We thank anonymous referee #2 for his or her comments, and for taking the time to review our manuscript. We've copied the referee's comments in blue. Our responses are in red.

The authors present a study using three different models computing SWE for the Hindukush-Pamir region, as well as snow stratigraphy for selected locations and are able to compare point scale simulations to large scale datasets. The topic is of great importance (as they point out, the region has seen very little attention in the direction of snow even though it is understood to be a major driver of streamflow and important to understand droughts/floods in both Pakistan and Afghanistan) and generally very little field data is available or accessible. The main advancement of the study is the availability of a large number of snow depth measurements from the region.

I find the study (a) in general of great value for the community and The Cryophere in its proposed scope, (b) covering an area that desperately needs more data analysis and (c) generally clearly written and of solid scientific quality. However I would like to see my major concerns I detail below addressed before I find it acceptable for publication.

Major concern:

I appreciate that getting field data as shown here is very hard to collect and once stored by on site staff, often difficult to impossible to obtain. That the authors managed to do so is impressive and I know the general hesitation of local institutions to provide any such sets. However I fear that this does not absolve authors of scientific studies to (a) ascertain the data quality and (b) if not provide then visualize the data to the reader. The authors themselves are cautious with the data quality (L216: 'appeared to be most reliable . . .'). This warrants an assessment of the quality, how that is assessed, what the 'good' data looks like (incl uncertainty) and eventually how this data quality affects model outcomes in ParBal/SNOWPACK. AKAH has excellent staff and can be considered among the most reliable institutes in this area, but many of the stations shown (like the one in the Little Wakhan or north of Ishkashem on the Afghan side) are in areas hardly accessible at all and I wonder how this data was recorded at daily intervals and quality checked internally. I have checked the earlier paper and the report on the avalanche program and equally fail to find details there.

We agree that the AKAH measurement quality concerns are valid. We extensively discuss the quality of the manual daily AKAH snow depth measurements in Section 5.4 where we cover all of the quality issues. Thus, we have already addressed point a). Regarding point b), we find that plotting the errors in the AKAH snow depth is unnecessary, as they are extensively described. For example, one of the most common errors was spurious drops with zeros instead of null values (l 230-232). We do not agree that these need to be plotted for the reader to understand what we are referring to. In terms of internal quality checks from AKAH, as far as we are aware, there were none. We will state this in a revision of the manuscript.

Further, in a revision, we will expand on what is stated on l 217; that is when total snow depth measurements were recorded and did not show spurious drops (there were no spikes – consistent with manual vs. automated acoustic snow depth measurements), that the data quality is likely high and the snow depth is accurate at that point. The only likely error from manually reading a total depth stake measurement is a transcription error, which would show up as a drop or spike.

We will emphasize in a revised manuscript that the errors in ParBal are completely independent of the manually measured AKAH snow depth, as ParBal does not use in situ measurements. The errors in ParBal are mostly based on the fractional snow covered area and radiative forcings, and are covered extensively in the cited prior publications (Bair et al., 2016; Rittger et al., 2016). By “earlier paper”, perhaps the reviewer is citing Bair, Abreu Calfa, et al. (2018), which focuses on the machine learning approach, but not on the sources of errors in the ParBal model ?

We will add uncertainty bounds on the red and blue SWE curves shown in Figure 4 in a revised manuscript. The range of SWE in the 9-pixel neighborhood around each AKAH station (l 295-297) will be used to create the bounds around the red ParBal curve. For the blue SNOWPACK curve, because we find the AKAH snow depth to be accurate (see above), we also find the accumulation season SWE modeled by SNOWPACK at that point to be accurate. We justify this assumption using results from Bair, Davis, et al. (2018), where it is shown that the peak SWE modeled by SNOWPACK at a site with little melt during the accumulation season (which seems to be the case for the AKAH stations based on measured depth) is mostly dependent on the precipitation forcing. A similar result is shown in Bartelt and Lehning (2002).

