**Herbert et al.**
**Submission 2, response to comments**

Again, we would like to thank the reviewers and the editor for their thorough review of this work. Here, we outline the primary alterations we made to the manuscript for this second submission. Line-by-line responses follow below.

## Organization

We have significantly reorganized the manuscript to improve the flow of the document. The majority of changes were made to the methods and results sections.

- The equation for relative snow depth was removed from the Introduction and moved into the methods.
- We have significantly reorganized the methods section.
    - Section 2.1 now focuses on data sources and spatial scales. We have expanded the subheadings of Section 1 to include: Station data, Lidar data, Landcover and topography data, Spatial scales, and Sources of point data.
    - Section 2.2 underwent significant reorganization. We added subheadings and organized the analyses in the order of the research questions they aim to answer. Subheadings include: Snow depth variability, Relative snow depth and representativeness, Landcover and topography analysis, and Consistency of RSD values.
- We reorganized the Results section. Our main goal was to ensure that the order of the results followed the order of the research questions and analyses described in the methods.
    - We split Figure 7 from the original manuscript into two figures. This figure illustrated the results of the expanded scale analysis (0.1-8 km scales) and showed trends of both site representativeness as well as the influence of landcover and topography. The part of Figure 7 that dealt with representativeness was added as its own figure to Section 3.3. The portion of the Figure that dealt with landcover and topography was left in the landcover and topography results section.
    - We moved the point snow depth comparisons to follow Section 3.3 (site representativeness). These results fall under research question 2, so this place is more appropriate in the manuscript. Second, it is a logical next step to show the point snow depth comparisons after demonstrating that different point snow depths yield different results in Section 3.3.

## Land cover and topography

Both reviewers highlighted that our landcover and topography analysis focused on elevation and not other influences on snow distribution such as vegetation and slope / aspect. This was because we only found significant relationships between relative elevation and relative snow depth and not the other factors. We updated the manuscript as follows:

- We re-phrased question 3 to specify that we examine relative landcover and topography variables to be more specific about the analysis we conduct.
- We have re-worked the first paragraph of section 3.4 to emphasize that southness and fractional vegetation do impact snow depth. To do so, we include the results from regressions of fractional vegetation and southness against snow depth at each site. These results demonstrate that southness and fractional vegetation have significant impacts at snow depth at ~90% of sites. We explain that

the significance breaks down when using the relative values and the reasons for this are further examined in the discussion section.

- We added subheadings to discussion section 4.2, addressing the impacts of elevation, vegetation, and southness on RSD separately. We expanded on the discussion of why we do not observe significant relationships between RSD and relative FVEG / southness.
- We include the figures displaying results from the insignificant landcover and topography variables (FVEG and southness) in the supplement.

**Temporal consistency analysis**

Both reviewers questioned how we decided upon the qualitative site groupings in Figure 9. We now group the figures quantitatively, with groupings based on the median RSD value. Delineations were made between sites with median RSD values of < -0.1m, between -0.1 and 0.1 m, and >0.1 m. The results were not changed when using these delineations as compared to the original method.

Second, we added an analysis of intra-seasonal variation of relative snow depth. To do so, we plotted RSD values against days to snow station melt out for each individual site. We included three example sites as Figure 11 in the manuscript. These sites are all from California. We note that this is required because California sites have the most Lidar surveys as well as surveys which span the greatest range of the snow season, making them more suited for intra-seasonal analysis. Sites from Colorado exhibit similar trends as the examples we provide, but they are not nearly as convincing given that most Colorado sites have only 2 surveys per snow season.

Reviewer: Hannah Besso

**General Comments:**
The paper constitutes an important contribution to the field. Snow station data are used for many applications in hydrology. This study adds to our understanding of these stations' representativeness of basin snow quantities and is an important addition to snow hydrology. The scale of the analysis and use of lidar sets it apart from previous studies. However, the authors should explain better and/or reevaluate the temporal analysis. They should also remove the landcover component from Research Question 3, since the author states in the discussion that the dataset used for this component of the analysis was inadequate. Additionally, the manuscript (especially the Analyses section) should be reorganized or condensed to make the story clearer.

