Dear reviewer,

We appreciate the thorough review and we believe the manuscript will improve as a result of your suggestions. Below we have addressed all your comments point-by-point and our reply is provided in *Italics*.

With kind regards,

Walter Immerzeel

General comments

The paper presents a UAV survey conducted in September 2015 of a mountain glacier in central Norway (Storbreen) to assess its geodetic mass balance. The authors present their survey, the challenges encountered and their quality estimation methods, then difference their DEM to a LIDAR DEM from 2009 and derive a difference of DEM (DoD) from which an estimate of the mass balance is done. Comparison between their data and stake mass balance is made and the authors conclude that the values are in agreement. The use of UAV to monitor glaciers is indeed very promising, though it is not undocumented (publications can be found easily as early as 2013, for instance [Whithead et al, 2013, doi:10.5194/tc-7-1879-2013]. One can also find records of AGU session dealing with it (AGU13 C41B ; AGU14 C31A C21D ; AGU15 C33G C41D)). The very same method has also been applied extensively to the analysis of other geomorphological feature that do not present different challenges (landslides for instance, see [Niethammer et al, 2012, http://dx.doi.org/10.1016/j.enggeo.2011.03.012] or [Lucieer et al, 2014, https://doi.org/10.1177/0309133313515293]; landslides are, photogrammetically speaking, slightly darker, slightly rougher surface than glaciers), so I do not believe that discussing the method of UAV survey without introducing a novel element in the approach is necessary.

Thank you. We are aware of most of those studies, but tried to keep the focus on the glacier applications. We will modify the introduction and include the suggested references. We do not go into far detail regarding the methodology and we mostly refer to existing studies. The scientific merit of our study is the combination with field observation, the comparison with the LIDAR dataset and the application on a fresh snow surface.

A number of comments and findings about the processing of UAV photography are not only not novel, but common knowledge taught in introductory courses in photogrammetry. For instance the statements about Ground Control Points (GCPs) in the paragraph starting line 311 : the proper number and spatial distribution of GCPs is not anymore a topic of scientific debate but the subject of blog posts (see for instance https://www.verticalaspect.com/gcp-improving-uav-survey-accuracy/). The paper provides the usual guidelines about GCPs in the conclusion (lines 394-405), but they were not applied during the survey itself.

I think the GCP statement and optimal planning of a field campaign is well suited to be part of the paper. It is important information for future field campaigns and the operational glacier monitoring of NVE.

Glacier field work is difficult and very different from remote sensing analysis and therefore any knowledge to make the field trips more efficient is useful and worth publishing to our opinion. Furthermore, blog posts and AGU sessions (as mentioned in the previous paragraph) are not considered peer-reviewed literature.

The invention of the concept of SfM photogrammetry is wrongly attributed (line 79) to Westoby, who simply applied it to geoscience and wrote about it and (line 165) to Szelisky, who wrote a book chapter about it (this is closer to be an adequate citation). A more correct citation would be (Snavelly et al,2006 doi:10.1145/1179352.1141964), this being the paper that presented Bundler, the first widely distributed SfM software.

Matt Westoby did a lot of work in this field, and to our opinion was one of the people who made the use of SfM mainstream in geosciences. The Snavelly paper was not related to geosciences, so something can be said for all of them. In the revised version we will add all three references.

Specific comments

Abstract, line 19 : change to “Routinely and accurately monitoring the outlines[. . .] is essential [for?]”.

The line will be removed.

Section 1, line 50: “Surveys can be terrestrial, airborne of from space” -> repetition from previous sentence, remove.

OK.

Line 81-82: What do you mean by “systematically tested” concerning the generation of DEM over snow surfaces? What would you suggest as an approach to perform such a test? Aren’t the papers cited lines 85-86 doing precisely that?

We will reformulate this.

Section 3.2 line 135 : What customized software? Your own homebrew? Magic Lantern?

The firmware is modified by the UAV manufacturer such that the camera can be triggered by the UAV and the pictures are geotagged. We will clarify this in the revised manuscript.

Line 138 : “Real world coordinate system 31 markers[. . .]” -> replace “real world coordinate system” by “cartographic projection” or “cartographic coordinate system”, or even indicate what projection you will use (UTM 32N or maybe the appropriate zone in the local Norwegian system NTM). Also, add a comma before “31”, I thought you were referencing the “Real world coordinate system 31”.

This will be rephrased.

In sections 3.3 and 3.6, the authors note that the analysis did not take into account the margins of the glacier, "due to steep slopes". A quick check of the GCPs positions and of the flight line indicates that some of the margin are neither well covered by the GCPs nor by the images (a safer approach would

have been to survey a bit more terrain on the edge, it would also help with georeferencing and with accuracy estimates). Best practices in DEM differencing over glaciers were laid out in [Frank et al, 2013, http://dx.doi.org/10.1016/j.rse.2013.07.043].

