The author's should consider each of the reviewers' comments when preparing their revised manuscript. However, to help expedite final acceptance, I note the following key points made by the reviewer that should be addressed. I have also added a couple of my own comments on a some points that I think could be made more clear. These additional comments are not meant as reviewer comments, but are easily addressed and would help improve the paper.

Both referees commented on the limited spatial range of the comparison, and very limited comparisons that can be made (only one point for ERA-I, and a few for AMSR-E). I agree with the reviewers that more discussion/qualification of the results with respect to the very limited comparison and atypical conditions for sea ice needs to be included.

Title: Agree with reviewer #1, this is a specific region, and so the title should reflect that and not generalize.

Initial conditions – As pointed out, the fastening date is not necessarily the onset of snow accumulation. The method used here could lead to some underestimation if snow had already accumulated – can you estimate how much it might have influenced results (though, SnowModel is biased high)?

I agree with the reviewer here wrt lines 454-456. While SnowModel as set up may require 200m resolution, you are not comparing at that resolution except with the in situ data in figure 4. At least based on your results, ERA-I does arguably better for CS-2 ice thickness, as reviewer #1 states (error range is lower in Figure 6) . So the value demonstrated by SnowModel here appears to be in matching the spatial pattern of snow distribution, and as you discuss based on physical reasons one would expect SnowModel to be better. But in the manuscript at least, there isn't evidence that SnowModel improves CS-2 thickness estimates (see also comment below on Polar-WRF). You should be clear in your discussion what your results demonstrate, and what they do not.

Figure 2 – agree with reviewer #1, the dots are hard to see.

Accuracy of results –I agree that the qualitative comparison between SnowModel and Figure 4 could be more informative if made quantitatively.

**Additional comments:**

Line 412-414 –It is worth clarifying that the difference in Pd here is because of the differences in predicted snow depth, not something physical, so what Pd does in the case of an incorrect snow depth is to compensate for getting the snow depth wrong, and does not necessarily indicate what the penetration depth really is. For example, equation 3 can be set up for a case where you have a true snow depth, and an estimated snow depth with some error. Then if you try to fit to match the ice thickness, then you get a Pd that is the sum of the true Pd and a correction (Pde) that results from an error in your snow depth estimate (Tse):

Pde = (pw-ps)/pw*(Tse) = 0.625Tse.

So, if you have overestimated your snow depth, your apparent penetration depth is corresponding larger than the true one, and vice versa.

Pd=0.5m seems too high, given Figure 2 shows a mean snow depth of ~0.1m swe (i.e. ~0.3m actual). How can you have Pd=0.5m in this case? You do say you cut off Pd at the snow depth, but I don't see any evidence that 0.5m is correct for any of your products or in situ data. The correction above would imply you'd need to be off by 0.8 m in Ts, which seems implausible.

I think it is important you clarify what is going on here, and be clear that these Pd values you calculate are not necessarily indicative of what is actually happening with the radar reflection. Your conclusions do properly reflect this and rightly only give the value based on the in situ comparison.

One thing you did not point out is that SnowModel takes as its input precipitation from Polar-WRF, which will be different from ERA-I. So the comparison between ERA-I and SnowModel and in situ (at least for the CS-2 comparison) mostly just shows that the retrieval is sensitive to errors in snow depth, and not which method is necessarily better (SnowModel would presumably be better where, as you note, snow redistribution matters, but you have not shown that this is a factor here).

476-478 - Note that while shot separation for ICESat-2 is 0.7m, you won't get a sufficient number of photons to get a reliable elevation until you sample something like 100 shots. It is unlikely you will be able to resolve meter-scale features. Might be better here to say that you might want to resolve a statistical distribution of features to capture snow accumulation rates in the presence of blowing snow. (not essential, but you might pick a more recent reference for ICESat-2 here, e.g. Markus et al., 2017).

Also note that different retrackers pick different interface positions. This introduces an error in addition to Pd and snow depth estimation error that could be mentioned.

For figure 5, you match in part based on the slopes. But I believe the thought behind incomplete penetration into the snowpack is due to some physical scattering horizon, either an icy layer or perhaps wicked brine. Then could it be that Pd is at different depths at different times?

**Minor/technical points:**

Line 61 – ICESat-2 has successfully launched now, so this statement should be updated.

Section 2.3 – as pointed out by the reviewer, it isn't clear if you have calculated snow depth yourself or used an existing product. If the latter, the dataset used should be referenced.

Line 158 – should be "in coastal Antarctic" I think.

Line 180 – "but precisely how *it* is dependent"

Line 269 – should be "see Hines et al., (2015)"

Figure 4 – you might consider narrowing the scale here, your in situ measurements go up to ~15 cm, but your SnowModel scale goes to 180! It would be more clear if these scales were similar.