The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC1
https://doi.org/10.5194/tc-2022-65-RC1, 2022

Comment on tc-2022-65

Anonymous Referee #1

Referee comment on "Spatially continuous snow depth mapping by aeroplane photogrammetry for annual peak of winter from 2017 to 2021 in open areas" by Leon J. Bührle et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-65-RC1, 2022

This paper assesses aerial photogrammetric digital surface elevations mapped during 'snow on' conditions. It also presents snow depths differentially mapped from the photogrammetric DSM and an airborne lidar DEM at peak snow depth spanning 5 consecutive winter seasons over an area that centers around the Dischma Valley/Davos, Switzerland. The focus is on assessing the ability of airborne photogrammetry to map snow surface elevations in open environments using check points, UAV photogrammetry, and bare ground comparison to lidar. It does not assess the accuracy of the horizontal alignment of the snow free and snow on datasets, or assess the accuracy of snow depth. One important outcome that limits broader adoption is the requirement of ground control points. Overall, this is a valuable and interesting data collection effort and kudos to the investigators for collecting and analyzing this dataset. That being said, the methods and motivation of the paper need work, at this time I cannot recommend publication without significant changes to the paper. Below major points are raised, followed up by more specific and line by line comments. I think the paper could be resubmitted after addressing the issues raised.

- There is a big difference between validating snow surface elevations and validating snow depths, the uncertainty of which is related to both the snow free and snow on dataset in terms of both horizontal and vertical accuracy. This work only assesses snow surface elevations, but seems to indicate that this is an assessment and validation of snow depth. Given the reported high resolution of the maps, and the number of manual and automated snow depth measurements in this region why was a snow depth assessment not also included?

- Relatedly, the title slightly misrepresents the work. If you use lidar snow free elevations, is it truly photogrammetric snow depth mapping? Uncertainties in photogrammetric elevations are higher than lidar (especially snow free conditions when low lying vegetation is present), and uncertainties in snow depth would propagate from snow free and snow on, therefore uncertainties would be higher if photogrammetry was used for both acquisitions - important to consider in areas without available snow free DEMs. Also, this study is only open snow mapping- outside of canopy, urban area, water bodies, glaciers...etc., and therefore is not truly spatially continuous. I would suggest the term 'open snow' or something similar.

-There have been numerous papers assessing the ability of photogrammetry to map snow elevation and depth in open environments at the UAV, airborne, and satellite scale. While the accuracy varies depending on the camera and platform, this is the case for any remote sensing method. At this point method papers focused solely on assessment are not needed; it's an established method, we can move past the point of trying to 'sell' it and move forward with using the datasets like any other snow depth dataset. This paper would be significantly more interesting and a more valuable contribution if the emphasis was flipped between elevation accuracy and snow depth distributions - include a section on snow surface and depth uncertainty, but make the focus on spatial and inter annual variability in peak snow depth. Right now the results on snow depth distributions are brief and essentially qualitative and not a valuable contribution to the paper.

-The 'almost automated' workflow is not justified. Would someone else be able to read this paper, take the datasets, and replicate them using published code, primarily hands off, from start to finish? I would argue having to process the images in Agisoft, then doing reproductions and difference calculation in ArcPro, then applying masks that require input from an expert is the very definition of manual, not automated, processing. It's also a very standard workflow for mapping snow depth and other volume changes, although the authors claim it is novel.

-Would not everyone get highly accurate results if area of poor performance and outliers were masked out and not included in the dataset? The results are not a true indication of the performance and ability of photogrammetry to map snow depth.

-The motivation behind the paper, natural hazards and hydrology, are not supported by the nature of the dataset (once per season, only for open snow) and primary results presented (validating DSM elevations). Ultimately this paper has a relatively narrow scope - it is essentially a validation of surface elevations when snow is present and the introduction makes it sounds much broader. Also, by limiting the dataset to where snow is most easily mapped the value for hydrology is reduced, because you are blind to the amount of water held as snow in areas that are more challenging for photogrammetry, which can be a significant amount of water. This is why lidar is so valuable and transformative for basin scale snow depth mapping in mountains. And for hazards, frequent flights capturing changing snow through the season would be very helpful- but how useful is just one map per season for avalanche or other natural hazard purposes? What does this tell practitioners that they don't already know? If it's depth distributions, independent of magnitude, wouldn't just one or two flights be enough? And wouldn't these flights needs to be near coincident with the occurrence of the hazard? (i.e. dry slab avalanches typically happen before peak depth, flooding typically happens after peak depth). The real value of these flights and datasets would probably be the utility for hydrologic modeling and bias correction of simulated snow depths, I'm surprised this isn't highlighted more strongly as opposed to hazards.