**Organization:** The Introduction includes a deep dive into several relevant papers, whose details are repeated later in the paper. These details should be removed from the Intro. The final paragraphs of the Intro, starting at line 91, would fit better in Methods. The Analysis section seemed to jump from one thing to the next and was confusing to keep track of what you were doing. Either introduce the whole section with a list (could even be bullet points) or make separate section headers with titles that describe what you do for each of these paragraphs/separate analyses. Then in the Results it wasn't always clear which part of the Analysis you were reporting on. It would probably be best to maintain the order in both sections, consistent with the order of the Research Questions.

**Vegetation Impacts:** The lack of a strong vegetation component is a big missing piece of this paper, and it should be highlighted as future work that should be done. I think the paper still stands without a veg component, because the impact of the paper is the bias-correction of the stations, not the reasoning behind that bias. However, understanding the impacts of vegetation on snow quantities at snow stations relative to the surrounding basins is important. Somewhere near the beginning of the manuscript you should acknowledge that vegetation has been proven to impact snow depth, but that your dataset was too coarse to adequately investigate the impacts. Or, if you want to include a vegetation component, you could come up with a simple metric such as distance from station to canopy edge (even just using imagery like on Google Earth). As the manuscript stands currently, the discussion provides good citations of others' work on snow-vegetation interactions, but your Results section reads as if you think there is no impact of vegetation on snow.

**Temporal component of the analysis:** How did you decide on the different groups (that were "typically" low, unbiased, or high)? I'm not fully convinced that these groups are distinct since there's so much overlap in Fig 9D-F.
Pflug and Lundquist (2020) (see Figure 6 of that paper) show that basin snow variability can change throughout a season based on snow covered area and whether it was a 'big' snow year or not. This seems relevant to your Section 3.1, where you argue that a larger range of snow depths increases the maximum magnitude of the RSD. So it would follow that there might be an inter- and intra-season temporal component to changes in RSD at a basin. And Figure 9A-C does show that stations can have a range of RSDs of up to about 50 cm, which seems large relative to your 10cm threshold. I don't find myself convinced that stations are so temporally consistent in their bias that it would be easy to bias-correct them based on just a few lidar flights. I think this would be a huge finding that would have implications for everyone who uses snow station data - I just want to see this proven/investigated a bit more thoroughly. A discussion also might be warranted of whether 3 years of data is enough to develop this relationship.

**Snow depth instead of SWE:** I've been told by people more familiar with CA data than I that snow depth measurements (especially at sites in California, managed by CDEC) are less accurate because they're not maintained or quality controlled as well as the SWE measurements are. I think your reason for using snow depth is valid (given that it's directly measured by the lidar data) but it's worth thinking about and maybe mentioning more than the brief description of your quality control method.

**Technical Corrections:**
Line 65: Define 'area-mean snow depth' since you use it throughout the paper.

This is now defined and consistent throughout the document.

Line 134: "provides no advantage" - higher accuracy of SWE vs snow depth measurements from CA snow stations.

Instead of "provides no advantage," we say "increases the potential error" (Line 256). We discuss the QC and potential issues of CA snow depth data in the paragraphs below.

SWE may have better quality control for the CA snow stations but this does not mean that converting all values to SWE would improve the analysis. Calculating density from the snow station requires both SWE and SD, which means we would be still relying on the CA snow depth data as well as adding the uncertainty of assuming a uniform snow density.

Lines 129 - 132: repetitive with the intro. I think you should remove these details from the Intro.

Deleted

Lines 150 - 152: Might deserve further discussion.

The explanation of quality control measures on snow depth data has been expanded.

Line157: “requires much less storage and computational expense to manage” in comparison, I assume, to the 3 m data set instead of the “range of larger scales” you reference in the previous sentence. Be explicit here.

Added ‘compared to the 3 m datasets’ to the end of the sentence.

Lines 154 - 159: How do they produce the 50 m product? Is it derived from the 3 m product?

The products are separate products produced from the Lidar point clouds generated by ASO.

Lines 161 - 164: DTMs can vary quite a bit in their RMSE, and errors can be spatially variable. For example, areas of a certain DTM with steep slopes or dense vegetation can have larger errors or even systematic bias than in areas that are flat with less vegetation. I think here you could just get away with reporting the error published in Painter et al., 2016. Especially because this is the only lidar dataset you use, so I don’t see any added benefit to making generalizations about other lidar products.

