Comment on tc-2022-191
Désirée Treichler (Referee)

Referee comment on "Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data" by César Deschamps-Berger et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-191-RC2, 2022

The authors present a study where they assess the potential of spaceborne laser altimetry from ICESat-2 to measure snow depth when compared to snow-free digital elevation models (DEMs) from various sources. These are compared to snow depth maps from airborne lidar data acquired close in time as a verification of their results. The study focuses on an extensive analysis of the residuals between the two datasets including considerations about spatial resolution, forest cover and surface slope. Accurate methods to measure snow depth from space would meet a great need within snow science. This study thus has the potential to be very useful for the snow community.

However, the manuscript currently has two main issues (outlined below) that limit its usefulness: 1.) With the focus on residuals, there are barely any actual snow depth results - which makes it rather too difficult for potential users of this novel data/approach to know whether this is useful for their purpose; and 2.) some of the methods/processing developed for the ICESat-2 approach leave a somewhat unfinished impression, i.e. some changes may greatly improve the quality of the snow depth measurements from ICESat-2. I suggest that the authors review their article with a focus on these two aspects.

This review was done in collaboration with a Ph.D. candidate in training who is specialising in this field.

Main/general comments:

1.) referring to the lack of snow depth results: As a reader, I would have expected an analysis of the actual snow depth measurements retrieved from ICESat-2 data, not only an analysis of the residuals with the validation dataset(s) where the snow depths
themselves are no longer visible. For example, users may want to know whether a single ICESat-2 ground track crossing their catchment can give them any useful snow depth data. You could thus provide a plot of a snow depth transect, or plot the snow depths of one overpass on a map. Given that individual measurements rather seem too inaccurate to be useful, users may wonder whether the ATL06-derived data accurately reproduces how snow depths vary in elevation or for different aspects. This would provide a much more approachable basis for the thorough analysis of the residuals that you already have.

2.) referring to methods/processing, this includes several points:
- from Fig. S2 it seems that coregistration between the DEMs and ICESat-2 was incomplete, with the remaining shifts causing elevation errors that are far larger than the assumed snow depths. The ASO DEM seems to be the most carefully curated and accurate reference DEM in the area and may be the better coregistration basis than ICESat-2. Given that all elevation datasets are patched together from several individual acquisitions, your results may improve greatly if you coregister smaller spatial tiles/units individually. You only analyse the residuals of the snow-on data points - how does the distribution look like for the snow-off data? Similar, narrower, wider? Do you see the same bias for snow-off points and snow-on points from the same overpass (if there are overpasses with both snow-on and snow-off data)?

- ATL06 data is an averaged product of 40m long track segments with photon returns spread ca. 11m in width (across-track). It assumes flat glacier surfaces, as you write. Although the segments are overlapping, resulting in 20m spatial sampling, I find it problematic to translate this into "20m resolution" given the potentially strong terrain differences within the 40x11m segments. You analyse the effect of the resolution of your snow-off DEMs, and by this you mean the cell size of the pixel the ATL06 centre data point happens to fall on - but one would assume that a correct representation of the ATL06 footprint including all DEM cells would make a much greater difference for the resulting snow depths, especially with a high-resolution snow-off DEM such as the ASO DEM? When analysing the effect of resolution, I would thus expect that you rather assess/also include the latter approach. In my opinion, this would strengthen the usefulness of your resolution analysis considerably.

