RC 1

The authors present a comparison of different downscaling methods to extract hyper-resolution snow cover information from widely available but (much) coarser spaceborne observations of fractional snow-covered area. The comparison is based on three published methods of increasing complexity and input-data demands, investigating the added benefits (and limitations) of using commercial data. Their results suggest that the most complex downscaling approach (based on a random forest model) generally outperforms simpler, either solely terrain- or climatologically driven methods when estimating snow cover evolution at a 3 m resolution. However, for a target resolution of 30 meters or more, these simpler methods that do not rely on additional commercial data may be sufficient. While the study focuses on relatively small alpine meadow domains, it can provide guidance on selecting appropriate downscaling methods depending on the intended application and data availability. The manuscript is generally well written: the methodology and results are presented comprehensively and in a logical order, and the Discussion addresses the relevant open questions and potential limitations of the presented findings.

My only general remark concerns the figures and their introduction in the text. First, the manuscript would benefit from guiding the reader more directly through the results, e.g., by explicitly introducing each figure and explaining what it shows. Currently, several figures are referenced only in passing, as a bracketed reference at the end of a statement. Moreover, not all subpanels (e.g., the right panel of Fig. 9) or supplementary figures (e.g., Fig. S13) are discussed in the text, and Fig. S20 (L567) is referenced before Fig. S19 (L604). Second, several figures (e.g., Figs. 1, 2, and 7) are small and densely packed, so that text, tick labels, and colorbars are difficult to read without zooming in considerably. I would therefore suggest enlarging these figures and increasing font sizes, splitting or removing less central panels where necessary (also consider whether all 22 supplemental figures are necessary), and ensuring that every panel shown in the main text is also discussed there.

> *We agree that the manuscript was often missing the explicit connections between the text and figures that were important for reader comprehension and the communication of the results. We will make revisions to both the text and figures to ensure that unnecessary information is not presented and connect the findings with the specific portions of the figure that demonstrate the result we discuss, especially when each figure is first introduced. We will also ensure that the figures are edited to make text and information more visible, reorder the supplementary figures, and only include the supplemental figures necessary to support our major findings. For example, while we find the comparisons of fSCA time series (Figures S10 – S12) and comparisons across land cover types (Figures S16 and S17) interesting, these had little relevance to the manuscript findings and should be removed.*

Overall, this is a relevant contribution to the field suitable for publication in The Cryosphere after minor revisions. I'd like to compliment the authors on their work. Please find my specific comments below, some of which may reflect personal stylistic preference.

We want to thank you for your detailed feedback. It was clear that considerable time and effort was dedicated to this, and we are excited about the improvements that will result from addressing your comments.

While certainly a stylistic choice, in my opinion, the use of "1) …, 2) …, and 3) …" makes the text more difficult to read in places.

- L151ff: The references to the corresponding subsections already separate the three items sufficiently. Moreover, the numbering conflicts with the Abstract, where 1) is the probabilistic method and 2) the RF approach.
- L190, L218f, L254ff, L478, L489, L563: With only two elements, numbering is unnecessary. At L190, if kept, it should read "resolved 1) the spatial pattern …, and 2) the overall bias …".
- L332ff, L640: The numbering interrupts the flow of the text rather than adding clarity.
- L424ff: Listing the core findings seems appropriate here, but consider "first, second, and lastly" instead.

  We agree that numbered lists were overused and interrupted the flow of the text. We will remove all the lists that you identified here except for L332 and L424. We are deciding to keep these since we believe the use of these lists more clearly highlight and separate the number of degrees of freedom influencing the accuracy of the downscalings, and the major takeaways.

Please avoid to have neighboring closing and opening brackets "(…) (…)" as in L121, L140, L141, L152, L214, L253

These double brackets will either be combined or removed. For example, "(ASO)(Painter and Bormann, 2020)" will become "(ASO; Painter and Bormann, 2020)", and "(Sect. 3.3) (Fig. 2)" will become "(Sect. 3.3 and Fig. 2)".

