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.

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.

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.

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.

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.

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

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

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.

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?

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.

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.

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.

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.

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.

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?

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

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.

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.

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.

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

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.

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.

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.

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?

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.

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?

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.

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.

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?

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.

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?

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

Technical Comments

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.