Comments concerning "Spatially continous mapping of snow depth in high alpine catchments using digital photogrammetry " by Buhler et al.

General comments

The paper shows a comparison between a digital photogrammetric approach based on the LEICA ADS80 digital camera with other alternative techniques (TLS, dGNSS, GPR, avalanche probes) to generate snow depth maps in two sites of the Swiss alpine area. In particular an accuracy analysis was done for the single ADS80 DSM (reference is DTM ALS) and for winter-summer ADS80 DSMs difference (snow depth). Performances were finally compared with ground observations obtained by the above mentioned techniques.

The topic of paper can be considered pertinent respect to the journal goals and interesting from a technical point of view, especially for the fact that used data (in particular the ones from ADS80) are new generation ones. Considering the ordinary scientific level and the attention paid to applicative aspects, I consider this as a technical paper (not a research one) focused on the validation of photogrammetric products for snow depth mapping. Unfortunately the case study is not perfectly designed to achieve this task (see forward for motivations).

It is thus my opinion that, to be accepted, the content of the paper has to be heavily revised; important deficiencies, in fact, can be easily recognized. In particular photogrammetric concepts, that should be crucial for the study (this is the focus of the title, isn't it?), denote that a technical photogrammetric skill in the research group is not present. Many considerations and information that for a photogrammetrist are obviously needed and necessarily to be reported in a technical work, are missing (see forward for details).

Probably as obvious effects of this fact, authors discussions about DSMs accuracy evaluation show critical points especially related to error propagation along computations and error distribution interpretation (see forward for motivations).

Finally, even if I'm not a mother language, I suggest to revise the English because some grammatical errors are present. Especially technical terms concerning survey and photogrammetry should be revised accordingly to the conventional ones.

Specific comments

Design and description of the case study

The central point of the discussion is not to generate a snow depth map of the two study areas; differently, its main goal is to evaluate performances of a photogrammetric approach based on ADS80 data. My opinion is that, if this is the real aim, the design of the experiment should be better defined and described.

In particular it's my opinion that the position and distribution of ground observations has to be better characterized. Horizontal position declaration (not present in the work), for this type of tests, is not enough as instrument performances are highly dependent from height (m a.s.l) and slope of terrain, that should be taken into account during the accuracy evaluation. Thus I would greatly appreciate if, for all the ground measurements from the different proposed technologies (probes, dGNSS, TLS and GPR), an histogram was

presented showing frequency distribution of points respect to height and slope classes. My sensation is that ground observations are poorly representative of the general conditions of test sites, because concentrated in a very little height and slope range. Moreover, some of the existing ground observations used for validation are badly positioned as the authors themselves admit for TLS (lines 9-11 page 3311) and for GPR (line 19 page 3312), suggesting that ground survey campaign was not well programmed. My suggestion for this last problem is to eliminate those inconvenient points preventively from the test set without spending words on it.

Another critical point is the demonstration that no significant changes occurred in the periods 2010-2012 and 2012-2013 as winter DSM from ADS80 is 2012 and summer DSM is from 2010 and 2013 flights. At this point authors has to provide some evidences of no significant terrain changes reporting for example some references to official documents or others.

Photogrammetric aspects

The first evidence suggesting a low experience in digital photogrammetry is immediately present in the title. The "Spatially continuous mapping" concept is quite redundant and improper as a map is always a continuous representation of an area. We can discuss about the level of discretization (or, if you prefer of continuity) but it is quite sure that if you map a place you are representing it in a continuous way. Otherwise you have just a set of measures and not a map. Thus, the title could be better sound like …"Snow depth mapping… " .

I see that another referee already stressed the strange statement of authors about the economical convenience of ADS80 acquisitions respect to the ALS one. I agree with his comment, because it is really difficult to guess where costs can be reduced. An airplane has to take off and fly, the instrument is not economical and processing is time consuming in both the cases .. thus?

