Response to Anonymous Referee#1

First we would like to thank Referee#1 for taking the time to review our paper. We appreciated your comments that have helped to clarify and improve our paper. We will address your comments in the order of the review below

First to address the general comments:

1) Validation of the assumption of July 1 Layer age as melt can continue into August.
   Yes we realize that melt can continue into August; however the majority of the warm temperatures are in late june and early july. The radar will cause the largest reflection where the density change is largest and hence we chose July 1 as this date. Early and late season melt event could cause a thin layer to form but it would not be the dominate peak in the radar return which would be caused by the larger summer-time densification. This same argument holds for hoar layers in the interior. Again we add a +/- one month error on this data to show the uncertainly as stated in section 4.2.

2) Conveying more detail on the MAR density model.- We have included the basic equation of the density model now in Section 3.2 for more clarity. Additionally we understand the reviewer was confused by how we were conducting our density comparison as we did leave out a very important sentence clarifying that our modeled and measured density profile were compared simultaneously in time. In Section 4.1 this sentence was added, “The comparison of measured and modeled density was simultaneous in time, meaning that the MAR density profile output on the day of the measurement was compared to the measurement.”

3) More appropriate cross over analysis comparing range bins-

   This has been changed to include both range bins and m w.e.

4) An improvement to the uncertainty analysis and description.
   We have added some clarification to Section 4.2 and below are calculations our calculations for the reviewer. First we have both correlated and uncorrelated errors as the density error is correlated. We take the now equation 3 and take the derivative as follows.

$$b = \frac{A\rho}{(\rho D+1)^{1.5}} \quad \text{where} \quad A = \frac{TWT(x)\cdot c}{2a(x)\rho_w}$$

$$\frac{\partial b}{\partial \rho} = \frac{A}{(\rho D+1)^{1.5}} + \frac{-1.5A\rho}{(\rho D+1)^{2.5}}$$

Taylor Series: to account for error, $\Delta\rho$,

$$b_{error} \approx \frac{A\rho}{(\rho D+1)^{1.5}} + \underbrace{\Delta\rho \cdot \left[\frac{A}{(\rho D+1)^{1.5}} + \frac{-1.5A\rho}{(\rho D+1)^{2.5}}\right]}_{\text{error term } \Delta_b}$$

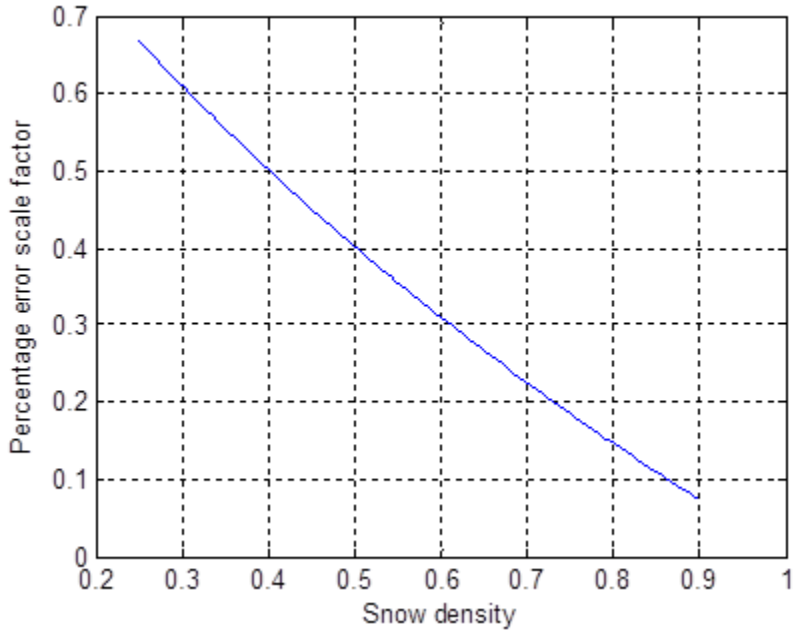
$$\frac{\Delta b}{b} = \frac{\Delta\rho}{\rho} + \frac{-1.5\Delta\rho}{(\rho D+1)}$$

$$= \frac{\Delta\rho}{\rho}\left[1 + \frac{-1.5\rho}{\rho D+1}\right] \text{//}$$

where $D = \frac{\epsilon_i^{1/3}-1}{\rho_{ice}}$. This equation relates

% error in $\Delta\rho$ $\left(\text{i.e. } \frac{\Delta\rho}{\rho}\right)$ with % error in b.

