Author's Response for the manuscript "Winter mass balance of Drangajökull ice cap (NW Iceland) derived from satellite sub-meter stereo images"

This text contains the author's response to the reviewers Ruzica Dadic and David Shean and the list of all relevant changes made in the manuscript. A major reorganization and proofread of the manuscript has been done as requested by David Shean. Since it is tedious to show the reorganization and proofread, we do not include a marked-up manuscript version in this document, although this can be added if requested. All the relevant changes made in the manuscript are however clearly indicated throughout the response to David Shean.

MS revised using an updated version of the Pléiades images

After submitting the TCD paper, we received an updated version of the Pléiades stereo-images. Further analysis revealed that some of the DEMs derived from Pléiades-1A images contained significant planar tilts. This is due to errors in the Rational Polynomial Coefficients (RPCs), introduced by the ground segment when trying to improve the geolocation of the Pléiades-1A images. The twin Pléiades-1B satellite is not affected. This is not the scope of this response letter to address in depth these technical issues. More details are available upon request to Etienne Berthier.

The new images have been reprocessed for the revised version of the Manuscript (MS), all calculations have been redone, and the results updated. The statistics, mass balance and validation with in situ snow thickness are virtually unchanged compared to the TCD paper, an indication of the robustness of our processing strategy.

A new correction for snow densification included

We also have realized that an important correction was omitted in the TCD paper. Indeed, we ignored the influence of the densification of the fresh snow that fell in the days/weeks preceding the acquisition of the reference DEM (October 2014 Pléiades DEM). An important message is that the calculations of geodetic winter mass balance should be carried out using a reference DEM acquired after the melt season and before the first significant snowfalls, so as close as possible from the beginning of the winter. This is now added in section 3.2.4.

Comment about the details in glaciological properties

Often the reviewer David Shean asks for detailed explanations on the corrections applied (e.g., the firn densification model, the vertical velocities derived from ice flow and from mass balance). We appreciate the reviewer's attention to detail and desire for a more thorough description of the corrections. While we replied to these comments individually, we feel that including in the MS a detailed explanation for each correction would result in a convoluted manuscript that would not focus on the most important messages/ methods we present. This approach is supported by the fact that the corrections only slightly adjust the calculation of glacier-wide geodetic winter mass balance, as we show in the manuscript. Additionally, while many assumptions and simplifications are made, we clearly state and acknowledge them in the manuscript. We also always use conservative error estimates when applying each assumption.

Each question of the reviewer is written below in black color. The author's response of each question is indicated in blue color.

Response to Ruzica Dadic

This paper is an excellent contribution towards continuous monitoring of glacier changes, especially in areas where data are scarce. While the authors acknowledge the uncertainties that accompany mass balance estimates from space, such as lack of density information, firn compaction and elevation changes due to ice dynamics, they also give clear guidelines on how to reduce those uncertainties and discuss the relative importance of them. This paper is well written and the methods are clear, easy to follow and repeatable. I look forward to seeing this paper published.
We thank Ruzica Dadic for her very positive feedback about our submitted manuscript.

Response to David Shean

This manuscript describes the application of repeat high-resolution DEMs to estimate the 2014-2015 winter mass balance for the Drangojokull ice cap in Iceland. Three high-resolution satellite DEMs from Pleiades and WorldView-2 sensors were processed, co-registered, and differenced to obtain maps of elevation change for two time periods. These measurements were combined with in situ snow cores, independent mass balance observations, a simple firn compaction model, and flow model velocity output to constrain estimates for glacier-wide average mass balance. The overall goal is to demonstrate that winter balance measurements can be extracted from remote sensing observations without extensive in situ observations. This is a challenging thing to do, and while there are several simplifying assumptions required, this paper outlines multiple viable approaches. The paper requires some significant revisions, but I believe it will be a valuable contribution.
We thank David Shean for his thorough review of the paper. His comments have greatly improved the paper reading and organization. The revised manuscript (MS) has been also read proof by Ian Howat and Leif Anderson, both native English speakers.

General comments:

The manuscript should be thoroughly proofread by a native english speaker (coauthor or colleague), as there are grammatical errors and confusing sentences that require attention. I corrected some of these errors, but stopped after a few pages, realizing the amount of work that would be required. This review should have occurred before submission.
Coauthors Leif S. Anderson and Ian M. Howat (both native English speakers) have now proofread carefully the manuscript, correcting grammatical errors.

The abstract needs some work. It includes too much detail on the methodology (the relative adjustment sentence is peripheral to the main purpose of the paper). The abstract mentions firn compaction and ice dynamics, but doesn't communicate that these contributions are estimated and accounted for to obtain mass balance numbers from observed elevation change. The abstract should mention that the DEM differences capture dh/dt spatial variability for large areas, while the in situ measurements provide calibration and validation.
The abstract has been edited and simplified, stating that in situ measurements are used for calibration (density & firn densification) and for validation (snow thickness).

The paper requires some reorganization. Methods, results and discussion are intermixed, jumping between different processes, different measurements and different assumptions. Firn densification and firn compaction are presented in two separate sections. Observational data (GPS, bedrock DEMs) are introduced well into the methods section. Results are presented throughout the methods section. One of the senior authors should provide guidance on organization.
Extensive work has been done in the revised MS to improve the organization. Among these changes, we have:

1) Added in the data section the reference for the bedrock DEM (this was a work in progress at the time of submission of the TCD paper), and mentioned the GPS data used for calibration of the ice flow model.
2) Added a paragraph explaining the structure of the method section (section 3)
3) Unified all bulk snow density used in a single section (section 3.2.2)
4) Merged the sections discussing firn densification (section 3.2.3)
5) Added a section describing fresh snow densification (section 3.2.4)
6) Rewritten the method of quantification of snow fallen and melted from simple degree-day modelling and precipitation scaling (now in section 3.2.4.)
7) Moved numbers to the result section (section 4)

The organization of the remote sensing methods section needs work - should describe the GCP identification, bundle adjustment and DEM generation methodologies before any discussion of co-registration.

Section 3.1 (methods in remote sensing) has been rewritten. We prefer to keep the DEM creation method as it was in the TCD MS because GCP identification is only followed in the Scheme A, and co-registration with pc_align is only followed in the Scheme B.

I suggest that the authors review the text and reorganize to clearly isolate:
-observations (elevation, elevation difference, volume change) and associated uncertainty for each
-inferred or derived values (mwe based on density assumptions) and associated uncertainty
Final uncertainty estimates should include cumulative error from both.

To address this, we added an introduction to the methods section delineating what is observed from remote sensing and what is inferred from field measurements. Table 3 has been also reorganized, and is captioned accordingly. The uncertainty estimates are calculated as a cumulative error of the four variables affecting the geodetic mass balance (i.e., elevation difference, density, firn and fresh snow densification).

The description of the in situ measurements should be cleaned up. Several sentences state that a data product "was used" but no information is provided about what it was used for, or why it was used. In general, I would like to see more background information about the various in situ measurements and how the "raw" in situ measurements were converted for comparison with the remote sensing data. In multiple sections it wasn't clear that an "apples to apples" comparison was being made.

We expanded the descriptions of the field measurement methods as suggested. We especially focused on the snow density and total snow thickness measurements. These adjustments should also clarify the comparison of in situ and remote sensing measurements.

Throughout the paper, it is sometimes unclear whether the in situ measurements are being used for calibration (correction of remote sensing data) or validation (comparison with remote sensing data, error evaluation). A number of different in situ observations are presented throughout the text, with some used for calibration and some for validation. Calibration discussion should come first, with comparisons and validation later. If the same measurements are used for both calibration and validation, this should be clearly stated.

We appreciate this suggestion and added a clarifying statement in the introduction of the method section and in the discussion section. We emphasize that the in situ measurements have two functions: (1) constraint for calculation of glacier-wide geodetic winter mass balance (by using measured density and the firn and fresh snow densification) and (2) validation of the remote sensing-derived elevation changes (by comparing elevation changes with snow thickness measured).

It seems that one of the goals with the two DEM processing schemes was to demonstrate that GCP identification and bundle adjustment is unnecessary if the output DEMs will be coregistered. This is good to know, but not a novel conclusion. The authors should state before any discussion that GCPs were used during bundle adjustment to improve absolute image geolocation prior to DEM generation.

This is a slight misunderstanding as the scheme A does not involve posterior co-registration (see Fig. 3 with workflow, where no ICP co-registration is carried out in scheme A). We rather want to prove that both schemes are useful for obtaining unbiased elevation differences, and that each of them has pros and cons as presented in the discussion.

There were some ambiguous units of meters - should carefully review the paper to make sure this is always clear. Meters of elevation change, meters of snow thickness, meters ice equivalent, meters water equivalent, etc.

We have checked and corrected the units in the revised MS.

The "in situ mass balance map" from 2013-2014 is described without any pertinent information about derivation and accuracy, and no citation. Since this is unpublished, I recommend that a figure be included with improved description about how the map was derived (e.g., interpolation method used).

This map is part of a work in process where some coauthors are interpreting all the in situ mass balance measurements available in Drangajökull. Therefore we cannot include this figure in this paper. We however acknowledge that further information is needed in this section, and we have included details in the interpolation method by adding a reference (Pálsson et al., 2012) which explains this on a similar work carried out in Langjökull ice cap.

A fundamental assumption throughout the paper is that the 2013-2014 annual mass balance is representative of long-term mass balance for the ice cap. The authors should provide additional supporting evidence for this assumption. Looking at Figure 7, the measured density values from 2013/2014 appear anomalous compared to other years in the record. How does this year compare to long-term records from the AWS measurements presented, and to other AWS observations in Iceland? What about reanalysis data (e.g., ERA-Interim)?

