We would like to thank the reviewers for their thorough reviews and constructive comments on the manuscript. The original comments are shown in regular black font. The responses to reviewer comments are shown in blue font, with text describing proposed additions and revisions of the manuscript shown in red font. Any original manuscript text is shown in gray font.

**Reply on Reviewer #1**

**MAJOR COMMENTS**

1. Our main concern is related to the motivation of the study. There are several snow covered area dataset available (https://lpvs.gsfc.nasa.gov/producers2.php?topic=snow). In particular NASA's MOD10 products provide similar information as the Snow CCI product and are available globally in near real time. The Snow CCI daily SCF version 2 dataset used in this study is available over the period 2000-2020 only and it seems that it is not updated (version 3 extends to 2022). We can think of some reasons but we recommend that the authors explain why they have chosen the CCI product among others.

Response: We appreciate this comment. This study presents a use case of the Snow CCI dataset by evaluating the use of Snow CCI snow cover fraction product to fill gaps in the Snow CCI SWE dataset. The Snow CCI product is part of the European Space Agency (ESA) initiative to generate historical records of essential climate variables that meet the requirements of the Global Climate Observing System. The Snow CCI project provides both global snow cover fraction and snow water equivalent (SWE) products. However, due to known limitations, a complex terrain mask is applied to Snow CCI SWE product. The choice of Snow CCI snow cover fraction product ensures the consistency with the overall dataset development framework and aligns with the objectives of the Snow CCI project. The MOD10 snow cover fraction product and others exist with potential differences in product characteristics. While further studies could explore these products, such a comparison is beyond the scope of this study. We acknowledge other products and emphasize that this work is specifically focused on the evaluation of the Snow CCI and that other products could be explored in the Introduction and Conclusions sections respectively:

In the Introduction (L.56-58), we propose to add a clarification: A snow cover fraction product derived from optical satellite data spanning multiple decades already exists within the Snow CCI program, which provides a natural connection to a mountain SWE product via the BSRF. We acknowledge the existence of other snow cover datasets, such as MODIS-based MOD10 SCF product (Hall and Riggs, 2016), which utilize different retrieval approaches. However, this study specifically focuses on evaluating the Snow CCI SCF product within the BSRF framework to maintain consistency with the overall dataset development framework and objectives.

In the Conclusion (L.581), we propose to add an acknowledgement: "While this study focuses on the Snow CCI snow cover dataset, future work should explore comparisons with other snow cover products, to assess their applicability and optimality for SWE estimation in mountainous terrain."

2. In addition, this study shows that a Landsat-derived SWE reanalysis largely outperforms the MODIS-CCI-derived reanalysis. Therefore, we are tempted to conclude that a global mountain snow reanalysis should be performed with Landsat fSCA. But the authors seem to implicitly consider that this is not an option. We believe that this should also be clearly stated and justified in the introduction.

Response: We appreciate the reviewer's comment regarding the performance of Landsat-derived SWE reanalysis and the potential of using Landsat fSCA for a global mountain snow reanalysis. The aim in this paper is not to indicate that SnowCCI fSCA should be used instead of Landsat fSCA, but rather explore the potential (and limitations) of using a fully consistent set of Snow CCI products. Future work aimed at a global reanalysis should ideally include all available fSCA products and their relative uncertainties.

While we acknowledge the advantages of Landsat data, there are potential challenges that make its use for a global reanalysis an open question worthy of future work beyond the scope of this study. Landsat's 16-day revisit period may be insufficient to capture dynamic snow cover changes, particularly in regions with frequent melt events or at high latitudes. Furthermore, cloud cover often leads to substantial data gaps, limiting its temporal completeness. Specifically, in Canadian domains, Landsat may not always outperform other products due to its coarse temporal resolution and frequent cloud cover.

In this study, due to the limited availability of spatially distributed verification SWE datasets, we generated a Landsat-derived SWE reanalysis, which performs well in the Western United States (WUS), to serve as a comparison reference. By comparing Snow CCI snow cover with Landsat snow cover, we were able to attribute SWE differences to snow cover differences and analyze the importance of snow cover fraction in SWE reanalysis.