So the % error in accumulation has a scale factor that depends on density. The scale factors dependence on density is as follows.



D = (3.15^(1/3) - 1)/0.917;
rho = linspace(0.25,1,101);
plot(rho,(1 - 1.5*rho./(rho*D+1)))

We choose the highest Percent error scale for our density measurements that rarely go below 0.3 giving a percentage error scale of 0.6.
Using sum of squares on uncorrelated density (12%* scale factor of 0.6) error to age (8%) error We get sqrt((12*0.6)^2 + 8^2) = 10.76 which we round to an error of 11%. If we assume the maximum age error of 10% as suggested by the reviewer. We get sqrt((12*0.6)^2 + 10^2) = 12.32 or 12%. We have changed the error to the higher error of 12% to stratify the reviewer and added clarifying statements in Section 4.2.

Specific Comments	**Response**
Justify comparing radar-derived and in situ measurements that are within 5 km of each other	Yes we realize accumulation can change on small scales, as shown by the ice cores in figure 12. Determining correlation length scales would vary considerable depending on the ice sheet region. While this could be done with our dataset it would be a very detailed study and beyond the scope of this paper. We choose 5 km as a scale that provided a few (2)

<table>
<tr><td></td><td>locations where we have both radar-derived and in situ measurements in relatively close proximity and in time. Choosing a smaller number like (1km) would only allow for overlap at 1 location. This is similar to other studies where ice cores are extrapolated over space to validate model e.g. Colgan et al., 2015. We realize it would be best to have ice cores directly under all IceBridge flight lines simultaneous in time but in reality few exist, hence, we set the distance at 5 km for this study.</td></tr>
<tr><td>Density comparison- Model evaluation</td><td>We have added additional equations and clarifications in sections 3.2 and 4.1. Again we are comparing SUMup Measurements on the date they were taken with the same profile date in MAR. If only a month was given we use the 1st of the month for comparison. We also state clearly in this paper in Section 4.1 that “We consider it beyond the scope of this study to investigate and explain why MAR underestimates near-surface density, therefore, here we assume that the firn density in the top 1 m is 0.338 g cm^{-3}. “ The reviewer is correct that much more needs to be done in understanding why the density model is not producing similar results to measurements in the top 1 m and Co Author Alexander is working on this exact problem for his post doctoral project and will be publishing more detailed results shortly.</td></tr>
<tr><td>Radar collection date to MAR density</td><td>In Section 4.1 we now clarify this with “the spatially-varying modelled density profiles are used for April 30” We are use density profiles from April 30 to calculate accumulation form the radar data which is approximately the mid-point for IceBridge flights.
We also note that in the MAR model during the spring time frame our choice of date would have little impact as shown in the table below for the different dates compared to the observed values from the PARCA cores.
<table>
<tr><td></td><td>Observed</td><td>MAR (Apr 15)</td><td>MAR (May 1)</td><td>MAR (May 15)</td><td>MAR (June 1)</td></tr>
<tr><td>0-1 m</td><td>338 ± 39</td><td>282 ± 40</td><td>280 ± 40</td><td>275 ± 45</td><td>277 ±52</td></tr>
<tr><td>1 – 2.5 m</td><td>381 ± 54</td><td>385±149</td><td>387 ± 149</td><td>386 ± 148</td><td>390 ± 148</td></tr>
</table>
</td></tr>
</table>