-The investigator seem to indicate that lower cost (relative to airborne lidar) is a major motivation for photogrammetry, but this is a flawed argument given that this data collection is obviously still very expensive - the investigators themselves can only only afford one flight per season - and their workflow relies on a lidar snow free flight. Airborne mapping and data collection is expensive, no matter the instrument on board. High accuracy lidar systems are a more expensive start up cost relative to a good aerial camera, but both systems require expensive GPS+IMU systems, and flying crewed aircraft and processing this type data will always be expensive, and for now, beyond reach for almost all practitioners and researchers. For something that costs tens to hundreds of thousands of dollars, arguing that something is slightly less expensive but still very expensive is splitting hairs. Also, citing a paper from seven years ago to justify this is outdated, the field is evolving very rapidly and now there are private companies doing lidar snow depth mapping.

Line by line

Abstract
(13/14) - Echoing comment above, broadly snow depth is valuable for natural hazards, but I'm not sure this is suitable motivation for the results presented in this paper.

(18) It's actually not considerably more expensive in terms of day to day operations, especially now that there are private companies like Airborne Snow Observatories, Inc streamlining data processing and collection. And again, not sure that's a reasonable argument in this case, when photogrammetry flights are so expensive only one per season can be afforded.

(24) It needs to be clarified here and elsewhere that this is open snow mapping - you can't call it a snow depth map when you explicitly exclude areas where there is snow.

Introduction

37-43 Not relevant background for the presented work - the temporal scale of one flight per season at peak is not all that useful for dry slab avalanches, that typically occur before peak snow depth, during the accumulation season, and you can't tell differences in distribution due to environmental processes with only one flight per season. This study also masks out urban areas, so I'm not sure why snow loads on buildings are relevant.

45-46 Again, relevance, this paper does not map catchment SWE. It maps snow depth in the open part of the catchment.

66 If you exclude vegetated areas, urban areas, etc you are also not mapping continuous snow depth distributions

68-69 Mentioning the need for a helicopter seems very specific to the cited study and out of place

73 is ranging -> ranges

80 Is importance the right word here? Or has it just seen wider adoption because the data collection and processing is simplified in comparison to traditional photogrammetry.

83 Is it true that the accuracy mainly depends on the sensor and GSD? What about GCPs (refer to your own manuscript, also Goetz et al)? What about collection conditions?

91 This statement needs elaboration as it makes it sounds like NIR is required for large scale snow depth mapping.

97-98 For the McGrath paper the RMSE was for the 8m output products and the 0.3m is the GSD

99 what is meant by 'stimulates the application'

100-101 This should say the actual spatial resolution and the RMSE (or what was used for error measurement)

104-107 Think these numbers have changed quite a bit in the last seven years with increased used of airborne lidar, also, these both are still considered 'expensive measurements'

100-115 Overall this discussion is confusing and inconsistent. It sounds like cherry picking results from different papers - i.e. Nolan did get excellent accuracy but over a very small area, and Meyer could have produced the maps at higher resolution with the point density but the maps covered a very large area - so it would be helpful to be consistent with the literature description. For example, what cameras were used in other studies, what was the area covered and resolution, what was the uncertainty. Otherwise it's very challenging to understand how these studies compare to each other and to this one.

120 Can seven years, not even a decade, be considered long term monitoring?

Data and Sensor

150 Can you provide more details? Is from a camera mounted GPS or from the airplane?

154 Did they actually end up being at peak? Although obviously easy to identify after the fact, it can be very challenging to predict the timing of peak. I assume this area has many automated snow depth sensors, so it should be easy to discuss how close to peak flights actually occurred.