- Among other methods you also assess the quality of snow depths derived from ICESat-2 data only, i.e. snow-on minus snow-off tracks. Not depending on other data sources would obviously be very useful but I have great doubts whether the approach that you chose in this study is sound, given the rough terrain in your study area. The ATL06 data are already spatially averaged over a 40m segment, and by converting the ICESat-2 data to a grid with a nearest neighbour approach, you effectively shift the ATL06 data around in space. To get snow depths, you again use a nearest neighbour approach (L126) with the same effect. The sum of this results in very high elevation uncertainty. The approach also seems to multiply the area in which ICESat-2 provides snow-off data by a factor of ten (25 km² snow-off data from 59% of the points, 1.8 km² snow data from 41% of the points), which is puzzling. This approach may be justifiable in polar areas (where tracks are repeated nearly exactly) and possibly also flat prairie landscapes but it performs very poorly in your area, as you show yourself, making it not convincing. It thus rather weakens the impression of the study as a whole and you may consider removing this part in this study (and rather publish it for a study area where it can be better justified?). If you want to include ICESat-2-only snow depths you could instead focus on cross-point comparisons for snow-off and snow-on tracks, as has been done extensively for ice
sheet/glacier changes for the predecessor ICESat data. Repeating that approach would be easy to justify and testing it may be appreciated by the community.

Minor/specific comments:

P1, L15: you choose the terms accuracy and precision for bias/uncertainty in your document. That's OK, but note that this may be less intuitive for some readers compared to other terms (uncertainty, bias). Please ensure that this is consistent/translatable throughout your document, e.g.
- mean absolute error for other datasets (L42f)
- bias / precision for your results (L184)

Introduction

P1, L20: seasonal snow / snow pack / snow mass (not cover)

P1, L23ff: Consider rewriting, as it is not clear from the context where the mentioned assimilated remotely sensed snow depths were coming from, and the second part of the sentence could be more logically connected.

P1, L26: Are these global efforts? Or US-focused? unclear

P1f, L29ff: Starting from here, the entire introduction of DEMs and how to translate them into snow depths would benefit from a more structured/complete approach to improve reading flow. For example, you could first introduce all the methods/carriers and then move on to examples (like the ASO). Consider:
- accurate DEMs can also be acquired from satellite data (you mention only airborne) or drones. Satellites pop up later (L36f) only.
- make sure the readers understand that the Airborne Snow Observatory is a US thing, not a globally available dataset.
- not all readers may know that you essentially mean the same thing with photogrammetry (L29) and stereoscopy(L36)

P2, L40ff: this is a very different method, consider a new paragraph. Rather use the term backscatter, not observations.

P2, L48ff: I suggest you add that the method by Treichler & Kääb (2017) required both spatial and temporal averaging, i.e., ICESat was best suited for average winter snow
depths over several years.

P2, L52ff: Consider re-structuring the introduction of ICESat-2 & specs and the level of detail that is suitable here or in a later section where you also introduce sensor/data parameters. You may also consider pointing to a figure that shows the ground track layout/availability (could be a map like in fig. 1). Some suggestions:
- L53: ...a strong and a weak beam EACH...
- L56: 20 m spacing, but 40 m spatial resolution - an important difference for rugged terrain
- ATL08 includes land surface and canopy heights (both)
- L60ff: the logical connection is unclear. I recommend to introduce the strategic off-pointing of ICESat-2 ground tracks earlier (together with the beam design): be very clear that the tracks are not repeated exactly, on purpose. Consider that most readers may not be used to the concept of profile data rather than spatially extensive maps/imagery, and they may also not find it obvious why you need a reference snow-off DEM.

P2, L63: New topic, start with "In this study" or something to help the reader.

P3, L67: unclear where the upper Tuolomne basin is -> map figure (see other comments)

Materials and Methods
--
- Data, study site and processing steps are a bit mixed in this section. I suggest you try to re-structure this and reconsider the sub-section titles so that they are representative for the respective contents. For example, it could be useful to introduce all important aspects/parameters of the ICESat-2 data first - and possibly all other datasets, too - and move all processing steps (L106ff) to a different sub-section.

- It would be very useful to add a map of the study area where you could include standard map layers and coordinates to show the location of your study site, the extents of all your datasets (now missing especially for Pléiades, L102), forest cover, elevations etc. You may want to embed this in a paragraph where you describe the study site (now spread across introduction and methods). The one sentence at L82f currently lacks context.