L12: Characterizing the fine-scale snow distribution as "… important for avalanche forecasting and biological refugia" might be too limited. Consider adding snow hydrological processes (and others you find relevant) as you did in the introduction.

Snowmelt onset timing and permafrost degradation will also be added to this list. However, because we are limited by the word count in the abstract, we plan to keep the mention of these applications and processes brief.

L20: Consider rephrasing the sentence "The approach used to …" as currently it is formulated slightly confusingly.

Thanks for pointing this out. We agree. "The approach used to downscaling influenced 3 m snow cover most" will be revised to "The downscaling approach influenced 3 m snow cover most".

L104: If "model architectures" here refers to the downscaling approaches, consider rephrasing (as it is the only location in the manuscript where this phrase appears).

Otherwise, please explain it to avoid ambiguity.

Yes, this refers to the downscaling approaches. We agree that this was ambiguous, and it will be revised to "different downscaling approaches".

L106ff: In contrast to the following paragraph, I could not find any information about the (average) revisit times or (average) number of observations per season for the PlanetScope data.

Thanks for pointing this out. We agree that this information is necessary for a fuller comparison versus the other remote sensing data sources in the following paragraph. We will add text to establish that due to the large constellation of PS cubesats, the frequency of observations is approximately daily. However, due to obstruction from cloud cover and retrieval quality errors like sun glint, Pflug et al. (2024) found that the exact date of snow disappearance was certain to within 3 days, on average, in these domains.

L116ff: Consider already hinting at the leave-one-out approach here. While you introduce the details below, the question about independent validation immediately came to mind while reading this sentence.

This is a great idea. Without this, it makes the training and evaluation versus PS discussed here seem unfair. We will add text to the first sentence of this paragraph to establish that we will use a leave-one-year-out approach where we assume the PS observations are available in all years except individual years withheld for model evaluation.

L125ff: Again, consider hinting at the method used to convert HLS data to fSCA here. Even though it is explained a few lines below, this sentence/paragraph left me wondering.

We will add a sentence to the end of this paragraph to specify that we used the NDSI calculation and the same NDSI to fSCA conversion used by MODIS and VIIRS for calculating 30 m fSCA.

L138: Consider removing "Finally", as it implies a sequential order of the data, which was not apparent to me.

"Finally" will be removed.

Figure 1: The figure is rather small, especially the text/sources in DAN and the tick labels. Removing the border outside the model domains might also improve visibility.

Thanks for the suggestion. We will revise this figure to increase the tick label size substantially. We will also zoom the domain extents to the white boundary (eliminate the outside border) and reformat the basemap attribution text to make it more readable.

L155: "in detail" is written without a hyphen.

The hyphen will be removed.

L162: Consider adding the corresponding reference (probably Pflug et al. 2024) to support the statement about the "high accuracy of these snow cover maps". Do the same for the second sentence in the Conclusion (L634ff).

Yang et al. (2023) is probably the most appropriate citation for both of these

passages. This citation will be added in both cases. We will also review the full manuscript to ensure that this citation is included in instances where we reference the high accuracy of 3m snow cover classifications.

L164ff + all equations: Multi-letter variables, e.g., FN, could be misunderstood as multiplication, like $FN = F \times N$. Consider reformatting these, e.g., using LaTeX. This applies to all equations and most variables (except k, z_0, $\underline{z}$, w) in your manuscript. At any rate, at least ensure consistency in variables, e.g., by also using italicized FN in the text and not only in the equations, using either $\mathrm{F1}_{\mathrm{score}}$ or "F1 score", etc.

This is a good point. We will revise the equations using variable formatting to ensure that combined-letter variables like "FN" aren't mistaken for multiplicative products. We will also revise their references in the text as you suggest (e.g., changing F1 score to $\mathrm{F1}_{\mathrm{score}}$*).*

L165ff: This paragraph could benefit from some improved phrasing and additional explanation.