Constant Density Assumption	Additional discussion is added in section 6, however, we note that the SUMup compilation of field measurements does not support the reviewer claim that surface densities should very by up to 30%. It is very rare to have surface measurement below 300 kg/m3 for Greenland. SUMup measurements, the largest compilation of publically available measurements that we are aware of, which are well distributed spatially on the GrIS (Figure 1) show a spatial variability of ~20% (12% std) spatially. In the paper we clearly state the assumption made and cannot address the spatial bias until models and measurements are in better agreement.
Accumulation rates and uncertainties : Age of first layer	This is defined in Section 5.1 the second paragraph and we added "We simultaneously compare the time represent by the layer to MAR estimates of accumulation." For clarification
Error Estimate	This is addressed in the opening comments.
Picking procedures. Smoothing.	We have added clarification to this section and Figure 3 is included. Changed smoothing to spline fitted for clarity. The data is not smoothed.
Results: Time frame	We went through results to make sure it was clear what time frame was represented as well as added time ranges to figure captions etc as suggested.
Annual Variations	Snow radar has previous been shown to detect annual layers (Medley et al., 2013 published by this journal) The layers here are annual as variation and not monthly variations as suggested due to the magnitude of the change. Shown in Figure
First layer	We chose to keep the analysis of the first layer. We provide the uncertainty estimates and the first layer is the most extensive across the ice sheet. Again we are comparing the 10 months represented by this layer to 10 months of modeled data so the comparison is valid.
Crossover Analysis	Included Range bins and clarifications as suggested. We do not do cross over analysis of deeper layers and there are not many locations to perform this analysis as Shown in Figure 6.
Comparison with model	Language has been toned down as suggested. We note again that the measured densities show less of a regional bias than the modeled densities so we would expect that using the average value decreases spatial bias over modeled values. We have clarified dates throughout as suggested.
Comparison with in situ data	We chose not to include the echograms we are using the pick at the closest radar trace for this analysis.
Discussion	We have added some to the Discussion but do not extrapolate to Greenland mass balance as that is future work. This paper is as the review suggests and introduction to this dataset and the description of how it was created with a preliminary comparison to MAR. Future work will

	expand its use.
Technical Corrections	
P: 6699 L20: remove “of ice” as it is implied L23: remove “being governed by” and “being dominated by” as it is redundant and awkward	Changed.
P: 6700 L3: “here after” should be “hereafter” L6: add “in number” after “limited” to clarify L11: comma after “(Benson, 1962)” L27: replace “and map” with “the lateral persistence of”	Changed.
P: 6701 L8: use of “to” after “penetrate” is redundant; consider removing “to” or rephrasing L10: comma after “frequency-modulated” L11: remove the comma after “radars” L25: comma after “preserved” and remove the commas around “, therefore,” or consider a semicolon after “preserved” and remove “and”	Changed
P: 6702 L16: comma after “Frequency-Modulated” ; also, I am not sure why “Frequency-Modulated Continuous Wave” is capitalized here and not on P6701, L10-11, so please be consistent. I suggest not capitalizing it. L17: change “when preserved” to “where preserved” Section 3.1: somewhere in this section there should be a description of the differences in the radar system for the different years, including its range precision.	Change, made consistent, range resolution is given in Section 3.1. We chose not to describe the radar changes here as they are given in the citations and not relevant to the work done in this paper. Additionally the radar changes are minor over this time period.
P: 6703 L3-4: remove “reanalysis” add “global atmospheric reanalysis” after “ERAInterim” L12: change “accumulation-rate” to “accumulation rates” as the former suggests you are using accumulation rate profiles from MAR, which seems awkwardly	Changed. To a depth of 15 m is clarified with the parenthetical information given in the paper (the depth to which MAR predicts firn densities). We clarified the sentence on the number of measurements and information and Figure 1 shows the number of locations. Changed sentence to: “which contains over 1500 measurements from snow pits and ice cores at 62 sites. At each site the number of measurements ranges in number between 8 and 170 and

phrased L20: Why only to 15 m? Is it due to the fact that no layers below 15 m are used? If so, please state it. L20: "1500 measurements" is misleading and really does not inform the reader of the value of the data set for comparison. I would prefer listing the number of sites, with a description of the range of measurements at each site. Something along the lines of "which contains measurements at ## sites, and at each site the number of measurements ranges in number between XX and XX and maximum depths of XX and XX." L23: change "measured" to "in situ" L26: change "additionally" to "additional" L27: The phrase "which includes additional cores to the SUMup dataset" is redundant because it was already made clear by the "additional" in the prior line.	maximum depths of 1 m to 15 m."
P6704 L5: The second half of the sentence is oddly phrased, please reword beginning at "we require: : :" L18-19: The sentence beginning with "We note: : :" needs to be appropriately cited as this is not common knowledge.	Change to "Because we seek to derive accumulation rates from near-surface radars across large portions of the ice sheet, we require firn density profiles that cover the GrIS." For clarity. Change to "Uncertainty in the top meter is assigned by the ±1σ variation in observed density (12%) which we assume is due to the natural variability in surface density."
P6705 L1-2: in the sources of error for derivation of radar depth, why is the actual density profile used not included in the list? The error from uncertainty in density is likely larger than based on the dielectric model used. L2-6: The description of the dielectric model evaluation is confusing, please clarify. Perhaps, begin with a statement explaining that you are evaluating X, Y, Z dielectric models because that	We choose not to add additional information on the dielectric models as we feel it is clear the models that were used from the references and the main point that there can be up to a 3% error is stated clearly. WE have rewriting previous eq 1 into Equations 1 through 3 for clarity please see section 4.2 as it has been undergone many changes for clarity. Added average.