In the Conclusion (L.568), we propose to add: "The Snow CCI reanalysis presented herein aims to provide a methodology to fill the mountain SWE gap in the Snow CCI SWE CDR. Rather than suggesting that Snow CCI fSCA should be used instead of Landsat fSCA, this study explores the potential and limitations of using a fully consistent set of Snow CCI products."

**MINOR COMMENTS**

Several acronyms were not defined (L14 WY, L69 fSCA, L11 CCI, L249 DOWY)

Response: Revised. fSCA is now defined in L82.

Fig 2: Because the tiles are defined in lon/lat angles, Fig. 2e merges tiles of different areas, giving more weight to tiles close to the equator.

Response: Cloud fraction is calculated for each tile based on the raw Snow CCI snow cover image. It is only dependent on time for each tile. We acknowledge that when aggregating tile-wise cloud fraction data without accounting for the varying physical areas of tiles, larger-area tiles (typically those closer to the equator) contribute more to the result. Through a rough calculation, normalizing the data by the actual physical area of each tile yields a cloud fraction threshold of approximately 0.55, which is close to 0.6.

Additionally, the WUS domains, located at lower latitudes, generally exhibit lower cloud fractions, resulting in a higher cloud fraction threshold compared to Canadian domains at higher latitudes. To ensure data quality, we apply a stricter cloud threshold for the Canadian domains to filter out more images with cloud cover.

L200 In this earlier study MODSCAG algorithm was used to retrieve fSCA and not SCAmod. Therefore there is no reason to specifically refer to this study to justify the 15% value. Other evaluations of MODIS-based snow products should be considered.

Response: The Snow CCI snow cover product includes an uncertainty layer that provides an unbiased RMSE estimate (that neglects systematic errors). For example, the average internal measurement error for the Snow CCI product is approximately 8% over the Tuolumne Basin. During the initial phase of this study, we conducted a sensitivity test on the measurement error and found that assuming a 15% measurement error for the Snow CCI (MODIS) product resulted in slightly improved performance, but without a significant sensitivity to the measurement error. The attached figure illustrates the comparison of posterior (post) peak SWE relative to ASO SWE under different measurement error assumptions.