We deleted the sentence on lines 162-162 and reduce the lidar error discussion to the citation from Painter et al. (2016).

182 - 183: Confusing sentence, I don’t understand what you did.

We now specify that we determine where a point falls within the cumulative density function of the distribution to improve clarity.

Line 184 - 185: Reword this sentence to be more clear. “We compared the snow depth from different data sources at each point location” or something. “Use different data sources to represent the point snow depth value” is confusing.

This sentence will be re-worded.

Line 185 - 187: Can you explicitly define SD? I understand the sentence but had to read it twice to make sure SD was defined.

We changed ‘snow depth recorded by the snow station’ to ‘snow station snow depth’

Line 187 - 189: Better suited when you introduce the datasets in 2.1.3.

Deleted these lines.

Paragraph starting with 192: Equation 1 and surrounding text would fit better here than in the introduction.

RSD equation has been moved to the analyses section.

Line 195: Why did you choose 10 cm? Also, "acceptable" is undefined. Why not say something like "this threshold will change based on the application, but here's why we chose 10cm".

'acceptable' was changed to 'representative'. The purpose of this section is to discuss that our choice of 10 cm is relatively arbitrary. We reference that this is similarly discussed in Meromy et al., 2013. We also add that our inclusion of probability density functions illustrates the distribution of RSD values regardless of what we determine to be representative.

Lines 212-213: Confusing sentence and probably unnecessary. This section would likely benefit from a summary sentence at the beginning or end that lists your analyses (see my above comments on organization).

The analyses section has been completely reorganized.

Line 215: "proportion of representative sites" was confusing, had to read twice. Make it more clear what you're referring to (I assume it refers to 2 paragraphs previous, where you talk about using each cell as the "station" location.)

Noted. We will change 'the proportion of representative sites' to 'the distribution of relative snow depth values.'

Figure 2: I like that you show the different scales using the boxes. But I want the boxes to be different colors than the SD and Elev scales. Also make the SD and Elev gradients different color scales.
For D can you plot the 50m SD and Station SD as vertical lines that intersect your different CDFs instead of points? Otherwise there are 3 points on this graph representing the same data.

- Instead of colored boxes representing the difference scales we use different line styles (solid, dashed, dotted) as to not interfere with the map colors.
- We now plot the point snow depth values on figure 2D as vertical lines instead of individual points.
- **The color of the DEM has been changed to be more distinct from the lidar SD.**

Line 229: Typo: extra d in "Dd Cumulative"

Fixed.

Figure 3: Same critiques as Fig 2.
Line 233: You use Cumulative Density Function in Figure 2 caption but CDF in Figure 3 caption. Be consistent.

Added an acronym definition for CDF in figure 2.

Line 234: What do you mean by “truncated”? Be explicit.

Re-worded to ‘cut-off’ to be more clear.

Line 245: Why include lidar flights that occurred when the study sites were mostly snow-free if this will skew your statistics?

These low-snow flights still provide valuable data on melt out timing and RSD near the melt out date. The proportion of low-snow flights is still only ~20% of the total data points. As such, we have discussed and elected to keep these flights in the analyses.

Note: you already defined a threshold of 10 cm magnitude RSD. Why so much emphasis on percentiles? This seems like a useful tool in characterizing site variability, but you say it yourself that it’s a problematic indicator of representativeness, so emphasize it less. Also, how does the timing of the lidar flights play into this quantile analysis see above comments about the temporal analysis? Do periods of ablation change this relationship?

The percentile we address here is the percent rank of a point value within the CDF, not the percent difference. We make this distinction clearer in the methods. Our results in this section also demonstrate why using the percentile with the CDF is *not* a good metric for representativeness due to the influence of the range of snow depths on the results.

We address the comment about intra-seasonal variation in the beginning of the document.

Line 264: stick to cm units for consistency

The document has been updated to use m instead of cm for consistency with the rest of the text.

Figure 4: Are the vertical lines on your CDFs supposed to represent the 5th and 95th percentile? They don’t look like they do (they’re all the same width just located in different places - maybe check your code for generating these). If they don’t represent those quantiles, I think they need to be labeled/explained.

The last sentence of the caption for figure 4 explains that the vertical black lines represent areas within +/- 10 cm of the median snow depth.

Line 289: what are “low sites”? Do you mean “low-biased sites”?

