rates generally exceed those simulated by RACMO2.3p2, by more than 30% on the lee side of the West AIS (Marie Byrd Land toward Ross ice shelf), and close to crests at the ice sheet margins. Here, a comparison with MAR could be interesting.*

We included a discussion of the results by Agosta et al. (2019) in the new section "Elevation dependent model drift", which we limited to the overestimation of RACMO results, because of their impact on our hybrid SMB map.

New section: "The observational SMB estimates by M14 indicate an elevation dependent drift of simulated SMB from RACMO. The authors find that RACMO underestimates the SMB at the high-elevation interior, which would also impact our ASIRAS–RACMO based estimates of total mass input. Indeed, this finding is also reflected in our data (see supplement S1) and suggests that the ASIRAS–RACMO based total mass input estimates are biased by the underestimated SMB contribution from RACMO. According to (Agosta et al., 2019), the opposite may apply for the ASIRAS–MAR based estimates. The authors observe a tendency for MAR to overestimate accumulation on Ross-Marie Byrd Land and conclude that differences between MAR and RACMO2 are very likely related to differences in the advection inland. Similar to our elevation dependent comparison between ASIRAS and RACMO SMB estimates, we find evidence of a drift in the MAR estimates with an opposite sign according to S1. We conclude that the best estimate for total mass input lies between ASIRAS–RACMO and ASIRAS–MAR estimates."

(mc14) - Lines 288: *The authors refer to the ASIRAS or the Hybrid SMB estimates, but line 296 they refer to the ASIRAS vs. the ME14 one, whereas they refer to the Hybrid estimate at line 298. The difference between these estimates is not clear. Why do the authors use ASIRAS / hybrid estimates in different parts of the text?*

We revised this section to include the MAR data and clarified our reference to considered hybrid estimates.

Previous version: "The Pine Island Σ_+ values are in agreement between all data sets within the estimated error margins. This is different for the Wedge outlines, where the RACMO Σ_+ estimates are between 35–40) % lower compared to the estimates of this study and ME14. Increasing the averaging time of RACMO estimates to the 1985–2009/10 period of the ME14 results yields an increase of Σ_+ by 2% for the Pine Island and 8% for the Wedge outlines. In comparison with the ASIRAS and ME14 estimates, the simulated total mass input for the Wedge outlines remains outside the error margins of both observations. Considering the further SMB properties according to Tab. 3, the hybrid SMB estimates show the largest variability and range for the Pine Island outlines. This is different for the Wedge outlines, where the hybrid SMB estimates appear to be larger for most of the basin area."

New version: "The Pine Island Σ_+ values are in agreement between all data sets within the estimated error margins. This is different for the Wedge area, where the RACMO Σ_+ estimates are between 35–40 % lower compared to the estimates of this study and ME14. Increasing the averaging time of RACMO estimates to the 1985–2009/10 period of the ME14 results yields an increase of Σ_+ by 2% for the Pine Island and 8% for the Wedge area. However, the RACMO based total mass input to the Wedge area remains below the observational error margins. With regard to the MAR estimates, we find that Σ_+ values are about 5% higher for Pine Island and 38% higher for the Wedge area. The higher MAR SMB compared to RACMO towards

the southern interior yields a 3% increase for hybrid SMB estimates based on complementary MAR estimates. Considering the additional SMB properties according to Tab. 3, the hybrid based SMB estimates of this study show the largest variability, except for the Wedge area."

(mc15) - Line 325: *"but also measured density and strainrate profiles suggest a mean annual SMB of 200 kgm-2yr-1 based on theoretical grounds, which both are in a better agreement with the collocated ASIRAS based results." => Is it possible to explain briefly the way this value of 200kg m-2 is computed?*

We replaced "based on theoretical grounds" with "based on the Herron and Langway stage 1 equation"

(mc16) - Line 352: *"secular" => decadal?*

See our response to mc1.

(mc17) - Line 395: *"Inspection of the artificial cluster highlighted in Fig. 5 revealed" => replace by fig. 8?*

We replaced "Fig. 5" by "Fig. 8".

(mc18) - Table 2: *please include site coordinates and elevation.*

We added latitude, longitude, and elevation to Table 2

(mc19) - Figure 7: *Please define NN*

We replaced "NN" with "Rmax=Nearest Neighbours in legend and revised caption: "PP-plots between SMB observations and estimates based on OK and OLK interpolation methods for varying thresholds of their maximum distance Rmax. [...] Average PP-distances (see main text for definition) and SMB values are shown in the legend."

Previous version: "PP-plots between SMB observations and estimates based on OK and OLK interpolation methods for varying maximum estimate distances with regard to the closest measurement locations. [...] Average PP-distances (see text) and SMB values are shown in the figure."

(mc20) - Figure 8: *I suggest to include the elevation contours. Is it correct to write Eastings when the x-axis is related to the north-south direction, and Northings for the y-axis when it is on the east/west direction?"*

We included light-toned 200 m elevation contours and replaced axis annotation to Polar Stereographic X and Y to avoid any confusion.

I don't understand why the kriging procedure induce strange vertical (or horizontal) lines of similar values in Figure 8a (this point is particularly visible near the glacier outlet)? Is it because northing, easting, and elevation are used as explanatory variables? If yes, is it not more relevant to use the distance from the coastline as explanatory variable? Or even the distance from Amundsen sea coast?

We added the following explanation to Line 259: "Furthermore, some streak artefacts are visible from the interpolation, which are mainly caused by the quadrant criterion of the OLK estimation. Increasing the number of nearest neighbours helps reducing these artefacts but at the cost of PP-agreement in terms of Fig. 7. We therefore kept the OLK settings according to Fig. 8 (a) hereinafter."