Reviewer 2

Manuscript Title: Characterizing Spatial Structures of Field-Scale Snowpack using Unpiloted Aerial System (UAS) Lidar and SfM Photogrammetry

General review: Cho and other authors measure the spatial heterogeneity of snow across a study plot in New Hampshire while evaluating the performance of UAS structure for motion (SfM) photogrammetry against lidar and in-situ observations during one snow season. This effort was conducted in New Hampshire, USA, and the study plot includes both a forested region and open field. The authors determined that the open areas tended to have deeper snow than forested areas and that lidar and SfM generally performed better, compared to in-situ observations and each other, in the open areas of the study plot. Static landscape variables, such as vegetation type and slope, impacted the distribution of snow consistently across the snow season. This project provides a new evaluation of UAS SfM as a tool to measure snow depth (a lower cost option compared to lidar). Much of the analysis, and thus manuscript text, includes a very clear description of the methodology used. However, the presentation of results in figures could be refined for clarity, and the results also require a more in-depth discussion around the performance of the select UAS tools, given the first two objectives of the study. The following line-by-line comments, which vary in "major" versus "minor" feedback, should provide more clarity and direction regarding this review, with the hope of better emphasizing the importance and value of this work.

[Answer] Thank you for your constructive feedback with specific comments on our manuscript. We have carefully revised our manuscript based on each of your comments.

Line-by-line comments

Abstract: Throughout, it is initially unclear and confusing what "spatial structure" is referring to. The words "patterns" and "spatial variability" are only used at the beginning of the introduction, which provide more clarity. Suggest briefly including a definition in the abstract, given the use of "spatial structures" in the manuscript title. Otherwise, suggest replacing "structure" with "variability, ""heterogeneity," or "distribution," which are more commonly used in the literature (including in the citations provided within this manuscript), whereas "structure" is often associated with the vertical microstructure of the snowpack.

We acknowledge the potential confusion around "spatial structure." The manuscript has been revised to clarify this term, either defining it briefly or substituting with "spatial variability" or "distribution" to align with conventional terminology.

Line 27-28: It would be beneficial and more complete to report, at the very least, the direction of the correlations.

Thank you for the suggestion. Since we did not explicitly conduct a collation analysis, we have revised the terminology and edited the text as follows.

"Within the field, the spatial distribution was primarily affected by slope and the shadowing effects of the forest canopy."

Line 36: Here "snowpack structure" is ambiguous, where, to some readership, the term insinuates the vertical, microscale structure of the snowpack.

The term has been replaced by "snowpack distribution".

Line 70: Unclear what "these transition periods" are referring to. Please define.

Additional clarification has been added to the text:

"However, investigating the transition periods between snow-on and snow-off poses challenges, primarily due to the snow becoming increasingly shallow and patchy, eventually revealing bare ground."

Line 73-75: It would be impactful for the authors to include why this type of forest/snowpack was chosen. For example, a number of the previously cited UAS works take place in other climates/forest types.

We conducted a review of Sturm and Liston's Climatology as well as Johnston et al.'s (2024) climatology and determined globally that ephemeral and transitional snowpacks cover large areas but are understudied. We are adding text to reference the climatological information.

Line 89: Reference error. There are a number of these throughout – also associated with figure references – thus I will only note this one.

This reference error, as well as others throughout the manuscript, have been corrected.

Section 2: Suggest further emphasis on why this area might be ideal for this type of study (shallow snow depths, type of forest, historical data, etc.). Obviously, there are many other locations which offer open versus forested regions.

In general, ephemeral and transitional snowpacks are understudied. The site has historical data and a rich history of forest ecology research.

Figure 1: It would potentially eliminate preemptive readership questions if the authors stated that the derivation of the variables shown in Figure 1b-g is explained in the following section (3).

Thank you for this suggestion. I agree that clarifying this in Figure 1 would be helpful for readers. I'll add a note to the figure caption indicating that the derivation of variables in Figure 1b-g is explained in Section 3 to provide clearer guidance on where to find this information.

Table 1: The concept of a mixed forest is not also shown in Figure 1b. Is this referring to a blend of both coniferous and deciduous trees? If so, it is unclear the fraction of coniferous and deciduous used to determine the mixed forest area (50/50?).

