Note: the author responses to comments will be posted in blue. Reviewer comments are in black.

We would like to extend our sincere gratitude to the reviewer for taking the time to review this manuscript. Their expertise and thoughtful feedback have been invaluable in improving the quality and rigor of our research. Here, we outline the substantive changes we will make to the manuscript based on the comments made by both reviewers. We address line-specific comments in the second section of this 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.

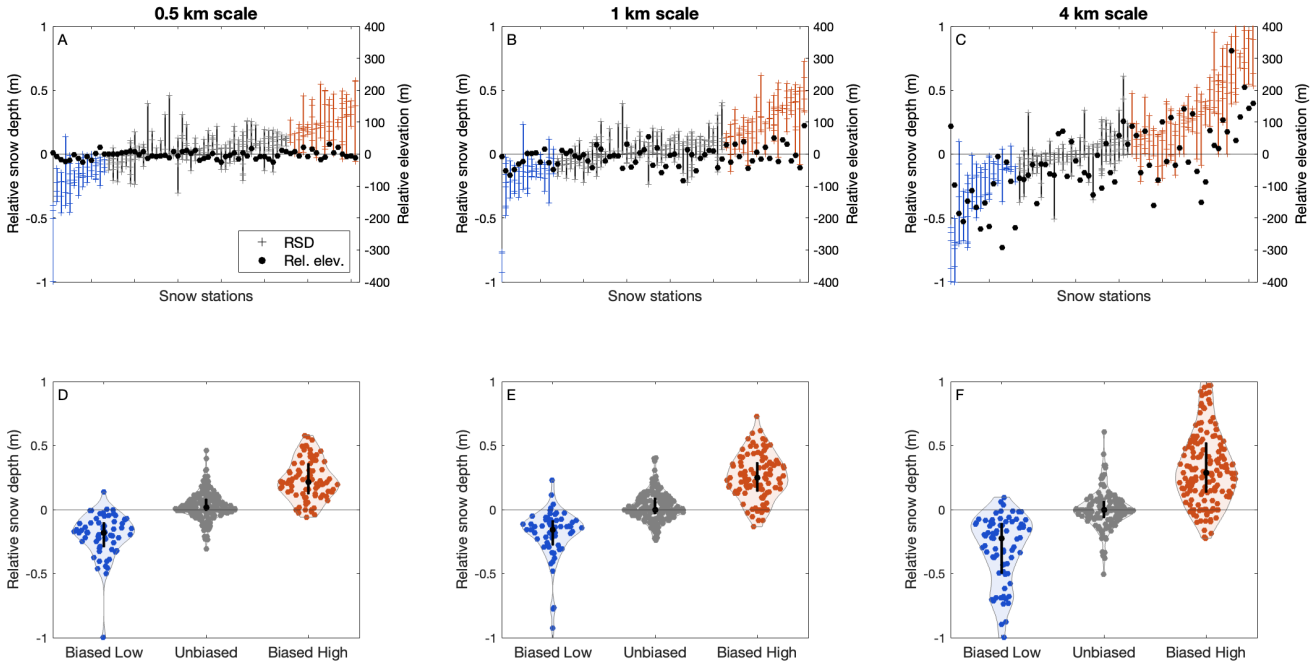
Organization

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

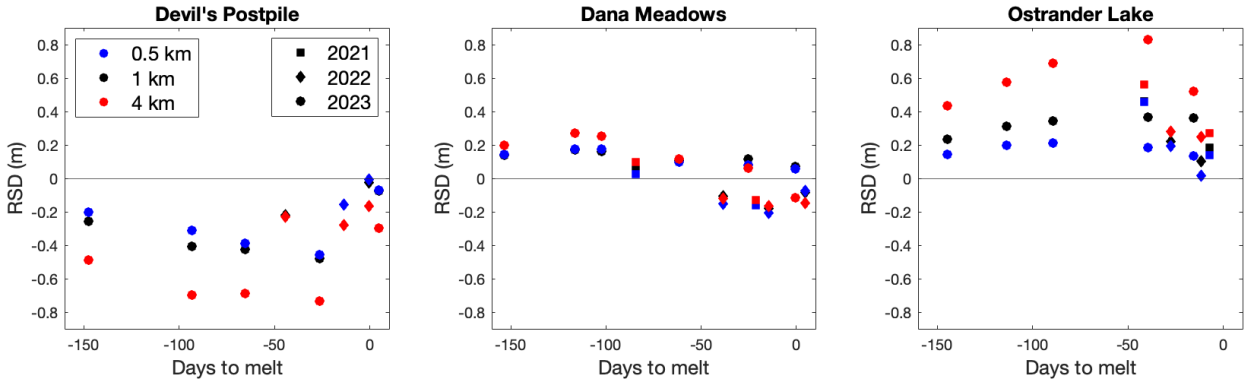
- The literature review presented in the introduction in the two paragraphs from lines 57-80 will be made more succinct. We will summarize the type of research that has been previously conducted in the area but will refrain describing the results in detail. Relevant results will be brought up in the 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.



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.

Will ensure to define this and use consistent terminology throughout.

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

Instead of “provides no advantage,” we will say “increases the potential error.” We discuss the QC and potential issues of CA snow depth data in line 150.

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. Either we would have to use the ASO SWE product (modeled, a black box), or calculate density from the snow station. 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.

We will delete the justification of our use of snow depth from the introduction.

Lines 150 - 152: Might deserve further discussion.

We will expand on our method of QC. We elected to use CA snow depth data because limiting sites to NRCS stations would reduce the number of basins in CA from 13 to 4. Visual inspection of post-QC station data and comparison with closest lidar SD pixels suggests that the CA data is viable.

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.

Will add ‘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. The 3 m product is higher resolution than necessary for most operational applications (see lines 187-189).

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 will delete 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 will 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.

To make our naming scheme more clear we will change '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.

We will delete these lines.

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

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' will be changed to 'representative'. The purpose of this section is to discuss that our choice of 10 cm is relatively arbitrary. We will provide further justification for our use of 10 cm as the threshold for representativeness.

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 end of this sentence will be changed to: 'both bilinearly resampled to 50 m to match the resolution of the Lidar data.' We will add subheadings to the analysis section to make the section as a whole flow better.

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 will use different line styles (solid, dashed, dotted) as to not interfere with the map colors.

We will plot the point snow depth values on figure 2D as vertical lines instead of individual points.

Line 229: Typo: extra d in “Dd Cumulative”

Will fix.

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.

We meant that the CDF is cut-off by the bounds of the x-axis. Based on the other reviewer’s comment we will expand the x-axis to include the entire bounds of the CDF, and remove the sentence about the truncated data.

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.

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 percentage rank of a point value within the CDF, not the percent difference. We will edit the text to make this more clear.

The purpose of the section is to 1) examine the distribution of snow depths surrounding a snow station, and 2) explain why percentile within the CDF is not an ideal metric for determining representativeness. We will update the first paragraph of section 3.1 to better explain why we conduct this analysis.

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

Line 264: stick to cm units for consistency

Will 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, this will be 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.

Will add: ‘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 will increase the transparency of the scatter plots in figure 6.

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

Will change 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.

We will update 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 will add a second sentence that describes how any bias correction would have to be site specific and require existing lidar data to do so.

We will re-word 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 presented in the results is presented to determine if the bias is 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 will add a section to the beginning of this paragraph stating something along the lines of, “Since snow depth bias is not caused by sampling error, what are the potential mechanisms which cause the high bias of 50 m SD values?”

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

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