Second Review for Herbert et al.
Hannah Besso
bessoh2@uw.edu

The authors have responded thoroughly to my review, improving the manuscript considerably. I don't need to see the manuscript again. It contributes significantly to our understanding of Snotel stations and snow variability in the Western U.S. I appreciate the extra analyses the authors have added since I last read the paper and I think the snow science community will benefit from this work. I have a list of suggested edits below, but none are required for publication. The one point I would like to emphasize for the authors to address is the framing around the motivation for using the 50 m Lidar pixel as a 'point' measurement. My opinion is that this value should be referred to as a 'pixel' instead of as a 'point', and that the motivation for this part of the analysis should be more clear.

Line 11: Sentence starting with "The nearest 50 m Lidar pixels had lower bias…" explicitly state what this bias is calculated in comparison to. This sentence makes it sound like you think 50 m lidar pixels are more accurate than snotel stations, so clarify what exactly you mean. A 50 m lidar pixel should be more representative of a larger area just by the nature of the 50 m pixel size as opposed to a point (or 3x3, as the case may be) measurement.

Line 14: Might want to mention aspect and vegetation here as the other variables you looked at.

Line 20: Maybe say 'comprises about half' instead of the 'majority'. The percentage is 53% according to Li et al., 2017, and the margin of error is likely such that there's a few percentage points of error on either side of that.

Line 86: Erroneous question mark after label for research question 4.

Section 2.1.2: I think you need to mention the relationship between the 3m and 50m ASO products. Is the 50m derived from the 3m gridded product, or are these independently calculating using the original lidar point cloud?

Line 178: It's unclear here how this second analysis of up to 8 km differs from the analysis explained in the previous paragraph up to 4 km.

Line 185: I think this needs slightly more clarification. Be clear about what the purpose of the 50 m Lidar pixel is, since the obvious comparison would just be the snotel point to the 3m Lidar pixel. So why do you use the 50 m Lidar grid cell throughout this paper? How does this help inform Snotel users? See my comment about this above, and my comment on the Conclusions.

Line 186: It seems a little odd to me to refer to point data at different scales. Point data is usually just a single point, not an aggregate of a 50m area. Is there a scenario where modelers would use a larger gridcell as a point measurement? If so, include something about this in your motivation for this section. This was an issue for me throughout the paper - if you feel you have

a strong reason for this then just be clear about it. Otherwise I don't think there's any downside to referring to this as a pixel instead of as a point.

Figures 2 & 3: For subfig A, I would recommend a non-diverging color scheme, like white to blue (using colors associated with snow), instead of one that emphasizes 0 values as very dark red.

Line 260: Is the 0.3 - 0.6 m depths true for all flights or just those near peak depth?

Lines 267 - 270: I'm not sure what these extra examples add. I don't think you need them. If you do keep them, I would include them in your reference on Line 265 where you reference Fig 4b,c and be more clear about what insight they add (are you saying they are additional examples of sites with a small range of values, or do they add something by being examples of sites with 'medium' ranges?).

Lines 264 - 277: I think a more clear way of illustrating this idea would be to combine these two paragraphs. I'm thinking along the lines of:

"The CDF plots demonstrate a range of possible scenarios created from different snow depth distributions. Sites characterized by lower snow depth variability (Fig. 4b, c, g, h) are less likely to have point snow depths far from the median due to the limited range of snow depths, while sites with higher snow depth variability (Fig. 4d, e, i, j) allow for greater differences between the median and point snow depth. For example, at the Michigan Creek Snotel site (Fig. 4b) the Lidar and Snotel point values correspond to the 7th and 95th percentiles, yet both values are within 0.1 m depth of the median value. Conversely, at Scotch Creek (Fig. 4e) the station SD (blue marker) is not representative as it is 0.46 m greater than the median snow depth value, near the 100th percentile of the areal distribution."

Line 294: I would change for clarity to: "Root-mean-square error (RMSE) is 0.46, 0.48, and 0.54 m for the same respective spatial scales."