P3, L75f: just somewhat beyond - or is it globally available? Why "sparse" - more sparse than elsewhere? Do you mean mountain "ranges"?

P3, L78f: 40m LINEAR segments? "land-surface photon returns": you may want to specifically mention that "land-surface" includes vegetation/forest photons as these will bias the mean elevation within the segment - something you later address in your forest analyses.
P3, L83: "granules" unclear

P3, L85: Explain what "sigma_h_mean" and "n_fit_photons" represent and where you get these parameters from

P3, L86ff: "...as snow and ice are highly reflective...": not a logical continuation of the sentence, needs a clear explanation. You may link this to the strong/weak beams mentioned earlier. Your snow on/off classification is a smart idea but the description here needs to become clearer.
- number of photons - per what?
- source/reference and resolution (spatial/temporal) of this dataset are missing. Please introduce it properly.
- explain what NDSI means, and justify your chosen threshold
- Please explain more clearly how you get from the photon and MODIS data to the photon threshold through Cohen's kappa, the readers probably don't know this approach. The curves in the supplementary don't help much there (how were they made? Adding a scatter plot with the number of photons vs MODIS fSCA may help).

P3, L96ff: See also main comment. Why 15 m (20/40m may be more logical)? A map with snow-off (and snow-on) points would be useful. Is the point2dem utility meant for a dataset of profiles with that many gaps in between?

P4, L103ff: What do you mean by "extracted"? Filled with miscellaneous external products: is this the case for your study area?

P4, L107: "the exact spatial scale of the ATL06 points remains uncertain": what do you mean? Please explain.

P4, L122: Did you harmonise the vertical reference systems of the datasets? Please mention this step in the processing since a universal vertical shift will not reconcile elevation differences due to different ellipsoids/geoids for larger areas, and users of your approach should be aware of this. The high values for the Copernicus DEM in the table in the appendix suggest that this dataset may be based on a different ellipsoid or geoid?

P5, Table 1: Note that the spatial resolution of the ATL06 data is 40 m, not 20. If included at all in this table, this dataset may need to be its own category (separated by a line?) given that it is not gridded and includes both snow-on and snow-off data. Also, the generated ICESat-2 snow-off DEM is missing in this table.
P5, L140ff: Is this really true for photon-counting lasers? ICESat was a full waveform sensor which is a very different system. The ICESat-2 ATL03 (photon) product has a spatial positioning uncertainty of ca. 5m for individual photons, and the along-track resolution is only 0.7 m - and in the case of snow, there are often several photon returns for each shot (i.e. several per 0.7 m). The resulting photon density represents terrain variations one order of magnitude smaller than the spatial resolution of the ATL06 product.

P6, L147ff: What does this dataset contain? Land cover, forest..? Could be introduced together with the MODIS snow dataset.

Results

P6, Fig. 1: This is the only figure that shows snow depths - for only one of the studied tracks. From the title and abstract I would have expected more.
a) this map is not enough to characterize the study site (but would work for introducing this particular overpass if there was a proper study site map, see comment above). Why are the ICESat-2 tracks cut in the North and West? What is the extent of the snow-off DEMs?
b) what is the regression line of this plot? It does not seem to be 1:1
c) axes labels? Does this include only data points (grid cells) where you have data from either dataset? There seems to be a lot of snow-free data points in the ASO data, does that mean that these were classified as snow-covered in your ICESat-2 dataset? Is this plausible?
d) axes labels? It would be useful to label median and NMAD (or accuracy and precision / bias and uncertainty). How does the distribution of the residuals of the snow-free ICESat-2 data points look like in this area or for the same overpass?

P6, L159f: the figure in the appendix doesn't tell how robust the photon threshold is.

P7, L165ff: Are these numbers for snow-on or snow-off data points?

P7, Fig. 2: the labels for the panels are swapped. What do the vertical histograms actually show - the distributions of the number of photons contributing to all the different ATL06 data points? Please add this in the caption. What unit is the MODIS snow cover area?