We acknowledge that Drangajökull ice cap has some year to year variability (as most glaciers) and it is fair to discuss the assumption of one individual year as representative for mass balance. We have analyzed the series of two in situ net mass balance with highest accumulation where the correction for firn densification and submergence velocities is the also the largest: V3 and V7 (Fig. 1) starting from 2004-2005, and calculated the firn densification and emergence rates as we did for the 2013-2014 campaign, observing that, while there is a clear year to year variability, the obtained mean value of firn densification and emergence rates are in good agreement with the results from 2013-2014 and the differences are well within error bars.

Table R1: Annual mass balance (b_n), and correction for firn densification ($C\{h_{Firn}\}$) and submergence velocities ($dh_{dyn\,bn}$) calculated for t_1-t_3 (October to May), at locations V3 and V7, from all available in situ net mass balance measurements. Some of the years are not included due to logistics and bad weather at the time of the field campaign.

Glaciolocal year	V3			V7		
	b_n (m$_{w.e.}$)	$C\{h_{Firn}\}$ (m)	$dh_{dyn\,bn}$ (m)	b_n (m$_{w.e.}$)	$C\{h_{Firn}\}$ (m)	$dh_{dyn\,bn}$ (m)
2004-2005	-0.03	0.00	-0.02	1.48	0.50	0.99
2005-2006	1.83	0.61	1.23	2.42	0.81	1.62
2006-2007	1.22	0.41	0.82	0.46	0.15	0.31
2009-2010	0.41	0.14	0.27	0.52	0.17	0.35
2011-2012	1.31	0.44	0.88	2.06	0.69	1.38
2013-2014	**1.74**	**0.58**	**1.17**	**1.32**	**0.44**	**0.88**
2014-2015	2.80	0.94	1.88	3.06	1.03	2.05
Mean	**1.33**	**0.45**	**0.89**	**1.62**	**0.54**	**1.08**
Mean-2013-2014	-0.41	-0.14	-0.28	0.30	0.10	0.20

Besides, the year 2013-2014 was chosen for corrections since it is the year where the most complete in situ mass balance data was collected, including GPR profiles that helped interpolation of the in-situ points. Previous years records were only based on two profiles of snow cores (V1-V9 in Fig. 1). We acknowledge that our firn model based on a single year of mass balance data is a simple approach by adding conservative (large) error bars.

Answering the two other questions of this point:

1) The density values from 2013-2014 are indeed lower than average (520 kg/m^3 in 2014 vs 548 kg/m^3 as average 2005-2015), but we attribute this difference to the early date of the field campaign in 2014, which was carried out by the end of March 2014. This is now stated in the revised MS (section 3.2.2).

2) ERA-Interim reanalysis data has already been examined over the Drangajökull Ice Cap for a previous study (Magnússon et al., 2016a) and the results are a strong underestimation of precipitation, probably due to the poor representation of the complex topography in a coarse weather forecast model. A sentence in this context can be seen in Magnússon et al. (2016a): "The modeled winter precipitation may, however, not be representative for winter accumulation due to excess lee drying in the modeled precipitation"

3) Some nearby meteorological stations are available at intermediate (~40 km) distances, but to use this data for such a scope requires extensive work in mass balance modelling, since our approach of degree day modeling and precipitation scaling is not valid on the long term mass balance (for example, the ddf needs to change through the year, considering different sun exposure angles).

The authors use DEMs derived from 3 different sensors (airborne lidar, WorldView-2 and Pleiades) and 4 different processing approaches (lidar gridding, ERDAS Imagine, ASP, and SETSM). Many of the comparisons between the different "schemes" are not appropriate given the different processing methods used for DEM generation (Imagine vs. ASP). I would exercise caution when making statements about relative accuracy from intercomparisons, considering the different processing approaches involved. With that said, the authors can potentially turn this criticism around and emphasize that their DEM differencing and mass balance methodology should work for DEMs with different source and processing methodology.

We agree and we have added a sentence about this in the discussion section (P15L8) of the revised MS: "The three different processing software (ERDAS Imagine, ASP and SETSM) proved satisfactory results in the resulting dDEMs."

It would be useful if the authors consistently specified the source of the DEMs throughout the paper, rather than just DEM date, especially when discussing the co-registration (e.g., "October 2014 Pleiades DEM" rather than "October 2014 DEM").

This is a good suggestion and the MS has been updated throughout.

It seems to me that a dynamic firn model like IMAU-FDM forced by RACMO SMB would be more appropriate than many of the empirical and scaling approaches used in section 3.2 and 3.3 in this study. See (Ligtenberg et al., 2011). There should be existing RACMO/FDM products over Iceland, although I have not personally evaluated their quality.

We appreciate this suggestion but we have a few reservations about using RACMO output in this case: 1) RACMO2.3 SMB has a spatial resolution of 11 km (Noël et al., 2015), which is too coarse to be useful for Drangajökull (with a ~140 km^2 area); and 2) Unfortunately the IMAU-FDM model has never been validated in Iceland. While it would be ideal for us to use a more complicated estimation scheme of firn densification we prefer to use our simple but robust approach and assert considerable, conservative (50%) error bars to honor the uncertainty inherent in our approach. We added a comment to the discussion suggesting that a dynamic firn model could be used in a similar way for larger glaciated areas (e.g., on the Greenland ice sheet).

The choice to mask areas with high slopes (>20°) and in shadows before computing statistics gave me pause. First of all, the 20° threshold is somewhat arbitrary. Why was this chosen? Why not 30°? Second, the 11 and 12-bit DN range for the Pleiades and WV-2 sensors should provide sufficient contrast in shadowed areas (see discussion of shadows in Shean et al, 2016).
The original DEMs are not presented in any figures, and the reader has no way to evaluate the effects of the masking or the interpolation routines used to fill resulting data gaps. See notes on Figure 2 for recommendation.

In the revised MS, we state and acknowledge that the statistics shown are valid for this specific scenario of gentle slopes <20° and without shadows. The mean slope of the ice cap is 6.2° and <1% of the ice cap has slope >20°. We consider this filter representative for our volume calculations inside the ice cap. The text in the MS has been edited and this information has been specified. We have also included the original DEMs as hillshades in a second panel in Figure 2.

Regarding the shadows, we obtain correlation in many shadow areas. For "dense" and large shadows we however end up losing the texture, despite the high DN range, producing gaps in the DEM. Below two examples, for both Pléiades and WV images. This only occurs in the October Pléiades images and the February WV2 images (and not in the May Pléiades images), likely due to the very low sun angle at these dates.

In addition, we observe a higher level of noise in shadow areas than in the sun-exposed areas. Since the big majority of ice cap is exposed to sun (<1% of shadows inside the ice cap in October; 3.8% shadows inside the ice cap in February), we consider that statistics including shadows would not be representative for the volume calculations inside the ice cap.
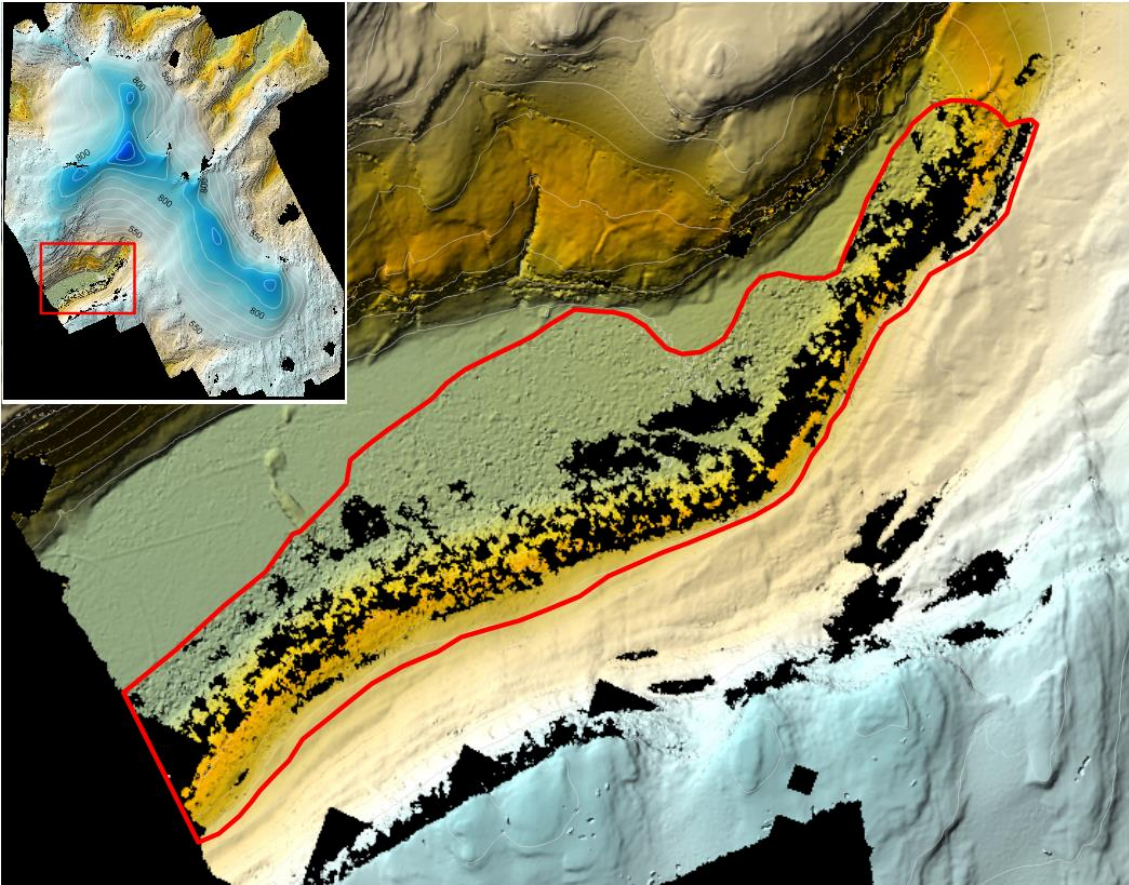
Fig. R1: 2014 Pléiades October DEM. Example of shadow areas outside the ice cap, in SW region of the ice cap. Red polygon delineates a shadow area at the north side of a cliff. These areas are affected by higher level of noise as well as data gaps.
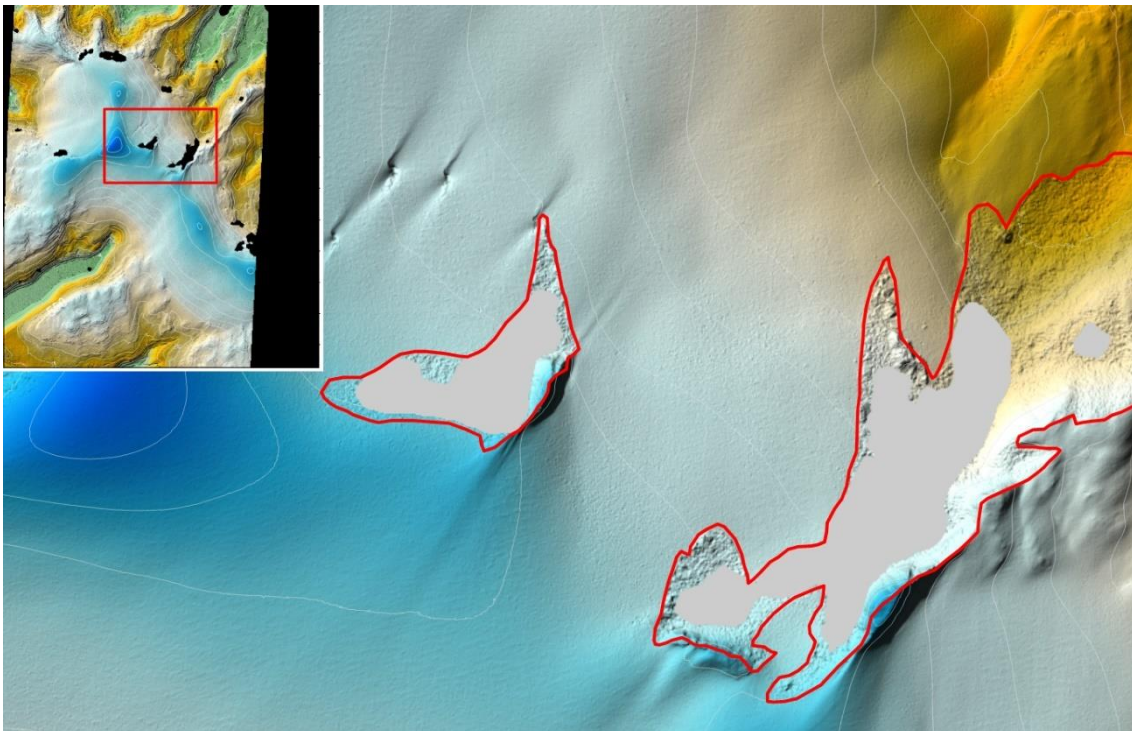

