Note: author responses to reviewer comments will be written in blue text.

We would like to sincerely thank the reviewer for their comprehensive assessment of this manuscript. We believe the alterations prompted by this review will improve the science and its presentation prior to publication. First, we summarize the substantive changes we will make to the manuscript. We respond to the line-specific comments in the second section of the document.

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 will update the manuscript to include more details about the landcover/topography analysis. We propose the following changes:

- We will add a section to the introduction discussing how the factors that control snow distribution vary with scale. The scales of our analysis are large (km scale) and thus are unable to capture all of the influences on snow distribution.
- We will add the correlation between FVEG/RSD and northness/RSD to the graph in figure 7C. While these relationships are insignificant, this figure will illustrate that we did examine topographic and landcover attributes other than elevation.
- We will expand our discussion to include more details about our expectations for landcover and topography on RSD. The influences of landcover and topography change with scale. Our hypothesis is that at kilometer scales, the effects of elevation are more pronounced than finer scale effects such as vegetation. The aggregation of a physiographic variable to a single value over a kilometer scale may hinder its utility to represent the complex dynamic of a snowpack.

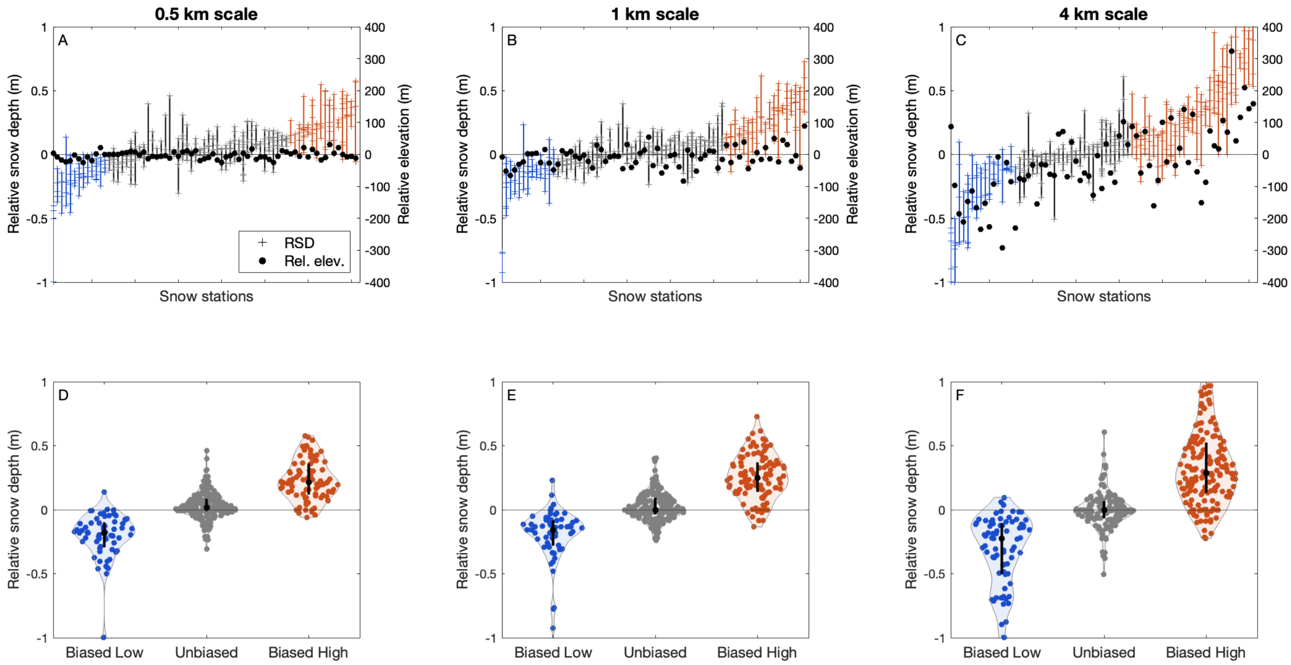
Organization

We will condense and re-organize certain parts of the manuscript.

