General Comments

The authors present a comparison of snow depths observed using UAV-based structure from motion. They explore the relationship of snow depth through a snow accumulation and melt season across various land cover types within a sub-arctic environment, paying special attention to the interactions between forest canopies and snow depth. The main comparison techniques include presenting differences between UAV-SfM observations and a single observing station, between land cover classes, and based on the distance from the canopy. I believe the time-series component mixed with both the spatial coverage and diversity of land cover types makes this work a valuable addition to the literature. I also agree with Reviewer 1 in that the tree masking was an interesting component and additionally think that the exploration into snow depth as a function of canopy distance was particularly informative. The use of a summary figure at the end of the manuscript was also valuable.

With that being said, I do believe there is also substantial room for improvement. The objectives stated in the introduction did not seem to align with the presented results. Also, while the dataset is impressive in its own right, the analysis and statistical methods were unclear or questionable at times. I have included some specific suggestions outlining where improvements can be made within the following sections. I think the paper is novel enough to warrant publishing with some revision.

Response: Thank you for the encouraging comments. Objectives were aligned more carefully with the results and the analysis, and the statistical methods were clarified. Detailed point to point responses are added below.

Major Suggestions/Comments

L84–87: These objectives don't seem to line up with what was addressed in the study. The first point seems to be the focus of the accompanying paper (not this one). The second and third points should be kept. Though, consider adjusting them to be more in line with the actual analyses done (1 – comparing spatiotemporal snow depth variability across different land cover types, and 2 – exploring the controls canopy has on this variability). There is also a large component in the way the results are presented that presents all observations relative to the point observation. Since assessing the spatial representativeness of the single-point site is such a focus of the analysis approach, I suggest adding a clear mention of this within the objectives as well.

Response: Thank you for pointing out this mismatch on the objectives. We reformulated the objectives accordingly and added sentence about single point measurement assessment.

L88–L93: "We compared the acquired snow depth data with manual snow course measurements and assessed the spatial representativeness of a single point snow depth measurement in relation to UAV-SfM derived data. The specific research questions were: 1) how spatiotemporal snow depth variability differ across forested and open mire landscapes, and 2) what canopy controls on this variability can be revealed with high spatial resolution UAS-SfM snow depth surveys."

The word 'variability' is used throughout when referencing the difference between the 5th and 95th percentiles of the snow depth distribution. This serves more as a measure of the range, not variability (like standard deviation/variance). Please revise your use of the word 'variability' throughout (or update the statistical method to better reflect variability). In most cases, it could be replaced with the word range.

Response: Done as suggested.

L230–244: This section and potentially the statistical approach should be restructured. Initially you mention that all classes are significantly different with high confidence (very low p-value), then process to counter this claim when using the smaller random samples of snow depth data. What is the takeaway here? I suggest selecting a single appropriate statistical test and sticking with it.

Response: Thank you for the good comment. In our opinion it's important to show the complete process for the reader because in using UAS-SfM method the amount of datapoints is usually very large and others could run in similar challenges with sample sizes that can be too large to fail. In the new version of the manuscript, we explain in a more detailed way why our original approach serves study better.

Clairifed in results. L249–L251: "However, using UAS-SfM method the amount of datapoints is very large, potentially making it difficult for the test to accept the null hypothesis. To address this, we reduced the sample size with random sampling to highlight the true differences between landcovers."

With such a considerable focus on comparisons between the ultrasonic sensor at Kenttarova and the UAV-SfM observations, there needs to be a more comprehensive discussion as to the land cover surrounding this site (i.e., distance from canopy, understory, Corine class canopy type etc.).

Response: Details of the location added. L154–155: "The Corine classification of Kenttärova snow depth sensor location is coniferous forest, the distance to the canopies is approximately 5 m and the understory in the sensor location is replaced with artificial green grass mat."

Can the vegetation classification (using Corine) be enhanced by using ortho mosaic data & your tree masks? As is, the resolution is somewhat limiting, and it is difficult to tell how effectively this captures the different canopy types. There would be considerable value in adding forest type & density information into the analysis, without adding much additional work.

Response: We used Corine dataset since it provides widely used and standardised method for canopy influence. Corine is also available for larger regions and thus allows further cross-checking with future studies. We see potential for more high-resolution canopy cover information gained from UAVs but this would be already another scientific study/manuscript.

Minor/Technical Suggestions

L24: First instance of using variability in place of range. Please adjust the terminology here (and throughout)

Response: Done.

L30–31: This point should be modified to reflect the fact that even if there is field data (collected in a classic way through point sites/snow courses) snow analyses are still limited. Doing so would make this more in line with the points made in the discussion and conclusions later in the paper.

Response: Done, sentence added. L31–L32: "…or where the available is data collected using classic point measurements or snow courses."

L49–51: "In forests,…." This sentence is a bit challenging to read, please consider revising

Response: Revised and sentence is split in two.

L59: remove ‘the scale of’, is redundant

Response: Done.

L68: Please specify more clearly that a more comprehensive review of UAV-SfM studies is included in the accompanying paper

Response: Done.

L71: For the likely audience of this work, an explicit definition of tree wells is unnecessary

Response: Definition removed.

L80: Just because the region is locationally different doesn’t necessarily make this work different. Please mention some of the unique considerations (like lighting, forest structure, snow properties) that make the subarctic region a unique study area.

Response: Done. Added sentence. L84–L85: “…consisting for mosaic of forested and peatland areas with challenging climate factors, such as variable light conditions and very cold temperatures.”