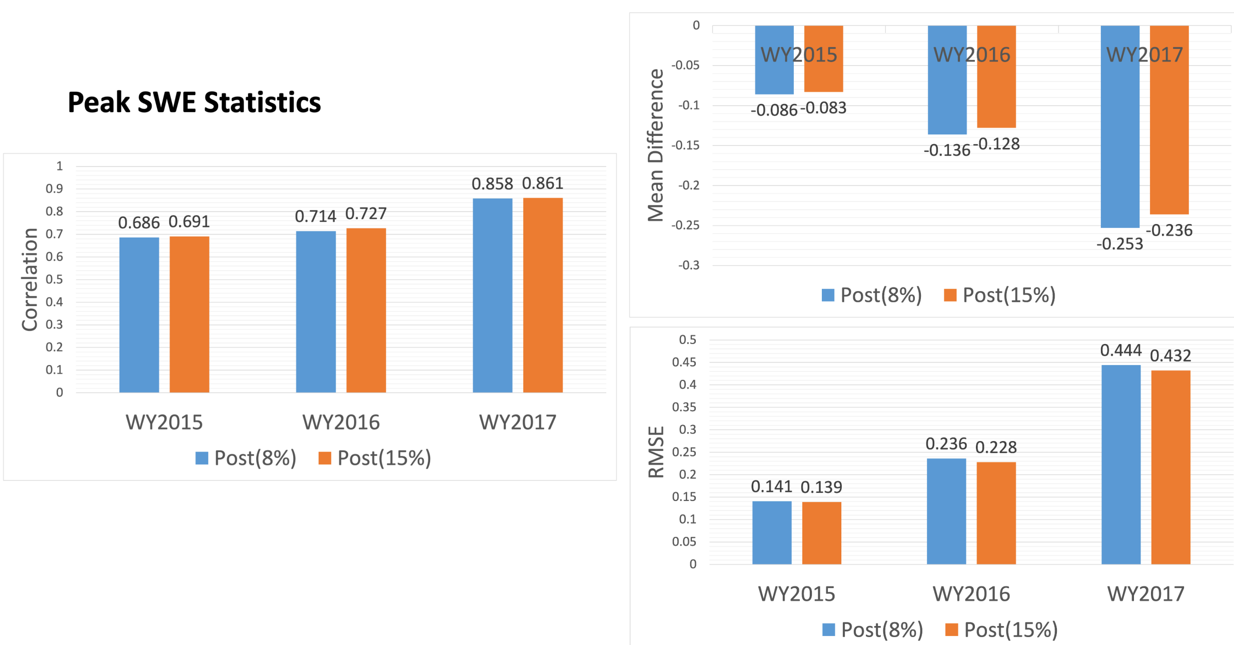



Figure R1. Bar plot of peak SWE statistics for comparison with ASO SWE in the Tuolumne domain, under different measurement error assumptions.

We acknowledge that a uniform and constant 15% measurement error is a simplification that should be studied more, where ideally a space-time varying estimate of the uncertainty is built into the Snow CCI product in a way that also accounts for potential biases.

In Conclusion (L.551), we propose to add: "Since Snow CCI fSCA is retrieved based on reflectance observations from the MODIS sensor, we applied the measurement error (standard deviation of 15%) used for other MODIS-based product applications of the SWE reanalysis technique (Margulis et al., 2019). We acknowledge that a 15% measurement error at nadir viewing is a simplification and that ideally the space-time varying estimate uncertainty built into the Snow CCI product would describe the uncertainty in the SCF estimates in a way that is meaningful for our data assimilation use case."

L206: "the weighting function $w(\theta)$ varies within (0,1] by its definition" Yet maximum MODIS scan angle is 55° hence w will never reach 0. It is difficult to understand how this weighting factor w was defined by Dozier et al. 2008. It would be useful to plot w as a function of the MODIS scan angle. In addition, from a more practical perspective, how were obtained the MODIS zenith angle values? It seems that the Snow CCI product does not provide such information.

Response: Thank you for your comment. To clarify, we propose to plot *w* as a function of the MODIS scan angle in a second y axis in Figure 3 for the revised manuscript (see below).

MODIS sensor zenith angle was obtained from the MOD09GA product, which can be downloaded from NASA Earthdata Search (https://search.earthdata.nasa.gov/search?q=MOD09GA&long=0.0703125).

On L.190, we propose to add a clarification: "… where the measurement error covariance $C_v^{Snow\ CCI}(\theta)$ is a function of the MODIS sensor viewing angle $\theta$, obtained from the MOD09GA product."