When describing spectral features of the camera (page 3302) provide information about wavelength of each available band of the sensor, and better focus on the importance of the NIR band to improve performance of ATE procedures. I see that this is the focus point, and not, like in many parts of the work is said, the 12 bits radiometric resolution. Snow, in fact, in the NIR band reduces its reflectance permitting a highest contrast of the image and consequently an improvement of ATE performance. However I agree that a further improvement comes from the 12 bit resolution as it improves the possibility of measuring littler radiance differences. In spite of this I retain not pertinent to spend words about this aspect without demonstrating by data the real improvement offered by the quality of the ADS80 data. I would limit the discussion stating that for this work the VNIR data from the leica AD80 digital camera were used.

When reporting methodology used to generate DSM from ADS80 data (page 3305 paragraph 4) it is very important to clearly indicate: a) the number and spatial distribution of Ground Control Points and, as more than one strip are used, the number and distribution of tie points; b) RMSE or similar metrics defining accuracy of adjustment (both horizontal and vertical), that is the one potentially affecting measurements made by stereo plotting or automatic triangulation from the adjusted stereo images ; c) GCPs and Check Points accuracy and source (do they come from GNSS ground survey? from an existing map or orthoimage? What else?..). please discuss this topic whose importance yourself recognize in the conclusion paragraph.

At page 2203 the statement concerning future Leica ADS100 is obsolete. ADS100 is now working. Moreover at this point the information is not important. Move this part in the Conclusion and further developments.

At page 3304, chapter 3.2.3 when describing TLS acquisition it is not clear at all the role of the coarse resolution respect to the final product. What is meant for "15 min" ? I think it is an angle measure (15') but it is not clear which is the full resolution of the system. Once the distance is fixed an estimation of n. points/m^2 is a better way to define the TLS resolution. Consider that this number can significantly vary depending on the shape of the imaged surface, thus just report an average point density. In the same paragraph in place of "scans which showed …" use "points which showed …" because the term "scans" is generally used to define a group of points obtained by scanning.

At page 3305 while speaking about Trimble (not Tribel!!) Geoexplorer authors use the acronym DGPS in place of dGNSS (like previously said) again showing a confused way to describe survey related topics. What do you exactly mean for dGNSS? A Virtual reference station acquisition (VRS) that is a RTK approach based on signal phases differencing or a post processing cod differencing approach? Trimble GEoexplorer can just manage code maesurements. Discuss better.

At page 3305 ch. 4, I suggest to avoid any general listing of parameters required by ATE. For each required parameter, the set value has to be reported.

At page 3306 it is not clear to me how DSM tile representing the same surface seen from different points of view can generate different terrain mean slope. Please clarify. The slope is referred to….? Terrain slope? Image tilt? What else .. ?

At page 3308 authors describe the way they used to evaluate ADS80 DSM potential accuracy by comparing 2010 and 2013 DSMs with a ALS generated one (2009). It is not clear why the authors used the acronym DSMs for the ones generated from ADS80 camera and DTM for the one the aerial laser scanning acquisition. Are they really a DSM and a DTM? I remind that DSM and DTM define two drastically different surfaces. The first one describe the whole of bare ground and of the above ground objects (where present), while DTM describe just the bare ground surface (cutting out overlaying objects). Even if authors state that they masked out vegetation and buildings I suggest to better face and discuss this topic. Moreover at this point I suggest to declare which type of data were compared (cloud points? Grid data?) And finally: where does the reference ALS DTM come from? Technical features? Please provide these infos.

In addition, at the end of paragraph 5, authors use the DEM acronym to probably define the same type of data. Please, try to be more rigorous and constant in your work. Otherwise the idea is that authors have confused ideas about this type of data.

In the same paragraph, again, is not reported if the DSM of the same area generated at different times were jointly adjusted (multi temporal block adjustment) or singularly. In this case GCPs remained the same? Accuracies of each adjusted stereo model? Discuss this.

In table 2 authors use the term of "correlated" and "interpolated" points to make the reader aware of the fact that some points generated by ATE module are not directly measured, but derived by spatial interpolation. I suggest to use the term of "measured points" in place of "correlated".

At page 3307 (and in the conclusions) authors say that the "final orientation accuracy is 1GSD" . This is a very unconventional way to state accuracy after image adjustment. Authors should report separately vertical and horizontal accuracy (of check points or the one resulting from a one-leave-out cross validation approach). This is basic to completely describe the data they are going to validate.

