We thank the referee for their meticulous analysis of our work and the numerous pertinent suggestions they have provided. We have conducted several additional analyses to address their comments, as outlined in the point-by-point response below. We believe we can propose a revised manuscript that would largely resolve their concerns.

Note : During the revision process we identified and corrected two minor issues:

1. A coding error in the conversion from NDSI to FSC which only affected NDSI values below 0.1 in resampled NASA products.
2. Three corrupted products in our VNP10A1 dataset, corresponding to tile h18v04 from December 3–5, 2023. These products correspond to clear-sky acquisitions following a major snowfall event affecting a large number of relevant pixel-to-pixel correspondences

As a result, the updated metrics show minor differences, such as a slightly higher overall RMSE and a slightly more positive bias. We confirm that this update does not affect any of the conclusions of the work. Hereby, we compare Table 3 of the original manuscript to its updated version.

Tab. 3 of the original manuscript

Product	**N pixels**	**% Snow Cover**	**F1-score**	**Commission Error**	**Omission Error**	**Bias [%]**	**RMSE [%]**
VNP10A1	4.33+06	17.7	0.93	0.01	0.07	0.07	9.82
VJ110A1	4.45+06	17.9	0.93	0.02	0.07	0.35	10.81

Tab. 3 after corrections

Product	**N pixels**	**% Snow Cover**	**F1-score**	**Commission Error**	**Omission Error**	**Bias [%]**	**RMSE [%]**
VNP10A1	4.48+06	18.7	0.93	0.02	0.06	0.42	10.09
VJ110A1	4.45+06	17.9	0.92	0.02	0.06	0.81	10.88

We propose to update all tables and figures of the manuscript to integrate these two corrections. The numerical results reported in this reply will include these corrections.

Hereafter, the point-by-point response. In italic, we report the commentary of the referee to address.

More text needs to be included that defends the use of a NDSI based fractional snow cover truth data set over a physically based model. Evaluating a product that uses NDSI makes sense, but comparison to a NDSI based fractional snow cover "truth" product concerns me given the existence of physically based models, including open source versions, that outperform NDSI for calculating the fractional snow cover of multi spectral pixels. NDSI does not contain information regarding the spectral signature of snow and various non-snow mixed pixels can have the same NDSI values as snow-covered pixels. Various studies have shown that NDSI is less accurate than spectral unmixing *techniques when estimating fractional snow cover. I'd prefer a physically based spectral unmixing truth data set for the highest fidelity truth set from Sentinel-2 and for the paper to have the most impact, but If the author wants to stick to a NDSI based truth dataset, they should better explain the choice. There are always trade offs with choice of validation data set, so I would like to see more discussion on the reason behind this choice and the known pros and cons (cons are omitted currently) so the reader can better contextualize the results.*

We agree that the choice of the retrieval algorithm for the reference dataset is relevant and needs to be further discussed. We specify that our goal is to use readily-available Sentinel-2 products as reference. This approach sacrifices some control on the reference data but allows for scalable, replicable, and straightforward ground truth generation. We apologize for the lack of clarity regarding this point. We propose to reformulate Section 6.1 to better explain our motivations and include some considerations on spectral unmixing and the choices of other authors.

Line 170 - Please explain and defend the choice to reproject and resample at the same spatial resolution as the original data set instead of coarsening to account for geolocation uncertainty. Coarsening is a typical approach to eliminate concern for geolocation accuracy that does not impact overall error analysis metrics (especially at the sample sizes of this paper), but can reduce the impact of geolocation errors on results.

We thank referee 1 for this comment. This choice is explained by our objective, which is to quantify the errors of VIIRS products for a snow modelling system that will assimilate VIIRS data at their native resolution. Our objective here is to evaluate a snow product (including all error sources) and not its snow retrieval algorithm. The finer resolution of VIIRS products (375 m) compared to MODIS products (463 m) is also a motivation for many users to move from MODIS to VIIRS, and, therefore, we believe that characterizing the errors at the original VIIRS resolution is also useful in this perspective. In the revised manuscript, we propose to clarify this motivation and make it clear that our assessment includes various sources of errors, including geolocation errors.