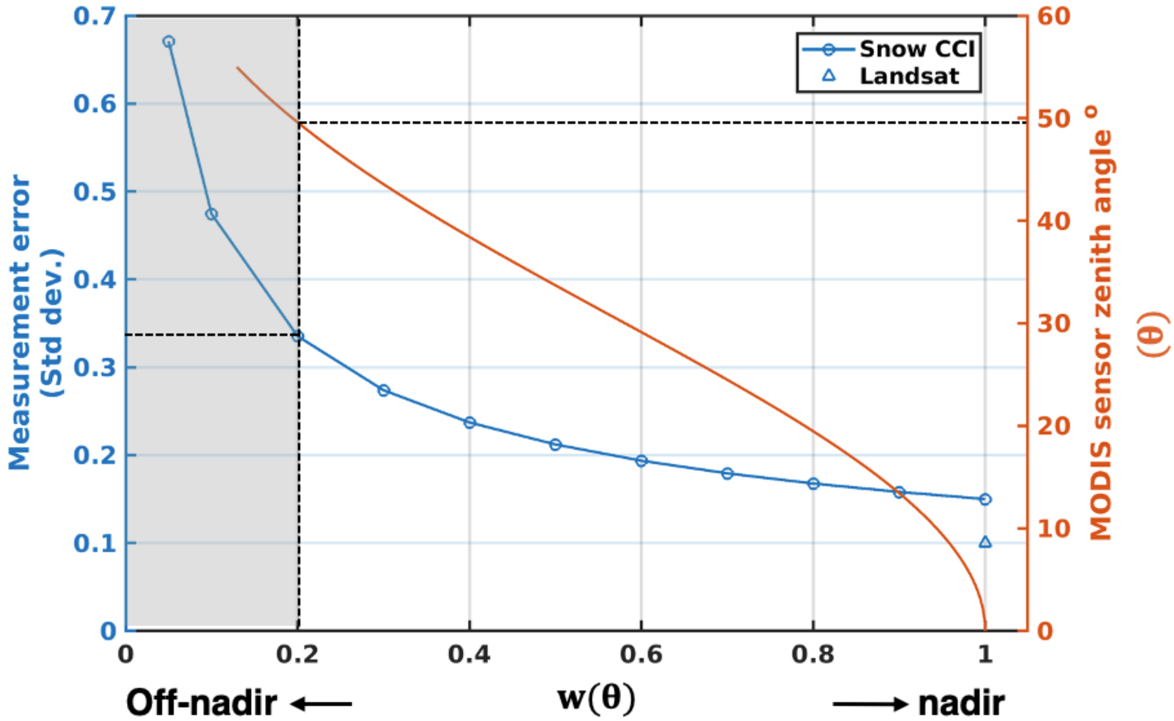



Figure 3. Left y axis: the impact of the w(θ) threshold on the accuracy, i.e. measurement error standard deviation, of Snow CCI fSCA for assimilation. Right y axis: the function of w(θ). Areas below the threshold of w = 0.2 are excluded from the assimilation. The measurement error of Landsat fSCA is represented by the red triangle (i.e., 0.1 at nadir).

L274. Cite the Vionnet et al. paper instead of the URL.

Response: Suggestion adopted.

L280. The interpolation method is first an "aggregation" and then a nearest neighbor interpolation. What means aggregation (average?). Why not resampling directly to the target grid in a single operation? Why a nearest neighbor interpolation?

Response: The ASO SWE data is available at a resolution of 50 meters, while the static topographic data (retrieved from NASA Shuttle Radar Topography Mission (SRTM) DEM) used in the reanalysis is available at a resolution of 30 meters. To harmonize these datasets, we first interpolate the ASO data onto the higher resolution 30-meter grid (SRTM grid) using the nearest neighbor method. This approach introduces minimal error, as the resolutions of the two datasets are similar. We then aggregate the ASO data to the coarser 0.01-degree resolution by averaging the subgrid SWE values.

L295. Why was the evaluation limited to peak SWE? There are many other ASO SWE products in the Tuolumne (e.g. 49 SWE products between 2012 and 2019, Sourp et al. 2024 https://doi.org/10.5194/egusphere-2024-791, data available online https://nsidc.org/data/aso_50m_swe/versions/1).

Response: We conducted evaluations on multiple days (using the additional data you cite) throughout the study period, and the performance on these days was consistent with the performance observed on the peak SWE day. The table below provides a summary of the statistics for all days when ASO SWE data was available. We propose to include this additional information in the Supplemental Information section of the revised manuscript.

The evaluation focused on peak SWE because it represents a critical metric for water resource management and hydrologic forecasting.