Yes, changed to ‘low-biased.’

Lines 291 - 294: I like this summary at the end of the section.
Figure 5: Explain what the gray vertical lines are.

Added: ‘The vertical grey lines at -0.1 m and 0.1 m represent the delineations between low-biased, representative, and high-biased sites.’ to the end of the figure caption.

Lines 305 - 308: See my above comments on the vegetation component.

We address this comment in the beginning of the document.

Lines 313-314: perhaps due to vegetation effects.

Certainly. We talk about the potential causes of snow depth bias in the discussion.

Figure 6: The different colors overlap such that they block each other. Is there a way to make both visible via a different type of plot or by using transparency? This is especially a problem at the 0.5 km scale where I think the pink is plotted on top of the blue. Also why is the .5 vertical line lighter than the others? I missed it at first.

We increased the transparency of the scatter plots in figure 6.

Line 349: "sensing scale"? Does this refer to remote sensing or something else?

Changed to 'spatial coverage' to be consistent with the rest of the manuscript.

Figure 8 caption: use consistent labels. "50 m Lidar pixel" vs. "50 m SD". Also, the 10 cm lines look gray to me instead of black.

Updated the figure caption to be consistent with the naming scheme we use in the rest of the document.

Figure 9: See my above comment about how you grouped the stations. There's a lot of overlap between groups (D-F).

We address this comment at the beginning of the document.

Line 384: I don't think "overrepresent" is the right word here.

We will change 'overrepresent the surrounding area' to 'yield snow depths greater than the areal-mean snow depth'

Lines 386 - 388: this fits better here than in the Intro.

We address this comment at the beginning of the document.

Lines 397 - 398: Rephrase. I think you're saying that any bias correction would need to be site specific. And do you mean "positively biased" not "oversampling"?

Yes, we added a second sentence that describes how any bias correction would have to be site specific and require existing lidar data to do so.

We re-worded this sentence to say: "Correcting the bias exhibited by snow station snow depths would mitigate this problem at some sites, but risks deteriorating representativeness at low-biased sites."

Line 400: Instead of a list you should present the infrastructure, flat terrain, etc as components of the location bias. Otherwise this conflicts with the other 2-component list you give in Results.

The list in the results is presented to determine if the bias in snow depth is a remnant of sampling error or a true bias in snow depth. The list in line 400 presents possible mechanisms for the bias in snow depth,

which we previously determined to not be a result of sampling error.
We now state: ‘There are multiple possibilities of why station location within a 50 m pixel causes a high-bias’ to be more explicit.

Lines 444 - 445: See above comments on vegetation component of the analysis. Line 465: “pixel”, not “point” for the lidar data

We deleted ‘closest 50 m Lidar point to represent the snow station SD’ and replace it with ’50 m SD’

**Review Summary**
Herbert et al. use continuous station data and repeat lidar data acquisitions to contextualize the representativeness of station snow depth to the surrounding areas at multiple spatial scales.
Through these mixed scale snow depths, the authors additionally work to identify differences in snow depth depending on the sensor and if there are temporal patterns across at each site between each of the lidar acquisition sensors. The primary results indicate that there was no significant difference between the snow depth point measurement and the 3 m lidar snow depth at the station, however significant variability in snow depth between the point locations at the 50 m lidar areal mean were present.
Generally, station snow depths area high or representative at 0.5, 1, and 4 km scales while at the 50 m scale snow depth is generally representative with some high. These results indicate that the point station snow depths are representative to the surrounding area, but stations tend to be placed in areas with greater snow depth than their surroundings.

The paper addresses a serious question that has been raised many times on the representativeness of current point measurements of snow across the US at SNOTEL or equivalent sites. These questions have become increasingly important to answer with the proliferation of modelling and remote sensing efforts that utilize the sites as a tuning parameter. The increasing availability of lidar data provides an intriguing opportunity to better define the representativeness of the sites.
While the paper needs refinement prior to publishing, the methods used are appropriate and provide great insight into the stationarity of snow depth point measurements at SNOTEL and CA DWR sites.

I recommend the manuscript for publishing with major modifications, and I provide comments that are necessary to address prior to publishing.