<u>Error analysis</u>

First comment concerns Figure 10 that shows correlations between snow depth measurement coming from ADS80 DSMs and other techniques. I wonder to see that the comparison for GPR is limited at the range 1-2 m (why at page 3305 do the authors say that GPR explore up to 2.70 m?) page , while other techniques explore a wider range of measurements 1-3. This makes the evaluations not comparable and not homogeneous. Discuss it.

I retain that the smoothing step (3x3 kernel) operated on the measured ADS80 DSMs is a critical point for a work that try to compare the accuracy of data. Once applied a filter changes the measured values thus making the following comparisons not reliable to demonstrate the potentialities of the adopted technique. If the authors' will is to maintain such an approach to recover a better continuity of the snow surface, they need to demonstrate that the filtering step introduce a deviations from the original measurements lower than the obtained accuracy (as defined during the adjustment/ATE).

At page 3308 authors present a comparison aimed at defining the accuracy of ADS80 DSMs versus an available ALS DSMs. I repeat here that it is mandatory to define all the technical features of the ALS DSM. Moreover, as horizontal spatial coherence between the compared DSMs heavily conditions height differences computations, while doing such a test the two compared DSMs should preventively suffer from a 3D least square adjustment (or ICP) to minimize displacement effects.

At pages 3309-3310 authors present some operations they did to exclude outliers from snow depth map. They assume that negative values higher than 0.5 m has to be considered as 0. My questions are following: Why did authors choose -0.5 m as reference value? Does it come from an accuracy assessment concerning the data ? In this case all the measurement having a positive value lower than 0.5 m should be set to 0 too.

The reference value defining which differences can be considered significant and which not can be obtained applying the ordinary variance propagation law. It states that if the accuracy of the compared measurements is known, the theoretical accuracy of their difference can be estimated as sigma(dh) = (sigma(h1)^2+sigma(h2)^2)^0.5. where sigma(h1) and sigma(h2) are the accuracy for the differenced DSMs.

The outlier problem instead is something different. Please try to justify through scientific motivations the reference values you choose for outliers (> 15 m and < -0.5 m). I personally retain that, dealing with mapping, the best way to recognize outliers is to proceed with neighborhood operators applied to the point clouds or grids.

At page 3311 (TLS paragraph) I cannot well interpret the sentence "Three negative deviations…". Can you clarify?

At page 3312 (hand measured plots paragraph) it is not clear the meaning in terms of practical aspects of the statistics MEAN and STD of the RMSE and NMAD values of the plots. I suppose that the mean value defines the uncertainty of the measure while the std value just demonstrates that the mean value is significant, that is appreciable (in fact the sensibility of the measure, in this case the difference, is given by the std value). Please discuss a little bit more.

At paragraph 5.3.4 authors recognize that the GPR survey suffered from some limitations. This seems to be mainly related to a bad design of survey. The only justification for this fact is that the authors used a set of measurements surveyed for a different task. Can you give some alternative reasons?

In the conclusions authors spent many words summarizing limits and potentialities of this approach. The main reference data at this point they refer their conclusions to is the ALS. My suggestion is to recover here the importance authors assigned to the role of the other survey techniques, otherwise the reader cannot appreciate the added value they gave to the paper.

In the conclusions authors once more stress the limitations the measurements suffer from in steep slope areas. My suggestion is again to complete their work by mapping test sites in terms of slope and demonstrating with statistics relating slope and errors that this is a limitation also for their case study.

Figures

Figure 6 is irrelevant. If authors want to better explain the effect of slope on measurements they have to present an horizontal map of slope where the distribution of measured and interpolated points can be observed.

Final comments

In my opinion the paper cannot be accepted in this form. The subject is interesting and the great variety of available data gives huge scientific potentialities to the paper. But, to achieve this task, authors have to heavily re-organize the paper. They have to better focus on the specific role offered by the different validation datasets, stressing weaknesses and strengths of the photogrammetric approach respect to each of them. Photogrammetric and survey (especially GNSS) aspects have to be heavily integrated, completed and corrected. At the same way error statistics should be better presented and managed.