Reviewer 3: Ross Palomaki

General Comments

The authors present a generally well-written and well-referenced study comparing two UAS-based snow remote sensing techniques (Lidar and Structure from Motion) across different landscapes in New Hampshire, USA. I appreciate that the study was conducted over relatively shallow snowpacks (~tens of centimeters), which I feel are underrepresented in the literature and pose unique challenges for obtaining accurate snow depth measurements. I also like that the authors incorporated the Relative Difference concept to analyze this time series of spatially distributed snow depth measurements. I was not familiar with this concept, but I think it led to an interesting framing of the results.

While the introduction and methods sections are generally clear and easy to follow, there is room for improvement in the presentation of results and subsequent discussion. As noted in the Specific Comments below, I feel that the discussion is lacking critical engagement with some of the more complicated findings from this study. In particular, the SfM results do not inspire much confidence for the technique overall. I appreciate that the authors did not try to hide these larger errors, but there is little discussion around the potential sources of those errors, or suggestions for how these errors might be avoided in future studies (if that is even possible). This study could be more impactful to the broader snow/UAS community if some of these topics were explored more deeply in the discussion.

[Answer] Dr. Ross Palomaki, Thank you for your constructive feedback with specific suggestions on our manuscript. We have provided our answers to each of your comments and carefully revised our manuscript.

Specific Comments

Line 21: I am a bit surprised to see a range of errors for SfM but not lidar here, since they both have ranges on the next line. Perhaps an accidental omission?

Based on feedback from other reviewers, the comparison of in situ techniques was removed from the manuscript. As a result, the ranges were removed, and this section now only includes the MAE for the UAS and magnaprobe comparison.

Line 42-49: This paragraph would benefit from some additional discussion/more precise language related to spatial variability (Line 42), spatial patterns (Line 43), and hydrologic patterns (Line 44). With multiple phrases present it is unclear if these are different concepts, and which properties of the snow are relevant. E.g. if snow stratigraphy is variable over some length scale but snow depth is not, is that spatially variable snow but not a hydrologic pattern? Please clarify here and implement similar changes throughout the paper, e.g. Sections 4.3, 4.4, 5.

Thank you for pointing this out with reasonable suggestions.

The spatial variability of a snowpack encompasses both static (e.g., terrain, vegetation) and dynamic (e.g., weather, microclimate) factors that influence snowpack characteristics across different spatial scales (Clark et al., 2011; Grayson et al., 2002; Mott and Lehning, 2011; Trujillo et al., 2007). In this context, spatial variability refers broadly to the differences in snowpack properties (e.g., snow depth, density, temperature, and stratigraphy) across the landscape, which can vary from small scales (e.g., microtopographic variations) to larger watershed or landscape scales. By contrast, *spatial patterns* specifically denote the distribution of these snowpack properties as coherent structures or repeated features that emerge across the landscape. These patterns are critical as they represent consistent snowpack behavior linked to landscape features, reflecting hydrological, ecological, or even biological functions within the snowpack. When we refer to *hydrologic patterns*, we are primarily concerned with features of the snowpack that directly influence hydrological processes, such as snowmelt timing, runoff generation, or infiltration. These hydrologic patterns tend to persist over time but may be altered by major weather events, temperature changes, or snow redistribution by wind. For example, a hydrologic pattern might include consistent snow retention in areas with specific shading or wind exposure, directly affecting water availability during melt periods.

The spatial patterns of the snowpack, particularly those that exhibit consistency or repeatability over time, are instrumental in applications requiring snowpack assessment, such as snowmelt forecasting, remote sensing downscaling, in situ observation integration, and data assimilation for model enhancement (Pflug and Lundquist, 2020; Cho et al., 2023). Recognizing these patterns also aids in understanding landscape features, biogeochemical cycles, and habitat conditions (Boelman et al., 2019; Pflug et al., 2023).

By combining our answers with feedback from other reviewers, we implemented changes throughout the manuscript, particularly Sections 4.3, 4.4, and 4.5.

Line 55: Please quantify (at least approximately) "small-scale snow patterns" – tens of meters? It would also be helpful if you could relate this phrase to "field or local-scale snow features" at the beginning of this paragraph (Line 50).

We have modified the text to consistently use local and field scales and have defined the scales in the text.