- "For each date, the F1 score is calculated twice, …" could be misinterpreted as "two times a day" rather than two different values, one for "snow presence" and one for "snow absence".
- Consider rephrasing "the number of snow present and snow absent cells", e.g., to something like "the respective number of cells with and without snow".
- To me, the sentence "The resulting F1 score varied between 0 and 1" implies this is particular to your case, while the $\mathrm{F1}_{\mathrm{score}}$ always ranges from 0 to 1.
- Consider starting a new paragraph when introducing "Cohen's kappa" on Line 173
- Similar to Line 183ff (i.e., "Put simply, …"), consider adding a short, practical explanation of both scores.'

These are good points. We will perform these edits to ensure that this text is as clear as possible. Specifically, we will clarify that the F1 score was calculated for each snow cover target classification (snow presence and absence) and move the point about the F1 score varying between 0 and 1 earlier in the paragraph to establish that this is a function of the F1 score and not specific for our setup. We will also separate the two paragraphs to establish the differences between the F1 score and Cohen's kappa. We also really liked your point about providing a short, practical explanation. We will add text to specify that we provide both of these statistics in tandem to provide a clearer picture of classification accuracy that is relatively insensitive to the total amount of snow cover (driven by the weighted F1 score), yet provides context about the accuracy that could have been achieved provided purely random guesses about snow presence and absence (driven by k, and analogous to the random downscaling approach).

L182: Consider removing the qualifier "that snow cover was downscaled from", as it seems unnecessary at this point.

This qualifier will be removed from the text.

L165ff and L182ff: These two paragraphs describing the evaluation metrics seem a bit misplaced here, in between a general outline of the Methods section and the definition of the three downscaling methods. Consider putting them in their own subsection at the

end of Section 3 (e.g., “3.5 Evaluation metrics”).

Thanks for your comment. We like this suggestion, and we agree that as currently written, the metrics may seem out of place. That said, since the focus of the manuscript is a comparison of these downscaling approaches (and since we find that the downscaling approach is a dominant driver of accuracy), we would prefer to keep the subheadings of Section 3 only linked to each downscaling approach. We also think that the metrics fit well with our discussion of the leave-one-out approach that should be discussed early in the methods. We expect this context to prime readers to understand how the downscaling approaches are compared when they reach those subsections. Therefore, we will add text to the first paragraph of Section 3 to specify that the section discusses evaluation metrics prior to the downscaling approaches.

Figure 2: Conceptually very nice figure outlining the three different downscaling methods. Consider reordering/grouping the panels so that it is instantly apparent that a) and b) refer to method 1, c) and d) to method 2, and e) to method 3. As for my general comment on your figures, Figure 2 is also rather condensed and small, and the text and labels are especially difficult to read without zooming in quite a bit.

Thanks for pointing this out. We will revise the organization of the plot so that the spatial plots for the SVI and percentage of days with snow cover are located on the left. Then, we will add arrows that specifies that the SVI goes into both the terrain based downscaling and RF approaches, and that the percentage of days with snow cover from PS informs the probabilistic downscaling. Subplots b, d, and e will be located on the right, and this reorganization will mean that the subplots will then read sequentially from left-to-right and top-to-bottom. Finally, we will make the black boxes indicating each subsection (corresponding to each downscaling approach) more visible. Additionally, this organization will make the figure a better size to fill the manuscript page and make all of the labels much more visible. Thanks again for your suggestions.

L196: This sentence makes it seem like the topographic downscaling method only looks for locations with above-average accumulation and thereby snow cover duration, while, from my understanding, it also targets locations with enhanced ablation and earlier melt-out. Consider clarifying this introductory sentence.

We agree that this text was misleading. Thanks for pointing it out. We will revise this topic sentence to specify that the topographic downscaling method is intended to identify the cells expected to have the greatest differences (both positive and negative) in both snow accumulation and melt (through the TPI and DAH, accordingly), and the resulting snow cover duration.

