

Review of „*Identifying Snow-Covered Areas from Unoccupied Aerial Systems (UAS) Visible Imagery: A Comparison of Methods*“ (Moradi et al. 2025)

General Comments

In their study, the authors present a method of mapping snow cover using drones with different sensor systems and land surface classification algorithms. The study's primary objective is to conduct an inter-comparison experiment to evaluate the performance of various snow classification methods in relation to different acquiring sensors, study sites, and training data availability. This is an interesting setup because it addresses common challenges associated with drone surveys and aims to provide recommendations on the most suitable methods for specific acquisition conditions or user applications. The manuscript is well written and mostly well structured. The scientific objects are within the scope of The Cryosphere.

However, I think the analysis of the differences between the methods is a bit shallow. On the one hand, the experimental setup is relatively simplistic (i.e. binary classification of “snow” vs. “no-snow”), but on the other, it may be flawed by some uncompensated differences in data acquisition (image dates), as outlined in my general comments below. Therefore, I believe that major revisions are necessary before this study can be published. Nevertheless, expanding the experimental setup could provide a much deeper understanding of the differences between the methods and their underlying causes.

- **Experiment images:** The dataset used is certainly interesting, covering a wide range of snow cover conditions. However, a significant limitation of the cross-sensor and cross-site comparison is that the images were not acquired on the same day. It is likely that there is some kind of date-related bias between the images because the properties of the snow, the illumination and the shadows are likely to be different. Ideally, images from different sensors at the same site would have been acquired on the same day, immediately after each other. As far as I can see, these potential sources of difference are not currently accounted for in the experiments. Were the images acquired at the same time of day? What about weather conditions (e.g. cloudiness)? To what extent are the observed cross-sensor and cross-site differences actually related to acquisition dates?
- **Experiment sites:** Similarly to the selection of images, I am not entirely convinced by the study area to which the image mosaics are masked (Figure 1). It looks like you have removed all areas except for the almost flat fields with sparse vegetation. While this certainly improves the separation of snow and bare ground, it means that you cannot draw any conclusions about the performance of the applied algorithms in more complex terrain, such as slopes and forests. Conversely, I suspect that the overall good performance of your classifiers (almost all approaches yield reasonable results) is due to the exclusion of challenging areas. This is made even more apparent by the fact that, in your manuscript, you mention that satellite snow cover struggles in complex terrain or forest areas compared to drone surveys (Line 59), yet your own dataset is restricted to areas where a satellite image would also most likely produce good classification results (at a coarser resolution of course). I think this is a missed opportunity, as your analysis could be much more informative if you assessed the accuracy of different methods in different landscapes.

- **Data availability:** Your data availability statement only refers to meteorological data which is available elsewhere. What about your survey data? The raw images and your intermediate classification results should also be made available via an open data repository.

Specific Comments

- Line 50-52: „Unoccupied aerial systems (UAS) are emerging as a reliable, low-cost method for collecting snow data at finer spatial resolution than satellites and are not limited by orbital revisit times.“ => While it is true that drones provide much better resolution than non-commercial satellite images, the logistical effort required depends on the use case. For example, the surface reflectance products of the Landsat/Sentinel missions have a revisit time of a few days. They are distributed free of charge by the ESA and NASA, and they include infrared channels, which are great for mapping snow. Conversely, drones require an operator in the field, surveys of large areas take time, and they are limited by weather conditions. So, all in all, there are specific applications where one system is superior to the other. If you want to compare satellites and drones in the introduction (which makes sense), you should briefly summarise the advantages and disadvantages of drones and satellites, and which data is useful for which application.
- Line 57-60: “... does not appear impacted to the same extent as satellite snow cover products in forested areas, complex terrain and shallow snow conditions ...” => add brief explanation why there is less impact
- Line 164-192: You provide a lot of technical details for each drone/sensor system. Instead, why not add a table that summarises the technical differences between the systems, such as sensor specifications, camera settings and flight line parameters? You could then focus the text on the key differences between the systems. However, if you decide to stick with the current version, that is fine too.
- Line 176 & 189: Please specify which DEM was used, and whether the same DEM was used to process each dataset.
- Line 208-209: “... most images collected during complete snow cover could not be effectively stitched together and were excluded from this study.” => I would imagine that it is not a significant issue for the purposes of your study, as separating snow from bare ground in images with continuous, thick snow cover is likely easier than in cases with thin, fractional snow cover (as you demonstrate later). Nevertheless, for an actual drone mapping campaign to monitor snow cover evolution, this would present a significant challenge, as the creation of orthomosaics would fail for certain acquisition dates. At the very least, I would mention in the discussion that your test dataset is not representative of complete snow cover applications.
- Line 209: What approximate snow depths do you mean by “uniform snowpack”?
- Line 220-222: “Field cameras were installed following the method used in NASA’s 2020 SnowEx field camera campaign ... “ => You refer to personal communication here, but do not explain what was actually done. Could you please provide a description of the approach that was taken?
- Figure 2 & 3 & 4: The images are relatively small, and it is not always easy to tell the difference between those with similar snow conditions. I suggest moving these figures to the

supplementary material. In the main text, you could show a figure with fewer, larger panels instead. For example, select one or two images with distinctly different snow conditions for each site and each camera. This would allow readers to compare the differences between sites and cameras with and without a large snow cover fraction.

- Line 250: "... the field area of interest process was isolated ..." => What do you mean?
- Line 251: "... forested areas was removed ..." => see my general comment about selection of areas
- Line 287-289: It is tedious for readers to have to look up the threshold value in the supplement. If the value is fixed, just mention it here and refer to the supplement for details.
- Line 302f: The results section is divided into subsections for each experiment. I suggest doing the same here, creating subsections for the 'base experiment', 'second experiment', and so on.
- Line 302f: To improve readability, you could replace "base experiment", "second experiment", "third experiment", "fourth experiment" with a more (mathematically) name convention, e.g. $SCA^{reference}$
- Line 404: "... is rooted in differences in snow conditions." => see my general remark about image dates and conditions
- Line 440f: In the conclusion you state that 12 images are the minimum requirement for a good training set (based on Figure 8?). Please elaborate why you conclude that 12 images is the threshold.
- Line 521: "... identify one or more classifiers that can rapidly ..." => Remove "rapidly" because you do not analyze processing speed.
- Line 685-688: How do your results validate "satellite snow cover products, converting satellite NDSI observations of fSCA, and downscaling satellite SCA observations"? Explain this or delete.
- Line 705: Data availability => Drone images and classification masks should be made available to reproduce results.

Technical Corrections

- Line 62: Here you could add a reference to large-scale snow cover products based on NDSI (e.g. <https://essd.copernicus.org/articles/11/493/2019/>)
- Line 67: replace „blue-channel thresholding“ with „blue-channel“ reflectance (or similar)
- Line 159: There is no citation of MODIS data.
- Line 226: missing reference „NOAA’s publicly available archive (ADD-REFERENCE)“
- Figure 1: The panel on the bottom right appears to be a Google Earth background. I think there should be a reference to this somewhere.
- Figure 5 and other bar charts: Do not include both red and green bars in the same graph.