| ASO basin | Water Year | Day of Water Year | Correlation | | | RMSD (m) | | |
|---|---|---|---|---|---|---|---|---|
| | | | prior | Landsat posterior | Snow CCI posterior | Prior | Landsat posterior | Snow CCI posterior |
| Tuolumne | 2015 | 140 | 0.58 | 0.82 | 0.7 | 0.07 | 0.1 | 0.1 |
| | | 156 | 0.61 | 0.86 | 0.76 | 0.09 | 0.09 | 0.1 |
| | | 176 | 0.53 | 0.86 | 0.78 | 0.08 | 0.06 | 0.07 |
| | | 185 | 0.48 | 0.83 | 0.75 | 0.08 | 0.05 | 0.06 |
| | | 191 | 0.45 | 0.81 | 0.72 | 0.12 | 0.07 | 0.09 |
| | | 197 | 0.45 | 0.82 | 0.73 | 0.11 | 0.06 | 0.08 |
| | | 209 | 0.44 | 0.79 | 0.7 | 0.11 | 0.07 | 0.09 |
| | | 213 | 0.47 | 0.79 | 0.7 | 0.07 | 0.04 | 0.05 |
| | | 240 | 0.62 | 0.81 | 0.66 | 0.07 | 0.05 | 0.06 |
| | | 251 | 0.7 | 0.72 | 0.52 | 0.03 | 0.03 | 0.03 |
| | 2016 | 178 | 0.66 | 0.9 | 0.84 | 0.41 | 0.26 | 0.37 |
| | | 183 | 0.64 | 0.89 | 0.85 | 0.37 | 0.22 | 0.32 |
| | | 190 | 0.62 | 0.9 | 0.84 | 0.36 | 0.21 | 0.32 |
| | | 199 | 0.65 | 0.91 | 0.85 | 0.34 | 0.18 | 0.28 |
| | | 209 | 0.68 | 0.92 | 0.83 | 0.39 | 0.24 | 0.34 |
| | | 222 | 0.69 | 0.93 | 0.86 | 0.36 | 0.19 | 0.29 |
| | | 240 | 0.56 | 0.89 | 0.87 | 0.28 | 0.15 | 0.19 |
| | | 282 | 0.77 | 0.82 | 0.6 | 0.04 | 0.03 | 0.04 |
| | 2017 | 154 | 0.59 | 0.88 | 0.8 | 0.45 | 0.3 | 0.48 |
| | | 183 | 0.54 | 0.92 | 0.87 | 0.6 | 0.27 | 0.46 |
| | | 214 | 0.6 | 0.91 | 0.85 | 0.58 | 0.29 | 0.5 |
| | | 282 | 0.31 | 0.82 | 0.72 | 0.28 | 0.17 | 0.2 |
| | 2018 | 205 | 0.59 | 0.86 | 0.79 | 0.3 | 0.19 | 0.4 |
| | | 240 | 0.51 | 0.82 | 0.73 | 0.17 | 0.11 | 0.35 |

L309-311. We find a bit confusing to use the Landsat posterior SWE as reference in section 3.1 especially in Figure 6 (where the colors indicate the residuals with respect to Landsat reanalysis). We could suggest to replace the right panel with another scatterplot showing the prior SWE instead of the Snow CCI posterior as a y-axis.

Response: Thank you for the suggestion. The bar plot in the right panel summarizes the statistics of points represented by different colors in the left panel. It highlights the pattern that when

Snow CCI SWE aligns with Landsat SWE at the basin scale, it demonstrates improved performance relative to in-situ SWE. In other words, the largest posterior Snow CCI SWE tend to be when it departs from the Landsat posterior estimates. For comparison, the performance relative to the prior SWE is illustrated in Figure 7.