L200: Why are you not using the already defined abbreviation “fSCA” here?

Thanks for noting this inconsistency. Here, and throughout the text, we will ensure that we use fSCA for consistency. That said, there is one use of “fractional snow cover” in the conclusion that we will keep to ensure that readers first reading the conclusions before the full text are not having to track down this acronym.

L204: Consider hyphenating south-southwest for better readability.

This will be hyphenated.

L203: It is unclear how the TPI values are normalized, as they have no predefined range.

We will add text to the first sentence of this paragraph (before Equation 7) to specify that the DAH and TPI were both normalized using minimum-maximum normalization provided the spread of values for each of these metrics in each domain.

L225ff: Consider clarifying that the "coarser-resolution footprint" refers to a single pixel of the respective sensor, and that the ranking of the SVI values happens independently for each sensor footprint/pixel. For example: "filling the snow cover in that order until the fSCA estimate *of the respective sensor* pixel was matched." I think this is pivotal for understanding the downscaling methods, and it was not immediately clear to me.

This is a great point. We will revise the text as you suggest, specifying that the SVI filling from lowest to greatest values is specific to fSCA from each pixel for each retrospective sensor. To address a comment from Reviewer 2, we will also change "footprint" throughout with "pixel" to avoid confusion.

L230ff: To me, it was not immediately and entirely clear what the term "frequency of snow cover" (used throughout the manuscript) refers to. This is especially true given the seemingly interchangeable use of "frequency", "relative frequency", and "probability of snow cover", as well as the use of different units ("percentage of days", "days"). Consider clearly defining the term here (or when introducing it on Line 84), or rephrasing it, e.g., to "number of days with observed snow cover". In general, the term is understandable from the context, but more precise wording would improve readability.

Thanks for pointing this out. We agree that the term "frequency" was used far too often to explain concepts and points that were sometimes slightly different. We agree that phrases like "percentage of days with snow cover" and "days with snow cover" are far more clear and should be used when describing the probabilistic downscaling approach and a lot of the figures (e.g., Figures 2, 3, 7, 9, and 10). This will be corrected in both the text and figures.

L291ff: As for my general comments on your images, consider introducing Figure 3 here more directly and guiding the reader by adding more detail (e.g., that the maps are based on VIIRS retrievals).

Thanks for pointing this out. This is a good example. We will revise the first paragraph of the results to explicitly mention how Figure 3 tests parameter sensitivity to snow cover downscaled from VIIRS and is a good example of the spatial differences in snow persistence provided different w values and terrain resolutions. We will also revise the text throughout the entirety of the manuscript, providing explicit descriptions of the results embedded in the figures.

Figure 3: Consider spelling out the "Three-meter snow cover frequency …" at the beginning of the caption.

Pertaining to your comment about our use of the term "frequency", this text will be revised to "The percentage of days with 3 m snow cover on dates with PS snow cover observations.

Figure 4: I suspect the light lines in panel b) represent the individual domains, but it is not

stated in the caption.

Thanks for pointing this out. We will revise the caption to specify that the light lines are for the different domains, and the darker lines is the combined distribution from all domains for each parameter and sensor combination.

L326ff: Consider rephrasing the sentence "This was driven by …" for improved readability.

We believe that replacing "driven by the" with "because" makes this sentence clearer. Then in the following sentence, we will add "rather than the explicit pattern of 3 m snow presence and absence" to the end. We believe that this combination of edits make it clearer that the downscaling matters little on dates that are nearly fully covered or fully snow-free.

L332ff: "decisions made by each downscaling approach about where to assign 3m snow presence and absence" seems like a very convoluted way of saying the downscaling approach itself.

We agree. This will be revised to "each downscaling approach".