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: A review. Hydrological Processes 9, 251–290. https://doi.org/10.1002/hyp.3360090305

We use the Blöschl and Sivapalan (1995) definitions of scale where point or local (~1 m) and Hillslope or field (~100 m).

Line 69-70: "e.g., forest and fields" – I think more than just these landscapes. UAS technology has also enabled rapid progress in complex/mountainous terrain for example. I suggest removing this parenthetical statement from the end of the sentence.

The statement at the end of the sentence has been removed.

Line 70: "these transition periods" – unclear, previous sentence mentions "the entire snow period"

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 75: Please clarify why you only investigate snow depths less than 35 cm. Is this just based on the datasets you collected or is there a particular motivation for snow depths below this threshold?

Thank you for the comment. We focused on snow depths less than 35 cm because this range reflects the specific conditions observed in our dataset.

We have updated the text for clarity, as follows:

(2) Conduct a quantitative comparison of lidar and SfM snow depths (< 35 cm) throughout the snow period, as this range of snow depth reflects the specific conditions observed in our dataset.

Figure 1: Consider widening the color bars in panels c-g. Panel b is wider and easier to discern the colors.

Agreed. We made the color bars widen in panels c-g.

Table 1 caption: The information about the sampling strategy in each 1x1 m grid cell seems better suited for the main text.

Based on feedback from other reviewers, Table 1 was moved to the supplementary material (now Table S1). The following text was also added to the main text (section 3.2) on Field Observations:

"At each point, nine, evenly spaced measurements were taken within 1 m x 1 m grid cells."

Line 149-152: With the in situ sampling strategy there are always 9 Magnaprobe measurements for the average snow depth. Is it possible to provide an approximate range of the number of lidar ground returns within each 1x1 m square? How does this number compare to the 9 Magnaprobe measurements, and is it relatively consistent throughout the study or does it change with time, landscape, etc?

A table summarizing the point counts by land cover for SfM and lidar was added to the supplementary material

Line 170: "following the same procedure as the lidar" – much work went into processing the lidar data. If you mean that you subtracted the snow-on and snow-off SfM maps to get snow depths, I suggest writing that explicitly here.

The statement was revised as follows:

"Snow depth products were derived following the same procedure as the lidar by calculating the difference between the ground classified snow-on and snow-off elevations within each pixel."

Line 180: Please clarify if the 9 measurements were taken in the same pattern at every grid cell (I am envisioning a 3x3 pattern hitting all 4 corners and the middle of the cell, but this is worth specifying).

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. As we mentioned above, Table 1 was moved to the supplementary material (now Table S1) and the following text was also added to section 3.2 on Field Observations:

"At each point, nine, evenly spaced measurements were taken within 1 m x 1 m grid cells."

Line 183: does the less accurate GPS (~centimeter scale) matter compared to the RTK- driven image geotagging (listed as sub-centimeter in Line 155)?

Thank you for this comment. The use of "centimeter-scale accuracy" for the RTK image geotags was incorrect. The authors have updated the text to read:

"The RTK system integrates a static base station that relays GNSS corrections to the UAS, enabling approximately 3-cm accuracy of image geotags."

Line 210-211: Did you factor in a potential change in the shadow hours between February 4 and March 7 based on changing solar angles? Maybe this is negligible for the results of this study but would be good to clarify either way.

Thank you for the thoughtful question. We agree that the change in solar angle between February 4 and March 7 could influence the shadow hours due to seasonal shifts in the sun's position, potentially impacting solar incidence and, thus, shadowing on the terrain.

For this study, we considered shadow hours as representative within the observation window from February 4 to March 7. Although the sun's angle does change slightly during this period, our analysis assumed a consistent shadowing effect throughout the day (7 am to 5 pm) for simplicity and practical applicability. The variation in solar angle over this period is relatively minor in terms of altering shadow duration or extent across the study area, particularly given the scale and resolution of our terrain model. Thus, we determined that incorporating adjustments in solar angles would have a negligible effect on the overall shadow hours and snowpack conditions measured here.