For the ablation season, there is considerable uncertainty in the blue SNOWPACK curve caused by the radiative forcings and the snow albedo parameterization. Given a lack of knowledge of the uncertainties in the radiative forcings for this region, we will use the 4-6% RMSE and negligible bias reported for our remotely-sensed albedos reported in a recent study (Bair et al., in press). These errors led to a 5% mean absolute error and 3% bias in ablation season SWE for a continental site in CO USA (Bair et al., in press).

I also find it problematic to sell the story as a ‘validation of models with measurements’, as the snow depth measurements are used as an input for one model (ParBal, equally dependent on other remote sensing data but which I think is being understood as the ‘validation data’ here) that is then compared to another (SNOWPACK/ALPINE3D).

As shown in Bair, Davis, et al. (2018) and Bartelt and Lehning (2002), given accurate snow depth measurements, which we believe to be the case for the cleaned AKAH stations, the accumulation season SWE can be well modeled. For the SNOWPACK and ParBal comparisons, we consider the SNOWPACK modeled SWE as legitimate validation data because it is the snow depth, not its density, that drives the variability in SWE. This has been pointed out repeatedly in The Airborne Snow Observatory (ASO) project (Painter et al., 2016) who show that the CV for depth is ~ 3-4X that of density. This has led to many studies using the ASO data for SWE validation, despite using modeled density (e.g. Broxton et al., 2019; Margulis et al., 2019; Oaida et al., 2019).

This makes it difficult to appreciate where the advancement via the new dataset is and whether the point measurements are not lost in the general uncertainty of the satellite data.

The literature review contains all the relevant studies on snow for the region, and as we show, they show wide variation with little to no in situ data used for validation, unlike this study. We therefore suggest that our two take home points discussed below are advancements in snow study for the region.

Additionally I find uncertainties of models inadequately addressed.

We will add a reference about the uncertainty of GLDAS snow estimates. For the other three models, we do not agree that model uncertainty is not addressed. We refer the referee to previous cited studies on the uncertainties in ParBal (Bair et al., 2016; Rittger et al., 2016) and on modeled SWE from depth in SNOWPACK (Bair, Davis, et al., 2018; Bartelt & Lehning, 2002). On l 208-210, we mention the results of Chen et al. (2014) who state that NOAH MP was able to model peak SWE in CO USA with a -7% bias.

The stratigraphy is shown as a single mean figure for all stations, with no further consideration of the spatial variability or at least a mention of it and I wonder whether the one line as an aggregate over all station locations does provide a trustworthy result.

Most of the spatial variability is caused by altitude, which is why we show two different plots for the AKAH stations (Figure 9ab) and the higher elevation 25 km pixel containing them (Figure 10ab). The two figures show very different snowpacks.

As a result the Conclusion becomes very brief and somehow lacking the essential take home message for further studies.

We will add to the Conclusion. We will emphasize the two take home messages. 1) ParBal does an accurate job of modeling ablation season SWE at the AKAH stations, validated with in situ measurements; 2) at the coarser 25 km scale, there is wide spread in the SWE across models, with ParBal on the low end.

And while I appreciate that a number of models were utilized, I think a clearer Figure 2 plus a longer Conclusion on the usefulness of all datasets/models would be prudent.

We will revise Figure 2 to be more clear, per the comments below.

Minor comments:

Figure 1: Here and in the text (L86) you write ‘stations flow into the Indus/Amu Darya’. I get the meaning but I doubt that is technically a sound expression. ‘Catchments including stations in xxx drain into the xxx . . .’

Ok, we will correct.

L93: ‘questionable’ – although I agree on the assessment, I think you need to explain your criticism rather than just throwing a lead without an argument if you want to mention this here

Ok, we will elaborate.

L103: replace ‘x’ with ‘times’ or reword

Ok.

L110: I don’t understand how limited climate data would explain why the remotely sensed data from Smith&Bookhagen2018 are too low.

At Salang Pass, Afghanistan, the WMO reports up 198 cm of snow depth (Bair, Abreu Calfa, et al., 2018) and we estimate up to 1000 mm of SWE (Figure S3, Bair, Abreu Calfa, et al., 2018). Thus < 100 mm of max SWE reported by Smith and Bookhagen is low by up to a factor of 10.