We agree that this would have been better; however the field reality was different. The margins of the glacier are close to steep slopes with strong upward winds and therefore it is not safe to fly the UAV near or over the margins of the glacier in many cases. This problem obviously does not occur when satellite imagery is used.

In section 4, line 254 to 263, the authors explanation on why the method yielded useful data over the snow covered areas which is somewhat misleading. Firstly, the SIFT algorithm (adding a citation to [Lowe D., 2004] wouldn't hurt) is used to find tie points, not to perform the dense reconstruction (this is the job of dense multi-stereo crosscorrelation). The SIFT algorithm may indeed in this case perform better than expected over the "pristine white upper part of the glacier", but actually visualizing the tie point distribution would be needed to assess this (in the software used here, Agisoft Photoscan : File/Export Points/ -> select Spare Cloud). Since the orientation was successful, one can infer that SIFT was indeed successful enough to tie all the images together. The high flight height gives a large footprint but a lower resolution, which might introduce problems to SIFT and to the correlation since contrast on pristine snow is mostly coming from small scale structures. Flying lower should actually help in that respect. However, high flights can increase the probability of each image to see part of the terrain presenting contrasted structure. Choosing the best flight height is therefore very dependent on the terrain and the local contrast present at the time of acquisition. The paragraph ends saying that "over saturated pictures" are not an issue, but they are, and if no issues was detected in this survey, it's because the images were NOT saturated (saturated = presence of large area where the image is 100% white -> no tie point and no accurate correlation).

In the revised version we will add the references and rephrase the paragraph and make the distinction between SIFT, point cloud densification and saturation more explicit.

The next paragraph discusses the DEM accuracy, something already discussed in part in section 3.3, the paper would be clearer without such repetition. I think the use of a long GNSS track to derive check point is clever in principle, but it doesn't help with the absence of GCPs, stake measurement of other checks in the southern, highest part of the glacier (where errors are probably highest).

In 3.3 we describe the method of the accuracy assessment, whereas in section 4 we describe the results of this assessment. Since the paper is for a large part methodological we propose to keep this as is.

As for the accuracy interpretation: we have masked out (Figure 7) those areas close to the margin or with an unrealistically high RMSE. Again the field reality dictated where it was safe to walk or acquire stake observations, and there is little that can be changed. We believe the amount of field data to validate the UAV data is already quite impressive compared with other studies.

Line 304-310 : The point cloud is not "geometrically corrected", it is subjected to a 7-parameter transformation (rotation, translation and scale). GCPs could be used to re-assess the camera calibration and their relative positions but it is not discussed here (and I do not know to what extent Agisoft Photoscan uses GCPs to re-evaluate the bundle adjustment).

A 7-parameter transformation of the spare point cloud is to our opinion a geometric correction, but we will rephrase this more specifically in the revised version. The GCPs are indeed used to correct the camera positions in AgiSoft.

The claim that the North-South gradient found in the off glacier, out of the convex hull of the GCPs has "no bearing on the on-glacier accuracy" is dubious at best, since it may be that the whole scene is tilted or domed compared to the real topography [James and Robson, 2014, http://dx.doi.org/10.1002/esp.3609], which would lead to local over/underestimations of the elevation change. Keeping the off glacier area and using them as additional control (by extracting GCPs from the LIDAR for instance), or by applying a global tilt to the data to fit the extremities might be better approaches.

We will include the reference in the revised version and we will add this caveat. We believe that the doming or tilting occurs mostly in areas which are not constrained by GCPs (James and Robson, 2014) and therefore the observed tilt applies to the off glacier area. For future applications we will include also GCPs in this area to validate this.

Line 311-322 : as stated above, the number and distribution of GCPs for a survey has well known rules and effects when the rules are not followed. The experiment choosing random points fails at showing the importance of the number of GCPs, but rather is an example that shows that the distribution matters most. Conducting the experiment while choosing the points so they are "the 5/10/20 best distributed points" would show more clearly the effect of an increasing number of points. However, this experiment does not need to be conducted since its results are already known.

OK, we will rephrase this and focus mostly on the distribution.

Page 15-16 : The method to extract mass balance is well described and well used, but however not novel. As stated in the referee comment #1, the analysis on that part is a bit short.

Ok we will elaborate on this part in the revised version.

Figure 7 is really hard to read (both in color and in black and white), the contrast should be higher and the DoD values less transparent. Maybe try to show the mask differently, like a hashed area for instance?

OK we will improve the figure and add more contrast and show the mask differently.