Fig. R2: 2015 WV2 February DEM. Shadow areas (red polygons) produce both gaps and noise in the shadow-covered surface.

Specific comments

Page 1

Line 13: swap order of "high-resolution" and "accurate"
Line 15: "as the source"
Line 16: "snow- and ice-free." Should be changed throughout paper.
Line 17: "The estimated accuracy" should be "The estimated relative accuracy"
Line 19: Remove "Bw ="
Line 19: Use past tense, "Winter mass balance was..."
Line 21: Remove "winter"
Line 22: "Cores at these sites show average winter snow depth of 6.5 m, while sampled DEM difference values
Line 23: There are many other factors that can contribute to the difference, including DEM error, density error, etc.
Line 30: "a continuous retreating and mass wastage" should read "retreat and mass loss"
Accepted and changed all the specific comments in page 1, including significant rewriting and reorganization.

Page 2

Line 1: Suggest rewording to "Seasonal (winter accumulation and summer ablation) records of glacier change, however, are sparse." I'm not sure this is an accurate statement.
Rephrased: "Seasonal records of glacier change, however, are sparse". Because accumulation can occur in Summer (see e.g., Nepal) and ablation can also occur in winter (see low elevation glacier tongue in Patagonia).

Line 2: "despite they contain" Changed.
I stopped detailed native english proofreading/grammar suggestions at this point in the review.

Line 7: e.g. should be in parenthesis. Done.

Line 7: "this method can be used to estimate" Changed. See (Fountain and Vecchia, 1999). Reference added.

Line 14: "access to any remote area of the world" suggest changing to "global coverage" Done - also, due to orbital inclination, these sensors cannot routinely image regions within the "polar hole" (e.g., 87-90°S)

Line 14-15: "mostly no saturation" suggest changing to "excellent image contrast" Done, also, this was detailed in Shean et al., 2016 Ref added.

Line 15-17: add Shean et al, 2016 for accuracy analysis of WV-1 and WV-2 Ref added.

Line 18: "mass turnover" - is this really the appropriate term to use? If net mass balance is nonzero, does "mass turnover" refer to input or output? I've typically used "annual discharge" for outlet glaciers and ice streams.
Term replaced for "mass-balance amplitude", as described in the UNESCO IACS mass balance terminology (Cogley et al., 2011)

Lines 20-24: There is no mention of "winter mass balance" here, just "snow accumulation" - should be careful about using these terms interchangeably. Changed to "winter mass balance".

Line 26: "Approximately 11000 km2 of Iceland is covered by glaciers..." Changed.

Line 30: "...the results provide glacier runoff estimates needed for water resource applications (e.g., hydropower)" Changed.

Line 31: "to record" Changed.

Line 32: "high mass turnover" - high relative to what? Rephrased & ref added.

Page 3

Line 5: How is the climatic situation different? Altitude, precip, temperatures?
Added "Due to its proximity to the Irminger Current" and additional references.

Line 14: Reword this sentence. Sentence rephrased. *Not really testing the images, but evaluating derived products (DEMs) with a footprint that covers most of the ice cap.*

Line 16: "...stereo images were acquired..." delete scheduled. Deleted.

Line 18-19: Just said the swath covers the entire ice cap at end of last paragraph. Suggest combining sentences. Combined and sentence removed.

Line 19: Delete last sentence - all Pleiades images should include RPCs. Deleted.
Suggest including a sentence here about how the images were orthorectified
Since this is a technical detail it seems more logical to us to include this in the methods section.

Line 20: Use proper ArcticDEM wording for credit here, images are available via NextView License, derived products (ie DEMs) have their own licensing. Suggest that Howat review this section carefully.
Coauthor Ian M. Howat has revised this section and we consider this an adequate credit to the ArcticDEM (there is no proper reference available for citation of this project).

Line 22: DEMs are posted at 2 m GSD (ground sample distance), but their true "resolution" may not actually be anywhere near 2 m.
Mentioned in the text. We refer as "grid size" instead of GSD.

Line 24: Last sentence should be reworded, should avoid starting a sentence with acronym.
Rephrased.

Line 27: "good geometry of steroscopy without compromising the coverage of the DEMs in steep areas" change to "excellent stereo geometry while minimizing occlusions due to steep topography" Rephrased.

Line 28: "one day after first significant snowfall" - how does this compare to Figure 5, which shows multiple precip events in the cumulative record before the Oct 14 DEM. While temps at the met station are warm, presumably there was snowfall at higher elevations during the Oct ~4-6 period (hard to say exactly as there are no ticks on the x-axis in Fig 5).
This previous snowfall is explained in the methods section (section 3.2.4), we think there is no need to mention it twice.

Line 29: I doubt that "all of the fine details" are still observed. "small boulders" - what are dimensions? The Pleiades images are ~0.7 m GSD, so the smallest boulders one should be able to resolve are ~2.1 m across (~3 pixels). What about variations in the fresh snow surface texture?
Rephrased. We however do not mention variations in the fresh snow texture, since previously it has been explained the excellent quality of texture in snow.

Line 31: "Solar illumination angle" - check terminology here, some angles are relative to nadir, others to tangent plane. Changed.

Line 32: With 12-bit images, there should still be plenty of contrast in the shadows. Previous studies with WV-2 11-bit images show that correlation within shadows is excellent.
See our response to your general comment about shadows (P6).

Page 4:

Line 2: "similar snow extent" - this is subjective, and just looking at the images in fig 2, I see some big differences between Feb and May snowcover.
Despite the appearance in the two quicklooks (highly different images due to the sun angle), a visual examination was done, observing that bare ground in February and May matched in most cases. Due to the low illumination angle in February, using the February orthoimage would require a tedious work to extract the bare ground, therefore for simplicity we used the May orthoimage as reference of the bare ground.

Line 5: should read "between 2008-2012" Changed.
Line 6: up to 10 km distance from the ice margin? Changed.
Line 7: A DEM with 2-m posting was produced from the point cloud. Changed.
Line 9: A vertical accuracy of only 0.5 m seems pretty bad for modern airborne lidar. Any idea why?
Changed. The accuracy analysis from Jóhannesson et al., 2009 states that the vertical accuracy for a similar dataset is "well within 0.5 m", based on a comparison with GPS profiles. Relative elevation uncertainty in the lidar DEM is typically 0.2 m to 0.3 m.

Line 11-12: rewording by native english speaker, don't use "have been" use "were" Replaced in several occurrences. Text now proof-read by native co-authors.

Line 13: "yielding the winter mass balance at each location" - how is this estimate extracted from the snow cores? Need more detail on this. Were the cores collected at the appropriate time of year to provide accurate winter mass balance? Detailed in the method added. Section rewritten.