168 Are you inferring snow free photogrammetric elevations are not accurate?

172 More details are needed here. For example, were the 20 pts per square meter all ground returns? Also its bold to create a 0.5 DTM from a point cloud that has 0.2m horizontal accuracy, especially since only snow surface elevations are assessed, the accuracy assessment ignores errors in snow depth that comes from misalignment of the snow free and snow on surfaces.

182 Can you quantify 'slightly positive temperatures' and what you mean by mentioning this, so it is not so vague?

187 This is a result, how the numbers were arrived at should be discussed in the results and not presented as a fact in the methods

191 - 192 Can you explicitly mention that these reference points, located in easy to access terrain, are not representative of more complex terrain and therefore would impact the overall accuracy of the maps (which cannot be known without assessment points in that terrain)

193-194 Can these also be shown on the map?

209 What were the other coordinate systems?

Methods

I'm not sure what part of this is considered automated as is mentioned in the abstract

217 You can set the level of filtering? You used aggressive in 1.6, why not just lower the level of filtering in 1.7? This part opens up more questions than answers and could be excluded.

4.1 Processing workflow - it is unclear on how the RMSE is arrived at. Since it seems like a confident result about only needing up to 5 points for 240 km^2, I think it would be worth spending the time to explain this accuracy assessment.

270 What was the point density of the SfM DSMs? Since it was presented for lidar, seems fair to present it here. Also 'is often' -> often being

287-289 Good and bad results are important, just masking out all the bad results does not make for good science. Can you go into more depth about the snow depth overestimation?

Figure 4. If I read the methods correctly, you are taking the geo-referencing results out of Agisoft at face value and not assessing any co-registration differences between snow on and snow free. I think this is pretty important given the high spatial resolution to at least have an idea on how much they are off set - it is critical for accurate snow depth mapping. Also, you are mixing masks from two different data sources at very high resolution, and offsets would also be relevant during this process.

312 Highlighting that any workflow that contains a 'determined by an expert' step cannot be claimed to be automated

4.2.2 Outlier Mask - Why does the UAS map use a value of 3x the confidence interval and here it is set to 10m? Consistency would be valuable. Also, why do you need this mask if the snow depths and error statistics have a ‘filtered’ and unfiltered category. Think this makes the unfiltered category look better than it is.

Table 3. A percentage map for all years would help to interpret the snow depth results and errors below

366-367 This sentence reads like a contradiction. Extensive accuracy assessment, but no ground reference dataset?

370 Why are satellite paper citations backing up this argument?

386 This seems like a huge number of a static surface.

389 - 390 Is this reasonable? The ALS has a vertical uncertainty of 0.1 m?

Results and Validation
5.1 Accuracy Assessment - Assessing the vertical accuracy is important, but what about the horizontal accuracy?

426 What is the definition of complex terrain? Vegetation? Steep areas? Could this also be affected by shallow snow depths?

Figure 8. Can you show location of GCPs? This looks like it could be systematic error that can be caused by the lack of well distributed points.

456 Are these the M1 check points? Those were stated as not well distributed. How can this give confidence?

457 Again, here, anything manually determined is an argument against the claim of 'almost automated'

468-469 Not the first study to show this, citations could be useful here

527 Overall I find the presentation of snow depth results too general and shallow - the results are obvious and can be found in every other paper looking at snow depth distribution, for example, more snow on north facing slopes

Figure 12. Is there something off with the color scale? Is the deepest red color zero?

570 Meyer and Skiles, 2019 established that fresh snow surface elevations can be accurately mapped with photogrammetry

571 This isn't a problem with Agisoft - if you can't see the surface you can't reconstruct it with photogrammetry

577 'approximately 0.1 to 0. 15 m (Fig. 10)' these are filtered values

580 - 581 This paper used the same data as Eberhard 2021? Seems like a circular reference, but also on Line 574, it said to have way better GSD. Why not better results?

582 -583 This is workflow is ...very standard? I'm not sure what this paragraph is trying to say. Everyone follows the same basic steps to get snow depth, and anyone can filter out poorly reconstructed areas to get good snow depths in open areas.

588 -590 So we always need roads? This paragraph does not say much.

598 I guess to summarize multiple of my previous comments here, I'm just not sure it can be considered unique to exclude areas where your mapping methods does poorly.

693 Earlir in the paper it was CHF, be consistent

697 (GCPs, distribution is not important) - > this statement is not supported by the presented results