Fig. 5: Looks like you lumped together results for CA and CO here. Do these distributions look similar between the two locations? In the other figures you separate the two, so maybe mention in the caption that results are for all sites.

Line 307: "when using 50 m SD as the point value" - noting the use of 'point' instead of 'pixel' in case you decide to change this terminology.

Line 324: Missing the word 'at'. Should read: "are that [at] the smaller scales".

Figure 6: Would prefer a more descriptive y axis title.

Line 346: Instead of 'higher snow depth', which is a weird phrase, I recommend: 'Thus, we conclude that the high bias reported by the station SD and 3 m SD is a result of true differences in snow depth at the station locations compared to the surrounding 50 m area.'

Section 3.4: I don't think you mention in the text that you use the 50 m Lidar pixel instead of the station, although this is in the caption for figure 8. Worth mentioning in the first paragraph.

Line 398: Do you mean higher magnitude relative elevation differences?

Line 408: When you refer here to the 'previous section', does this refer to the previous paragraph about timing or the rest of the results section? If it's referring to the previous paragraph, change the wording to 'the above paragraphs' since they're within the same section. If you're referring to the other results in the results section, make 'previous section' plural.

Line 424: Rewrite this sentence - it's ~50% of stations that exceed areal-mean by more than 10cm, and the percentage of stations that are not spatially representative are more than this. So be more precise with this wording perhaps by removing 'were not spatially representative'.

Line 425: Likewise, be more precise here: It's the finding that snow stations tend to be high-biased that is not unprecedented, not the tendency itself.

Line 435: Used 'point' to refer to the 50 m lidar pixel again here.

Section 41: Consider using another word instead of 'elevated', which makes me think it's related to elevation, and the term 'elevated snow depth' doesn't really make sense. Maybe instead say, for example on line 445: 'suggested that deeper snow at stations compared to the surrounding area was a result of flat terrain…'

Line 447-448: Fix the usage of commas.

Line 457: First sentence of 4.2: could use a few more citations for this broad statement.

Line 492: You mention the need for high resolution - could be helpful to restate the resolution of the vegetation product you used so we can compare that to the <5m resolution needed.

Line 493: Edit this sentence for clarity. Perhaps something like 'Here we used relative fractional vegetation as the metric to describe vegetation dynamics, which reduces these dynamics to a single value. This value may be insufficient to…'

Line 505: To be more succinct, use 'it is possible' instead of 'it is a possibility'.

Line 510: Use a hyphen: 'within-season'. Or just say 'inter- and intra-annual consistency'.

Line 530: Be more precise. The term 'When using the 50 m SD' leaves ambiguity. State what you are comparing more clearly.

Lines 530 - 538: I think what you're saying is that the 50 m Lidar pixel is a better representation of large-scale snow depth than the snotel data is, so if we can bias-correct snotel to better match the 50 m lidar pixel, then the bias-corrected snotel data will therefore better represent large-scale depth as well. I didn't feel like this motivation was clear throughout the paper (if I am indeed interpreting this correctly). I was unsure throughout the paper as to why you were using a 50 m Lidar pixel as 'point' data in addition to the snotel data. I think the framing around this needs to be better: you're proving what we intuitively know - which is that when you incorporate data from a larger area of the basin (50 m lidar pixel instead of 3m-scale snotel point), you will get a better estimate of the depth throughout the basin. Then you show that there is a temporally-consistent offset at the snotel point compared to that 50 m pixel - so that we can bias-correct the snotel data to better represent snow depth within a larger area. Throughout the paper, I felt like you were building a case for people to use 50 m lidar pixels instead of snotel data, which of course we can't do in places/times without lidar. Regardless of whether my interpretation of this is correct, please add clarity throughout the paper about the motivations here. Section 2.1.5 would be a good place to add detail.

Lines 551 - 554: This might be journal-specific, but it would make more sense to me if this was moved to the Data Availability section.