Line 13: Five additional points - what kind of points, more snow cores? Rephrased.

Line 15-16: rewording required by native english speaker Reworded.

Line 17: What records of snow density (density profile at X cm intervals, bulk density?) and how were they used? Added "bulk density".

Line 18: How was this map produced, and can it be reproduced as a figure in the paper? It sounds like it is an interpolated product, which is not an in situ measurement. As a reviewer, how am I supposed to evaluate the accuracy of this product?
Added details on the interpolated map of 2013-2014 mass balance based on a reference for a similar case-study (Pálsson et al., 2012).

Line 22: delete "the entire years", reword "Daily precip and temp data...", provide a reference/source for the IMO met data, are these public? Rephrased.

Line 26: The description of Pleiades processing needs some work.
Rephrased. This is only introducing the following sections, which explain the Pléiades processing.

Line 28: "...scheme A used" past tense, change to "lidar-derived" Changed.

Line 28-29: The description of the two schemes is incomplete and somewhat confusing Section rewritten for clarification.

Line 29-30: Already presenting a result here, and the reader still has no idea about what was actually done. Also, this sentence should be reworded. Sentence deleted (mentioned in section 3.1.4).

Page 5:

Line 3: Why downsample the output DEMs here? - the co-registration process shouldn't require downsampling. This should be listed as a separate step that involves resampling all co-registered DEMs and images to a common grid for differencing. Also, suggest that information about the common projection used for analysis is offered, as the input products likely have different projections.
We still don't have any information at this point about how the Pleiades DEMs were produced, and there is already discussion about co-registration.
Section reorganized and information added about the projections, and resampling is now explained in the WV2 section. During co-registration of the WV2 DEM, since the master DEM is in 4 x 4 m, we consider appropriate to use the same resolution for the slave DEM. The reprojection and resampling of WV2 were performed both at once with the GDAL libraries.

Line 9: "adequately spread horizontally and vertically" is subjective. Added: "surrounding the ice cap and on two of the nunataks exposed within the ice cap". *Also, the Nuth and Kaab requires an e.g.* changed.
Finally, we find out how the Pleiades DEMs were actually generated. This should be described much earlier.

Line 13: OK, so the RPC model was refined. Any way to assess the magnitude of the changes?
Both ERDAS and ASP show statistics of the bundle adjustment and residuals from the least square adjustment can be obtained. For ERDAS (used in Scheme A), the residuals of the adjustment are <1m in ground XYZ and <0.2 px in image xy.

Line 13: "pixelwise" - what does this mean?
Replaced to pixel-based. Pixelwise is a common type of stereo matching, as we can talk about feature-based matching or area-based matching.

Line 14: replace "raw" with "native" and explicitly state that the correlation was performed on images resampled to ~1.4 m GSD. Changed. *In my experience, doing the correlation at lower resolution is not worth the compromise in quality/accuracy, esp. for snow-covered regions w/ limited texture.*
This does have an impact in the processing time, which can be a limiting factor in some studies. Besides, in other types of stereo images, e.g. aerial photographs, this level of pyramids offers much better results in matter of coverage of the DEM.

Line 16: What does "linearly interpolated into gridded DEMs" mean? Was a TIN created, then sampled at regular grid spacing? Clarified.

Line 21: delete "only". Deleted.

The comparison of the two schemes has a major uncontrolled variable – two very different software packages were used to generate DEMs. If the goal of this experiment is to evaluate the improvement offered by including GCPs, then use the same software package (pick either ERDAS Imagine or ASP) to produce all DEMs. If the goal is to evaluate the different software packages, run both packages with or without GCPs (ASP can also ingest GCPs and do bundleadjustment).

We have added this in the discussion on results from the two schemes. We have done the test using ASP with the same GCPs as used in ERDAS, and the results from ASP are worse in terms of RPC refinement, producing elevation biases in areas where no bias was found running the process in ERDAS. Below we included a figure with the comparison. It is however beyond the scope of this study to investigate why ASP does not do a better job when GCPs are provided.
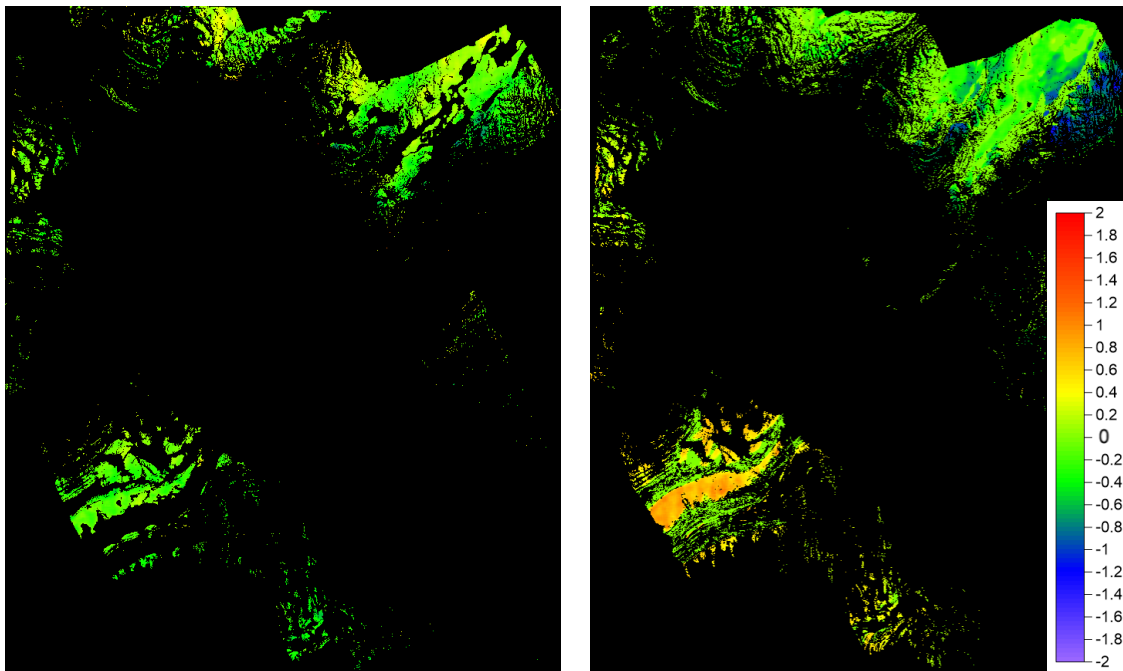

Fig. R3: Comparison ERDAS vs ASP using GCPs, based on elevation difference between the October 2014 Pléiades DEM and the May 2015 Pléiades DEM in snow- and ice-free areas. Left: Results from ERDAS Imagine. Right: Results from ASP. Red color (S-W) and blue color (N-E) indicates remaining biases in certain areas when images are processed with ASP, despite the use of the GCPs.

Line 26-27: This sentence is confusing. I still don't understand what was done. Areas with sparse cloud coverage? Clouds like water vapor clouds obscuring the surface, or clouds like point clouds? Were these areas masked? What initial DEM and orthoimage??? Rephrased.

Line 27-28: Reword this sentence. Gradual correlation is not the right term to use. Pyramidal correlation could work. Both ASP and eATE use a pyramidal correlation scheme, with previous correlation results seeding the next (higher resolution) level. To clarify, you stopped the eATE correlation, but let ASP continue to the full-resolution images?
Rephrased. Yes, ASP was run in full-resolution images and eATE in Pyramid1.

Line 30: This is confusing. For a single stereopair, the DEM and orthoimages showed an offset? Or for multiple overlapping stereopairs, the DEMs showed an offset? Was the orthoimage produced using ASP? Or is the comparison between the ASP DEM and the eATE orthoimage? Orthoimages won't really display a visible offset in vertical positioning, just horizontal offsets; DEMs can display both horizontal and vertical offsets.
Rephrased for clarification. The orthoimages (processed in ASP) revealed horizontal offsets, and the dDEM revealed the vertical offset and planar tilt.

In my experience, only a small subset of WV DEMs processed with ASP display a noticeable planar tilt (<5%). Was the need for a tilt correction here determined using exposed bedrock surfaces in the lidar? Otherwise, solving for a tilt could actually introduce more error, as it is liable to over-fit when reducing errors over limited control surfaces.
Details about the planar tilt and a test using ICP with 3 parameters is described two comments below in this page (P13).

Line 32: Suggest using Shean et al (2016) reference for ASP implementation of the Pomerleau ICP routine, as the ASP ICP is a "value-added" geospatial version of the generic ICP. Done.

Page 6

Line 2: Shouldn't this be a 12-parameter transformation? Also, why not assume scaling is correct and limit ICP to a rigid body transformation? How robust is this estimate of planar tilt, as this was calculated for limited snow-free control surfaces. In this situation, I would limit the ICP to solve for a simple translation rather than a solution with many poorly-constrained parameters, to avoid over-fitting.
Read first our general answer "MS revised using an updated version of the Pléiades images". In our revised MS, all calculations/statistics are based on the newly Pléiades-1A images and the strong tilt reported in the TCD paper disappeared. However, we compared the statistics on the residual elevation difference over snow- and ice-free areas using 3 parameters (option --compute-translation-only in ASP) and 6 parameters (3 translations + 3 rotations). The latter is a robust method used often in photogrammetry (Miller et al., 2008) and we do not think this over-parametrizes the transformation, since both DEMs behave as "solid".

Below a table (Table R2) showing the residuals after co-registration using both methods. The spatial coverage of residuals also shows slight tilts in the co-register dDEMs, which disappear when applying the 6 parameters co-registration. We can see this especially affecting the NMAD, decreasing from 0.31 m to 0.23 m for the comparison May 2015 Pléiades DEM minus October 2014 Pléiades DEM, and decreasing from 0.43 m to 0.35 m for the comparison February 2015 WV2 DEM minus October 2014 Pléiades DEM.