Also, in Section 4.4 I recommend only reprojecting the truth data from Sentinel-2 to the VIIRS projection. Please explain the reasoning to choose a third projection, a non-equal area projection to reproject all the data into for a fractional snow cover area analysis. The point is made that the errors (in terms of area size) introduced are small, but why not keep the VIIRS data in their original projection and only retroject the high spatial resolution truth data to the VIIRS projection for error analysis? This choice to reproject both datasets is error inducing. I strongly recommend keeping the dataset that is being evaluated in the same projection that it is delivered in, and coarsening prior to analysis.

We acknowledge the referee's point that the least error-prone approach is to evaluate VIIRS data in their native projection. However, we evaluate here two VIIRS snow cover products with different coordinate systems: NASA (MODIS sinusoidal grid) and Météo-France (geographic coordinate system). We initially chose to (re)project VIIRS datasets to the same UTM spatial reference system as the Sentinel-2 products to minimize the number of reprojection operations. In fact, most users reproject moderate-resolution sensor data for their applications, with UTM being a common choice. Based on preliminary tests, we assumed that this reprojection step would not significantly affect the results. However, following the referee's comment, we assessed this assumption by reprojecting the reference data to the VNP10A1 projection. We illustrate the differences by reporting Table 4 of the original manuscript in the two cases. Please note that the results have minor differences from the original manuscript for the reasons stated at the beginning of this document.

Table 4 in UTM projection

Reference FSC	Landcover	Aspect	N pixels	% Snow Cover	F1-score	Commission Error	Omission Error	Bias [%]	RMSE [%]
1-99 %	Forest	N	4.99e+04	39.58	0.71	0.20	0.28	-1.38	29.72
1-99 %	Forest	S	1.28e+04	43.83	0.80	0.15	0.20	-1.86	22.97
1-99 %	Open	N	6.58e+04	68.07	0.90	0.17	0.12	-1.74	23.06
1-99 %	Open	S	5.76e+04	65.63	0.93	0.17	0.05	3.31	18.01
0-100 %	Forest	N	3.18e+05	8.41	0.76	0.02	0.24	0.42	13.63
0-100 %	Forest	S	2.02e+05	4.32	0.87	0.01	0.14	-0.13	6.42
0-100 %	Open	N	2.44e+05	36.10	0.95	0.02	0.07	-0.64	13.00
0-100 %	Open	S	3.69e+05	18.27	0.96	0.01	0.03	0.55	7.43

Table 4 in VNP10A1 projection

Reference FSC	Landcover	Aspect	N pixels	% Snow Cover	F1-score	Commission Error	Omission Error	Bias [%]	RMSE [%]
1-99 %	Forest	N	4.61e+04	40.12	0.71	0.20	0.28	-2.01	30.72
1-99 %	Forest	S	1.23e+04	42.05	0.80	0.15	0.19	-2.01	23.79
1-99 %	Open	N	6.17e+04	69.00	0.90	0.18	0.12	-1.51	23.72
1-99 %	Open	S	5.38e+04	64.33	0.93	0.18	0.05	3.13	18.86
0-100 %	Forest	N	2.99e+05	8.50	0.77	0.02	0.23	0.31	13.97
0-100 %	Forest	S	1.91e+05	4.32	0.87	0.01	0.13	-0.16	6.66
0-100 %	Open	N	2.30e+05	37.36	0.95	0.03	0.06	-0.61	13.34
0-100 %	Open	S	3.49e+05	18.14	0.96	0.01	0.03	0.49	7.72

The differences are in the order of 0.5 to 1 % and do not alter the conclusions. Hence, we propose to keep the UTM results in the manuscript, and to add a comment in Section 4.1 to explain the motivations and the little impact.

Line 220 - "We binarize the FSC values into "snow" and "no-snow" classes using a threshold of 50% for both the reference and evaluation datasets." I disagree with this choice. Pixels with substantive snow cover are classified as snow free pixels for analysis. Why was the binary snow cover threshold not set at FSC>0? It is possible to accurately detect snow in Sentinel data well below 50% snow cover and I'd expect the truth dataset to be a binary mask of all pixels that include snow. This is also a reason for such a high representation of TN's in the overall dataset. It seems that the snow covered pixels are a significant portion of the TN dataset with the 50% cutoff. I disagree with this decision and would like to see the snow cover fraction threshold that binarizes between snow and not snow set at a FSC>0 or similar. This choice, alongside the choice to reproject both the truth and VIIRS data, concerns me on the quality of the output analysis. The results in the first few rows of Table 4 seem to invalidate the decision to use a threshold of 50% for snow / no snow cover as you see significantly worse performance in this range of valid snow covers from 1-99% FSC, which is an artifact of the choice to call snow covered pixels with less than 50% snow cover, snow free pixels, alongside the reprojecting and lack of coarsening to account for geolocation errors in the analysis