L113: Please clean this up a bit, the use of parenthesis is excessive and challenging to associate the numbers mentioned with the cited works

Response: Cleaned.

L149: I assume ‘high-quality and moderate depth filtering settings’ are specific to the software used. Can you please provide a reference to what these parameters mean, or briefly add description herein

Response: These are indeed software specific settings.

The quality parameter is a 5-step (Ultra High, High, Medium, Low, Lowest) setting specifying image downscaling during the depth map generation. Ultra high quality utilizes original photos, while each following step downscales preliminary image size by a factor of 4. The Ultra High quality setting has considerably longer processing times, and more importantly, requires a very large amount of random access memory for large datasets.

Depth filtering is a 4-step (Disabled, Mild, Moderate, Aggressive) setting specifying the aggressiveness of outlier filtering during the dense cloud generation. Aggressive filtering is generally recommended for aerial data processing but Mild filtering might be more suitable for example in the case of poorly textured roofs, etc. Our testing with the December 2018 dataset indicated that Moderate depth filtering produced all around best results.

Reference:

Agisoft Metashape User Manual: Professional Edition, Version 2.0. Agisoft LLC. Available online: https://www.agisoft.com/pdf/metashape-pro_2_0_en.pdf, 2023.

L165: You introduce DoDs here, then quickly shift to referring to them as snow depth maps. Am I correct that they are the same? If so (or if not) make sure it is clear that they are interchangeable terms (or not).

Response: Yes, they are interchangeable here. Short sentence added. DoD here is interchangeable with snow depth map.

L178: What does ‘i) approved vegetation classification’ mean in this context?

Response: We added reference for European Union level Corine datasets and its classification and changed approved to harmonized.

L199: Please re-iterate that the ‘point site’ refers to the single automated station located within the forest

Response: Done.

Table 1: Please clearly indicate which depth is subtracted from which in the figure caption. Also, while presenting only the differences between depths at the courses and to the observing site is interesting (and relevant to the spatial representativeness question), I think this table would greatly benefit from the inclusion of the actual median depth values by class. For example, include the median depth followed by the difference relative to the point site (i.e., 55 cm (-10 cm))

Response: Actual median values added. UAS-SFM derived snow depth is subtracted from the single point snow depth, added to table caption.

L216: Please clarify ‘point measurements’ – is this plural to represent the time-series at the single observing site, or does it refer to the snow course obs. Please try to make this clear throughout the paper.

Response: Refers to the single point measurements. Corrected throughout the manuscript.

Table 1 -> Section 4.2 (and other locations): Please make sure your use of units is consistent (pick m or cm)

Response: Done throughout the manuscript.

L219–L223: I was confused by the statistics here. Why not just present the median depths?

Response: We improved this section to be crisper. Presenting the ranges in addition the median snow depths gives more information about the landcover effect on snow depth variability.

L224: “difference in variability” (?)

Response: Corrected

L225: ‘difference’ compared to what? The point reference?

Response: Revised. Compared to the single point reference.

Table 2-3: Again, in my opinion, there is an unnecessary added layer of complexity here. I suggest presenting the true depth values (5-95%) within each field and timestep, then presenting the observations for the same timesteps at the point site & the range/median of the snow courses. One idea… a time-series figure (or 1 per plot) with the data bounds may be a good additional way to visualize changes occurring across each class (& their relationship with the observations). -> it also would capture similar information to what is shown in Figure 5

Response: Table will be edited as suggested.

Figure 5: This is a great figure, please ensure it is referenced/discussed sufficiently in the text

Response: Discussion will be added.

Line 263: Again, variability needs to be clearly defined

Response: Clarified.

Figure 6-8: Since these plots are showing similar information, I think their number can be reduced. It is not clear to me what Figure 8 adds to the paper. Consider revising.

Response: We would prefer to keep the Figure 8. It shows that the number of data points gets scarce further away from the canopy. This figure shows that the statistical uncertainty gets substantially higher with less data points. We will add mode details to the manuscript.

Section 5.1: This section is well written but lacking a bit in terms of the actual findings regarding spatiotemporal variability during the accumulation and melt season. Consider adding to the discussion using relevant statistics from the paper relevant to the spatiotemporal variability

Response: Thanks for the good suggestions. We will modify the paper as suggested.

L308, 310: careful using 'variability' here

Response: Changed to range.

L313–314: sentence here is a bit wordy

Response: Revised.

L326: Note the magnitude of the variability/range

Response: Done.

L342: Do you mean 'similar' types here? Or 'different'?

Response: Similar as the ultrasonic station is also located in forest. Made small change.

L344–345: Snow courses are good, but I think it would be useful to reiterate that they are limited in their ability to describe the types of canopy interactions observed in this work (10's-100's of obs, vs. thousands)

Response: Sentence added. L363–L465: "Snow course measurements are limited in their ability to describe detailed canopy interactions with their low number of observations (10's-100's) compared to UAS-SfM method (1000's)."

L365: '….by the limited canopy effect,…'

Response: Corrected.

L380: errant '('

Response: Corrected.

L388: completely up to the authors, but I think a more reasonable range of magnaprobe survey observations is in the 1,000-10,000 range. 100,000 seems a bit high from my experience. This also helps to make the value of the UAV-SfM clearer

Response: Changed. 100000 might be an overkill.

L411: flip local & medium

Response: Flipped.

L419: consider removal of 'illumination conditions'. While mentioned in the part 1 paper, this does not seem to be something addressed in this manuscript

Response: Removed.