P7, Fig. 3: The data in this figure is barely readable as the very narrow data stripes disappear, both digitally and in print. The elevation colour scale is indistinguishable with the dark grey background. If you want to show the location of the snow-off points, it may be better to provide a scatter plot (with larger markers). Coordinates, map layers etc. missing.
P8, L178ff: Readers might still think 12 days is rather long given the large variations in elevation/aspect/vegetation cover in the study area. You could provide met station time series in the supplement.

P8, L184ff: Are you comparing the same data points for all datasets, i.e., only the area where you also have the Pléiades reference DEM? Otherwise the bias/precision values may not be directly comparable and could be an artefact of the different spatial sampling with different terrain/forest/elevation characteristics.

P8, L188: The figures in the supplement indicate that the poor performance is caused by a persistent (and spatially consistent?) spatial shift that looks like it could be removable with the Nuth/Kääb approach (?). Give it one more try with a tiling approach? Likely, many readers would have greatly appreciated if this approach worked with the COP30 DEM, as many readers might not have a snow-free DEM of ASO quality available.

P8, L205: It would be interesting to discuss the slope-dependance of the performance of the different datasets - especially because it differs for the different datasets (increasingly positive bias for ASO, increasingly negative for Pléiades. I could not find this in the discussion section). Any ideas why it is different for these two datasets?

P9, L215f: The mode requires binning of the data. What is the bin spacing in your case? Coarse bin spacing may result in remaining vertical shifts/bias.

P9, L226ff: you introduce several new parameters and an entirely new processing aspect here, and it's difficult to understand why you did this and what effect it would have had on the results. This should probably be moved to the methods section (and the parameters properly explained).

P10, Fig. 4: Why are some boxplot classes missing in some plots? E.g., b) IC2-IC2 has only five boxplots. You chose to plot the individual data sets according to the delay between ICESat-2 and ASO acquisition, but there seems to be no detectable signal/dependency on the delay, and this is nowhere discussed. Personally, I would have preferred a timeline or any other way of arranging the boxplots that shows the acquisition dates, as the timing within the snow season provides a lot of meta information - it affects which elevations (and thus vegetation cover, slopes etc.) are included, and whether this is during the accumulation or melting season with their rather different snow pack.

L240: You wrote earlier that the Pléiades DEM is not covering the entire study area?

Discussion
P11, L260: What leads you to the conclusion that the reason for your unsuccessful coregistration is the "coarse" pixel size of 30m? The spatial resolution of the dataset should not matter for the coregistration approach you used (Nuth/Kääb with interpolated elevations) - you could test this with a deliberately shifted 30m version of the ASO DEM. I am convinced the reason for the poor performance lies somewhere else than in the spatial resolution of the dataset.

P11, L267f: "This approach could not be applied in the general case" - unclear, what do you mean here?
What do you mean by "sufficient accuracy"? Resolution? Or would the data be better in Europe? How about testing other US/global DEMs (e.g., the SRTM DEM)?

P12, L279: Shean et al. (2022) is missing in the reference list

P12, L288ff: This analysis shows that the ICESat-2 tracks are not representative for your particular catchment (which is fairly large). But are the ICESat-2 data points accurately reproducing the snow depths/elevation gradient of the corresponding locations/pixels in the ASO map? If yes, then the ICESat-2 data may very well allow an accurate estimation of the snow volume in the catchment, given a smart spatial extrapolation.

P12, L296f: "Each ICESat-2 ATL06 snow depth point is informative over a small sampling area...": I disagree with this conclusion - see main comment about the resolution of the data. In addition, the snow depth uncertainty of individual ATL06 data points seems far to large for such a statement.

Conclusions and subsequent sections
--
P13, L310ff: This sentence is hard to understand (in particular because a "while" connects the two parts). Please rephrase.

P13, L317: there is no author with initials NLM. JLM?

P13, L335: this publication is from 2022, not 2021