However, we acknowledge that incorporating dynamic solar angle adjustments could be beneficial in studies where seasonal or monthly changes in shadowing may significantly impact outcomes, particularly in cases involving longer observation windows or areas with high topographic relief. We have clarified this in the revised text to address potential reader questions on this point:

"*The shadow hours, representing the number of hours from 7 am to 5 pm experiencing shadowing, were calculated using the unfiltered UAS lidar digital terrain model and a static sun incidence angle based on average conditions for the study period. Given the minor variation in solar angles from February 4 to March 7, any change in shadow hours was considered negligible for this study.*"

Line 230-249: Please clarify in this paragraph if USCRN precipitation values refer to snow depth or snow water equivalent. It would be helpful to specify in the first mention of the station (Lines 190-192) what sensor is available to measure precipitation and how you convert that to snow depth (assumed density?) if there is no dedicated snow depth sensor on the station.

Thank you for this comment. The precipitation depth from the CRN station represents a water equivalent (measured by weighing precipitation gage). The description of the CRN station at the end of section 3.2 was updated with this information and clarification was added to the Figure 2 caption.

Line 242-246: This is probably fine since we expect a fair amount of variation across the transects with a fairly shallow snowpack overall. But how did the field camera measurement compare with the closest 1-2 magnaprobe grid cells? I think that would provide more meaningful information than a comparison with the average across the entire transects. Similar comment for the forest site (Line 251-253).

Thank you for this feedback. Based on comments from other reviewers, the comparison between the in-situ measurement techniques was removed from the manuscript. As a result, this information was removed, and the text instead focuses on the comparison of UAS techniques.

Figure 2: I'm having a difficult time understanding what's going on here with 4 shared y- axes. Some points of confusion:

Cumulative precipitation doesn't start at 0, which implies it's cumulative from some date earlier than the start of the field campaigns. Maybe the start of the water year? But the field cameras at both sites imply 0 snow depth at the beginning of this timeseries. To me the cumulative precipitation does not provide any meaningful information in the context of these field campaigns. It also looks like the same curve in both subplots so I suggest at least removing one of the redundant curves, if not both.

I assume precipitation (mm) is in reference to snow water equivalent when the air temperatures are below 0 C. I suggest stating this explicitly somewhere. Is there a snow depth sensor on the USCRN station?

Caption states "UAS-based measurements represent average of all samples" but it also looks like some error bars are included, which are often covered up by different colored error bars from a different location. I suggest either removing the error bars completely and just stick to an average, or find some way to stagger the different sampling locations to prevent overlap. Also specify if the error bars represent IQR, +/- 1 standard deviation, etc.

One way to make this information clearer (with fewer shared axes) might be to have one subplot for temperature and precip (since the data are the same at both sites) and then separate subplots for the field and forest snow depth measurements.

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 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 measured by the CRN station (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 precipitation depth from the CRN station represents a water equivalent (measured by weighing precipitation gage). The description of the CRN station at the end of section 3.2 was updated with this information and clarification was added to the Figure 2 caption.

The second subplot 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. The authors feel that the error associated with these techniques is sufficiently described elsewhere in the manuscript (e.g., Figure 4) and so average values in Figure 2 are shown without error bars to improve visualization.

Line 258-259: "All snow observing methods were able to distinguish that the average snow depth was slightly deeper in the forest than the field." Is this a mixup of forest and field?

Thank you for this comment. The text in section 4.1 were revised to better describe the updated figures. This statement was removed and more detail was provided on the difference between measured snow depths in the field and forest.

Compare reported snow depths in Line 236-238 as well as results in Section 4.3.

We have updated section 4.1 to now include this comparison.

Figure 3: Please note in the caption the different axis limits between the subpanels.

Based on feedback from other reviewers, the authors have updated both of the subplots in this figure to a range from 0 – 50 cm on the x and y axes (this range matches the change made to figure 4). The field camera data were removed from this figure and the updated version now compares the two UAS measurement techniques against the magnaprobe measurements. The colors and marker shapes have been updated to reflect this change.

Figure 4: I like the subpanel showing the color coding of the different fields, but perhaps remove the forest outline as it took me a minute to realize those data are not included on the left figures. Also it is difficult to discern the difference between the solid Lidar lines and the dashed SfM lines in 4c.

Thank you for these suggestions. The forest outline has been removed from the reference map. Based on comments from other reviewers, we decided to reduce the x and y axes limits on these plots to 0 – 50 cm. The figure has been divided into eight subplots showing the forest and field sections individually to improve visualization of this data.

Figure 5: Personally I am not a big fan of the snow depth color bar and I'm not sure that it will be colorblind friendly. Did you try using a simple white -> blue gradient for snow depth? Your choice in the end, this is just a suggestion. The red -> blue gradient makes sense for the difference maps.