**General Comments**
The manuscript has lots of great details and presents significant scientific findings, however the lack of structure reduces the clarity of the methods and findings. The paper would significantly benefit from additional sub headers within the methods and results sections that guide the reader through the four questions being asked, this would allow the reader to link each of the methods/results to the questions and help guide the reader through the scientific story.
Additionally, while there are lots of great details in the introduction and methods, much of it is repetitive and I suggest the author carefully consider what is included and what distracts from the main points of the paper and should be cut. I recommend the authors condense significant portions of the paper but additional details are needed, specifically in portions of the methods where key data decisions and assumptions are made. The paper would also benefit from consistent use of terminology when referring to snow depth measurements at the varying scales. Finally, there are a few methods that need to be further explained or clarified prior to resubmission.

**Line Review Comments**
33-34 – Tying in the purpose for this paper at this point of the paper is confusing to the reader and would be better suited for the end of the introduction.

Deleted. This paragraph has been adjusted to include more information on the strategic placement of snow stations.

43-89 – It would be beneficial to the article if you revisited these paragraphs and condensed them down. There is a lot of overlap and while the content of each is useful to the reader, it could be condensed to the most relevant information for the paper (use of SNOTEL sites as model validation, assimilation datasets, and extrapolation). Then you could point out the known flaws of using point snow data to represent larger areas and begin to tie in how lidar is a useful tool to fill this knowledge gap in the literature.

We have re-worked this section to increase conciseness to the overarching conclusions made from the relevant existing literature. Though, we believe that this section provides crucial background information to the reader in that it 1) provides context for the uses of snow station data and 2) discusses the results found in previous investigations which assessed representativeness of point snow data. We only include key results that are relevant to this investigation.

91-104 – By simplifying the earlier parts of your introduction, I think you will be better able to set up this study, why it is different and important. Leaving the reader with a clear understanding of what you plan to accomplish in the study and the questions you will be answering.

96 – It might be best to leave the equation for the RSD until the methods, so it is clearer for the reader when specifically describing your methods than having to reference the equation here.

The equation for RSD has been moved to the methods.

103-104 – While I think this is a fine method to take for this research, I do think that this assumption needs to be explained further in the methods.

We added "See section 2.1.2 for further explanation on this decision" at the end of the paragraph to reference where we discuss this decision further.

109-110 – I do not believe this sentence is needed here, it should be mentioned in the introduction that basin wide aerial lidar is becoming more prevalent (ASO, etc.) but it distracts from the study site and data used in the paper when included in this portion of the paper.

Deleted.

123-126 – Do the CA-DWR sites use the same snow depth instrumentation as SNOTEL? Add a citation.

Yes, CA-DWR depth sensors are the same as NRCS sensors based on personal communication. We added a sentence describing that CA-DWR does not provide precision values, but they should be comparable since they use similar hardware.

128-137 – Refer to comment on lines 103-104.

This section provides our justification for using the ASO snow depth product over the SWE product. We will add: "Lidar SWE products use modeled density (Painter et al., 2016), increasing the uncertainty of the SWE product compared to snow depth" to follow the first sentence of the paragraph. This provides an initial justification for our employment of snow depth. The following sentences describe how an assumption of uniform snow density across the landscape provides no advantage over using snow depth as the key variable.