Figure 5: Consider grouping the boxplots in panels b) and c) by the sensor instead of the downscaling approach. This would align better with the text, which focuses more on comparing improvements between different downscaling approaches (from the same sensor) than between sensors. However, both possible visualizations/groupings are interesting and simply highlight different parts of your results.

Thanks for your suggestion. We will investigate reorganizing the subplots this way. Since we found that the downscaling approach has a bigger impact on accuracy statistics than the sensor, the general rule-of-thumb to make plots as interpretable as possible is to organize groupings by the first-degree driver and sub-groupings by the secondary drivers. This is how we currently have Figure 5b organized. This is reversed in Figure 5c which makes interpretation much more difficult, but this was done intentionally to maintain a consistent x-axis grouping across the subplots. We will test multiple reorganizations of this plot going forward.

L343: I think this should be Fig. 5a, if I am not mistaken.

That's correct! Thank you for catching this. This reference will be changed to Fig. 5a.

L384: Consider guiding the readers' attention more directly by referring to "rightmost top and bottom panels in Fig. 6".

This is a great idea. We will explicitly reference the difference in total snow cover that you can see between the leftmost and rightmost panels for both MODIS and HLS in this figure.

L393ff: Consider giving both R^2 values and reordering the bias values (MODIS first, VIIRS second), matching their introduction in the sentence before (L 390).

This sentence will be revised as you suggest to report both R^2 values (0.01 when using MODIS and 0.06 when using VIIRS), and then add parentheticals after -15% and -4% (in that order) to specify MODIS and VIIRS, respectively.

L399ff: Consider adding a sentence about the decrease in R^2 in the RF-based downscaling of VIIRS data compared to the probabilistic method.

Thanks for the suggestion. We will add a passage after the sentence noting forest obstruction and geolocation issues to specify that where VIIRS fSCA was most low-biased in this domain aligned fairly well with where PS snow cover was reduced, resulting in a relatively high R^2 despite an overall low bias for the snow cover downscaling using the probabilistic approach.

Figure 7: The "spatial plot" is very small, and its colorbar is illegible.

We agree. Overall, while the information in the spatial plot was interesting, we found that the point we made using this plot was not entirely necessary and potentially distracting. We will revise this plot to only present the binned scatterplots, like we have presented for the other domains in the supplementary.

L430: As I understand it, domain, year, downscaling method, and sensor are all available degrees of freedom. Wouldn't that make the sensor the "least important" descriptor of variance rather than the "next-strongest"?

That's correct. We agree that the term "least important" would be clearer here. This text will be corrected.

L434ff: Consider adding a reference to where to find these values (i.e., Table 1 and Figure 5).

This is a good suggestion. We will explicitly point to Table 1 and Figure 5b to demonstrate this.

L461ff: This is a very interesting and relevant paragraph. Maybe you can indicate the mentioned scale breaks (or ranges) in the left panel of Figure 9. (A8)

We're glad you enjoyed this analysis. This is a good idea. We will add shading to the semivariogram to correspond to the range of scale breaks noted in the text (30 – 60 m).

L469: Missing "of" in "large portion of the snow cover pattern variability"

Thanks for catching this typo. This will be corrected.

L479: Missing "the" in "This reduced the impact …". You could consider rephrasing this (and previous/subsequent) sentence(s), e.g., to avoid the repetitive beginnings "This reduced …" and "This was opposed …"

Thank you for catching this typo. It will be corrected. Also, we liked the suggestion about rephrasing. We will reorganize the following sentence so that it starts with specifying that subgrid information about the likelihood of snow cover is more valuable since coarser scale footprints were more likely to have fSCA that averaged more variable snow conditions (essentially flipping the sentence).

Figure 9: It is unclear where/if you refer to the "spatial plots" in the right panel in the text.

Thanks for catching this. This text was unintentionally removed during the manuscript preparations. We believe that the spatial plots are valuable for visualizing the points that we make about the spatial autocorrelation in snow cover. We will add a sentence early in the second paragraph of this section to explicitly discuss these spatial plots.