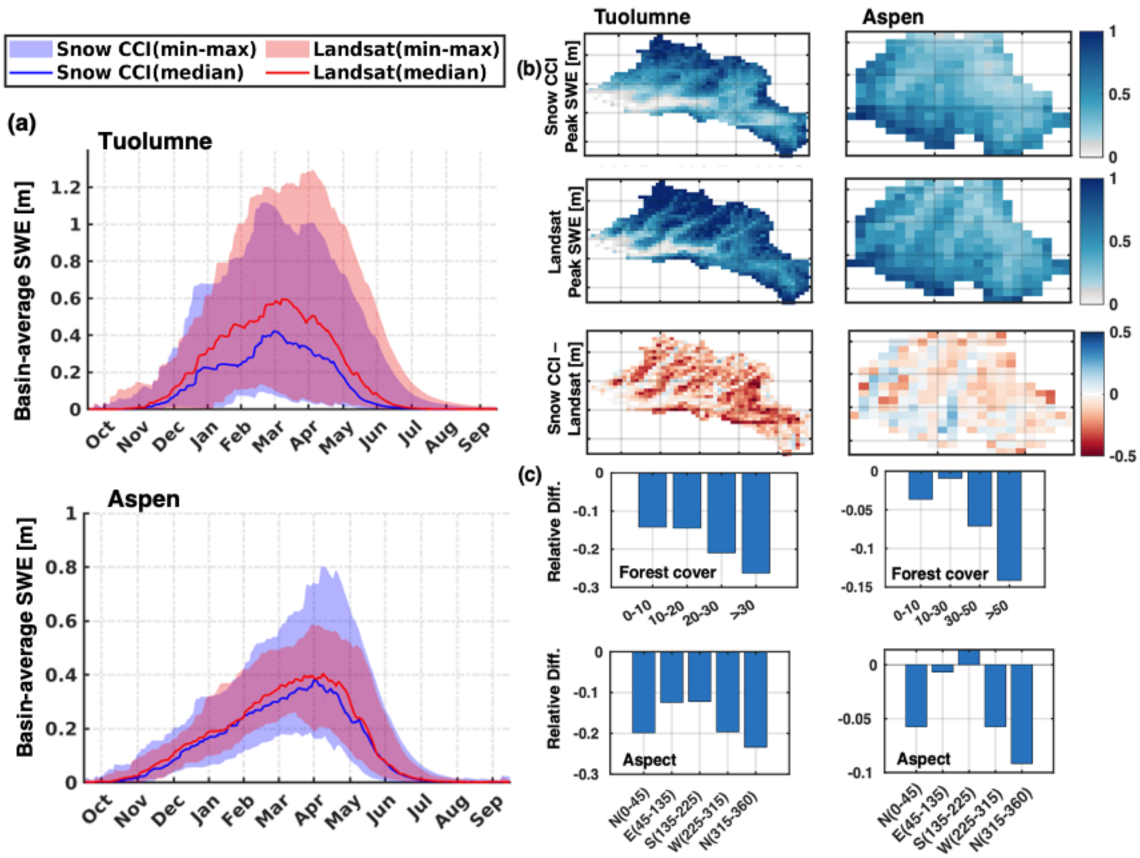
However, we find Figure 7 very informative and well designed. "30 out of 59 sites-year show improvement relative to prior" Does it suggest that assimilating the Snow CCI product was not beneficial on average?

Response: Thank you for catching this. While 30 out of 59 (in situ) site-years show improvement relative to the prior, it does not necessarily imply that assimilating the Snow CCI product was not beneficial on average. The results depend on factors such as in-situ site characteristics, variability in snow conditions, or specific site-year combinations. For example, assimilating Snow CCI product improves the correlation across all snow pillow sites in the Tuolumne in 2007, whereas the improvement in 2003 is less significant. Furthermore, it is important to note that available in-situ SWE measurements are relatively sparse and may not always be representative for comparison with gridded SWE estimates at a 0.01-degree resolution.