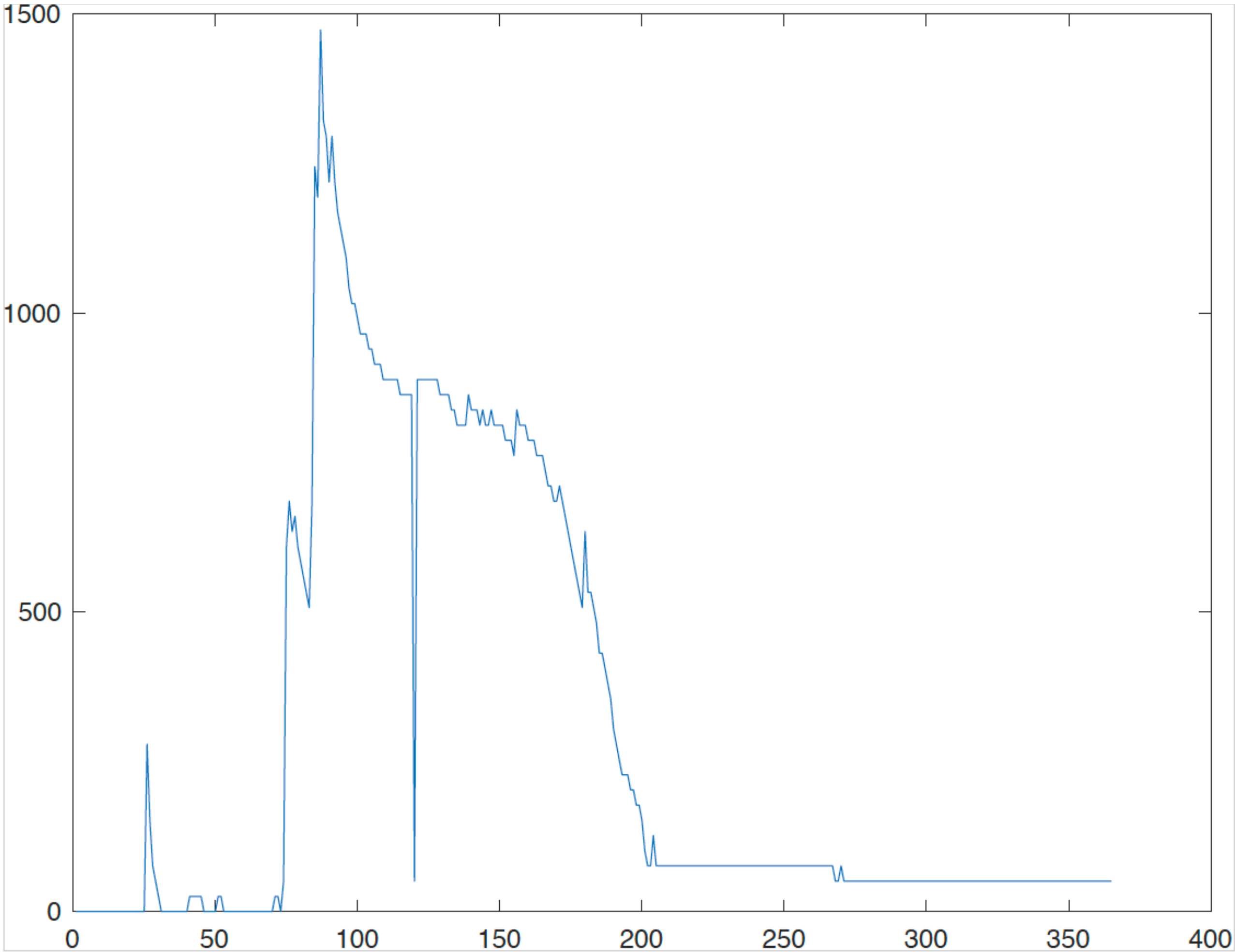
141-149 – What is the positional accuracy of Google Earth imagery? It could be worth working with the NRCS and CA-DWR to ensure you have the correct coordinates for the station and an idea of where the depth sensor is located at each.

We state that Google Earth imagery is available to the fifth decimal place in decimal degrees (~1 m resolution). We utilized the official coordinates provided by the NRCS and CA-DWR (i.e., the most accurate available data). We found that the location data was not always precisely located on the snow station in Google Earth, necessitating updated coordinates in certain cases. We only used sites that we could visually confirm using satellite imagery.

150-153 – Why does the SNOTEL data not also need further QC? NRCS typically only corrects the daily snow depth data (midnight). What QC methods were taken other than discarding data with a greater than 50 cm difference between hours? Why 50 cm? That seems like a large upper limit that could be tightened to a lesser number that would be more typically seen as a realistic hourly snow accumulation/ablation amount. Would it make sense to use the daily data for this work instead of the hourly data? Assuming that lidar was collected on clear sky days, can we assume that there was minimal snowfall/snowmelt during the day? Do you use the hour of the flight to compare between the station and the lidar?

The existing text contains a mistake. We *do* employ daily snow depth data, not hourly snow depth data. We switched from hourly to daily data prior to the first submission of this manuscript and did not update the text to reflect the change. We will update the text to specify that we use daily data for NRCS and CA-DWR sites.

