Title: Advances in mapping sub-canopy snow depth with unmanned aerial vehicles using structure from motion and lidar techniques
Authors: Phillip Harder, John W. Pomeroy, and Warren D. Helgason

Paper Summary:

The authors show a comprehensive comparison between snow depth derived from UAV structure from motion and UAV lidar. They compare both datasets in forested areas, shrub areas, and in open/smoother terrain to manual snow depth measurements that are geolocated with GNSS systems. The authors show that UAV lidar can provide information beneath the canopy. This allows the user to look at snow depth variability and snow-vegetation processes with lidar. The authors clearly show issues with UAV SfM. The authors also nicely show a cost comparison stating that lidar is more accurate but costs ~15,000 dollars per additional cm of accuracy. The paper is well written and it discusses many caveats and issues that remain with lidar. The paper is a nice demonstration of the accuracy of UAV lidar, its utility, and remaining limitations.

The authors do not just evaluate the two techniques. The authors show how lidar can capture fine scale variability, such as tree wells, and detect fine scale processes with prairies. This shows originality and significance. I recommend the paper be published pending minor revisions.

General/Major Comments:

No major comments. Mostly, nit-picky comments. Enjoyed the paper, particularly Figure 7 and Figure 10 and their ability to capture tree wells and their changes throughout time.

Specific Comments:

Title sounds like a review paper. Perhaps consider something like, UAV lidar improves observations of sub-canopy snow depth variability over UAV SfM.

Line 7: I would disagree that techniques are lacking. You might say something related to that they don't always exist; satellite remote sensing is difficult. Airborne lidar captures this. So does TLS. This has been shown.

Line 26: Traditional remote sensing methods is vague. What's traditional to you might to be traditional to someone else.

Line 35: I would just say test processes

Line 38: I don't think Painter et al. 2016 initialized or validated a model. Andrew Hedricks recent WRR paper (Hedrick et al., 2018) would be better suited, which uses ASO data to update iSnobal (reinitialize).

Line 66: Leading to variably, I think you mean variability

Line 70: It would be great to reference (Currier et al., 2019) here. Table 1 in their paper reviews this and they provide their own evaluation metrics of ALS in a forest and open area. I would also reference (Mazzotti et al., 2019). They showed a comparison of lidar in Switzerland to snow depth transects in forested areas as well.

Line 75: TLS was used in the forest in (Currier et al., 2019). Yes, the TLS did not go all the way into the entire forest but from an evaluation perspective of airbone lidar or SfM there's little difference from being 300 meters in a forest as long as there are consistent trees overhead that would inhibit returns from the laser. Also, their paper did not explicitly show that TLS couldn't be used further in the forest, it just gets more complicated.

Line 90: Could add that (Zheng et al., 2016) lidar to understand vegetation processes effect on snow. They particularly note bias that might occur due to tree wells. (Currier & Lundquist, 2018) used lidar to understand the snow-vegetation interactions in multiple climates. (Mazzotti et al., 2019) also used airborne lidar data to improve the understanding of snow depth related to the forest in Colorado and Switzerland.

Line 190: I would mention here that the code is provided on your github page. Great job with providing this.

Line 205: Trees typically are taller than 50 cm. Most people consider a tree to be at least 2 m tall. Why did you choose 50 cm? This is inconsistent with what the caption shows in Figure 4.

Line 230: What is estimated and what is observed? I'd say UAV-derived Snow Depth and Snow Depth Probe Manual Observations, or something more specific.

Line 235: Yes, the reported error metrics are inflated when moving into the forest. It'd be worthwhile mentioning that the sample size is much less. Some lidar points do great. In the methods the GNSS mentions a ±2.5 cm accuracy, how was that determined. Is it possible that this is inflated when in the forest? If not, mention that. Are these errors from how the point cloud was processed and points were classified? Is ±2.5 cm true for both horizontal and vertical accuracy?

Line 238: I'd start a new paragraph when introducing the error metrics with SfM.

Line 245: The authors should be using Digital Terrain Models instead of Digital Surface Models throughout.

Figure 6: Cool analysis. I would consider adding a black dashed line for 2.5 cm. This plot supports the results of Currier et al. 2019, that the airborne lidar is more likely to penetrate the shrubs than the TLS observations. What's the scientific name for the shrubs found at these locations?