Thank you for this feedback. Based on feedback from other reviewers, Table 1 was moved to the supplementary material (now Table S1) and now only includes sites which were visited during all of the sampling campaigns. Extra samples collected on 2/4 and 2/24 were removed. The caption for Table S1 was updated to add further clarification of how these forest types were determined:

"Table S1 Number of magnaprobe sample locations (1x1 meter grid cells) by land cover type for each snow-on UAV flight over the field campaign period in 2021. Each grid cell was comprised of nine, evenly spaced Magnaprobe snow depth measurements and one snow tube SWE measurement. Forest type was determined based on the binary Green Leaf Index (GLI) output within a 10x10m grid around the sample point (deciduous < 40% leaf on, mixed 40 – 60% leaf on, coniferous > 60% leaf on)."

In addition, Figure S1 was added to the appendix to show the coarsened GLI output for these sites.

Line 131: Can the authors provide insight on which ground conditions resulted in more returns versus less?

Thank you for your suggestion. Generally, ground conditions such as vegetation cover and surface roughness influenced the number of lidar returns. We applied this into the text.

Flights produced between a total of ~70-140 million returns per mission, depending on site ground conditions (e.g., vegetation cover, surface roughness, etc.).

Line 174: Suggest a Sturm citation for the magnaprobe.

Sturm, M. and Holmgren, J., 2018. An automatic snow depth probe for field validation campaigns. *Water Resources Research*, *54*(11), pp.9695-9701.

Line 180: As written, it is unclear if the 9 in-situ measurements were at random within the 1x1m or consistent across each survey. And can the authors please elaborate on why full sampling was not conducted during each flight (time/personnel constraints)?

Thank you for your comment. The 9 in-situ measurements were evenly spaced throughout the 1x1 m grid cell. Locations were relatively consistent, with small adjustments made between sample dates to avoid disturbed snow from previous sampling campaigns. The full sample campaigns were conducted on two of the dates as part of a prior study on comparison of in-situ snow depth measurement techniques (Proulx et al. 2023). Upon further discussion, the authors determined that the inclusion of both the full and partial sample data for this work was likely to cause confusion and potentially bias the statistics computed for each date. As a result, we removed the extra samples collected on 2/4 and 2/24 and retained only those samples which were included in nearly all surveys. Table 1 was moved to the supplementary material (now Table S1) and the caption was revised to specify what the numbers are referring to (i.e., number of locations where a grid of 9 samples were collected).

Additional text was added to section 3.2 to describe the sampling campaign and provide explanation for dates with missing samples.

"In-situ snow depth sampling was conducted in the field and forest using two methods: a Snow-Hydro LLC magnaprobe (Sturm and Holmgren, 2018) and three Moultrie Wingscapes Birdcam Pro Field Cameras. The magnaprobe sampling followed a single long transect (18 points) and two short transects (3 points each). The long transect was approximately 145 m long and laid out from east to west (Figure 1). From east to west, the transect started in the open field area, then transitioned to the coniferous, then mixed, and finally, deciduous forested areas. The two short transects were located in the open field; one in the northwest portion and the other in the southeast. At each point, nine, evenly spaced measurements were taken within 1 m x 1 m grid cells. It is worth noting that some dates were missing sample points due to disturbance of the sample area, either by collection on previous days or recreational use at the site, or due to personnel and equipment limitations (Table S1)."

Additional text was added to the Table S1 (formerly Table 1) caption in the supplementary material:

"Table S1. Number of magnaprobe sample locations (1x1 meter grid cells) by land cover type for each snow-on UAV flight over the field campaign period in 2021. Each grid cell was comprised of nine, evenly spaced Magnaprobe snow depth measurements and one snow tube SWE measurement. Forest type was determined based on the binary GLI output within a 10x10m grid around the sample point (deciduous < 40% leaf on, mixed 40 – 60% leaf on, coniferous > 60% leaf on)."

Figure 2: It is currently challenging to determine the main takeaway of this figure – is it simply to observe the timeseries or to compare across the field versus forested areas? For example, it appears that air temperature and precipitation/cumulative precipitation are the same, which would make sense given data availability, but is thus redundant. It is particularly challenging to follow the 3x y-axis labels on the right side of each figure. Suggest reformatting as a sequence of timeseries – with only 1x precipitation/cumulative precipitation panel, 1x air temperature panel, and potentially 2x snow depth panels for each area (forest versus field), including the in-situ observations.