Re: QC. We performed QC on all station data. NRCS stations did not require any data deletion, but CA-DWR stations did. The most common error in the data was a rapid shifts of snow depth in one direction followed by a shift in the opposite direction. The figure below illustrates an example of such a shift. We found that accounting for 0.5 m multi-directional shifts in snow depth reduced obvious errors in snow depth data. We updated the text to make this clearer.




174 – Additional sub headers of the different analyses would be beneficial for the reader to understand what methods you are using for each of the separate research questions.

We have reorganized and added subheadings to the analyses section.

186-187 – I think it would be beneficial to the reader if you continued to call these three point measurements by the names listed here throughout the paper.

We ensure consistency with the terminology throughout the paper.

187-188 – This line about 50 m resolution has been included multiple times. I recommend you remove repetitive text throughout the paper unless it is a key result that deserves repetition.

We will delete the justification for using 50 m data in lines 187-188. This sentence is repetitive after the justification we provide in the data section (156-159).

194 – The idea of identifying the representative of the stations to the surrounding area is an important topic, but I wonder if you could use a percentile-based approach that would get the user a percentage of accuracy at each of the scales that they could then judge what "acceptable" is for their use case. This approach would remove the limitations you mention in line 195, although as you mention there are limitations to percentile based analyses.

The limitations that we mention in line 195 are meant to showcase that an acceptable range of error is dependent on the application for which snow station data is used. This issue would persist for any metric we use to determine representativeness. Meromy et al. (2013) use an acceptable error range of 10% for RSD, but state than any cut-off for acceptability is relatively arbitrary.

In the preparation of this manuscript, we have used absolute magnitude, percent difference, and percentile (rank within the cdf of all snow depths) for classification of RSDs. Meromy et al. were able to use percentile with more success because their analyses did not include surveys near the snow melt-out date (i.e., the range snow depth values was relatively constrained). Our analysis includes data from many different sites, years, and times within the snow season, meaning we have a wide range of snow depth magnitudes (~0-6 m). This range of magnitudes makes a percent difference approach difficult to interpret between data points.

In figure 4 we demonstrate the difficulty of using a percentile-based approach. Areas with uniform snowpack may yield high percentiles for snow depths which are close in magnitude to the point snow depth (e.g., Figure 4B, lines 251-254).

Due to the complications with using percentage and percentile approaches for RSD classification, we elected to use an absolute magnitude approach for RSD classification.

206 – For topography, why are you only analyzing elevation or is aspect also included in the topography analysis? Elevation and landcover are two key portions of a complex relationship between snow depth and mountainous terrain.

We address this comment in the beginning of the document.

212 – This section could be trimmed down to better describe the analysis to the reader succinctly, but additional detail on what you are doing here could also be beneficial. Are you trying to identify the distance at which sites become unrepresentative to the surrounding area? I think this analysis would benefit from some more spatial statistics like variograms (Anderson et al. 2014) or assigning the percentile of representativeness to each scale. Additionally, when stretching to 8km scales, were major topographic features (i.e. ridgelines with low snow due to high winds, topographic basin directions which lead to preferential snow fall, etc.) accounted for in the analysis?

This paragraph has been re-worked with the rest of the analyses section. Figure 6 in the updated document displays how the percentile of representative sites changes with each scale.

221 – Are the other analyses only conducted at one survey date, averages of each date, or are those results also temporal?

We conduct all analyses with all available coincident lidar-snow station data. For each lidar flight in CO and CA between 2021-2023 we find all available snow station data spatially and temporally coincident with the flight and calculate relative snow depth for that date/location. We have updated the text to make this clearer.

Figures 2 and 3 – Although the x-axis scales would be different between figure 2d and 3d, I do think it would be interesting to show the full cdf for both sites.

We have updated the axes in Fig 2d, 3d to extend to 3m so they now capture most values in Spratt Creek (aside from the highest values at the 4 km scale).

245 – Are there surveys available for periods when the stations have snowpack still? Is this the SNOTEL point data of the RSD distributed data? If RSD, at what scale? Can you drop the

surveys with no snow.

Yes, we include all available Lidar data. Some flights occurred in the late snow season, while others were flown closer to peak SD. We do not delete the low snow flights because snow melt-out at the 500 m scale may not reflect snow melt-out at the 4 km scale.

261-264 – How often were the 50 m and SNOTEL measurements ~30 cm different? Keep the units the same (cm or m) when comparing two values.