Table R2: statistics obtained after running ICP co-registration with 3 parameters (option -compute-translation.only) and with 6 parameters. The statistics shown are filtered from shadows and slopes >20º as done in the Table 2 of the MS.

	Scheme	N (x10^6)	Mean (m)	Median (m)	SD (m)	NMAD (m)
3 Param Filtered	B –ICP *Oct 2014 Pléiades DEM – May 2015 Pléiades DEM*	1.7	-0.09	-0.01	0.73	0.31
	WV2 ICP *Oct 2014 Pléiades DEM – Feb 2015 WV2 DEM*	0.9	-0.03	-0.09	0.56	0.43
6 Param Filtered	B –ICP *Oct 2014 Pléiades DEM – May 2015 Pléiades DEM*	1.7	-0.07	-0.02	0.66	0.23
	WV2 ICP *Oct 2014 Pléiades DEM – Feb 2015 WV2 DEM*	1.0	0.08	0.01	0.54	0.35

All tables and figures have been updated with the updated Pléiades files, without any significant change in the results.

Line 9-10: So ice-free surfaces were used for ICP co-registration, but there was snow on the ground during the February and May images, right? Do you mean "snow-free"? Otherwise, using snow-covered surfaces for co-registration could be problematic. Even if there is only ~10-20 cm of snow, the spatial distribution will be non-uniform, and can lead to errors in the resulting co-registration.
Yes, snow-free (and not ice-free) areas were used, this is now corrected. See comment above (P10, 3rd comment), there is a good matching between the coverage of snow-free areas of February 2015 and May 2015. We therefore use this source for co-registering the WV2 data, after manually shifting the May 2015 orthoimage to align with the WV2 orthoimage. We do acknowledge, however, that the reference DEM (October 2014 Pléiades DEM) contains a thin layer of snow in the snow- and ice-free areas of May 2015 (described in uncertainty section, 4.1). This thin layer is neglected since most of the ice-free terrain lay on low elevations, and maximum snow accounted outside the ice cap would be c.a. 20 cm.

Line 11: Include better description of "original DEM" Changed.
Does the fact that both "slave" DEMs had a tilt mean that the "master" DEM was actually the problem?
Text reorganized. For the TCD paper, the problem was actually both DEMs: master and slave. With the new version of the Pléiades images this tilt has nearly disappeared. We still do apply correction for tilt (a 6 parameter co-registration) as it improves the statistics and eliminates small-scale tilts.

I would think that a better approach for Scheme B would be to co-register the summer DEM to lidar, then co-register the later DEMs to the summer DEM.
One of the aims of using scheme B is to run a whole chain of processing without using external data, only the satellite stereoimages, therefore we prefer not to co-register to the lidar DEM.

Line 16: "binarized" - I think this means the following: created a binary mask using thresholding, so that any pixels with DN <819 were considered to be "valid" exposed rock control surfaces. Could consider first scaling the DN values to top of atmosphere absolute reflectance (0.0-1.0), which is less arbitrary than DN values (which were scaled for sensor gain/offset during image acquisition and subsequent calibration of Level 1 images).
Rephrased. Indeed, working with top of atmosphere absolute reflectance is appropriate too. In fact ASP provides the orthoimages in these units, we only rescaled it for the text (this was however rescaled for the statistics with ERDAS, since the output orthoimage ranges 0-4095 ND).

Line 23-25: reword these sentences Done. *This approach is somewhat questionable - some might consider it manipulation to improve accuracy numbers. " in a negligible amount over the ice cap" - how negligible, or what is the actual percentage masked over the ice cap? The values and percentages in Table 2 do not appear negligible.*
Percentages over the ice cap have been added in the text and detailed information has been added as justification. While samples for statistics are reduced from Table 2, the spatial distribution of the samples remains unchanged. We have added a figure in the supplementary material (Fig. S1) showing the residuals off-glacier before and after filtering.

Again, I am surprised that results are not better within shadowed regions - there should be enough contrast in these areas for successful correlation. I would also expect that north-facing alcoves (where shadows are present) are pretty important when it comes to snow accumulation, and a simple interpolation approach across these gaps will underestimate true snow depths.
Substantial rewording and clarification has been added to this section. Details on justification and close-ups of shadow areas are shown in the general comment about shadows and slopes.

Line 30: Is this a form of bootstrapping? Yes.

Page 7

Line 1: The equations presented are for winter balance, assuming there is snow over the entire glacier. Suggest explicitly stating "geodetic winter mass balance" rather than "mass balance" as some might confuse this with annual net mass balance. Changed for all occurrences.

Line 2-3: should use commas here, not semicolons Changed.

Line 5: Recommend using "bulk density of snow" since density variations within the snowpack are not considered. Changed for all occurrences.
Why use glacier-wide average dh/dt values a dhfirn/dt values here rather than summing values at all pixels? Sentence clarified.
Should there be an area term in Equation 1?
The definition of the limits of the ice cap is often subjective, and so is the case of Drangajökull, especially since it contains snow patches attached in the southern and western parts of the ice cap. For this reason, we prefer not to discuss effects on area delineation. We have emphasized in the main text that the area is consistent with the areas defined by Jóhannesson et al., 2013 and Magnússon et al., 2016a. All the calculations carried out are accurate for our defined area of Drangajökull.

Line 7: Firn compaction is not occurring over the entire ice cap, only in the accumulation zone. Compaction of seasonal snow, on the other hand, is occurring wherever there is snow on the ground. It is important to consider the two separately, and may require changes in variable names.
Sentence rephrased. An important modification of the calculation of the geodetic winter mass balance has been done, involving the correction due to snow densification, which in our case is rather small (20 cm extra to $\bar{h}_{dDEM}$).

Line 8-12: I would introduce the uncertainty discussion after introducing all of the relevant terms. At this point, we don't know anything about chFirn, and its uncertainty is already discussed. Also, suggest listing values for each uncertainty component after the equation (delta_rho_snow = 5 kg/m^3)
This section has been substantially reorganized. Based on the reviewer suggestion we clearly indicate each of the variables affecting the geodetic winter mass balance, divided into observations and inferred values. A new section (3.2.5) has been added explaining the error propagation once all the variables have been adequately described. We finally refer to Table 3, which summarizes each value and its uncertainty.

Lines 13-19: Seems like this section should go with earlier discussion of masking.
This section is confusing and needs work. Also, the nomenclature in the text (dDEMt3t2) is different than the nomenclature used in Table 2 - recommend using consistent nomenclature throughout paper.
This section has also been reorganized and now is under a new section, 3.2.1 (Average elevation change). Nomenclature is different because we want to distinguish between $\bar{h}_{dDEM}$ (average inside the ice cap) and dDEM (just difference between two DEMs).

Line 16: So the gaps were filled with a constant value? The constant value was the average of all values within ~1 pix of the gap? Yes. Clarified in the text.

Line 16-17: "virtually no effect" - what does this mean? Rephrased.
I don't understand what the linear relation and what "average elevation difference at the overlapping areas" mean. If interpolating across 8-10% of the ice cap, it is important that the gap-filling procedure is well documented and understandable.
Rephrased and clarified.

Line 23: "digitization" Changed.

Line 23-24: Just because images are high resolution doesn't mean that there aren't significant errors in the areal extent due to subjective delineation of cap margins. I would rather see the inclusion of some uncertainty estimate for the total glacier area in the overall error analysis.
Explained in previous comment (P15). We prefer to avoid entering in the subject of what is the definition of the ice cap. We clearly state now in the text that all results are valid for our interpreted glacier margin.

Line 26-27: I think this should be "If this is not taken into account..." Changed.

Page 8

Line 2-3: OK, the percentages are 64% and 58%, but do these areas have the same spatial distribution? Again, the reader has no idea about what this mass balance map looks like.
Rephrased.

Line 4-5: This is a fundamental assumption that requires more justification. Based on the precip and T measurements, is 2013-2014 actually representative of long-term mass balance? The amount of material moving across the firn/ice transition (ie the layer with density of ice, 917 kg/m^3) should be approximately equal to the long-term accumulation rate.

Did the authors consider regional climate model SMB results (ie RACMO or MAR) - I believe Iceland is included in the Greenland simulations. These should provide monthly SMB from 1979-2015 and can be used to estimate the long-term SMB.

This is discussed in the general comment above and a paragraph has been added in the discussion section (5.2) in this regard. We consider that this approach is not appropriate for Drangajökull ice cap, since the regional climate model SMB results are too coarse (11 km, i.e. one pixel for the entire ice cap!) and have not been evaluated over Icelandic ice caps. They may not adequately represent the complex topography of our study area. We consider that a more complex analysis of the firn is out of the scope of this paper, and we provide generous error bars in our estimates.

Line 5: "vertically integrated" Changed. *The net annual surface elevation change* Changed.

Line 8-10: Why discretize only two firn layers? If density profile is known, should be able integrate all layers to get expected surface elevation change?

(Ligtenberg et al., 2011) has a nice figure and disucussion of the different components involved with surface elevation change and firn compaction.

We don't have information about the density profile of the firn in Drangajökull (snow cores do not reach further than the last summer layer). Besides, only the top and bottom firn density are required to calculate the integrally averaged densification and under the asserted assumptions (Sold et al., 2013).

Line 12: The near-surface firn compaction rate will vary seasonally due to T and accumulation variability, especially if there is meltwater percolation in the firn. Deeper in the firn column, this is less of an issue, but I recommend correcting this statement. Corrected and reference added.

