The revised version of the paper “Using drones to map snow depth variability at cm resolution: an evaluation at peak accumulation “ by De Michele and others introduces the application of UASs to map snow depth in alpine settings. The author present a qualitative accuracy analysis of the summer DEM based on a topographic map and an evaluation of the accuracy of snow depth by comparing to few probe measurements. And they compare snow volumes derived from different interpolation techniques to the volume measured by UAS. The topic is very relevant and the next years will probably see many applications of UAS for snow cover mapping.

Most of the comments of the two reviewers have been addressed in the new version. An analysis on the aggregation of the data sets from cm to 100m cell resolution has been included which clearly improves the content of the paper by providing indication on required sample resolutions. Still, the implications of these results should be discussed in more detail.

However, the main shortcomings of the study which is the sparse data set used for the snow depth validation are still existing and cannot be solved in a simple review. A robust data set for validation would include many more points, more than a single winter survey and possibly even more study sites. In my opinion more of these surveys should have been performed before submission. This shortcoming reduces the impact of a very promising paper to a very preliminary piece of work (as the authors correctly state). I do not want to reject the paper as it is in principal very valuable. But I strongly suggest to perform more field campaigns in the coming winter and to resubmit an extended version of the paper in the next spring. In the meanwhile the “preliminary” results are still available to the public in the TCD manuscript. The design of additional field campaigns could benefit from the suggestions of the reviewers, it could even be considered to include state of the art areal snow depth measurements such as laser scanning for validation.

Specific comments:

P4 Study area: Please provide some more characteristics on the study site such as aspect, slope angles, vegetation cover and their variability

L179 why were three different resolutions extracted? What are advantages/disadvantages of the different resolution (processing)

L186: If a total station was already at site I wonder why no additional measurements of the snow surface were performed for the validation of the UAS data set. A total station allows to collect very accurate surface data (much more accurate than snow depth measurements by probing) and would have been a good supplement to the probing (such measurements have been used for the validation of terrestrial laser scanners by Prokop et al. 2008 and Grünewald et al. 2010)

L188 please add that the “horizontal” accuracy is meant

L207-2010; The authors state that there are guidelines in literature to define the number of samples but they do not apply them (also suggested by Reviewer 1). I wonder why. Of course, it is too late now but it could at least be tested and commented if the 12 samples were reasonable number or not.

L2018 The formulation is unclear, what is meant by “worst case” and why was it chosen?

L271ff the analysis of the contour lines just provide qualitative accuracy estimation. Only the comparison with the DEM gives some quantitative measure. This measure should be mentioned in the text and possibly a figure of the comparison could be added.

L284-286: what would be the consequence if no “brown” areas would be available. Would it reduce the accuracy? This is very important as one cannot assume to find such contrasts when mapping snow. It could be considered to calculate, show and discuss the differences between calculating the maps with “brown areas” and without such areas (only using the parts of the map which are white snow.

Table 1: 1) Root mean squared error or the absolute mean difference should also be provided; 2) A scatter plot would be more illustrative than a table. 3) observations 3 to 6 seem to be outliers in comparison to the other values. This illustrates one of the problems with only 12 measurements: More points would enable to clearly detect outliers . Is there an explanation for these outliers? Where are they located?

L339-41: This statement cannot be reasoned from the analysis: It can be said that 20cm resolution is not much worse than 10 cm but one does not know if the “trade off” could be even higher (e.g. 25, 50 or 100cm). What is the consequence if 20cm can be used instead of 10? How far does this reduce processing? Or does it imply that a higher flying altitude would be adequate (which would be good because one could cover larger areas)?

L361-62 I do not understand this sentence

L366f: Please reason why minimum and maximum appear to be constant. One would expect that also at the smaller cell sizes some degree of smoothing should result from the aggregation.

L370-371: this is a very interesting finding and it should be discussed in more detail what it means for future studies (e.g. that a sample resolution of 1m might be enough)

Technical comments:

L33 sun

L44 an

L123 $30*10^4$ m^2 poor diction write 30000 m^2 or 0.3 km^2 instead

L284 I think “interested” is the wrong word

L297 thickness

L361 cell size instead of “maps resolution”