Response to J. Brown's review

General comments: In this manuscript the authors estimate the mass balance of Pine Island and Thwaites glaciers by comparing surface accumulation estimates with estimates of ice discharge. Accumulation estimates for the glaciers is derived from airborne radar surveys tied to a dated firn core from Pine Island Glacier (PIG). The method of estimating surface accumulation employed by the authors is widely accepted as an accurate proxy for large scale accumulation variation. Ice discharge estimates provided in the manuscript are based on radar estimates of ice depth, satellite estimates of surface elevation, and surface velocity estimates from InSAR and speckle tracking. Regional average accumulation rate maps are derived from interpolation of inline data. Elevation-dependent accumulation rate derived from this study is compared to relevant climatologies and reanalysis and climate models. A comparison of the gridded accumulation product and the widely-used RACMO2 accumulation map for the region is also included in the manuscript. The interpolated average regional accumulation maps are used in conjunction with the catchment scale ice discharge estimates to constrain time varying changes in the regional mass balance for five catchments in the Thwaites/PIG area. Errors for each measurement are assessed as are errors for the raw and interpolated data products. Also, the methods used in this study are explained thoroughly and precisely, I compliment the authors on their work.

Overall, this is a well written manuscript with scientifically interesting results that show significant deviation from the model based estimates of the large scale accumulation pattern in the region. The method of determining spatial variations in accumulation rate from airborne radar employed in this study is appropriate and fairly thorough. The upscaling of the accumulation estimates from the radar profiles to a gridded accumulation product for the region is carefully and thoughtfully achieved. The limitations of the methods as well as the product are discussed which adds validity to the final estimates of the catchment mass balance.

The only major concern that I have relates to what I find to be a relatively small error assigned to the accumulation estimates derived from the airborne radar data. The source of my concern is the constant density profile used to determine the traveltime/depth relationship of picked internal horizons as well as the total integrated mass above the horizon. I think that the justification of the single density profile is valid, as is the justification of the error assigned to the accumulation estimates, however, I do not think that the 1 standard deviation error estimate attributed to the density profile and propagated through the twt/depth and cumulative mass/depth calculations (figure 3) is conservative, as the authors claim. As I argue in my comment on Section 3.4 below, just by taking into account the variation in mean annual temperature over the region the calculated error in the density profile is greater than the 1 standard deviation estimate given by the authors, potential errors from initial surface density and annual accumulation rate are greater still. I would like to see a more conservative error estimate given to the radar derived accumulation estimates, especially in light of the large region surveyed in this study.

That being said, I think that this is a quality paper and I would recommend publication after addressing the following comments:

The authors would like to express sincere appreciation of the thorough and constructive comments by J. Brown and his positive support of this work. We respond to each of his comments below in bold.

Section 1: Page 956 Line 19-21: Here the authors correctly state that "in-situ measurements are inadequate for mass balance studies because recovery over inaccessible regions, such as highly crevassed areas, is not possible". This implies that airborne radar surveys are appropriate for highly crevassed areas. However, these regions will also have large disruptions in internal reflection horizons (i.e. vertical discontinuities in layer TWT, highly dipping layers over snow bridges, reflection hyperbolas from crevasse walls, etc.; in general these regions are a total mess in ground based radar measurements). Please explain how a single horizon can confidently be tracked through these regions, especially as it pertains to an along track spatial sampling rate of 500 m.

The reviewer makes a very fair point: it appears that we are suggesting that accumulation rates can easily be derived over highly crevassed areas using the radar data. As the reviewer mentions, tracking radar horizons over highly crevassed regions is quite difficult and perhaps impossible. We removed the reference to "inaccessible regions, such as highly crevassed areas," and instead mentioned the limitations of field measurements in terms of scale: "…recovery at the catchment-scale is not possible."

Section 3.1: Page 960 line 7: Please be specific as to which cores were used to determine the ages of the mapped internal horizons.

On page 960, we mention in the sentence following line 7: "All horizon tracking began at the PIG2010 site where the horizons were dated." We added to that sentence: "…using the PIG2010 depth-age scale." to make it clearer to the readers.

Section 3.2: Core sites were selected based on the radar data (Page 960 lines 11-12). Please specify what criteria were used in site selection for the cores. Presumably, the criteria for the site selection affects the range of variability observed in the core density profiles.