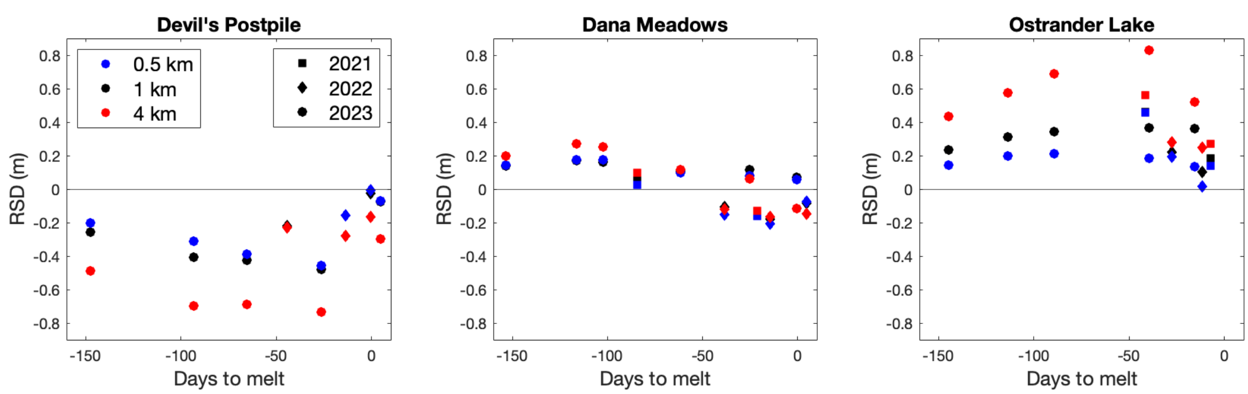
- The literature review presented in the introduction in the two paragraphs from lines 57-80 will be made more succinct. We will summarize the type of research that has been previously conducted in the area but will refrain describing the results in detail. Relevant results will be brought up in the discussion.
- We will remove the equation for RSD and place it within the methods section.
- We will add sub-headings to the analysis section to improve clarity and readability. We will ensure that the order of the analysis section matches the order in which we present the results.

Temporal consistency analysis

Both reviewers brought up the qualitative site groupings in Figure 9. We will group the figures quantitatively, with groupings based on the median RSD value. Delineations will be made between sites with mean RSD values of < -0.1m, between -0.1 and 0.1 m, and >0.1 m. An updated figure 9 is presented below:



Second, version 1 of the manuscript did not examine the intra-seasonal variation of relative snow depth. We will add a figure which highlights the intra-seasonal variability of RSD at three selected sites (see below). This figure will be included in the temporal analysis and be included as Figure 10 in the manuscript. This figure 1) highlights that RSD varies seasonally, peaking near peak SWE and reducing in magnitude as melt-out date approaches, and 2) it provides an additional illustration that many sites do have a consistent sign of RSD at all three scales.



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.

We will delete the last sentence of this paragraph.

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 will condense this section to improve conciseness. Though, we do 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 will reduce the section to a general summary of the existing literature and only include key results.

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 will be 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.

Regarding the decision to use snow depth and not SWE we will add “See section 2.1.2 for further explanation on this decision” At the end of the paragraph.

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.

Will delete this sentence. The increased availability of lidar data discussion will be left to earlier in the introduction.

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 have been unable to find documentation on CA-DWR snow stations. If we are unable to find any documentation we will email people at CA-DWR to get confirmation regarding the instrumentation used in DWR snow stations.

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.