We updated the document to ensure that all units are reported in meters, not cm. We address the differences between station SD and 50 m SD in later sections (Fig 8).

Figure 4 – It might be helpful for the reader if you added a median line to the plots.

We attempted a median line in Figure 4, but the additional horizontal line makes the figure too busy.

282 – A more thorough comparison of station SD and 50 m lidar SD at the same site could be very informative and help the reader better understand the differences in results between each of the RSDs and the "point" measurement. Jumping right in without this context makes the differences in RMSE hard to understand.

We have moved the point snow depth comparisons up in the document to make the results flow better.

287, 302 – What are the virtual snow stations? What are the Sim. Sites? These all need to be called the same thing if they are, or better explain each in the methods.

The use of 'Sim. Site' in Table 1 was a mistake. We updated to 'virtual site' to be consistent with the rest of the text.

305 – I really appreciate you bringing the question back to the reader, would be helpful if you did this for the other sections as well.

Figure 6 – I think it would be helpful to plot the linear trend lines that you have defined to draw the readers eye to the trends/lack of trends within each. Additionally, are blue points being blocked by pink? It would be interesting to see the difference between the states since we know Colorado and California snowpacks act very differently.

We added transparency to the markers and add trendlines to figure 6.

321 – Why are the sites more likely to have a higher magnitude relative elevation? Is this because sites are typically at low elevation compared to their surroundings?

We added: "due to the increased range of elevations values with scale" to line 322. The reason for increased relative elevation with scale is simply because stations are located in heterogeneous terrain, meaning larger scales are likely to be more different in elevation than smaller scales.

335 – Figure 7c, I am not sure what this is adding, we know that SNOTEL sites are located in mountainous areas with complex terrain features, leading to significant variability (increasing STD) in elevation over larger areas.

The purpose of this is to demonstrate the point in the comment above: larger scales have larger ranges of elevation. These larger elevation ranges lead to increased relative elevation values, in turn decreasing the proportion of representative sites at larer scales. We have split figure 7 into two parts: the parts relating to representativeness and the parts relating to elevation (as described in the beginning of the document).

338 – This section needs to be the first of the results. It sets the stage for all of the other analysis. Then you can identify why you use the point measurement you choose (50 m SD) to complete the remainder of the analysis. Why did you choose to use the 50 m SD for the other analysis?

We have moved this section to just after the representativeness results (Section 3.3) to address the causes for differences in results when using different point snow depths.

Figure 8 – Adding the 1:1 line while making the +-10 cm lines darker would be very helpful for the reader. Again, transparency or making CO and CA their own plots would also be informative.

**We increased the transparency on the scatter dots and made the +/-10 cm lines more pronounced.**

357-365 – This analysis is confusing to me and I think needs further explanation. Does the site need to be "high" at all three scales to be in the "high" grouping or are the groupings different based on each spatial scale? Why are the groups the same size, shouldn't the sites determine if they are negative/zero/positive? Are the RSDs below zero in Figure 9d-f due to the station being grouped into the three groups across all scales not for each scale individually? Line 375 starts to answer this but needs to be better included in the body of the paper.

We address this comment in the first section of the document.

393 – Are these completely independent data sources, or does ASO do any QC/shifting of the point cloud based on the SNOTEL site snow depth?

SD is an independent metric. ASO may alter their density values based on in-situ data, but this is reflected in the SWE product, not the SD product.

398 – Removing this bias systematically? Or by each site? I think this would have to occur on a site by site basis since each SNOTEL will act independently from its surroundings.

Yes, we will explain this further. Adjustments would have to be made on a site-by-site basis. This would only be beneficial to sites that have Lidar data. A catch-all adjustment would be more useful since it could work for sites/periods without lidar data, but this method risks deteriorating sites with low RSD vals. We will update the text to clarify this.

435 – Could this be a scale issue? As you mentioned, it is well documented that there is less snow under canopy than open areas depending on the time of year. Is this due to the way RSD is calculated?

We address this comment in the beginning of the document.

465 – another name for the 50 m point measurement, please coordinate these throughout the paper to simplify for the reader.

We ensure consistency of terminology throughout the document.

**Technical Comments**

These technical comments have been addressed.

229 – typo "d" should be a ")"?

247 – missing units on the snow depth range 331

Figure 7b, missing ")"

395 – “location bias.”