Winter mass balance of Drangajokull ice cap (NW Iceland) derived from satellite sub-meter stereo images
Belart et al.
TCD doi:10.5194/tc-2016-241

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.

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.

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 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.

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.

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.

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.

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.

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 co-registered. 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.

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.

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).

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)?

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.

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").

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.

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.

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"

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.
Line 2: "despite they contain"
I stopped detailed native english proofreading/grammar suggestions at this point in the review.
Line 7: e.g. should be in parenthesis
Line 7: "this method can be used to estimate". See (Fountain and Vecchia, 1999).
Line 14: "access to any remote area of the world" suggest changing to "global coverage" - 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", also, this was detailed in Shean et al., 2016
Line 15-17: add Shean et al, 2016 for accuracy analysis of WV-1 and WV-2
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.
Lines 20-24: There is no mention of "winter mass balance" here, just "snow accumulation" - should be careful about using these terms interchangeably.
Line 26: "Approximately 11000 km2 of Iceland is covered by glaciers..."
Line 30: "...the results provide glacier runoff estimates needed for water resource applications (e.g., hydropower)"
Line 31: "to record"
Line 32: "high mass turnover" - high relative to what?

Page 3

Line 5: How is the climatic situation different? Altitude, precip, temperatures?
Line 14: Reword this sentence. 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
Line 18-19: Just said the swath covers the entire ice cap at end of last paragraph. Suggest combining sentences.

Line 19: Delete last sentence - all Pleiades images should include RPCs.
Suggest including a sentence here about how the images were orthorectified
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.
Line 22: DEMs are posted at 2 m GSD (ground sample distance), but their true "resolution" may not actually be anywhere near 2 m.
Line 24: Last sentence should be reworded, should avoid starting a sentence with acronym.
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"
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).
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?
Line 31: "Solar illumination angle" - check terminology here, some angles are relative to nadir, others to tangent plane.
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.

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.
Line 5: should read "between 2008-2012"
Line 6: up to 10 km distance from the ice margin?
Line 7: A DEM with 2-m posting was produced from the point cloud.
Line 9: A vertical accuracy of only 0.5 m seems pretty bad for modern airborne lidar. Any idea why?
Line 11-12: rewording by native english speaker, don't use "have been" use "were"
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?
Line 13: Five additional points - what kind of points, more snow cores?
Line 15-16: rewording required by native english speaker
Line 17: What records of snow density (density profile at X cm intervals, bulk density?) and how were they used?
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?
Line 22: delete "the entire years", reword "Daily precip and temp data...", provide a reference/source for the IMO met data, are these public?
Line 26: The description of Pleiades processing needs some work.

Line 28: "...scheme A used" past tense, change to "lidar-derived"
Line 28-29: The description of the two schemes is incomplete and somewhat confusing
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.

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.
Line 9: "adequately spread horizontally and vertically" is subjective. Also, the Nuth and Kaab requires an e.g.
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?
Line 13: "pixelwise" - what does this mean?
Line 14: replace "raw" with "native" and explicitly state that the correlation was performed on images resampled to ~1.4 m GSD. 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.
Line 16: What does "linearly interpolated into gridded DEMs" mean? Was a TIN created, then sampled at regular grid spacing?
Line 21: delete "only"
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 bundle-adjustment)
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???
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?
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.
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.
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.

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.
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.
Line 11: Include better description of "original DEM"
Does the fact that both "slave" DEMs had a tilt mean that the "master" DEM was actually the problem? 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.
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)
Line 23-25: reword these sentences
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.
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.
Line 30: Is this a form of bootstrapping?

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.
Line 2-3: should use commas here, not semicolons
Line 5: Recommend using "bulk density of snow" since density variations within the snowpack are not considered.
Why use glacier-wide average dh/dt values a dhfirn/dt values here rather than summing values at all pixels?
Should there be an area term in Equation 1?

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.

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)

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.

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?

Line 16-17: "virtually no effect" - what does this mean?

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.

Line 23: "digitization"

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.

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

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.

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.

Line 5: "vertically integrated"

The net annual *surface* elevation change

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.

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.