Thank you for your feedback on how to revise Figure 2. We have combined your feedback with that from the other reviewers and created a new version of this figure with three subplots. The purpose of this figure is to show the reader how typical conditions at the field site changed throughout the sampling period. We hope that the context provided by this time series also aids in the interpretation of the raster timeseries shown in figures 5 and 6.

To reduce redundant information, the first subplot shows the meteorological data which illustrates the seasonal conditions (precipitation liquid equivalent and average air temperature). Cumulative precipitation has been removed since the plot only shows the field season rather than the entire winter season. The second shows the snow depth evolution in the field from the field camera (continuous measurement) and the median snow depth value for the two UAS measurement types (SfM and Lidar). The third subplot shows the snow depth evolution in the forest from the two cameras and the median snow depth value for the two UAS measurement types.

Figure 3: Are N-values the same across the two panels? Please add.

Thank you for this suggestion. N-values for the samples shown within the bounds of the plot axes have been added to each plot. The updated figure caption describes these N-values and also includes the number of points excluded from each plot.

Line 278-279: Suggest including the sub-areas of the field when introducing the field and an explanation as to why there the authors created a division here (e.g., what led to the decision making for a NW vs. W vs. E sub-area of the field?).

Thank you for this suggestion. The sub-areas of the field were selected based on a prior study at this site (Cho et al., 2021) and were not part of the original sampling design. As a result, the authors would prefer to introduce these field sections along with the results rather than in Figure 1. Additional information on the selection of these regions was added to section 4.2:

"The field was divided into three areas based on an early study (Cho et al., 2021) showing distinct topographic and soil characteristics in each section."

Figure 5: Suggest a more intuitive color scheme for snow depth. The difference color scheme makes sense (negative = red vs. positive = blue). For just snow depth, suggest purple leading to blue and then red (or something similar where red is not a color in the middle of the color bar).

Thank you for this suggestion. We have reviewed other journal articles which present spatial snow depth data measured by UAS and updated Figure 5 to a sequential, cool-toned color bar similar to that used elsewhere in the literature. The difference color scheme was kept nearly the same.

Figure 8: It is unclear what 1-5 (low to high) represents – this should be stated in the figure caption. Further, are there any statistical differences? If so, please note here and in the paragraph above with relevant p-values.

The 1-5 for each boxplot represents the relative range of each physical variable showing the histograms below (e.g., for slope, 1: 0-5 %, 2: 5-10%, 3: 10-15%, 4: 15-20%, and 5:20-25%). As suggested, they were specified in the figure caption.

To quantitively examine the statistical significances, the authors are analyzing them with a Kruskal-Wallis test, which is suitable for comparing medians across multiple groups and does not assume normal distribution. This test indicates that there are/are not statistically significant differences in MRD values among the physical variable groups. In response to your suggestion, we will also explore the use of notched boxplots as a visual indicator of median differences between groups. The notches provide an additional visual cue for significance; when notches between two groups do not overlap, this suggests a statistically significant difference between the medians at a 95% confidence level. These notched boxplots are now included in the revised Figure 8 to visually support the statistical findings.

Line 338: Here and throughout the manuscript, the terms “modestly” and “higher” read subjectively and would be more impactful if numerical values accompanied them and/or if there was a defined threshold for what the authors considered “modest” versus “high.”

Agreed. The sentence has been revised to include numerical values as indicated below. Additionally, the authors have reviewed and revised similar subjective language throughout the manuscript.

“Compared to in-situ measurements, SfM experienced higher error in the field (MAD: 3.5 cm for lidar and 4.0 cm for SfM) and notably higher errors in the forest than lidar (MAD: 6.3 cm for lidar and 31.4 cm for SfM)”

Discussion: From the results section (e.g., Figure 4a, Figure 5), readers are led to believe that using SfM for snow depth generally is not a feasible option (without significant uncertainty) except for the west side of the field, and there isn’t much of an explanation as to why. It is unclear what makes the west side of the field different from the rest of the field? Differences to the forested portion of the study plot are perhaps more obvious but are also not stated explicitly. Further, what might the authors suggest doing differently to reduce the numerous erroneous SfM measurements? The only plausible explanation currently provided is insufficient number of point clouds. It is stated in the introduction that this methodology is still an emerging one, thus this seems like an opportunity provide insight into how UAS SfM for snow depth measurements may still evolve.

