Author responses to Referee #1

Major Comment 1 (MC1) “On line 49 there is mention of "adjust"ing layers at crossovers.”

We agree that the term “adjusting” is not properly describing the manual layer “correction”, which we did during the layer tracing.

“How was this accomplished?”

While we initially used the annual markers in the density profiles from the Neutron Probe (NP) measurements to trace the same isochronic reflection layer throughout the entire flight survey, we recognised crossover points where the traced layer did not match between the intersecting profiles. This was typically the case for profiles which are subject to complex stratigraphy or poor signal returns. As stated on Line 128, we performed a visual inspection of the vertical reflection layer sequence to guide the layer bridging wherever it is required. Intermediate reflection features or low signal-to-noise may challenge the visual inspection, which also depends on the experience in the manual layer tracing. We therefore considered layer mismatches at crossover sections as indicators for systematic errors that occurred during the manual bridging at difficult radargram sections. Typically, retracing the adjacent layer of one profile could fix the mismatch at one crossover point, but may yield a new mismatch at another crossover point. Hence, we were iteratively retracing challenging profile sections until we achieved a plausible intersection of the same layer at all crossover points.
We can clarify this point according to the explanation above in the revised text and replace the term “adjusting” with “correcting”.

“Will this impact the results?”

If we selected the reflection layer based on the annual markers at iSTAR sites only, we would shift the ASIRAS results towards a better agreement with the iSTAR results at their closest approach. However, this would certainly result in inconsistent trace jumps at crossover points. Hence, the applied correction makes the ASIRAS estimates independent from single iSTAR sites, which we also point out on Line 134.

“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”

It is difficult to reconstruct the impact of these corrections to the results because they were continuously applied during the manual layer tracing with the aim of producing a consistent trace for the selected reflection layer as the final product.

“Are only the layer traces impacted or are the layer ages adjusted to the adjusted layer depth?”

The layer age of the final trace is estimated based on the depth age scales at all iSTAR sites (Table 2). It is correct that the allocated layer age changes with the new depth of a retraced (i.e. corrected) layer. We think that this point becomes clear, once we explained the manual correction as discussed above in the revised text.

MC2 “Regarding the analysis of the NP density profiles: please clarify in the Fig. 2 caption an 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?”

We agree that the stated number of 43 NP profile scan lead to some confusion, as also questioned in mC8 of Referee #2. According to our reply, we count the adjacent profile sections at each site (apart from site 2) separately on Line 89. These profiles were concatenated to a single profile per site (no concatenation necessary at site 2). Indeed, Figure 2 already shows the concatenated 21 profiles in addition to the single profile of site 2, which we shall clarify 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 suppose that local noise in the stratigraphy is one obvious explanation for the observed differences between reflection layer and density peak depths. Keeping in mind that we are comparing point measurements with spatially integrated radar profiles (4.5 m along track bin, according to Table 1), which are separated by several 10s or 100s of metres (with one even being 2 km), the resulting dating uncertainty from the ensemble of site matches (Table 2) is rather good and my not be expected from areas with low accumulation (Line 314).

Another explanation is a difference in the timing of the formation of the reflection layer and density peak, i.e. the dielectric contrast generated by the formation of thin ice layers (Arcone et al., 2004 and 2005) may not coincide with density peaks.

This question is closely related to MC1 of the Referee #2 and will be considered as discussed above in the methods section.

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."

Figures 4 a and c show two different results according to Line 182-185, where we say that we consider the error estimates based on Fig. 4 (c and d) for the measurement uncertainty estimation. We can add this to the caption.

"It is not entirely clear why there is a "picking" uncertainty in the age uncertainty as well as an evaluation of the 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 term "picking uncertainty" is confusing with regard to the additional evaluation of layer intercepts at iSTAR sites as pointed out above. Indeed, we shall use "layer dating uncertainty" instead and clarify the reasoning for this uncertainty in the text as follows.