L513: It is unclear how the distributions in Figure 5 were used to identify outlying poor

performances.

We agree that this was not sufficiently described. We will revise the text to make it clear that we used the outliers from Figure 5b to identify the downscaled snow cover maps with poorer than average performance, specifically F1 scores that were worse than the lower quartile F1 score by at least 1.5 times the F1 score interquartile range.

L518ff: Where is the "terrain-based downscaling benchmark" defined?

The term "benchmark" was unnecessary here and likely confusing. We were attempting to specify that this downscaling was the base for the comparison with the random forest downscaling. The revised text will remove "benchmark".

L536ff: The "speckled patterns" in the terrain-based downscaling are interesting. If I understood you correctly, this happens when HLS grid cells have a fSCA close to, but not at, 100%, and hence a few (of the lowest SVI-ranked 3 m pixels) remain snow-free. Consider refining your formulation and how the RF method avoids such issues.

Thanks for the comment. You are correct, and we agree that this was not described as clearly as it could have been. We will revise this text to clarify first that the RF approach tended to learn where and when HLS fSCA that is near but less than 1.0 (full snow cover) should have actually been equal to 1.0, based on the historical PS observations. However, the terrain based approach is limited to filling snow cover in the order of increasing SVI values to match HLS fSCA. As a result, when HLS fSCA is low-biased (and even by small amounts), the cells with the highest SVI values were not assigned snow cover. This is particularly evident when you look at results like in Figure 10, which sums the number of days with snow cover and HLS fSCA low biases appear often.

Figure10: Consider adding the labels of the domains directly in the figure (not only in its caption) and coloring the titles of the second and third columns of the right panel "HLS + Terrain-based" and "HLS + Random forest" blue and red, respectively (to establish a direct visual link to the left panel). Also, the white dashed lines representing the median in the left panel are difficult to see. (A9)

These are great ideas. This figure will be revised as you suggest to present the information more clearly.

L564: It is not fully clear how you bias-corrected the MODIS fSCA data. Consider explaining this in more detail.

Thanks for pointing this out. In the revised text, we will make it clear that the RF downscaling approach was first used to estimate 3 m snow cover from MODIS. Then, the 3 m binary snow cover maps were coarsened back to the MODIS grid to calculate fSCA, which was adjusted as a function of the annual fSCA biases learned by the RF downscaling. This was then applied as the input to the probabilistic downscaling approach. Performing this test and then comparing to the original RF and probabilistic downscalings, allowed us to measure the degree to which the bias correction being done by the RF downscaling improved results relative to the combined impact of the bias correction and spatial distribution of snow estimated by the RF downscaling.

L583: Consider specifying the somewhat generic statement about "implicitly accounted

for spatial autocorrelation features and alternative snow depletion and accumulation processes".

Our intent was to specify that the best performing SVI maps, which used 15 m-resolution DEMs and a heavier weighting of melt effects instead of accumulation effects may have partly been a result of snow cover duration patterns that corresponded to processes that the DAH and TPI were not directly intended to represent. We will revise this passage to spell this out more clearly.

L593: It is unclear to me how Figure 3 supports this statement.

This was an additional reference to the findings that we discussed in Lines 515 – 518, where we talked about how important the RF approach was for correcting fSCA biases in the Northernmost portion of GIN where forest cover made snow cover difficult to observe. However, we feel like this was already discussed, and may be distracting to reiterate here. Additionally, the remainder of the paragraph shifts the focus to sensor viewing angle issues. This sentence will therefore be removed in revisions.

L630: It is unclear how this statement relates to Figure 2.

This reference was intended to support how common inputs were used by the different downscaling approaches. However, we agree that it might be a stretch to state that this figure is a good representation of this point. We believe that this is set up well by the text, and this figure reference will be removed.

L636: I think this statement should be in the past tense: "… we tested …"

Thanks for catching this. This will be changed.