We added several sentences concerning our reasoning behind each core site, which I include here. "Specifically, the PIG2010 site was selected because of its location at the convergence of several flight surveys (see Figure 1). The DIV2010 was selected based on its proximity to the coast and the presence of a sequence of relatively flat radar horizons. The THW2010 core was selected on the westernmost flight path to ensure we could date the radar horizons in case we were unable to track horizons continuously from the PIG2010 site." As the reviewer states, these criteria could impact the variability in the observed density profiles. Climate (accumulation and temperature) was not a factor in our selection, thus we show no potential bias towards any density regimes. While PIG2010 and DIV2010 had similar accumulation rates, DIV2010 had a much larger surface density and is located at a much lower elevation and experienced much warmer temperatures than PIG2010. THW2010 receives less (< 3/4ths) accumulation than PIG2010 and DIV2010 and is the highest elevation site (and the coldest).

Also, the cores were obtained the year after the radar survey (Page 960 line 12-13) and the vertical resolution of the radar is too coarse to image annual stratigraphy (Page 959, line 4-5). How were the horizon depths corrected to adjust for the accumulation during that year?

Here, we did not adjust the horizon depths to accommodate for the additional accumulation but instead adjusted the depth-age profiles and removed the new layer of snow that accumulated since the radar data collection. This process was done simplistically (essentially zeroing the depth-age scale at the beginning of 2010). We realize that this does not account for the enhanced densification experienced over 2010-2011 and will result in depths biased low in our depth-age scale. Using the Herron and Langway (1980) model to estimate the amount of total column compaction experienced between the surface and a horizon of 25 years in age (approximately the age of the horizon used in our study). We find that at an accumulation rate of 0.4 m w.e. yr^{-1} (approximate accumulation at the PIG2010 site where the layers were dated), approximately 56% of total column compaction occurs (assuming steady state densification rates). Estimating the total column compaction using the formula $b*(1/\rho_0 - 1/\rho_i)$, where b is the accumulation rate in water equivalence, and ρ_0 and ρ_i are the initial and ice density, respectively, yield total column compaction for one year of 0.56 m. This means that over a one year interval, the firn between the surface and a horizon of 25 years age compacted 0.31 m, indicating our depth for 25 year old firn is ~0.3 m too small and that our derived age will be too old. This magnitude of compaction is much less than the vertical range resolution of the radar system (50 cm in ice and 62 cm in firn).

Therefore, we do not believe this bias is substantial because the range resolution of the radar is much greater than this bias. We include a digitization error of ± 1 range bin and assign an error in the age of ± 1 year, values when combined are much greater than the impact of a depth error of ~0.3 m. Additionally, the fact that the date match well between the cores supports the fact that our method likely did not bias our results. Due to the different climate regimes, the total compaction that occurred between the surface and 25 year old firn varies. Even with the different biases at the different core sites, the ages are relatively well-constrained. Thus, we feel the simplistic method of accounting for the additional year of accumulation in the depth-age scale justified.

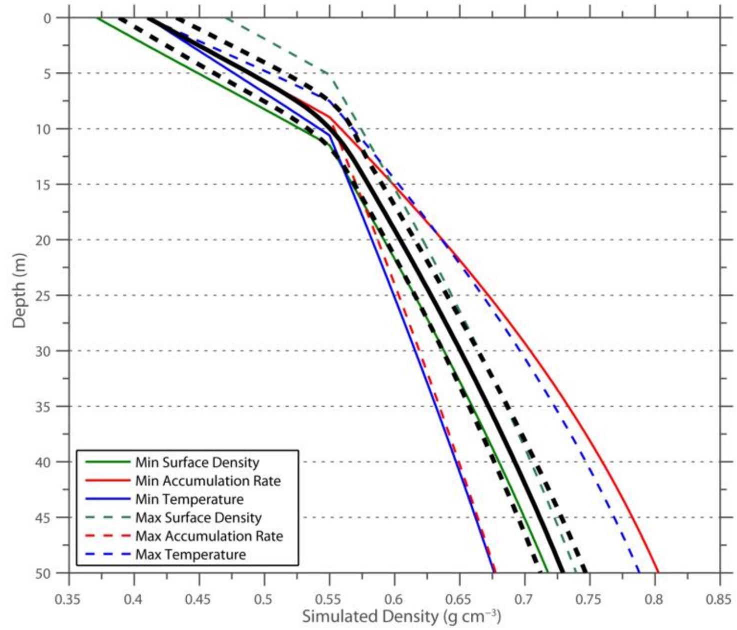
Section 3.3: The sentence of this section states: "Spatial variation in the depth to a given horizon is a consequence of variations in the accumulation rate". The structure of this sentence indicates that the only factor in variations in layer depth is accumulation, which is not true. There are other factors that influence the depth to a given horizon which include lateral variations in vertical strain rate and lateral variations in firn density. Please change this statement to better reflect the complexity of the problem.