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 6/7: I suggest switching the order of these figures. The MRD map is a slightly easier concept for me to grasp and leads nicely into the individual RD maps. Plus it's nice to see the larger, detailed map before the smaller subpanels in the current Figure 6.

Thank you for this suggestion. We agreed on this point. The two figures were switched.

Figure 8: I really like this layout. Keeping five consistent boxplots is helpful across the different variables, and I appreciate the distributions below showing how they divide into the different boxplots. However, the discussion in Sections 4.4 and 5.2 would be strengthened if you could bring in some measure of statistical significance, e.g. Line 321- 323 "In the combined areas, the MRDs seem to decrease with increasing the Ksat values, except for the highest Ksat group, there are no significant patterns of MRDs when field areas are analyzed only." – how can you be certain there are no significant patterns without a statistical test? Perhaps look into notched boxplots as a starting place, but there are other possibilities here.

Thank you for the insightful suggestion. We agree that incorporating statistical significance would strengthen the interpretations in Sections 4.4 and 5.2, particularly regarding trends in MRDs across the physical variable groups.

To address this, 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.

We will update the text in Sections 4.4 and 5.2 to reflect the addition of these statistical tests and incorporated wording to clarify the observed trends in MRDs across the physical variable groups, based on both statistical testing and visual interpretation.

Section 5.1: To me this discussion is lacking critical engagement with some of the more complicated findings from this study. I'm not sure I agree that "It is clear from the results of this study and previous ones that both UAS SfM and lidar techniques provide a viable method for monitoring snow depth change across many land cover types." (line 358- 359) based on the SfM results in Figures 3 and 4 where the SfM depths are anywhere from 2-10 times larger than the in situ measurements. I doubt there are many applications where errors of this magnitude are acceptable. Additionally it doesn't seem feasible to rely on the SfM technique in forested areas based on all the missing data in Figure 4. Can you expand upon either of these? You briefly mention overcast skies possibly affecting SfM data collection (line 337) but this doesn't seem to explain why the SfM snow depths in the western field had much better agreement than the E and NW fields (Figure 4). What was the vegetation like in the fields? Was it fully buried by snow or partially extending above the snowpack? Could there be GPS/processing errors affecting the final results? Including individual photos from the SfM photosets could help illustrate some of the challenges.

Thank you for this feedback. The authors agree that the statement on line 358 – 359 did not accurately represent the findings of this study. We have updated the statement as follows:

"While it is apparent that the accuracy of SfM-derived snow depth estimates cannot match that of lidar, the results of this study indicate that both techniques provide sufficient accuracy for monitoring the median change in shallow snow depths over time in flat, unforested land covers when there are a sufficient number of unique characteristics for SfM post processing. It is clear from the results of this study and previous ones that compared to in situ data, UAS lidar techniques produce lower errors and fewer data gaps than SfM, especially in forested land cover and over homogeneous snowpacks (Bühler et al., 2016; Bühler et al., 2017; Harder et al., 2020; Revuelto et al., 2021b; Miller et al., 2022)."

Additional discussion on the limitations of SfM in the forest and over homogeneous snowpacks was also added to section 5.1. 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 field portion of the SfM map on 2/4.

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).

Section 5.2: Similar to a comment above, this section would be stronger if the relationships between physical variables and snow depth could be quantified statistically.

Thank you. We have updated this section to reflect the addition of the statistical tests to statistically quantify the observed trends in MRDs across the physical variable groups.

Lines 409-422: In the description of the in situ data collection you noted that one SWE sample was collected in each grid cell. Did you try any analysis with those measurements?

We determined the density and SWE and briefly looked at differences between the field and forest. However, because the snowpack was quite shallow, we believe that there was considerable error in those measurements and are not comfortable using them for research purposes.

Technical Corrections

Line 26: all areas → both areas? Complete

Line 88-89: Missing reference Corrected

Line 122: acronym IR not defined Corrected

Line 154: acronym CMOS not defined Complete

Line 182: remove superscript formatting from “antenna” Complete

Line 185-186: Typo in personal communication date? Data for this study collected in 2021 but personal communication listed as 2023 Corrected to 2020

Line 200: Missing reference Corrected

Line 200-201: “snow-off” Corrected

Line 231: Missing figure number Corrected

Line 242: Missing figure number Corrected

Line 277: Missing reference Corrected

Line 283: Missing reference Corrected