only became apparent at the end. Eq1: Why is a dependent on x? The age of a layer should not be dependent on location as the layers are assumed isochronous. The equation might need further clarification because variables should be dependent on x, but also on depth (or on the layer number). I suggest stating the equation is for a given horizon to eliminate the additional complexity. L13: The phrase ": : : is cumulated snow/firn density at depth: : :" is confusing. I suggest adding "average" after "cumulated" because otherwise it sounds as if the densities are just added together. L16: The same issue arises here as with the previous comment. The use of "cumulative" suggests adding together all the densities below that depth, which in an integrative sense would produce a cumulative mass (kg m-2). Perhaps, reword or add "average" again.	
P6707 L2-3: If vertical traces are tossed out if it appears the surface is not properly picked, how is the stacking procedure done? If a few traces in a row are tossed out, you would not want to average the now spatially separated traces. L7-8: Why not stack a different number of traces to end up with similar along-track spacing for all years? L13: change "in" to "from" L24-26: please rephrase the sentence beginning with "Layer indices are: : :" because I find it difficult to understand what is meant by the "partial overlap that can exist between layers." A graphic of the procedure is really necessary.	The stacking procedure is described a few lines down with "The radar data are then horizontally averaged (stacked) 10 times to an along-track spacing of ~50 m, in 2011 and 2012, and ~10 m, in 2009 and 2010" Yes vertical traces are removed because the surface is not always pick correctly. This common with radar data and is generally due to 2 differenct causes 1) there is not a strong return form the surface or 2) the planes altitude adjusted quicker than the radar setting and the radar data switches Nyquist Zones. In either case we do not include the data in our dataset. We keep the same number of stacks to keep the same processing scheme and averaging of the radar data for signal to noise consistency. Changed in to from. The Layer indices sentence was rewritten from clarity and figure 3 is cited for a graphic representation.

P6708 L14: Insert “the” before “accumulation rate”	Changed.
P6709 L3-9: Consider moving to the picking section as it seems more appropriate. L16-18: It is not clear which cluster in the crossover analysis show rates off by a factor of two, so perhaps circling it on Figure 8 would make it easier.	We prefer to keep the section on layer numbers detected in result of our procedure. We chose not to add a circle as there are few locations where the factor of 2 is apparent for instance at 0.25 and 0.5. We have changed this figures as suggested later in this reviewers comments so hopefully that will make it clearer. Additionally as shown by the scatter plots these possible errors are not extensive so there are not many in the scatterplot. Also shown in the statistics of Table 1.
P6709 L24-26: Consider applying a threshold number of radar measurements for comparison with the MAR grid cell to eliminate comparisons that are likely not as representative.	We keep the comparison as is and not that in all of the grid boxes we have multiple radar-derived measurements. Previous comparisons with ice cores set a precedent that one measurement per grid cell is sufficient. Ie Burgess et al., 2010; Colgan et al., 2015.
P6710 L7-9: The larger differences are associated with areas of higher accumulation. A more informative comparison would be as a percentage. Otherwise, the details in the low accumulation areas are lost. L17-20: The strong statement of “These values are not well correlated: : :emphasizing that further improvements in accumulation-rate modeling are needed: : :” should be reworded because the measurements are not without fault, so putting the blame on the model is risky. L27: consider changing “closely” located” to “nearly co-located”	We did not change to percentage difference as we feel the accumulation value is more important for SMB studies. We have change the figures to be clearer as suggested. Reworded to “emphasizing that further improvements in accumulation-rate modeling and measurements are needed, particularly over the southeast and northwest GrIS.” Changed.
P6711 L20: consider removing “the” before “large portions” L20-23: Again, this is a very strong statement. It should be changed to state that while these are useful for model evaluation, we must still consider the assumptions that go into the radar-derived measurements. Such a statement would give way for a discussion of the new data needed to reduce those uncertainties.	Removed. We left sentence as “The pattern of radar-derived accumulation rates compares well with known large-scale patterns and clearly shows that these accumulation-rate measurements are useful for evaluating model estimates.” As the radar estimates do compare well with large scale patterns and are useful for evaluating model estimates. We address the uncertainties in the radar-derived measurements throughout the paper and again note that our assessment of error is very similar to error assigned by Medley et al. (2013) averaged out to less than 5% (10% and 15% also given) and Das et al. (2015) between 6% and 17% in total SMB.