Thank you for this feedback. The authors reviewed the SfM rasters and workflow to determine if there was a possible explanation for the difference in performance in the western field. It appears that there were a greater number of "low confidence" tie points in the eastern field on 2/20 and 2/24, potentially contributing to the greater difference between techniques on these dates (Figure 5). New versions of figures 4 and 5 were produced using updated rasters which exclude these low confidence tie points. Low confidence points are likely a result of the fresh snow in this portion of the field lacking a sufficient number of unique characteristics for the SfM algorithms to stitch images together. This is also the likely cause of the large gaps in the SfM map on 2/4.

Options for improving the performance of SfM in the forest are limited due to the obstruction of the ground by branches and leaves. Mixed forests like the one in this study are also challenging due to the messy, repetitive features in the forest canopy which present challenges for finding valid tie points between images. While flying lower and slower can improve sensing in more open areas, lidar presents a much better alternative for measuring snow depths in forested regions.

While pixel-based comparison showed considerable scatter between the two techniques (Figure 4), both were able to capture the differences in snow depth over time (Figure 5). The authors have revised section 5.1 to include more discussion on the relative performance of the two techniques and the limitations of SfM for the study region. We added more information on the greater potential for SfM tie point errors in forested/vegetated areas, over homogenous snowpacks (e.g., fresh snow), and in sub-optimal lighting conditions. Citations are presented for each of these conditions. In addition, further discussion was added to describe problem areas in the field at the Thompson Farm site (e.g., northwest field, east field) and potential causes for the larger differences in these areas (e.g., lack of unique features, fresh snow).

Line 344: It would be helpful to connect the western portion of the field in this study to the subsequent sentences on past studies – e.g., does the western portion have a different vegetation type or other static landscape characteristic/combination of note (nothing particularly stood out in Figure 1)?

Thank you for this suggestion. We have added the following text to section 5.1:

"We also found that the SfM snow depths did not consistently agree with the lidar snow depths over the entire field and on all dates. On most dates, the difference in UAS-measured snow depths was close to 0 cm in the field. The best agreement occurred in the western portion of the field while the southeast and northwest portions had a larger amount of variability in measured values. The shadow hours and land cover type in the east and west fields are similar, however, the eastern field has a more gentle and less variable slope and fewer unique features (e.g., access road, USCRN station, pond, dirt piles, footprints) than the western field. The relatively homogenous features in the eastern field indicate that the difference between techniques is likely due to a lack of sufficient valid tie points for SfM. It is not clear what caused the differences between SfM and Lidar in the NW field that were not present in the other field areas. Unique features in the NW field are prevalent drainage patterns and shadowing that could be investigated further in the future."

Line 353: Can the authors indicate what likely caused the erroneous values of 150+ cm of snow depth as measured by SfM?

SfM point clouds are generally noisier than lidar because they are derived through calculation and not a direct measurement, however, the authors do not have an explanation for the exact cause of these erroneous values.

Conclusion: Suggest restating the error values when referring to "lower error"

The error values were restated for UAS techniques compared to in situ in the field and forest.

"Lidar demonstrated superior performance compared to SfM when evaluated against in-situ observations, exhibiting lower errors. Both UAS techniques exhibited lower errors in field settings (lidar MAD = 3.5 cm, SfM MAD = 4.0 cm) than in forested environments (lidar MAD = 6.3 cm, SfM MAD = 31.4 cm). Though, as expected, differences between lidar and SfM snow depths were more pronounced in forested regions (MAD = 55.7 cm), with SfM often registering anomalously deep snow depth values."

Line 431-433: Suggest explicitly restating the relationship – e.g., x vegetation type leads to deeper [or shallower] snow depth.

Now: The spatial structures of snow depth captured by lidar remained consistent throughout the study period. For the entire study area, deeper snow was found in the field, in locations having shallow slopes and low soil organic matter. Within the field, snow deepened with increasing shadow hours.