Referee comment on "Mapping snow depth and volume at the alpine watershed scale from aerial imagery using Structure from Motion" by Joachim Meyer et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2021-34-RC2, 2021

The paper entitled “Mapping snow depth and volume at the alpine watershed scale from aerial imagery using Structure from Motion” by Joachim Meyer et al. presents an evaluation of photogrammetrically derived snow depth mapping from a manned airplane with LiDAR derived snow depth from the same airplane. Even though the topic of snow depth mapping is of high interest and the basic setup of the data available is promising, there are in my opinion several major problems that have to be carefully answered and solved before this paper could be considered for publication:

1. Similarity with previously published research:

The two main authors published the paper “Assessing the Ability of Structure From Motion to Map High-Resolution Snow Surface Elevations in Complex Terrain: A Case Study From Senator Beck Basin, CO” in the Journal Water Resources Research (Meyer and Skiles, 2019). In my opinion this publication is clearly too close to the WRR publication, applying the same sensor and the same reference data just in another area. In times where thousands of papers are published per day, it gets more and more important that publications, in particular within sound journals, present novel and innovative approaches. Applying already published procedures to other areas is not one of them. I am well aware that many PhD students are under a very high pressure to publish. However, the scientific system depends on a rigorous peer-review assessing the novelty and the scientific contribution of publications.

Therefore, the authors have to generate a clear offset to their own and further previous publications presenting novel aspects. There is some potential to do so. For example, the influence of different vegetation classes could be assessed. Or the photogrammetric performance over different snow surfaces. Furthermore, strategies could be presented to improve the performance in steep terrain or over shallow snow packs. The paper should
therefore only be published with a clear additional benefit.

2. Incomplete state of the art:

Mapping snow depth from manned airplanes is not as new as the authors suggest. SFM is also photogrammetry, producing the different viewing angles by motion. Already in the 1960ies attempts have been made to do so (Smith et al., 1967). More recently Lee et al. (2008) and Ledwith and Lundén (2010) made further attempts. And finally Bühler et al. (2015) published snow depth maps of a high alpine catchment (150 km$^2$) in this journal validating their results with independently acquired Terrestrial Laser Scanning (TLS), Ground Penetrating Radar and manual snow depth mapping. These investigations are not mentioned in the current manuscript even though they are highly relevant for the topic. These works have to be included into the state of the art but also the discussion.

3. Validation:

The section 4.2 Co-Registration is confusing. I do not fully understand what the authors do here. However, they write “An added advantage of co-registering of the SFM point clouds to the ASO lidar point cloud was the implicit alignment with the ASO snow depth product”. In my understanding, the authors vastly reduce the value of the validation with this. If you align the point clouds of the method to test with the point cloud of the validation dataset you apply, you can proof that the point cloud alignment works well but not that the method to test can be applied independently. Therefore, the subsequent measures to evaluate the accuracy are at least flawed. In my opinion, a meaningful validation is not possible like this. How would the validation values look, if the SFM point cloud is not co-registered to the ASO point cloud? This would give a more meaningful assessment of achievable accuracies. Maybe I did not correctly understand what the authors do here, in that case they would have to present this step much clearer, maybe also with figures.

Another issue is the downscaling of the reference dataset from 3 to 1 m. A validation should always be done on the native resolution of the validation dataset or a resolution worse than that. The original point clouds of the ASO DSM should therefore be higher. As Mrs. Bormann of the ASO team is also co-author on this paper she should be able to help with that.

On page 12 line 274 the authors state that the accuracy of the ASO snow depth map at 3 m spatial resolution is 5 cm (ASO, personal communication). This is a very important value and should have been soundly investigated. There must be published investigations proving this value.

Throughout the paper I often see a sloppy treatment of references. I am also a friend of short and concise publications but relevant publications have to be properly referenced.
The authors often use the term cm scale. This is not defined. What do they mean by that? Better than 1 cm, 1 cm or 1 – 10 cm? This is important.

4. Vegetation:

An important issue mentioned is the problem of photogrammetry on vegetated areas. But the authors only use the class “vegetation”. Inside this class there must be a very large range of coverages ranging from grass to bushes and trees. And within the tree class there are firm standing trees but also trees that are bent to the ground by snow such as green alder or knee pine (at least there are in the Swiss Alps). They have very different effects on under- or overestimation of photogrammetric snow depth mapping as reported by several authors before, e.g. (Harder et al., 2016; Bühler et al., 2016; Harder et al., 2020). Here the authors could produce additional value by further investigating this problem.

5. Lack of meaningful figures

Most of the Figures are not very informative or meaningful, starting with Figure 1 where not a lot can be seen. In Figure 3 the legend bar does not allow more than distinguishing where there is a lot of snow and where only a bit of snow. The authors could choose another color scale to make more details visible. Furthermore, they should add a difference map and some subsets, where details get visible. In Figure 5 the SfM and the ASO plots seem to have different resolutions. Is that coming from the downscaling of the validation data? In Figure 6 only the negative SfM values are assessed. What about potential wrong positive SfM values? They are harder to identify but are like present, the picture seems to be very incomplete to me.

6. Weak conclusion

Finally, the presented conclusions are weak. There should be more specific points on what to improve on and how this could be tackled. Like presented at the moment, the conclusions are not helpful for the scientific community.

I am aware that my review is skeptical. But I hope that it helps to improve the current submission and to make it ready for publication.
REFERENCES