The reviewer makes an excellent point here. We have revised the sentence as follows: "At shallow depths, the spatial variation in the depth to a given horizon is largely driven by variations in the accumulation rate and, to a lesser extent, the density profile and firn compaction rate, which are also dependent on that accumulation rate."

Section 3.4: The Herron and Langway model is fit to the retrieved density profiles (Page 959 Line 20-24). This steady-state model has three fitting parameters: 10 m temperature, accumulation rate, and initial (surface layer) density. The fit to this model is then used to calculate the depth to the internal reflection horizon and the mass of the overlying firn. The error analysis described in this section is appropriate and very reasonable for the data presented in this paper however, I take exception to the statement that the error estimate is conservative. Calculated depths for layer H1 range from 4.3 m to 36.9 m (Page 960 line 1), calculated accumulation measurements range from 0.13 to 1.37 m w.e. yr-1, and average surface temperature (equivalent to the 10 m temperature in the H&L model) likely ranges by at least 15 degrees C (as indicated by the ALBMAP gridded data for the region) over the ∼1700 m elevation range sampled by the airborne survey; any one of these large variations will cause accumulation estimates based on the H&L curve used in the study to vary by more than the estimated 1 std. used as error bounds in the paper. I have uploaded a figure which shows the variation in density profile calculated with the H&L model for the ranges of the accumulation quoted in the paper (red dashed curves), measured surface density variations (350-450 kg m-3) taken from figure 3 in the paper (blue dashed curves), and a 10 degree mean annual surface temperature variation (black dashed curves); the solid black curve is calculated with the values for accumulation and surface density given in the paper and a mean annual temperature of -23 degrees C. The point here is that the range of values given in the paper do not support that the error estimate given in the paper is conservative, it is more likely a very non-conservative estimate of the errors inherent in the method.

We appreciate the detail provided by the reviewer concerning our statement that the errors presented are likely a "conservative" estimate because we use the mean and standard deviation of density profiles from nine firn cores from the region. His point that the density ranges found in the ice cores likely do not encompass the variations in temperature, accumulation, and surface density over these basins is valid, but there are additional things to consider. The reviewer's plot shows how if you vary these three variables over a range of values likely observed over the catchments, you can generate large differences in the density profile. This analysis, however, assumes that these three variables are independent of one another, which is not true in this area. In fact, all three are highly correlated with the others spatially.

Using RACMO2 1979-2010 average surface mass balance, temperature, and surface density from cells with centers within 13.5 km of our radar measurements (> 200 cells), we find the correlation coefficients for accumulation-temperature, accumulation-surface density, and temperature-surface density of 0.78, 0.93, and 0.92. Now the question becomes: how do these positive relationships translate into impacts on the density profile? Here, we show a very similar plot to the reviewer's (see figure below), but instead show different line types to represent an increase (dashed) or a decrease (solid) in the surface density (green), accumulation (blue), and temperature (red). Here, we used the minimum/maximum values for each variable from the RACMO2 subset described above. Most interesting is the comparison of the density profile variations from accumulation and temperature. An increase in accumulation results in less dense firn at a given depth (below 550 kg m-3), while an increase in temperature results in denser firn at a given depth. These almost entirely offset the other in this region. If temperature and accumulation were inversely related spatially (higher accumulation, lower temperatures), the climate variations in this region would generate huge variations in the density profile. Because they are in fact directly related, their impacts essentially offset each other.



We calculated the density profiles using the RACMO2 climate data from each cell in the subset described above and took the mean and standard deviation of the values and plotted them in black. (see figure above). This analysis allows us to look at realistic climate combinations for the region. As you can see the actual spread is much less than when investigating any of the individual components because they are in fact not independent of one another. Therefore, we believe that our error estimates are valid because the variables controlling the firn densification are actually positively correlated spatially.

Section 3.6: The first sentence of this section indicates that the along track measurement interval is small even though the quoted interval is 500 m. This interval is small in relation to the size of the basin but it is very large compared to the Nyquist spatial sampling intervals for the radar frequencies used in the survey, which are much less than 1 m. I recognize the need for such large sampling intervals in airborne data acquisition, but 500 m sampling interval for data with a vertical resolution of 50-62 cm is not considered small in a geophysical sense. Please revise the statement.

