We thank anonymous referee #1 for his or her comments, and for taking the time to review our manuscript. We've noticed discrepancies in line numbers and figure references, however we believe that we've identified all of places the referee is referring to. We've copied the referee's comments in blue. Our responses are in red.

In this study, snow in the upper Amu Darya and Indus basins is modeled using multiple techniques and a newly available snow depth data record. These data and results provide insight into the likely range of snow depths and snowpack properties in a region where there was previously very limited understanding. The study is well researched and designed and the paper is well written. I think this is a good contribution to the snow research literature for HMA, and provides evaluation of some potential tools that could be used for avalanche prediction. I have some minor suggestions and comments for consideration.

Line 87: Instead of "snow on land melt" use "terrestrial snow melt"

We assume this is referring to l 87. The term used in cited Armstrong et al (2019) reference is "snow on land", so we'd like to keep this as is for consistency.

Line 143: Move the reference for APHRODITE to the earlier section (lines 110-111) where all the other global precipitation products are referenced.

Again, the line reference appears transposed. We assume this is referring to l 134. And yes, we will move the Yatagai et al reference up.

Table 1: It is not clear to me why Alpine3D and NOAH MP are run at 25km resolution. Why not run at a higher resolution for direct comparison?

This was driven by the GLDAS and NOAHMP model resolution. GLDAS is available at 0.25º or 27.7 km in the north-south direction. For NOAH MP, the forcings used were from MERRA-2 at 0.625º x 0.5º. 25 km the highest resolution we felt comfortable comparing.

Section 4.4 It's not clear how the AKAH stations are combined with SNOWPACK. Are the observations directly inserted and the model is used to estimate other snow properties (i.e. density, grain size, etc)? Ok, I see in the appendix that it is used as precipitation. I think that should be mentioned in the paper.

We assume you are referring to Section 5.4. In Table 1, under the SNOWPACK row, it says that the AKAH snow measurements were used as forcings along with downscaled forcings from ParBal. This is shown graphically in Figure 2. We will however, reiterate this in the text since it was not clear.

Line 266: should this date be "2018-4-1"?

On l 264 and elsewhere throughout the manuscript we were asked to use the international date format by the Editor, which is "1 April 2017" for that date.

Line 316: remove "are" after values, so it reads "median values between years are a week apart"

Ah, thanks for spotting. We will remove the typo.

Figure 3: I would be interested in seeing a spatial comparison of the individual stations, particularly in 2017. Are there certain stations where most of the disagreement occurs, or it is similarly biased at all stations?

We are not clear about potential biases in this figure. Perhaps you are referring to Figures 4 & 5, not Figure 3? In any case, we didn't find any spatial trends. Figure 5 summarizes the error distribution and shows that most stations are slightly negatively biased (blue box is 25^{th}-75^{th} percentile).

Line 355-356: you say, "The ParBal results are confounding given that the agreement between the modeled SWE from ParBal and SNOWPACK at individual AKAH stations is much better for both 2017 and 2018." The agreement for 2018 shown in Figure 3 is good, but in 2017 there is quite a large bias – similar to what is seen in Figure 5 for the whole study area.

The differences between the four models in 2017 at 25 km shown in Figure 6 were much greater than the differences between ParBal and SNOWPACK at the point scale, shown in Figure 4. In 2017, for the four models, NOAH MP showed 234% of ParBal, Alpine 3D showed 195% of ParBal, and GLDAS-2 showed 146% of ParBal. In contrast, SNOWPACK showed 128% of ParBal at the point scale in 2017. Thus, we still find that the agreement between ParBal and SNOWPACK at the point scale was much better than between ParBal and the other 3 models at 25 km.

Figure 6: The Alpine 3D images shows considerably less SCA than the ParBal estimate. Since the Alpine 3D estimate is using the gage data I assume where it shows no snow, the snow depth reported at those sites was zero. Is that correct? The ParBal estimate is using MODIS data, which is presumably showing snow in those areas on that date. I'm surprised there is such a large difference. In fact, looking at the movie there is consistently a large difference in snow extent between those two estimates, even at the lower elevations where there are more stations. Why do you think that is?

We assume this is in reference to Figure 7. Because of the austerity of this region, none of these models use gauge data because it's generally not available, hence all of the reanalysis-based forcings. We agree that ParBal (the only model driven by remotely-sensed snow measurements) shows more SCA that persists longer into the melt season than Alpine3D. There are many factors that could cause this, but we suggest the most likely is inaccurate precipitation extrapolation, as we discuss on l 351-352. Thus, too much precipitation is placed in some cells, and too little in others. This extrapolation problem stems from the fact that all of these stations are located in valley floors, which are lower than the average grid cell elevations (l 333-335).

Line 404: Is there a reference you can add for the statement "Compared to a previous effort. . ."? Not sure what this is referring to.

Section 4 "PREVIOUS WORK WITH AKAH SNOW MEASUREMENTS" contains two paragraphs about this.