Referee comment of egusphere-2025-1297 manuscript entitled "UAV LiDAR surveys and machine learning improves snow depth and water equivalent estimates in the boreal landscapes" by Ylönen et al.

General comments

This contribution by Ylönen et al. provides a comprehensive assessment of UAV-based LiDAR combined with machine learning techniques for improving snow depth and snow water equivalent (SWE) estimations in boreal landscapes. Data collection was carried out in two study sites in Northern Finland. Overall, the study is scientifically robust, methodologically sound, and contributes some novel insights into snow hydrology and closing the observation gap between local, in-situ and regional-level snow depth and SWE mapping. It successfully integrates high-resolution UAV LiDAR data with ground-based measurements and advanced clustering methodologies to generate spatially detailed snow characteristics. Overall, the manuscript is well-structured. However, to enhance comprehensibility, some sections could be revised (see comments below). I also found several typos and formulations that should be double checked by a native-speaker / typesetter. Some sections are somewhat difficult to understand due to a convoluted and partially unclear sentence structure. Figures in the manuscript are generally well presented and organised, captions are clearly written and comprehensible.

Specific comments

I realise there is up to now no operational implementation of satellite-based snow depth monitoring, but I feel this point should be further elaborated on in the introduction, rather than just mentioning it in passing.

I missed a mention of the publications on ALS-based snow depth monitoring in the introduction, e.g. the ASO (Deems et al.) – feel this should be included here.

The terms ‘snow course network’ or ‘snow course measurements’ is maybe slightly misleading – in-situ snow depth measurements? Also, I did not find any reference to AWS in the introduction. At least in the Alps, AWS are traditionally the main drivers of spatially explicit snow depth maps – maybe something that could be included in the introduction too.

Overall, I would advise to outline more clearly, how the integration of UAV LiDAR with machine learning significantly improves upon traditional remote sensing methods. This improvement could be emphasized more explicitly to distinguish the presented approach from previous methods.

Regarding the comparison with the GCPs: In which study area was this comparison carried out? I take it the GCPs where present for all flights during a given campaign, as the May campaign is stated as being the one with the poorest results? For anyone not familiar with the mentioned processing methods, a brief explanation of the routines and the employed methodology would be helpful. Were multiple echos available for processing? How come the R-methods were outperformed? Also: Part of the paragraph, where this is first described (2.2.1) is redundant to part of chapter 2.3.1. Structure-wise it would seem clearer to me, that the method-section of 2.2.1 was moved to the ‘data analysis’ section below. Regarding the issues with the overlapping points clouds mentioned in the former section, I feel this could also better be moved to and discussed in the analysis chapter.

Adding to the comments on chapter 2.2.1 above, what was the point cloud density of the recorded ULS-scans? What kind of accuracy is stated by the manufacturer? Generally speaking, the term DTM is used somewhat confusing to me in the manuscript, since it usually refers to the terrain, i.e. surface of the bare ground or vegetation. Thus, I suggest either using the term *DSM* for the snow-on datasets and DTM for snow-off, or the generic term *DEM* for both.

In 2.3.1: How was the choice of the parameters for ground classification made? Was this same parameter set used for all campaigns, i.e. were both bare ground and snow cover identified with this setting? I guess at least the parameter 'steepness' would change with the snowpack build up and terrain features being smoothed out?

In chapter 2.3.3 I'm having some trouble grasping the method – I advise restructuring and rephrasing this section to improve clarity. Overall, it remains unclear to me how the k-means and random forest methods were connected and how the random forest model was parameterised.

In chapter 3.1 (linking to the comments above on 2.2.1), I take it the accuracy of the RTK GNSS measurements was recorded? This could give some indication, whether the varying accuracy is in fact the cause for poorer May results as indicated in the manuscript.

Generally, I suggest providing justification for selecting three clusters despite methodological indices suggesting varied optimal numbers. Clarify the criteria beyond "simplicity and comparability" for site comparisons.

In the results chapter, the discussion of errors, particularly for the May campaign in Sodankylä, could be expanded. How exactly do flooding conditions influence LiDAR returns? Provide more explicit details or references that explain why these conditions cause significantly larger errors.

The discussion chapter could more explicitly show how the presented clustering method generalises to other boreal or similar ecosystems, particularly considering interannual variability. Are there landscape conditions or climatic contexts where this method might face significant limitations?

Consider expanding on how the presented findings specifically support operational hydrological forecasting and climate adaptation strategies. Providing more context on potential applications would strengthen the impact of the manuscript.

Technical corrections (non-exhaustive)

- Title: Maybe the title could underscore the connection between ULS and ML clearer, e.g. '<u>Combining</u> UAV LiDAR surveys and machine learning…'
- Line 13: regions <u>are</u> experiencing
- Line 18: and <u>a</u> machine
- Line 26: patterns ~~in~~ <u>at</u> the
- Lines 29f: management<u>,</u> ~~and~~ offer<u>ing</u> new
- Line 30: forecasting<u>,</u> ~~and~~ climate
- Line 34: ~~in~~ <u>at</u> high latitude<u>s</u> and <u>in</u> mountainous
- Line 35: cover~~, the~~ timing
- Line 35: distribution <u>directly influences</u> ~~on~~ climate
- Line 42: this passage could be more specific re the mentioned impact on the snowpack (i.e. volume, extent, snow depth, season?)
- Line 44: flood monitoring/early warning/prognosis (?)
- Line 52: However, <u>for the</u> accurate
- Line 67: re lighting conditions, maybe add that this applied to snow in particular
- Line 83ff: Not sure I understand the point that is being made here
- I'm confused, why the manuscript sometimes mentions two (e.g. line 107) and other times three study sites/areas (e.g. line 117).

- Table 1: Instead of ‘LiDAR extent’, maybe ‘Area mapping with ULS’ → potentially a good move to introduce ULS (UAV Laser Scanning) as an acronym in the manuscript for brevity (see related publications)
- Figure 2 caption: ~~UAV drone~~ ULS
- Line 175: ~~GPS~~ GNSS
- Throughout: Ensure consistent use of abbreviations (e.g., SWE) once introduced to avoid unnecessary repetitions.