The reviewers comment is well-taken, and we revised the sentence as follows: "While the along-track measurement interval (500 m) is small relative to the catchment size, there are large data gaps (up to 150 km between flight paths; Fig. 6a)"

Section 3.7: Equation 1: Is the assumption that all of the ice flow is due to sliding at the bed valid for the entire catchment? Would the final mass balance estimate change drastically if large portions of the basin were frozen to the bed? Please justify the pure sliding assumption and estimate the error added to the full mass balance of the catchment due to frozen-on portions of the bed.

Since we are measuring flux across the grounding line, the degree of shearing is only applicable at the gate through which we measured the flux (i.e., what happens in the rest of the basin is

irrelevant). We computed the ratio of column averaged to surface speed near the Thwaites grounding line and obtained typical values of greater than 0.99 (often above 0.995), indicating a bias of less than 1%. We have addressed this in the text by adding the statement: "Our velocity measurements are made at the ice sheet surface but we assume they are equivalent to column average velocity due to the high-degree of sliding at near the grounding line. Based on analysis of estimated deformation velocities that are internal variables of a temperature model (Joughin et al., 2009), any biases introduced by this assumption are less than 1% and are not included in our error analysis."

Section 5: Page 975 line 18-20: This sentence needs a date range for the accumulation range, I assume that it is 1985-2009 but this should be stated explicitly here.

We added "for the 1985 to 2009 interval" to the end of the sentence.

Table 4: The quoted value for the total mass balance for 2010 does not work out mathematically (257.4-158.5 does not equal 96.1). This erroneous value is also quoted in the text (page 976 line 24). Please double check the rest of the table values for similar errors and change any erroneous values in the text.

Thanks for the correction. We have checked all the values presented in the tables and have fixed the values quoted in the text.

Figure 1: Many of the flight lines in this figure indicate that none of the three internal horizons are traced for portions of the flight line, yet they are tracked further along the profile. How can the H1 horizon, which varies with depth and absolute reflectivity, be positively identified over large data gaps and thus be considered isochronal on each side of a ~50 km gap in horizon continuity? Please expand on how H1 was identified over gaps where neither H2 or H3 were tracked, specifically at the core site THW2010 which is not continuously connected to the PIG2010 core, yet the H1 horizon is used to compare ages in the two cores.

The largest gaps occur in the snow radar measurements from Medley et al. 2013. These measurements were derived by counting annual horizons in the radar data. Thus, gaps are possible because this method does not require continuous tracking of the horizons and local depth-age scales can be derived where a clear sequence of horizons exists. Looking at the accumulation radar measurements, few gaps of up to 20 km exist. These jumps were made only where the radar horizon sequence (or fingerprint) was clearly matched on both sides of the gap. To reach the THW2010 site with the H1 horizon, we tracked H1 beginning at the PIG2010 site outward from the several flight paths that initiate there. Many of these flight paths intersect the flight path covering the THW2010 site. We again matched the horizon sequences at these intersections to continue mapping H1 in the THW2010 survey. While H1 was lost just before the THW2010 site, we were able to match the horizon sequences to begin tracking it again and reach the actual core site. The THW2010 depth-age scale was not used to determine the H1. Therefore, dating H1 with the THW2010 depth-age scale is an independent test of the age from PIG2010.

Figure 4: The solid red line that indicates H1 completely obscures the horizon in the data. Please change this line to a dashed line so readers can assess the continuity of the layer better. Also, please indicate where the H2 and H3 horizons are in these figures.

We have adjusted the line type on Figure 4 to a dashed line to help the reader see the continuity in the horizons. We also added H2 and H3 to the PIG2010 echogram. H2 and H3 were not tracked to the other cores since H1 was easily tracked to them.

Figure 6: Two of the ITASE core locations are absent from this figure (01-6 and 01-3), please add them.

We went through all the maps and ensured that all the cores were displayed properly.

Again, I compliment the authors on a thorough and well written paper.

We would like to thank J. Brown again for his insightful, interesting, and constructive comments on our paper

References

Herron M.M. and Langway, C. C.: Firn densification: an empirical model, J. Glaciol., 25, 373-385, 1980.

Joughin, I., Tulaczyk, S., Bamber, J. L., Blankenship, D., Holt, J. W., Scambos, T., & Vaughan, D. G. (2009). Basal conditions for Pine Island and Thwaites Glaciers, West Antarctica, determined using satellite and airborne data. *J. Glaciol.*, *55*(190), 245-257.