Figures: I would change the easting northing to the total number of meters within the domain, or start at 0 and show ticks from 0 m. I don't know the projection information, and if I did the numbers aren't that meaningful. If the location is important, please provide the UTM zone. But still it's a bit annoying to do the subtraction each time to get a sense of scale. I would just make it easier for the readers, if possible. Otherwise the figures are great.

Line 317: This seems like an appropriate time to re-mention UAV lidars ability to capture tree wells.

Line 321: Confusing sentence. Deems reported errors in the forest larger than 14 cm? Why is 14 cm mentioned. Figure 5 reports RMSE of 0.15 and 0.16.

Also, in the previous sentence. Studies have masked out the forest? Studies have looked at airborne lidar accuracy in the forest.

Line 355: Really cool figure and analysis

Line 375: Green polygons look cyan when zoomed out, might choose a different color. Furthermore, the near infrared data seemingly comes out of nowhere – maybe provide some more context within the section for it and why it needs to be mentioned. Provide a citation for NIR serving as a proxy for albedo.

Line 435: "The accuracy and resolution demands mean that bare surface classification techniques suitable for airborne platforms that efficiently resolve topography and hydrography at watershed scales from last returns will be unsuitable for resolving the snow depth around a particular shrub from a dense point cloud for example"

The paper did not show that using the last returns was unsuitable. The classification technique used something similar to last returns. Previous studies have showed using the last returns resulted in a generally unbiased snow depth estimate, and provided a reasonable approximation of the variability. I am not sure what this sentence is attempting to say.

Line 465: A discussion referencing the difficulties with modeling in Mark Raleigh's paper seems appropriate and a better citation then Tom Painters 2016 paper. Furthermore, when mentioning snow pack density variability, mentioning Karl Wetlaufer's paper seems appropriate (Raleigh & Small, 2017; Wetlaufer et al., 2016).

Line 479: "The UAV-lidar metrics consistently exceed the UAV-SfM metrics and are better than previously reported results in the airborne-lidar and UAV-SfM literature."

This isn't true. Metrics are similar but not better than. Please note line 69.

References

Currier, W. R., & Lundquist, J. D. (2018). Snow Depth Variability at the Forest Edge in Multiple Climates in the Western United States. *Water Resources Research*, *54*, 1–18. https://doi.org/10.1029/2018WR022553

Currier, W. R., Pflug, J., Mazzotti, G., Jonas, T., Deems, J. S., Bormann, K. J., et al. (2019). Comparing aerial lidar observations with terrestrial lidar and snow-probe transects from NASA's 2017 SnowEx campaign. *Water Resources Research*, 1–10. https://doi.org/10.1029/2018wr024533

Hedrick, A. R., Marks, D., Havens, S., Robertson, M., Johnson, M., Sandusky, M., et al. (2018). Direct Insertion of NASA Airborne Snow Observatory-Derived Snow Depth Time Series Into the iSnobal Energy Balance Snow Model. *Water Resources Research*, *54*, 8045–8063. https://doi.org/10.1029/2018WR023400

Mazzotti, G., Currier, W. R., Deems, J. S., Pflug, J. M., Lundquist, J. D., & Jonas, T. (2019). Revisiting Snow Cover Variability and Canopy Structure Within Forest Stands: Insights From Airborne Lidar Data. *Water Resources Research*, *55*(7), 6198–6216. https://doi.org/10.1029/2019wr024898

Raleigh, M. S., & Small, E. E. (2017). Snowpack density modeling is the primary source of uncertainty when mapping basin-wide SWE with lidar. *Geophysical Research Letters*, *44*(8), 3700–3709. https://doi.org/10.1002/2016GL071999

Wetlaufer, K., Hendrikx, J., & Marshall, L. (2016). Spatial heterogeneity of snow density and its influence on snow water equivalence estimates in a large mountainous basin. *Hydrology*, *3*(1). https://doi.org/10.3390/hydrology3010003

Zheng, Z., Kirchner, P. B., & Bales, R. C. (2016). Topographic and vegetation effects on snow accumulation in the southern Sierra Nevada: A statistical summary from lidar data. *Cryosphere*, *10*(1), 257–269. https://doi.org/10.5194/tc-10-257-2016