## Reply to reviewer#2

The manuscript "Glacier topography and elevation changes from Pléiades very high resolution stereo images" of Etienne Berthier and others provides detailed technological and methodological information for glacier thickness change determination from data of the very new Pléiades satellite. This concerns in particular DEM extraction and data post-processing from these very high resolution stereo images. It is of high interest for glaciological related work with this very new available data and underlines the high potentials of this sensor for coming research activities. Overall, this study provides results that are of high interest for people that are dealing with Pléiades DEM extraction and their post-processing as well as for coming glaciological research activities. It emphasizes the suitability of Pléiades for geodetic mass balance estimates from optical stereo data in very different topographies.

This is clearly a competent and amazing study, and I would recommend this manuscript for publication after revision process. Please consider in this regard my general and specific comments:

**GENERAL COMMENTS**

- For me it is not 100% clear what the stated objective of this manuscript exactly is. It is great in regard of Pléiades data processing and comparison for the determination of glacier elevation changes, but it is submitted to a journal that is more interested in glaciological findings. Apart from the high quality of methodological and technological description, what are the core glaciological findings and conclusions that result from your work? What could moreover be of particular interest in regard of the scope of this journal?

  It is true that the TCD readership is more interested in thematic (e.g., glaciological findings). So we tried to put slightly more emphasis on this aspect in the revised MS, in part by computing an additional Pléiades DEM covering the southern part of the Mont-Blanc area and thus allowing us to estimate the region-wide mass balance for this entire area for the first time (a glaciated area covering more than 160 km$^2$). We agree that our study could alternatively have been submitted in a remote sensing journal (RSE, ISPRS, etc…) but our primary target audience is glaciologists. Thus we preferred to choose an open-access journal, widely read by our community. Our understanding (after discussion with two TC editors) is that *The Cryosphere* is OK to consider papers in which methodologies useful for the glaciological community are the core of the study. Like in the "Instruments and Methods" section of the *Journal of Glaciology*.
  In agreement with reviewer#3, we did not push further the glaciological interpretation of these results.

- The introduction focuses on geodetic mass balances from various remote sensing data, so I would expect the same for this paper with Pléiades as the core result. Why was the study not conducted to an end for more glaciers in the difference image of Pléiades to SPOT-5? Two mass balances were calculated, but there are more glaciers in this region. The section of geodetic mass balance determination is in general pretty short. You have great data and results, so readers would be surely interested in further glaciological results.

  The goal of the study is to demonstrate the potential of a new family of satellites, not to compute geodetic mass balance everywhere (anyway Pléiades and SPOT5 stereo

pairs are not available everywhere). The revised MS now includes the geodetic mass balance for the whole Mont Blanc massif, and the 10 largest glaciers in this area.

- Please overwork the structure of this manuscript. Make it more clearly by re-arranging and shortening sections chapters. Cross-references to subsequent text passages make it hard to read. Sentences particularly at the beginning are often long and phrasing in such cases is complicated. There are often long and multiple parentheses even in one single sentence. Please try to avoid too many parentheses in the text if possible. I would prefer shorter and more precise sentences to easier extract your information
(e.g. P4852, L15-22).

The structure and writing of the paper has new been revised following the suggestions of the three reviewers. We stress that the submitted MS was proof-read by a native speaker. Many multiple parentheses were removed.

- Several parts of this study are extensively described (e.g. about the NMAD and the settings of PCI), but other important parts are in my opinion too short. This concerns for example DEM post-processing with quality assessment and outlier detection to the final mass balance.

The procedures to account for data gaps and derive the final mass balance are now better described.

- You did not fill DEM voids, but there was no statistical evaluation of extracted terrain values conducted in order to exclude DEM pixels of poor quality. How was outlier detection employed? This is surely an issue in glacier accumulation zones were terrain extraction might be hampered due to low contrast. I would still expect areas of poor elevation estimates in snow covered glacier areas despite the sensors high radiometric resolution (12-bit). The study sites "Tungnafellsjökull" and "Astrolabe" show in Figure 2 low contrast alterations. A hillshade of your extracted DEMs in general, but particularly at these areas would be interesting to see. Low contrast alterations might be also an issue in the DEM of SPOT-5 that you used for differencing. You can generate an additional score channel image when extracting a DEM with PCI which provides information of the correlation coefficient for each extracted DEM pixel. Wouldn't it be advisable to use a correlation threshold for the exclusion of poor quality terrain? I wonder how good this correlation coefficient would be in snow covered areas of Antarctica.