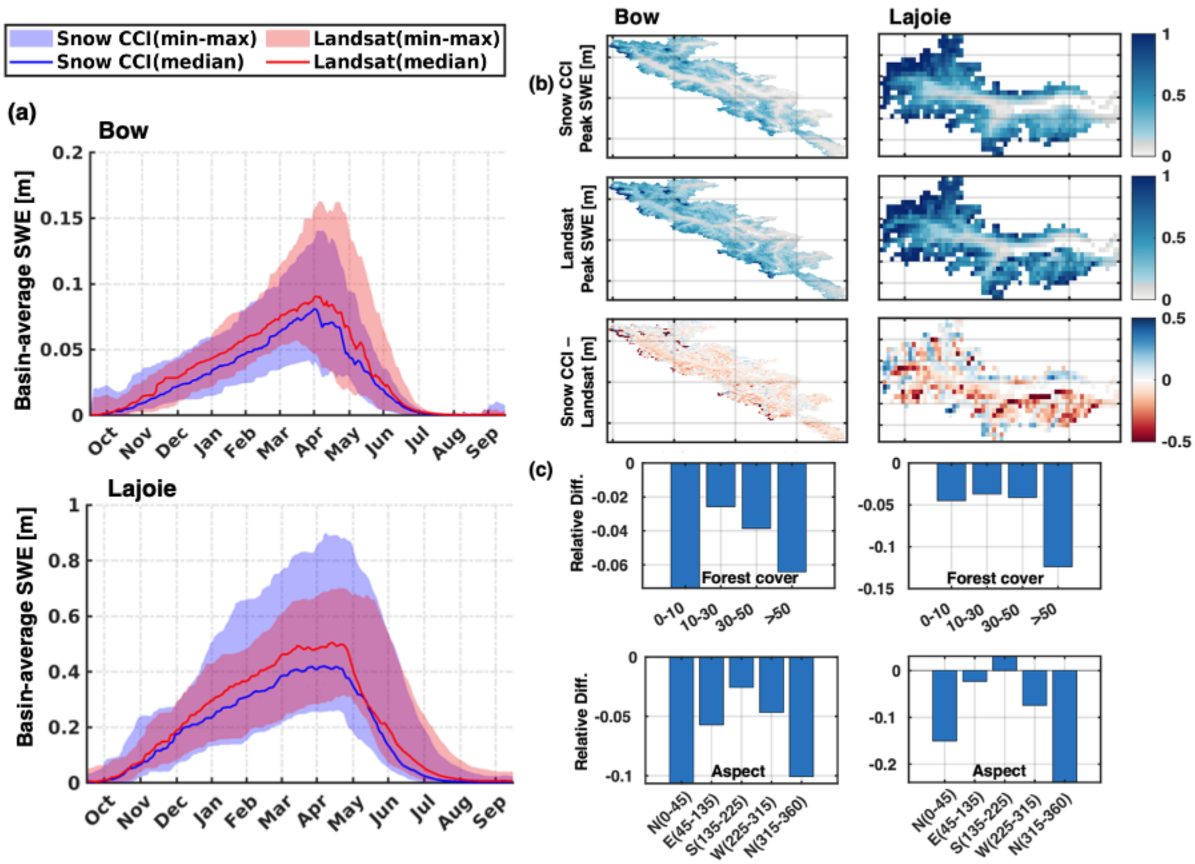
For this reason, we would like to clarify that the ASO comparison (Figure 8 and Table 3) provides a more representative assessment since it is a spatially-distributed product. This comparison demonstrates that assimilating Snow CCI fSCA improves the correlation relative to the prior, highlighting the overall benefits of the Snow CCI product.

Fig. 9: 1%-99% percentiles are usually taken to represent large sample size, here there are only 20 values.

Response: Thank you for your comment. We acknowledge that this range is typically applied to large sample sizes. To address your concern, we explored alternative ways to represent variability, such as using the full range (minimum – maximum) to ensure clarity for the given sample size. We propose to revise Figure 9 to:




Similarly, we propose to revise Figure 14 to:




L470. Figure 11 suggests that the thresholds of cloudiness and *w* discussed earlier in the paper could be revisited. This could be discussed and ideally a sensitivity analysis to these thresholds would be useful (but it may be a lot of computation to ask).