Line 15-17: It seems like the firn compaction correction should only be applied over pixels with firn, rather than averaged over the entire area. If not, should explicitly state why in text. Also, the rate of firn compaction is spatially variable, based on the long-term accumulation values at each pixel. I'm not sure that the approach using glacier-wide average observed dh/dt and firn compaciton corrections will correctly account for this spatial variability.

Sentence rephrased (P9L23 in revised MS). This correction ends being "glacier-wide", although it does have different magnitude depending on the location in the accumulation area, and indeed $\overline{C_{t_1}^{t_f}}\{\bar{h}_{Snow\ t_1}\} = 0$ in ablation areas. However, for simplicity in the calculation one can just scale the mean obtained in the accumulation area, into the 58% which is what the firn covers around the ice cap. That way we avoid adding the firn densification mask into the DEM.

Line 18: "equal amount of net accumulation occurs every year" - I'm not sure what this means. Equal to what? Should this be "constant"? Rephrased.

Line 23: Would be useful to provide some metric for spread about this average density, maybe standard deviation of the 8 values. Added SD.

How was density determined from these snow cores? Presumably the cores were drilled from the surface to the previous summer surface? What is the density distribution with depth?

Further information and references have been added in the data section (s 2.4) about how the drilling is carried out and how density is determined. This is a standard procedure for Icelandic glaciers. We do not include a profile of density with depth since this is out of the scope of the paper (we work instead with bulk snow density). We however can learn from the density profile (Fig. R4) that there can be very different ranges of density through the winter. This is an additional proof of the bulk snow density chosen to calculate the mass balance October 2014 – February 2015: anything ranging 450 kg/m^3 to 550 kg/m^3 can be possible.
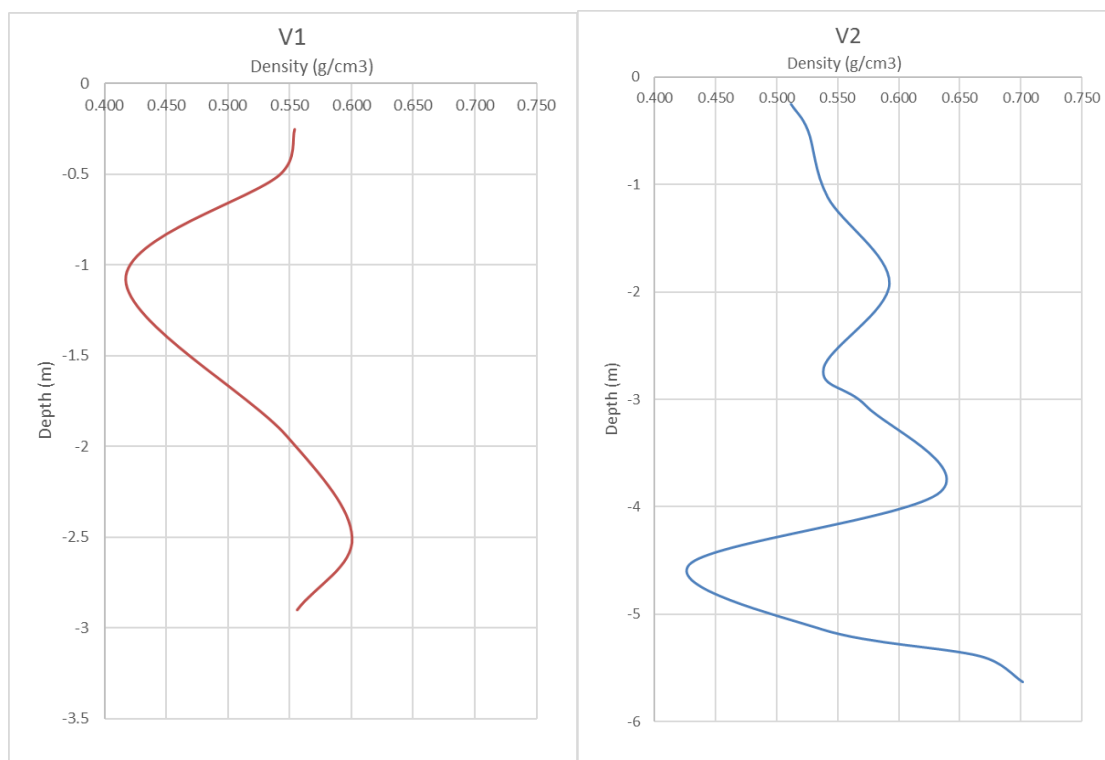

Fig. R4: Two examples for the density profile over depth, for the field campaign of 19 June 2015 at locations V1 (291 m a.s.l.) and V2 (668 m a.s.l.).

Recommend using the term "bulk snow density" where applicable. Done.
Should we expect constant snow density across this range of elevations?
No. Density will likely change with elevation as we have multiple variables dependent on elevation, such as rain events in low areas which are snow events in higher areas.

Line 26: "year to year and point to point variations of the snow density" - I don't understand what is meant by this. What steps were used to determine the +/- 27 kg/m^3 value? Rephrased.
Line 29-30: Why is fresh snow earlier in the season be less dense than fresh snow later in the season? Is there a previous citable study that demonstrates this phenomenon?
Bulk snow density is less dense earlier in the season because of the snow densification over time. This is observed in Drangajökull: 520 kg/m^3 were measured in end of March 2014, versus the 554 kg/m^3 measured in June 2015 (mentioned in the paper).

Line 31: How does Figure 7 show this? Are there specific years we should be considering? State the years in the (Fig 7) reference. Done.
Where does +/- 50 kg/m^3 come from, and does "this density" mean the 554 or 500 kg/m^3 value? Clarified.

Page 9

Line 1: Why not use a density estimate for dDEMt3t2 and calculate directly? How different is the direct calculation compared to the value calculated as the difference between the two periods?
The calculation of mass balance of the last sub-period is challenging with our equation of geodetic winter mass balance (Eq. 1 and 2). An extensive description of this effect is now found in section 3.2.4 (Fresh snow densification in the reference DEM). The snow load of the second period induced densification of the snow fallen between October and February, such that the February surface is deforming. Using Eq. (1), the estimated bulk density needed for getting $Bw_{t_2}^{t_3}$ would be ~675 km/m^3, which is unreasonably high. To avoid applying this correction we use the October 2014 Pléiades DEM as reference, since this leads to a substantially smaller correction. The discussion section also addresses this issue (P15L19 in revised MS), and for this kind of study it is highly encouraged to use a snow-free DEM for reference, as close as possible to the beginning of the winter.

Line 10: How does this equation relate to Equation 1? It seems like it might make more sense to lead with an equation that introduces all of the relevent terms, then provide the equation for glacier-wide mass balance, explaining why the ice dynamics term is not considered.
We think that it is easier to explain in section 3.2 all the methods related with the geodetic winter mass balance and in section 3.3 all the methods related to validation and comparison with snow thickness. The fact that ice dynamics are not considered in the glacier-wide geodetic winter mass balance is explained in the discussion.

I am confused about the chtOct and chtMay&Jun terms here. After reading later sections and rereading this section, it started to make sense. Suggest an improved introduction of these corrections. Sentence reworded.

Line 16-18: Needs rewording. I can't understand. Recommend starting with statement about when the in situ measurements were made. Then in a new sentence, restate when the DEMs were acquired. Sentence reworded.

Line 24-25: How can a melting event be identified from low-resolution MODIS data? Or is this inferred from some other source? An explanation would help.
This was a simple visual inspection and no measurements were done with these datasets. MODIS images of 250 m GSD were used for finding snow events. This sentence has however been removed from the revised MS.

Line 28-29: Suggest stating what is meant by "recorded total winter precip" - what is the source? Source (IMO) specified in the data section (s 2.3).

This scaling factor is directly related to the local in situ measurements. Rather than state a difference in scaling factor, why not lead with a description of the variations observed in the snow cores? I'd rather see actual values so that I can assess the precip spatial variability.
Values of maximum and minimum winter mass balance added in this sentence.

Page 10

Line 2: Is this lapse rate constant over seasonal timescales?
Most likely this lapse rate has seasonal variations, still we think this is out of the scope of the paper for reasons previously mentioned (e.g., will likely have a small effect on the results and main conclusions of the study).

Line 5: Wasn't value of 500 kg/m^3 used in a previous section for fresh snow? Should provide a citation for the 400 kg/m^3 value typical for Iceland.
We have restructured and simplified density values for calculation of snow accumulation & melt, now it is clearly indicated in section 3.2.2 (bulk snow density). Density 400 kg/m^3 is used for fresh snow of 3-14 October, both for snowfall and for snow melted. Density 515 kg/m^3 is used for snow accumulation & melt between 22 May (Pléiades images) and the in situ measurements. The above values are considered more coherent, as fresh snow 400 kg/m^3 is usually found at close dates from snowfalls. No citation is available for this value but this is based on extensive field experience from coauthors Finnur Pálsson and Thorsteinn Thorsteinnsson. The latter value, 515 kg/m^3, is obtained from the uppermost part of the measured densities during the 2015 campaign.

Line 5: "converted to snow depth" or thickness. Changed.

Lines 4-16: Lots of detail here, without much explanation for why this is necessary and how these corrections were actually applied. Another paragraph could help. This has been substantially rephrased and reorganized (now in section 3.2.4.).
The values from T scaling are 4-6 cm, but values from precip scaling are 15-30 cm? Are these supposed to be equal?
Rephrased and reorganized (moved to section 4.1 in the revised MS). This correction is done for estimating snow existing in the ice-free areas in the October Pléiades DEM. For this estimation, we are only using a single precipitation event (1-2 days before the October Pléiades images), and the closest in situ locations to the ice cap margin. This is important to stress that the results shown are slightly affected by this snow, but this snow layer is maximum 27 cm thick, and most of the data used for statistics is derived from low elevations. (P12L24).