L243: 'hr' to 'hour'

The Cyrosphere (https://www.the-cryosphere.net/for_authors/manuscript_preparation.html) requests that SI-accepted units (https://www.bipm.org/utils/common/pdf/si-brochure/SI-Brochure-9-EN.pdf) be abbreviated in conjunction with numbers, which would be "h" according to the link above. We'll wait for the Copy Editor to clarify.

L269: the 'Nuristan avalanche' is not a commonly known event. You place a citation later in the text, consider placing that here instead with the description of the impact

Ok

Figure 2: I find the salad of arrows unnecessarily confusing – a number of them could be straight but have corners for no reason making it very hard to decide which path to follow to really get the main aim

Ok, thanks! We were looking for a way to improve the readability.

Figure 7: Remove 'The white letters are . . . codes'. Just the acronyms plus actual names are fine.

Ok.

Figure 7: The video supplementary material is valuable, however I would also expect a discussion of the consistency/variability between the models in space, i.e. have a corresponding map that shows the average over the complete period or the total number of days the models actually simulate any SWE.

Ok. As the other reviewer pointed out, there are significant differences in snow covered area between the models. We will discuss these differences further. We still find that the video itself is the best way to illustrate these differences, as they vary over space and time.

Figure 9/10: It is explained in the text but I believe the Figure still needs a y-label

Ok, we will insert a label with "HS/fraction of facets/# of critical layers"

Bair, E. H., Abreu Calfa, A., Rittger, K., & Dozier, J. (2018). Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere, 12*(5), 1579-1594. https://www.the-cryosphere.net/12/1579/2018/

Bair, E. H., Davis, R. E., & Dozier, J. (2018). Hourly mass and snow energy balance measurements from Mammoth Mountain, CA USA, 2011–2017. *Earth System Science Data, 10*(1), 549-563.

Bair, E. H., Rittger, K., Davis, R. E., Painter, T. H., & Dozier, J. (2016). Validating reconstruction of snow water equivalent in California's Sierra Nevada using measurements from the NASA Airborne Snow Observatory. *Water Resources Research, 52*, 8437-8460.

Bair, E. H., Rittger, K., Skiles, S. M., & Dozier, J. (in press). An Examination of Snow Albedo Estimates From MODIS and Their Impact on Snow Water Equivalent Reconstruction. *Water Resources Research*. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019WR024810

Bartelt, P., & Lehning, M. (2002). A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model. *Cold Regions Science and Technology, 35*(3), 123-145. http://www.sciencedirect.com/science/article/pii/S0165232X02000745

Broxton, P. D., van Leeuwen, W. J. D., & Biederman, J. A. (2019). Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements. *Water Resources Research, 55*(5), 3739-3757. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024146

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., et al. (2014). Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study. *Journal of Geophysical Research: Atmospheres, 119*(24), 13,795-713,819. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD022167

Margulis, S. A., Fang, Y., Li, D., Lettenmaier, D. P., & Andreadis, K. (2019). The Utility of Infrequent Snow Depth Images for Deriving Continuous Space-Time Estimates of Seasonal Snow Water Equivalent. *Geophysical Research Letters, 46*(10), 5331-5340. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL082507

Oaida, C. M., Reager, J. T., Andreadis, K. M., David, C. H., Levoe, S. R., Painter, T. H., et al. (2019). A High-Resolution Data Assimilation Framework for Snow Water Equivalent Estimation across the Western United States and Validation with the Airborne Snow Observatory. *Journal of Hydrometeorology, 20*(3), 357-378. https://journals.ametsoc.org/doi/abs/10.1175/JHM-D-18-0009.1

Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (2016). The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment, 184*, 139-152.

Rittger, K., Bair, E. H., Kahl, A., & Dozier, J. (2016). Spatial estimates of snow water equivalent from reconstruction. *Advances in Water Resources, 94*, 345-363. http://www.sciencedirect.com/science/article/pii/S0309170816301506