Review #2 for The Cryosphere Discussion - 2020-09-09
Snow depth mapping with unpiloted aerial systems lidar observations: a case study in Durham, New Hampshire, United States

Jacobs et al. provided substantial improvements in the manuscript structure and scientific content. Adding the comparison between forest type is especially interesting (Figure 6 and S3). However, some parts are still a bit confusing and would need some work to ease the readers' comprehension of the work. I think the article can be recommended for publication if the following minor points are addressed.

The two main points are :

1. The comparison with the work of Harder et al. 2020 seems a bit light. I think it is beneficial for the community to have several validation of similar methods published, but it would be interesting to highlight what are the main differences between the two study methods (in the introduction) and between their results (in the discussions).

2. I appreciate the clarification brought by the authors about the terms « precision » and « accuracy » as requested in my previous review. However the use of the «within cell variability » term is still a source of confusion for me. The « local scale variability »(L349)/« within cell variability » (L351) is measured with the « standard deviation of the lidar elevation values ». What is this standard deviation ? Please provide the formula or a clear description of it in the methods. Is there one standard deviation for each DTM or is it a pixel-based metric ? Is it related to sigma_on and sigma_off used to calculate the one-sided confidence interval (equation 1) ? If they are related, please make clear why one is the combination of the lidar accuracy and surface elevation variation (one-sided confidence interval, L234) and the other is the local scale variability of the ground surface elevation (L349).

Below are minor points.

Title : Why was « unmanned » replaced with « unpiloted » ? I cannot find the term « unpiloted » in the snow depth UAS literature cited in the article (De Michele et al., 2016, Bühler et al., 2016, Adams et al., 2018, Harder et al., 2020, Eberhard et al., 2020).

L61 : the citation of Deems 2013 and Lopez-Moreno 2017 suggests that they made a review of remote sensing methods for snow depth mapping while these article focus respectively on ALS and TLS. I would remove these citations as the list of the remote sensing methods is given in the next sentence.

L62 : ALS is airborne laser scanning, fine. L196 : « UAS laser scanning » becomes ALS. Confusing, especially in discussion : « the UAS lidar surveys presented in this study have key differences from previous ALS surveys » L458 and in the following paragraph. I would use two acronyms for airborne laser scanning (ALS since at least Deems et al., 2013) and UAS laser scanning (ULS ?…)

Deems JS, Painter TH and Finnegan DC (2013) Lidar measurement of snow depth : a review. *J. Glaciol.* **59**(215), 467–479 (doi:10.3189/2013JoG12J154)

L76 : This sentence is too long. Split or reduce. Name the « snowpack features » ?

L100 : now we want to know more about this paper...what do they conclude ? What does your work add ?

L141 : interesting information, but please explain the « to 3.6 V per cell » Cells are never mentionned before.

L150 : please expand a bit on the geo-referencement. Is it adjusting one point-cloud relatively to the other ? Or are they completely individually geo-referenced ?

L174 : How are the boundaries of the field and forest « known » ? External dataset ?

L175 : Maybe remove the second « (Figure 1) ».

L173-185 : I am a bit confused : « intact canopy » means with needles/leaves in winter ? Is it that the needless branches backscatter less the laser, making the deciduous tree canopy appear lower ? Add a sentence or two to explain that, although Figure S3 provides a good visual intuition of what is happening.

L186 : Why extract 5000 points and not use all available points ? Even at the highest resolution (0.1 m) the 0.1 km² raster should represent ~10^7 points which is not computationally unbearable. To be considered in future work.

L186 : I think you can simplify. « Once the vegetation forest type was classified, the raster binary image was vectorized. » is not really necessary. Something like : « Three sets of 5000 points were extracted respectively in the field, in the eastern forest and in the western forest. At each of these random points... »

L191 : Has this test ever been used in geophysics studies ? It would be good to cite other papers using it. It is otherwise really hard to assess the relevance of this test without extensive statistical background. If it is relevant, this is a very welcome technic to estimate significant snow depth differences.

L229 : Consider giving the one-sided confidence interval equation.

L229 : « of the snow on and snow off elevation » is confusing. Could you remove it ?

L233 : « of the snow-on and snow-off **ground** returns » ?

L264 : «**absolute** low bias ». In absolute yes, in relative no.

L266 : « mean height profiles » does not seem right. The heights are not averaged, they are normalized. It is rather a distribution of the normalized elevation, or something along that.

L272 : bring « respectively » earlier in the sentence.

L321 : confidence interval **decreases**

L346 : « the ability to capture the mean snow depth » I disagree: one-sided confidence interval still includes natural infra-pixel variability. You might perfectly capture the mean of a pixel with high infra-pixel variability and still have a large confidence interval.

L359-L363 : this paragraph is a bit confusing. It would benefit from i) defining better the standard deviation (see main comment 2. above) and ii) explain the relationship between point density per cell and cell size. The last sentence of the paragraph implies that reducing the cell size reduces the ground return density. This should be explained.
L359 : **point cloud density** instead of « the point per cell »
L360 : **snow depth map resolution** instead of « spatial scales »

L387 : « reduced accuracy of the GNSS » : in forest ?

Paragraph from L410 : This paragraph is a bit hard to read. It should be simplified and the conclusions better highlighted. See some suggestions below.

L412 : « While this result is not entirely surprising » : little added value of this sentence. I suggest removing it.
L414 : « in snow depth due to pockets of duff and woody debris, and due to higher variability in subnivean terrain in the forested areas of the study site » are duff and woody debris not part of the subnivean terrain ?
L414 : is it really related to « variability in snow depth » ? or variability in snow-off terrain since this is the only thing discussed after.

L416 : Long sentence. Consider shortening with « drive higher confidence interval **in these areas.** »
L416 : High relief terrain is defined by high elevation variability over short distances. Also, I am not sure what this sentence is stating.
L418 : remove « of the study site »

L419 : Interesting result. Please repeat which is higher than which (deciduous, coniferous, field). Cite literature showing the canopy interception effect. Is it not surprising that deciduous and coniferous trees have the same impact interception ?

L426, L440, you mention large UAV and heavy payloads. L372 the UAV seemed to be light (<25 kg) and small in size. Please homogenize.

L463 : SD ?

L475 : please say what area is typically covered by airborne laser scanning campaign.

Figure 4 : In a and c a certain amount of points classified as non-ground is centered on 0 m height. This seems to be points which are wrongly classified. Is this assumption true ? Are they at specific locations ? Consider adding something about this in the results.

Figure 6 a and b : there are some negative snow depth. This is possible due to the uncertainty of snow-on and snow-off DTM but should be presented in the results.

Figure 8. a.: See main comment about the standard deviation of lidar elevation. Here it seems to be a general metric : one standard deviation per snow depth map.