Line 18: "This process" - what process? Sentence deleted.

Line 18: 3.2.1
This section has been deleted, the introduction of 3.2 refers now to the previous section, which explains this phenomena.

Line 20: "presented" word choice Changed.

Line 25: "full Stokes" Changed.
The modeling section seems to come out of nowhere (no real introduction) and receives limited attention - is there another reference that presents the modeling results for this particular ice cap in more detail? What about sliding?
There is an upcoming publication which will use the full Stokes ice flow model for further objectives, we therefore prefer not to explain many details of it, while it still fulfills its objective in this paper. Sliding is described in the discussion.

Line 28: "evenly distributed" - this is somewhat subjective and the reader has no way to evaluate. Replaced with reference to the bedrock map.

Line 29: this is the first mention of GPS observations??? How many, where are they located?
GPS observations are now described in the data. This is just a handheld GPS measuring the location of the in situ snow cores.

How were these used to calibrate A? Are there data assimilation routines in Icetools/Fenics?
A grid search over the viable range of flow law parameters was performed. The A parameter was optimized to the surface velocity measurements using a root-mean-square-error approach. Because we cannot isolate the effect of sliding from the surface velocity measurements the A parameter is tuned to account for the effects of both ice deformation and sliding.

Line 31: Do the GPS measurements show constant flow velocities? Should we expect significant seasonal velocity variations due to changes in subglacial hydrology?
The GPS measurements are carried out annually, or at most twice a year (in spring and in autumn), and therefore cannot resolve such a signal. There are most likely seasonal velocity variations in the ice cap, as in most glaciers in Iceland where there are good records of GPS measurements in daily or hourly basis. This is neglected in our study for simplicity.

Would it be possible to include a map of the velocity vectors and emergence/submergence velocities? The reader is only presented with values for the in situ sites in Table 4.
Since there is ongoing work with the ice flow modelling and this is a very peripheral part in our study of winter accumulation from satellite data, we prefer not to keep a big focus on this data. Results are however clearly indicated in the table.

Page 11:

The intro stated that this ice cap is not in steady state, but is losing mass at -0.26 mwe/yr. How does this affect a steady state assumption?
This is commented in detail in the discussion section. Indeed, the ice cap is not in balance, and there are surge-type glacier outlets. For some areas of the ice cap however this assumption is less erroneous, e.g. points in the southern part, which are not affected by surges and have shown much less changes during the last century (Magnússon et al., 2016a).
Regarding the mass loss of -0.26 $m_{w.e}a^{-1}$, this is only valid for the total average 1945-2011. An updated period of geodetic mass balance shows that for 2011-2014, Drangajökull is nearly in balance with $b_n = 0.0 \pm 0.3$ $m_{w.e}a^{-1}$ (Magnússon et al., 2016b).

Again, please provide better justification for the use of 2013-2014 as representative for long-term mass balance.
Both the comment above, and the general comment in this regard should provide justification.

Lines 10-12: Looking at Table 2, I see sample counts decreasing by 64% (2.2x10^6 to 1.4x10^6). That's pretty substantial. Rephrased. *Not really surprised that masking these areas reduces SD and NMAD - past studies have documented this relationship out to much higher slope angles (Müller et al., 2014; Shean et al., 2016).* References added.

Page 12:

Lines 1-2: These are uncorrected dDEM values, right? So some of this difference is likely due to spatially variable ice dynamics and firn compaction.
Clarified. This has some influence (<0.5 m) from firn and fresh snow densification. However, these values are not affected by ice dynamics, since they are compensated glacier-wide. This also applies for the mean $\bar{h}_{dDEM}$ for East and West sides as they are split in different glacier catchments.

Line 3: How much of this two thirds is due to the timing of large precipitation events, ongoing compaction of this fresh snow, and the timing of DEM observations? I'm not sure how useful this metric is given the sparse temporal sampling of the DEMs.
Snow (and firn) densification are very low in comparison to the total $\bar{h}_{dDEM}$ in our study and therefore we prefer to keep this as a valid metric. This effects however need to be acknowledged and can be of bigger extent in other cases, e.g. if multiple DEMs are analyzed through one winter with short time interval between each other.

Are there any rain on snow events here?
Most likely there are. In addition, the February DEM surface will be deformed by the load of the February-May snow.

Line 11: Explain why this is expected. This is explained (and referenced) in Fig. 5.

How were the DEMs sampled at the in situ sites - direct extractions of value from nearest pixel, a median value for a window (say, 3x3 pixels) around the in situ location, or some other approach? How much local variability is observed around each site?
Sentence added in the introduction of the section 3.3, explaining how we extracted the elevation at the 2015 in situ locations. We tested different ways of sampling (not shown in the paper for the sake of brevity): (1) bilinear interpolation, (2) a 3x3 averaging window and (3) bilinear interpolation of Gaussian filtered dDEM. The three cases revealed highly similar results (cm differences), which indicate that the dDEM is smooth with minor spatial variability at local scale.

Was there any effort to account for the fact that these in situ locations are moving horizontally through a spatially and temporally variable accumulation field? What are horizontal surface velocities from GPS and flow model results - if velocities are small, then this is less important.
Values of dh are extracted from the in situ location in June 2015.
The time difference between in situ and RS data is c.a. 1 month, and the maximum horizontal velocity in the ice cap does not reach more than 10 m a^{-1}, i.e. less than a meter in one month. The impact of this over the dDEM can be maximum ¼ of the pixel size of the dDEM, and as explained in the previous comment, the dDEM is very smooth. We therefore neglect this effect.

Is sampling performed at estimated in situ locations for each DEM timestamp, or at fixed locations for all timestamps?
Added details in intro of section 3.3.

Lines 17-19: What is causing these differences and why are they important?
This is explained in the discussion, section 5.3.

Page 13

Line 3: At this point, I don't think it is fair to say that this is entirely from remote sensing, as the methodology involves model results, assumptions about firn compaction, and independent mass balance measurements/maps for corrections. Sentence rephrased.

Line 4: Should this be glacierized rather than glaciated? I can never keep these straight. Check other occurrences in paper.
Opted for glaciated. Also consistent with other works, e.g. Noh and Howat "Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over **glaciated** regions".

Line 7: "The use of external reference data for bundle-adjustment prior to stereo correlation" –
Changed. *also this really only applies to the sensors used in this study. For some sensors (ie Hexagon historical imagery) working with ground control before stereo reconstruction is essential to properly constrain interior orientation.* Rephrased.

Line 10-11: This statement seems a bit overreaching. I can't think of any off the top of my head, but there are likely some older ICESat or satellite radar altimetry studies of winter accumulation for the ice sheets and/or ice caps. There are also GRACE observations of seasonal mass change.
Rephrased. "Optical satellite-based".

Line 18: I don't understand why the data gaps should differ between Scheme A and Scheme B, except for the fact that two different correlation routines were used. If the same bundle-adjusted Scheme A images and the uncorrected Scheme B images were run through the same correlator, should end up with the same gaps. Rephrased.

Agree that Scheme A with GCPs should reach the same amount of gaps, although we have experienced that some deformation is produced in the resulting dDEM when refining the RPC with GCPs in ASP (see Fig. R3).

Line 19-20: The ICP method is pretty robust. While well-distributed static control surfaces are desirable, the Shean et al (2016) paper demonstrated that this co-registration is possible with only a few sparse lidar flightlines. What really matters is the distribution of slope and aspect over the available control surfaces. As a thought experiment, consider co-registration of two perfectly planar DEMs - there is no unique solution, no matter the distribution of control surfaces.

Sentence rephrased. We have included in the supplement S1 the figures with the spatial distribution of the residuals in snow- and ice-free areas to show that there is adequate surface for running the ICP method.

Lines 30-31: Ice dynamics do affect this number if the glacier is not in steady state.

This can have small implications for a glacier which is not in steady state, as the retreat/advance of the glacier front can be affecting the $\bar{h}_{dDEM}$ at the glacier margins. This depends on the pace of retreat/advance and the tongue geometry. Over a period of months this effect is very likely low.

For example, what if the DEMs happened to capture a surge?

In this case, while being difficult to differentiate what is due to snow accumulation and what to surge effects, the glacier-wide average can still measure a general mass gain due to the snow accumulation.

Line 31: Yes, as long as the glacier extent/length remains constant and crevasse dimensions and distribution remain constant. This is probably OK for time spans of months, but not years.

Changed, remarked that this is valid for glacier-wide geodetic winter mass balance.

Page 14:

Lines 12-13: This is a big claim. I'm not sure that a general statement like this is appropriate – if the few density measurements are not actually representative of snow density for the full glacier, this doesn't work.

It is indicated that they need to be adequately selected. Otherwise, indeed, this can cause erroneous results.

Line 14: Should this be adds "4% uncertainty"? Changed to "increases by 4%".

Lines 33-34: Where is this result presented? The reader is never presented with maps of emergence/submergence velocities, and how do we know there is an underestimate (comparison with GPS data???) This is a misunderstanding, sentence rephrased for clarification.

The comparison in Table 4 shows that the results from ice flow model at the southern locations is rather small.

The lack of basal sliding in the flow model could be a fairly significant limitation. Presumably this ice cap is temperate? Rather than using GPS measurements to calibrate A, what if A is estimated based on temperature (a common approach), then use the flow model to calculate expected velocity due to deformation alone, then use the GPS measurements to estimate the component of observed velocity due to sliding? Also, when calculated flow parameter A is mentioned earlier in the text, should include a value in parenthesis.