P6712 L9: consider changing "resolves" to "will resolve" L13: the phrase "constantly varying flightlines" is unclear as to what is varying, please reword	Changed. Changed to " Spatial extrapolation between the flightlines, which vary in position from year-to-year, will be left for future work, as additional data are collected and made available to fill in gaps."
Table 1 Please state in the caption what time interval is used from MAR (July1-April30 or July1-May31). Consider adding a column of the mean accumulation from the crossover points for each year.	Added date clarification. Adding the mean accumulation from the cross over points is likely not a useful number as it is spatially dependent and the crossover are not consistent in space from year to year. We did not add.
General figure comments Please change the color intervals used in Figures 4 & 5 to be more meaningful: e.g., 0.2-0.3, 0.6-0.7. The values are non-traditional, making it difficult to quickly interpret the patterns. The black background does not add to the meaning, and is a little ink heavy.	The color bar and numbers are held consistent with that of Burgess et al., 2010 and were not changed. We also choose to keep the Blue Marble as the background image.
Figure 1 Is there overlap between the density measurements (red) and ice core accumulation measurement in blue?	Yes in some locations there are. Added Echogram locations to map.
Figure 2 Please change depths to positive numbers since a depth is positive moving downward. The caption should be very descriptive as to what the differences existing in the timing of the measurements and what model timing is used. This relates to the statements in the beginning on explaining the details of the comparison. For instance, if the average April 30 density profile from MAR is used, please state it. Please do something similar for the measurements as well.	Added 1 "and the measurements and modeled profiles are contemporaneous." For clarity. Depth changed to positive numbers.
Figure 3 Please change the Distance values along the x-axis to more appropriate intervals (26, 78, etc. are odd values). An inset map of these transects would be beneficial. They could even be added to Figure	We left the distance values as is and feel they are clearly labeled. The locations of the radagrams were added to figure 1.

1.	
Figure 4 Please state that only the accumulation rates from the top layer is plotted for each year in the caption.	Added "representing the top layer in each year (July 1 to April 30). "
Figure 5 Same as with Figure 4, state the time intervals represented here (May1 – April30?). Consider overlaying the radar-derived measurements for comparison	Added" (representing July 1 to April 30 to match the radar-derived estimates)."
Figure 6 The intervals in the legend should be changed to not have overlap: 1, 2-3, 4-6, etc.	Changed.
Figure 7 These values should be plotted as percentages rather than absolute values because the crossovers in regions of low accumulation are lost. Also, as described above, the crossover analysis as done here is only a measure of the ability of the picker, so the maps shown here would be better off showing the differences in range bin picks, not in total accumulation. Please be sure to use appropriate intervals for the color bar, if the mean crossover difference was 0.03 m w.e., then majority of them would fall into the first interval.	Percentages must assume that one pass is more valid than the other which we are not able to do. We left figure 7 in m w.e. and changed figure 8 to range bins so the reader is given all of the information. Also change in Table 1.
Figure 8 Similar to Figure 7, this plot should be comparing the picked range bin rather than accumulation rate.	Changed to range bins.
Figure 9 The color bar should be a gradient between two colors, reaching white in the middle in order to appropriately show regions where the model is less than or greater than the measurements. There are too many colors here, making interpretation difficult. Also, be careful with the value intervals making sure the center interval straddles	Changed.

zero evenly (e.g., -0.05-0.05). This way people can easily see the transition between more/less accumulation difference. A histogram of the differences would be a useful addition that can be inlaid onto each map.	
Figure 10 There are a few interesting features here that could be further discussed in the paper. For instance, the 2011 (blue) dots appear to have a linear feature at 0.75x and at 1.5x suggesting the picker detected the 2nd layer rather than the 1st. All the previous plots were broken down by year, it might be useful to do the same (4 plots) to see the details of each as the values <0.5 m w.e. get lost. I would suggest showing the best fit line to the data as well to ease interpretation.	We do not see in the data that there are mispicks of second layers, there a very few. The differences are likely due to discrepancies between the measurements and the models.
Figure 11 It would be useful to have the echograms from each year shown as well, so the reader can see the differences in the data between years. It would also lend insight into whether the very large accumulation from the radar in 1995 is due to the picker missing a layer, which is especially interesting because the 2011 data end in 1996.	We did not include the echograms as the data is taken at a single radar trace for this comparison and changed figure 11 as suggested by reviewer #2.