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 correclty account for this spatial variability.
Line 18: "equal amount of net accumulation occurs every year" - I'm not sure what this means. Equal to what? Should this be "constant"?
Line 23: Would be useful to provide some metric for spread about this average density, maybe standard deviation of the 8 values.
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? Recommend using the term "bulk snow density" where applicable.
Should we expect constant snow density across this range of elevations?
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?
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?
Line 31: How does Figure 7 show this? Are there specific years we should be considering? State the years in the (Fig 7) reference.
Where does +/- 50 kg/m^3 come from, and does "this density" mean the 554 or 500 kg/m^3 value?

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?
Line 10: How does this equation relate to Equation 1? It seems like the 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.
I am confused about the chtOct and chtMay&Jun terms here. After reading later sections and re-reading this section, it started to make sense. Suggest an improved introduction of these corrections.
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.
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.
Line 28-29: Suggest stating what is meant by "recorded total winter precip" - what is the source? 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.

Page 10
Line 2: Is this lapse rate constant over seasonal timescales?
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.
Line 5: "converted to snow depth" or thickness
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.

The values from T scaling are 4-6 cm, but values from precip scaling are 15-30 cm? Are these supposed to be equal?
Line 18: "This process" - what process?
Line 18: 3.2.1
Line 20: "presented" word choice
Line 25: "full Stokes"
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?
Line 28: "evenly distributed" - this is somewhat subjective and the reader has no way to evaluate
Line 29: this is the first mention of GPS observations??? How many, where are they located? How were these used to calibrate A? Are there data assimilation routines in Icetools/Fenics?
Line 31: Do the GPS measurements show constant flow velocities? Should we expect significant seasonal velocity variations due to changes in subglacial hydrology?
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.

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?
Again, please provide better justification for the use of 2013-2014 as representative for long-term mass balance.
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. 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).

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.
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.
Are there any rain on snow events here?
Line 11: Explain why this is expected.
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?
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.
Is sampling performed at estimated in situ locations for each DEM timestamp, or at fixed locations for all timestamps?
Lines 17-19: What is causing these differences and why are they important?

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.
Line 4: Should this be glacierized rather than glaciated? I can never keep these straight. Check other occurrences in paper.
Line 7: "The use of external reference data for bundle-adjustment prior to stereo correlation" - 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.
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.
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.
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.
Lines 30-31: Ice dynamics do affect this number if the glacier is not in steady state. For example, what if the DEMs happened to capture a surge?
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.

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.
Line 14: Should this be adds "4% uncertainty"?
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???)
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.

Page 15:
Line 13: Relative DEM accuracy (for slopes <20°)
Not really surprising that the relative elevation change accuracy is similar for both Scheme A and B, because both were eventually co-registered, right?

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.

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.

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. 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).
Why is bias corrected SGSim only given for two of the 6 rows?

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.

Table 4:
Recommend describing each variable in caption. Why isn't Bw in m.w.e.?

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 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.

Figure 4: There shouldn't be any firn in the ablation area, correct? It might be useful to show a snow core in the June diagram, labeled with the variables from the text for measured snow depth. 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.

Figure 5:
Add ticks/labels to x axis.
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?

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.
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.
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? 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?).

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?

References used in review text:
Fountain, A. G. and Vecchia, A.: How many stakes are required to measure the mass balance of a glacier?, Geogr. Ann. Ser. Phys. Geogr., 81(4), 563–573, 1999.

Höhle, J. and Höhle, M.: Accuracy assessment of digital elevation models by means of robust statistical methods, ISPRS J. Photogramm. Remote Sens., 64(4), 398–406, doi:10.1016/j.isprsjprs.2009.02.003, 2009.

Ligtenberg, S. R. M., Helsen, M. M. and van den Broeke, M. R.: An improved semi-empirical model for the densification of Antarctic firn, The Cryosphere, 5(4), 809–819, doi:10.5194/tc-5-809-2011, 2011.

Müller, J., Gärtner-Roer, I., Thee, P. and Ginzler, C.: Accuracy assessment of airborne photogrammetrically derived high-resolution digital elevation models in a high mountain environment, ISPRS J. Photogramm. Remote Sens., 98, 58–69, doi:10.1016/j.isprsjprs.2014.09.015, 2014.

Shean, D. E., Alexandrov, O., Moratto, Z. M., Smith, B. E., Joughin, I. R., Porter, C. and Morin, P.: An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery, ISPRS J. Photogramm. Remote Sens., 116, 101–117, doi:10.1016/j.isprsjprs.2016.03.012, 2016.