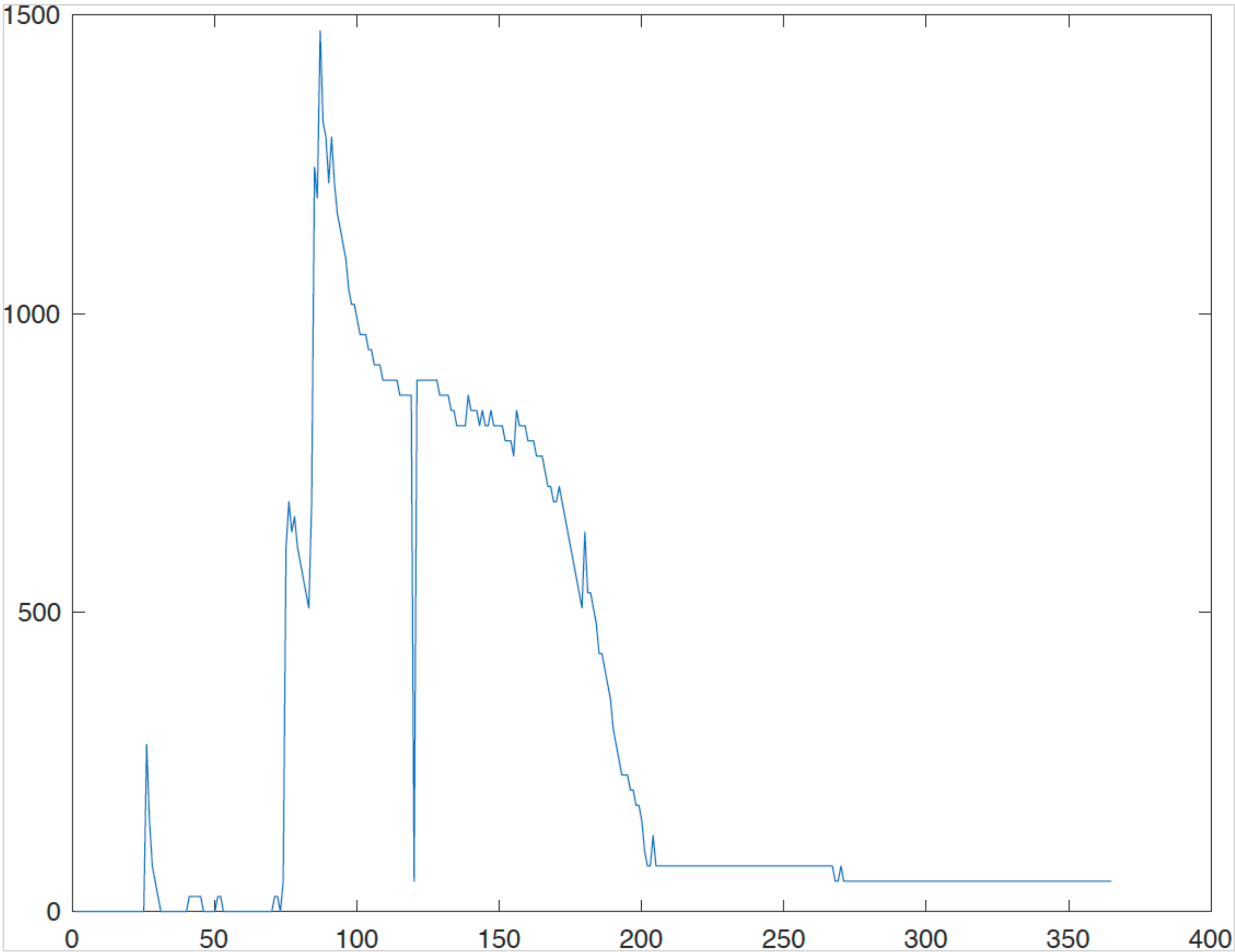
141-149 – What is the positional accuracy of Google Earth imagery? It could be worth working with the NRCS and CA-DWR to ensure you have the correct coordinates for the station and an idea of where the depth sensor is located at each.

Lines 144 and 146 describe 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 at ~120 on the x-axis. We found that accounting for 50 cm multi-directional shifts in snow depth reduced obvious errors in snow depth data.



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 will add sub headers to this section to improve clarity / readability. Additionally, we will ensure that the order of the analysis section follows the general order of our presentation of the results 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 will make sure to be consistent with this 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?

Yes, the purpose of this section is to assess how representativeness, as well as the influences of landcover and topography change with scale. We will edit the text where we describe the qualifications for inclusion in the analysis (lines 216-220) to be more concise.

Figure 7C shows how the standard deviation of elevation changes with scale. This is essentially a variogram conducted in grid space as opposed to the transect version presented in Anderson et al. (2014). Figure 7D shows demonstrates representativeness changes with 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 will update 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 will update the x-axis in figure 3 to include the entire cdf of snow depth.

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 will update 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 do this comparison later in the manuscript (Figure 8).

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 will update 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 will add 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 will add: "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. This may be more clear if we change standard deviation of elevation to the 5-95th percentile of elevation in figure 7C. We are trying to show that with increasing scale 1) the range of elevations increases and 2) the correlation between relative elevation and relative snow depth increases, and 3) the reason for decreased representativeness at larger scales is likely due to increased influence of relative elevation.

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 use 50 m SD for most of our analyses because it ensures that there is no bias caused by differences in sampling methodology (in-situ versus lidar). Comparing areal-mean lidar to point lidar allows for a direct analysis of the location of the snow station within the landscape.

We will highlight this further in the analysis section when describing the point data. Second, we will reference our point snow depth comparisons in section 3.2 to inform the reader that we do examine causes for differences in 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 will increase the transparency on the scatter dots and make 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 will address this.

Technical Comments

These technical comments will be addressed.

229 – typo “d” should be a “)“?

247 – missing units on the snow depth range 331

Figure 7b, missing “)”

395 – “location bias.”

References

Anderson, B. T., J. P. McNamara, H.-P. Marshall, and A. N. Flores (2014), Insights into the physical processes controlling correlations between snow distribution and terrain properties, Water Resour. Res., 50, 4545–4563, doi:10.1002/2013WR013714.