It is common practice to compute snow-covered area by applying a threshold of 0.4 to NDSI values (Härer et al., 2018), which roughly corresponds to 50% FSC. However, we understand the referee's concern and acknowledge that, following the definition of the NDSI_Snow_Cover variable of V[NP|J1]10A1 *(Riggs and Hall, 2023)*, 0 % is indeed a more appropriate threshold. Therefore, we recomputed the confusion matrix using a threshold of FSC>0 to binarize both target and reference datasets. The F1 score decreases by 2 to /3 % for V[NP|J1]10A1 products, with different commission and omission error estimates. Also, the confusion table becomes more balanced with a TN gof 73 % instead of 81 % (not shown as correctly guessed by the referee). We emphasize that the new binarization threshold only impacts the F1 score, commission and omission errors, while the RMSE and bias calculation remain unaffected. This means that the conclusions regarding the higher uncertainty found in mixed pixels remain unaltered. We propose to update the manuscript with these new methods and results. We hope that this revision will address the referee's concerns regarding the quality of the analysis. Finally, the choices about reprojection and lack of coarsening and their impact have been motivated earlier in this response and will not be addressed in this paragraph.

Use of accuracy metric - I disagree that this is a valuable measure of algorithm performance given the unbalanced nature of the data set and relative ease algorithms have in distinguishing snow free pixels from pixels with factional snow cover, FN and FP are far more challenging issues that are important for understanding the value of a snow cover product. I'd rather only see F1 score in the graphs in figure 4 and the detailed performance analysis.

We thank the referee for this suggestion and will remove the accuracy metric from the manuscript.

Line 312 and use of VIIRS to validate MODIS - I disagree with this approach. The same truth data set (Sentinel) should be used to access MODIS and VIIRS and make determinations on the relative performance of the two. The current chained approach to error analysis introduces unnecessary error propagation into the analysis.

We fully agree with the suggestion and confirm that the manuscript already implements what is proposed. We apologize for the lack of clarity in our original formulation. We propose to add a revised and clearer formulation of our methodology to assess MODIS in the manuscript.

Discussion Section - Many results are presented directly in the discussion section. Recommend moving these important results into the results section and explaining the approach to these analysis in the methods section.

We assume that the referee refers to the comparison of VNP10A1 vs. MOD10A1 performances (Figure 6) and possibly the comparison of VNP10A1 NDSI vs. Sentinel-2 FSC (Figure 7). We acknowledge that Figure 6 can be included in the results section. Hence, we propose to move the paragraph about the MODIS product evaluation to the results section and adjust the rest of the manuscript accordingly. We thank the referee for this suggestion that indeed allows us to highlight a result of interest for the community. We think that Figure 7 belongs to the Discussion, as we use it to investigate the reasons for the mixed performance of the products in mixed pixels, and it is more a perspective of our work.

Technical corrections:

- *Suggest rewording sentence from line 18-20 for clarity.*

We propose to reword this sentence.

- *Line 35 - recommend changing “guaranteed” to “operationally planned” or similar, there is risk with any operational system that it does not meet expected mission lifetime and there can be a gap in coverage before a replacement is fielded.*

We accept this remark and propose to include it in the manuscript.

- *Line 67 - contradicts with Table 2 - reviewed and ASO snow depth data, not SWE was used for the validation. Suggest changing “which provides lidar based snow water equivalent (SWE) measure” to “which provides 3m spatial resolution lidar based snow depth measurements that were converted into binary snow masks.”*

We propose to correct the manuscript and apologize for the oversight.

- *Line 227 - after “we also report the F1 score “ suggest adding “which is the harmonic mean of omission and commission”.*

We propose to include the referee’s suggestion and reorganize section 4.5 presenting first commission and omission errors and then F1 score in order to provide more meaningful insight on the proposed metrics.

References

Riggs, G. A. and Hall, D. K.: NASA VIIRS Snow Cover Products User Guide v3, 2023, doi:10.5067/VIIRS/VNP10A1.001

Härer, S., Bernhardt, M., Siebers, M., and Schulz, K.: On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales, The Cryosphere, 12, 1629–1642, https://doi.org/10.5194/tc-12-1629-2018, 2018.