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 timeseries 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.

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?

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.

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).

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.

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

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?

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

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

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?

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.

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).

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

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.

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.

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).

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.

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

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

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.

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.

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.

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.

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.

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.

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?

Technical Corrections

Line 26: all areas → both areas?

Line 88-89: Missing reference

Line 122: acronym IR not defined

Line 154: acronym CMOS not defined

Line 182: remove superscript formatting from “antenna”

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

Line 200: Missing reference

Line 200-201: “snow-off”

Line 231: Missing figure number

Line 242: Missing figure number

Line 277: Missing reference

Line 283: Missing reference