While other studies such as Medley et al. (2013) consider an uncertainty of +/-1 month for the formation of the annual reflection layer, we extend this uncertainty range to one year to account for systematic errors in the picking process which may arise from ambiguous reflection features or bridging of difficult sections along the flight track (see also our response to MC1, Referee #2). Hence, we are using the same uncertainty range for the parameter $\overline{\delta t}$ = +/- 1 year (Line 158) as in Konrad et al. (2019) but for a different reasoning (i.e. +/- 1 year uncertainty to account for the 1 year offset between GPR and firn-core data).

"What does the small-scale variability in snow accumulation look like?"

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 stratography 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 proposed alternative sounds reasonable, but we would 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."

Yes, we shall clarify this in the text.

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 pessimistic 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 Comment (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."

Yes, we can add the range of N to Table 2 to give a better sense of the considered sample sizes.

“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.”

In short, “extrapolated” means that the reflection layer exceeds the depth of the deepest annual marker in the NP density profile at this site, hence, the layer age has been estimated based on extrapolation of the depth age scale. We discarded these measurements because we cannot assess the robustness of these extrapolated values. “Noise” means that no stratigraphy was visible in the profile section, e.g. due to exceeding rolling angles of the plane. For the erroneous NP dating, we refer to Line 320-326.

We can include the explanations above in the revised Table 2 caption.

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.”

We confirm that we only consider the standard deviation of N age estimates near each site to estimate the layer uncertainty from the spatial displacement of the ASIRAS and NP measurements. We agree that any uncertainty in the TWT to depth conversion translates to an additional uncertainty in layer age. However, in contrast to the considered uncertainty for the SMB estimates (Eq. 5), we are not using the fitted regional density profile from all iSTAR sites, but we use the local iSTAR density profile for the layer dating at each site (Line 140). Thus, we do not have to consider the spatial uncertainty of the composed regional density profile and expect that the error in the TWT to depth conversion is dominated by the digitization error, which we expect to have a minor impact on the resulting uncertainty.

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 can include the suggested clarifications for the caption of Fig. 4 in the revised manuscript.

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.”

We have to test the suggested correction first and may consider its impact on total mass input in the revision. Depending on the impact of MAR results on the discrepancy above, we may also consider its impact on total mass input (see response to MC4, Referee #2).

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 y / 10.1 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 y / 25 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 basinwide 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% temporal uncertainty 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 Line 5: add a comma after "scheme" Line 6: replace "the same main" with "similar" Line 17: remove "the" in front of "upwelling" Line 18: consider replacing "stimulating" with "initiating" Line 28: replace "in particularly" with "in particular" 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. 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"
Note, there are several minor fixes needed beyond section 1, which will require further refinement by the authors.
Figure 3: add a righthand y-axis with depth equivalent.

We appreciate the suggested minor edits and consider them for the revision.

Arcone, S., Spikes, V., Hamilton, G., & Mayewski, P. (2004). Stratigraphic continuity in 400MHz short-pulse radar profiles of firn in West Antarctica. Annals of Glaciology, 39, 195-200. doi:10.3189/172756404781813925

Arcone, S., Spikes, V., & Hamilton, G. (2005). Phase structure of radar stratigraphic horizons within Antarctic firn. Annals of Glaciology, 41, 10-16. doi:10.3189/172756405781813267

Konrad, H., Hogg, A., Mulvaney, R., Arthern, R., Tuckwell, R., Medley, B., & Shepherd, A. (2019). Observations of surface mass balance on Pine Island Glacier, West Antarctica, and the effect of strain history in fast-flowing sections. Journal of Glaciology, 65(252), 595-604. doi:10.1017/jog.2019.36

Medley, B. et al. (2013), Airborne-radar and ice-core observations of annual snow accumulation over Thwaites Glacier, West Antarctica confirm the spatiotemporal variability of global and regional atmospheric models, Geophys. Res. Lett., 40, 3649– 3654, doi:10.1002/grl.50706.