The value of A does not have very much physical meaning and it is rather a tuning parameter, since the model does not include sliding. The tuning however compensates for the sliding at the GPS stakes by making the modelled ice softer than it actually is. In this study, we originally chose an A parameter based on literature in temperate glaciers (A = 2.4e-24), and this was afterwards tuned to best fit the GPS velocities, obtaining a new A = 4.35e-24. We opted for skipping this from the main text as it can distract the reader from the main topic, since the 3D ice flow modelling is not a major subject of this paper. Furthermore, changing the value of A has little effect on the emergence velocities and therefore the main conclusions of the study.

Page 15:

Line 13: Relative DEM accuracy (for slopes <20°) Added.
Not really surprising that the relative elevation change accuracy is similar for both Scheme A and B, because both were eventually co-registered, right?
Scheme A was not co-registered (see Fig. 3 with workflow), as the use of GCPs is already a method of co-registration.

Line 24: This needs more careful wording. Also careful about going from absolute values (0.2 and 0.4 m.w.e.) to percentages (4%). Suggest picking one, or better yet, present both, and be consistent throughout the paper.
Reorganized and rephrased.

Page 16:

Line 2: What is the magnitude for each? An order of magnitude (e.g., 0.1 vs. 1 m vs. 10 m) comparison is not really useful here.
Clarified "ranging centimeter to meter scale".

Tables:

Table 2:
I don't recall seeing a description/explanation of the "trim mean" in the text. This seems somewhat arbitrary - taking the center 90% of errors. Trim mean deleted. *Already presented std (68%). Recommend using 68% and 95% for robust equivalent of 1 and 2-sigma, respectively. See metrics in (Höhle and Höhle, 2009).*
Std and NMAD kept in the revised MS. We consider that these two metrics of dispersion are sufficient for the statistics.

Why is bias corrected SGSim only given for two of the 6 rows?
SGSim is done following the steps described in Magnússon et al. (2016a), and therefore the data needs to be filtered first. Thus it is not presented in the first 3 rows. It is not applied for the WV2 DEM since, despite the good matching between the coverage of snow-free areas of February 2015 and May 2015, it is likely that some small snow areas remain in the statistical analysis. While NMAD and median exclude these outliers, the SGSim does the opposite and produces unrealistic deformations in the simulated errors of the DEM.

Table 3:
This is mass balance for the entire ice cap. Recommend stating this in caption.
I'm still not entirely clear about why the snow density values for the two periods are different and how this was determined.
Changed. About the density, for the period t_1-t_2 (14 October 2014 – 13 February 2015) it was crudely constrained based on bounding, extreme snow densities between 450 kg/m^3 to 550 kg/m^3. For the second sub-period, t_2-t_3 (13 February 2015 – 22 May 2015), it has been now clarified in the MS that this is a challenging calculation that should not be calculated from Eq. (1) but from difference of mass balances using DEM$_{t1}$ in the calculation.

Table 4:
Recommend describing each variable in caption. Added some of them in the caption. *Why isn't Bw in m.w.e.?* Changed to h$_{\text{Snow in situ}}$ (snow thickness), expressed in m.

Figures:

Figure 2:
I would really like to see the original DEMs in a second row below the images. Ideally, there would be a 2nd row of "raw" color shaded relief maps and a 3rd row of "masked" color shaded relief maps so the reader knows the spatial distribution of valid DEM pixels used in the analysis.
We need some way to evaluate where the interpolated dDEM values are located.
Figure updated according to reviewer's suggestion. For evaluating the pixels used for the statistical analysis we have added a supplementary figure (S1) with detailed distribution of the pixels used in the statistical analysis, as well as the histograms.

Figure 3: Am I correct that rectangles are processing steps and parallelograms are products? If so state in caption. If not, might be good to distinguish the two.
Exactly. Sentence added in the caption.

Figure 4: There shouldn't be any firn in the ablation area, correct? Correct, that's why the lower half of the figure does not include the tern h{Firn}. *It might be useful to show a snow core in the June diagram, labeled with the variables from the text for measured snow depth.*
Extra labels h$_{\text{Snow in situ}}$ added to the figure, and a sketch of the drilled snow core has been included in the June diagram.

It might also be good to have a thick line of constant color that represents the surface (h) that is recorded by the DEM/lidar. Added the line indicating $\bar{h}_{dDEM}$

Figure 5:
Add ticks/labels to x axis. Changed.
Would be useful to see two separate panels here:
A) Temperature
B) Precipitation (individual precip events plotted as spikes) with cumulative precip in background.
Would it be more useful to show scaled T at lowest and highest elevation sites rather than the AWS T, which are much warmer?
We did not change the structure of this figure, we prefer simply to show the raw data from the AWS, which makes it simpler to understand. We do add the minimum and maximum scaling rate in the main text (the highest elevation does not show the highest scale rate, since the accumulation is affected by wind redistribution).

Figure 6:

Looks like the small outlet glacier in the valley on the SW margin of the cap shows some anomalous dh/dt signals, likely due to ice dynamics and not snow thickness. Might be worth noting this.

This area is known to have large snow deposition, likely explained by wind drift. This can be observed from optical images, this area always remains with snow even in late summer.

Panel D) This looks like a cross-section showing ice thickness above bedrock. Definitely had me confused at first. Different colors would help, but it might be better to break into two panels rather than sharing an axis. Could potentially plot submergence/emergence velocity magnitude along this profile using a different color ramp.

The panel D has been edited accordingly, and the colors are now more simple to avoid confusion. We however prefer to keep both surface height and h_{dDEM} in the same X axis, as this allows a better understanding on the snow differences between east and west of the ice cap.

Does "Snow Acc. (m)" represent snow thickness after all of the corrections described in the text (firn compaction, ice dynamics), or is this just simply the observed elevation difference? This has been replaced for elevation difference h_{dDEM}. *Make sure this is clear in the caption and figure labels. Also, should specify a datum for absolute elevation in the caption (height above WGS84 ellipsoid or some geoid model?).* Added elevation datum.

Figure 7:

Would be useful to include a legend so that the reader knows which color circle represents each location in Fig 1. Why are there two light blue circles in a given year? It would be useful to have horizontal dashed gridlines at 10 kg/m^3 intervals, so we can compare year to year.

Why don't we see the same pattern for density values at each site from year to year? For example, in 2006, the highest density is the maroon site and lowest at orange site, but other years highest density is at light blue site and lowest at purple site?

Figure edited based on reviewer's suggestions. These changes in the measurements are likely due to the date of the survey (e.g. in 2015 all measurements are relatively higher because it was carried out in mid-June). It should be also noted the possibility of having some of the density values as outliers (e.g. V1 in the 2014 campaign).

References:

Cogley, J. G., Hock, R., Rasmussen, L. A., Arendt, A. A., Bauder, A., Braithwaite, R. J., Jansson, P., Kaser, G., Möller, M., Nicholson L., and Zemp, M.: Glossary of Glacier Mass Balance and Related Terms, IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2, UNESCO-IHP, Paris, 2011.

Jóhannesson, T., Björnsson, H., Pálsson, F., Sigurðsson, O. and Þorsteinsson, Þ.: LiDAR mapping of the Snæfellsjökull ice cap, western Iceland, Jökull, 61, 19–32, 2011.

Jóhannesson, T., Björnsson, H., Magnússon, E., Guðmundsson, S., Pálsson, F., Sigurðsson, O., Thorsteinsson, T. and Berthier, E.: Ice-volume changes, bias estimation of mass-balance measurements and changes in subglacial lakes derived by lidar mapping of the surface of Icelandic glaciers, Ann. Glaciol., 54(63), 63–74, doi:10.3189/2013AoG63A422, 2013.

Noël, B., van de Berg, W. J., van Meijgaard, E., Kuipers Munneke, P., van de Wal, R. S. W., and van den Broeke, M. R.: Evaluation of the updated regional climate model RACMO2.3: summer snowfall impact on the Greenland Ice Sheet, The Cryosphere, 9, 1831–1844, doi:10.5194/tc-9-1831-2015, 2015.

Magnússon, E., Belart, J. M. C., Pálsson, F., Ágústsson, H. and Crochet, P.: Geodetic mass balance record with rigorous uncertainty estimates deduced from aerial photographs and lidar data – Case study from Drangajökull ice cap, NW Iceland, The Cryosphere, 10(1), 159–177, doi:10.5194/tc-10-159-2016, 2016.

Magnússon, E., Belart, J. M. C., Pálsson, F., Anderson, L.S., Gunnlaugsson, Á. Þ., Berthier, E., Ágústsson, H., and Geirsdóttir, Á.: The subglacial topography of Drangajökull ice cap, NW-Iceland, deduced from dense RES-profiling, Jökull, 66, 2016.

Miller, P. E., Kunz, M., Mills, J. P., King, M. A., Murray, T., James, T. D. and Marsh, S. H.: Assessment of Glacier Volume Change Using ASTER-Based Surface Matching of Historical Photography, IEEE Transactions on Geoscience and Remote Sensing, 47(7), 1971–1979, 2009.

Noh, M.-J. and Howat, I. M.: Automated stereo-photogrammetric DEM generation at high latitudes: Surface Extraction with TIN-based Search-space Minimization (SETSM) validation and demonstration over glaciated regions, GIScience Remote Sens., 52(2), 198–217, doi:10.1080/15481603.2015.1008621, 2015.

Pálsson, F., Guðmundsson, S., Björnsson, H., Berthier, E., Magnússon, E., Guðmundsson, S. and Haraldsson, H.: Mass and volume changes of Langjökull ice cap, Iceland, ~1890 to 2009, deduced from old maps, satellite images and in situ mass balance measurements, Jökull, 62, 81–96, 2012

Sold, L., Huss, M., Hoelzle, M., Andereggen, H., Joerg, P. C. and Zemp, M.: Methodological approaches to infer end-of-winter snow distribution on alpine glaciers, J. Glaciol., 59(218), 1047–1059, doi:10.3189/2013JoG13J015, 2013.