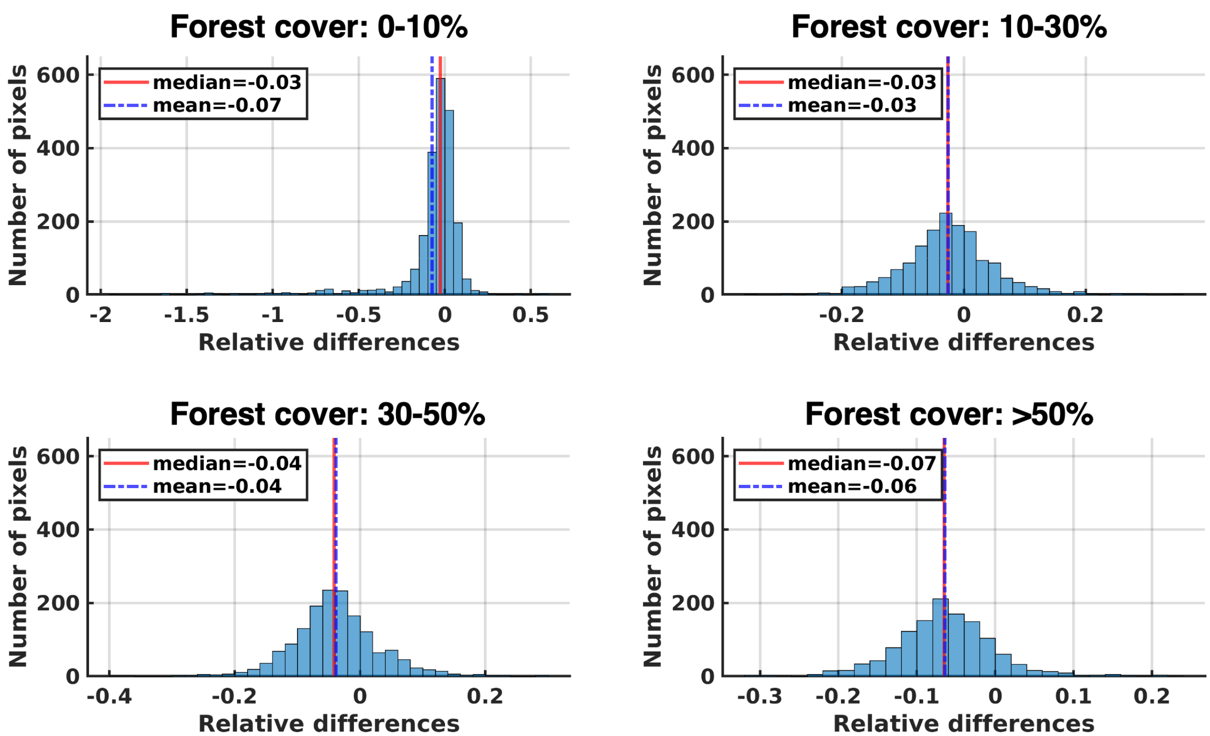
Response: Thank you for your suggestion. We agree that revisiting the thresholds for cloudiness and sensor viewing angle (through *w*) could potentially improve the number of good-quality snow cover fraction image. While a full sensitivity analysis would indeed be valuable, it is beyond the scope of this study due to the computational resources required. We propose to revise the discussion of Figure 11 (L.467) and conclusion (L.551) to: "Future work could explore sensitivity tests of the relative differences regarding the thresholds of cloudiness and sensor viewing angle discussed in Sect. 2.3."

L526. Fig14 How to interpret the poor performance (i.e. the large difference with Landsat posterior estimates) in Bow domain for forest cover 0-10% in comparison with 10-50%?

Response: Thank you for pointing this out. Upon further examination, the poor performance in the Bow domain for forest cover 0-10% compared to 10-50% appears to be attributable to outlier data points. To address this, we have added a new histogram plot showing the distribution of relative differences across all forest cover bins. For the 0-10% forest cover range, the mean value is notably smaller than the median, driven by a few pixels with significant negative relative differences. These pixels are rare and represent outliers in the distribution. When considering the median value, the performance for the 0-10% forest cover range is comparable to that of the 10-30% range.

In the revised manuscript, we propose to clarify this on L.526:

"…, Bow forest cover 0-10% excepted. Note that the bar plots show mean relative differences in each bin as functions of forest cover and aspect. The poor performance in the Bow domain for forest cover 0-10% is likely due to outlier data points. However, when considering the median relative differences in each bin, the performance for the 0-10% forest cover is comparable to that of the 10-30% range."




L568. 0.01°

Response: Suggestion adopted.