See response to other referees and Fig. R4 and R5 regarding snow-covered areas. Shaded relief images of the DEMs are now shown for Astrolabe Glacier, the Mont Blanc area. We do not find the "low contrast alterations" mentioned by reviewer#2. To compute the geodetic mass balance and avoid outliers, a classical 3-sigma filter is used in each elevation band. This is now described in the revised MS. We also explain in the MS how data voids were treated (hypsometric extrapolation) and their error bars. Geomatica is actually using a threshold on the score channel to exclude unreliable pixels. The score channel has values ranging from 0 to 100. All pixels with a correlation score below 40 are excluded in the standard settings and replaced by the no data value in the DEM. For the Icelandic study site, we verified that, for the pixels above this threshold, there was no relation between the value in the score channel and the precision of the DEM. A clear illustration of this is provided by the comparison of the Score Channel on and off the ice cap. On the Tungnafellsjökull study site, the mean

score channel is 64 on the ice cap whereas off the ice cap it is 82. However, Table 4 shows that the precision of the DEM is higher on the ice cap than outside of it. Thus, no simple relationship exists between the score channel and the DEM precision. Thus we do not think it is worth complicating our procedure by generating systematically the score channel and applying a site-specific threshold to determine the outliers. The threshold will necessary be arbitrary and may exclude some reliable DEM pixels.

- In comparison with Pléiades, you used a lot of different data at very different study site. Once with LIDAR, once with SPOT-5, sometimes with and sometimes without GCPs... This can be confusing for the reader and it is not always easy to correctly relate the data and sites to each other. So you should try to make these things a little more clear in your text what is probably not easy.

A dataset section has been added with a description of all the data used in our study.

- Horizontal co-registration: Why did you not follow horizontal co-registration according to Nuth and Kääb (2011)?

This is because, based on (Rodriguez et al., 2006), we have developed our own procedure to co-register two DEMs (Berthier et al., 2007). We have verified that this procedure leads to similar horizontal shifts as the (Nuth and Kääb, 2011) algorithm. The latter is used however when a DEM and point-wise measurements (Lidar, GNSS) are compared.

- Vertical co-registration: Figure 3 indicates that your Pléiades DEM is sort of tilted related to your reference surface. Same is in my opinion still visible in Figure 6. In this regard you mention spatially-varying elevation changes that are however low (P4859). Instead of reducing the mean offset of elevation difference, why did you not calculate a linear trend surfaces to evaluate and remove your tilt as probable result of satellite attitude parameters? I am not sure, but maybe a polynomial trend surface of second order might be also suitable to correct for eventual further systematic influences that caused these offsets.

This suggestion was followed for Figure 6. We note that the influence of this additional correction on the geodetic mass balance is negligible (less than 0.02 m $a^{-1}$ w.e.), mainly because the Mont-Blanc area is located in the middle of the study region. We also tried a second order polynomial fit to the elevation difference. Here again the reduction of the standard deviation off glacier was minor (standard deviation lowered by only 0.2 m or 3%) and the region-wide geodetic mass balance changed by only 0.02 m $a^{-1}$ w.e., well within error bars. For the Icelandic study site, our goal in Fig. R3 is precisely to show those spatially varying biases, not to remove them. Here again, because the biases are small (about 0.1 m at most) and the ice cap is located in the middle of the study area, removal of this tilt would probably not alter significantly the elevation differences on the ice cap.

- The chapter of "Pléiades stereo images (2.2)" is very informative, but quite long, can you shorten it and make it more precise? It is quite technical for the scope of the journal, but surely of interest for glaciologist that intend to work with this data. I also ask myself if part of this information should not be better discussed in chapter 5, "Discussion and conclusion", which is by the way relatively short compared to the other chapters. Particularly your text from Line 21(P4858) to Line 8 (P4855) has not much to do with Pléiades imagery itself, but

with the specific data which was used in this study and which is well explained in Table 1. Maybe make a new section for it. Line 9 to Line 19 on this page (P4855) is about uncertainty estimation and should be placed elsewhere.

See our re-organized paper. We believe that it is very important to describe the Pléiades images because they are the heart of the paper and, as stated by the referee, this information is useful for glaciologists.
Text from Line 21(P4854) to Line 8 (P4855) is now included in a separate sub-section "2.3 Validation data: GNSS and Lidar"

- Captions of tables and figures are generally too detailed, please provide such information somehow in your manuscript text in order to make the captions more short and precise

We disagree. We believe that self-understanding figures and tables are much more convenient for the reader who wants to skim through a paper.

- In regard of your GCPs, how was their distribution in the scene? Isn't this an important influence factor how equally well distributed these GCPs are in your DEM? Where all of them clearly visible in the data?

Distribution of the GCPs shown for the Mont-Blanc and Icelandic study sites (Figure R6 and R7). In all cases we tried as much as possible to obtain the best GCPs coverage. Of course all of them where visible in the data, otherwise they would not be GCPs.

**SPECIFIC COMMENTS**

P4851:

L7: What kind of validation was employed? Rather study sites?

"validation sites" replaced by "evaluation sites".

L10-11: For what study sites you used GCPs?

These details are provided elsewhere in the paper, and they would make the abstract too long and distract the reader from the main results

L13-14: What do you mean with "around these biases"?

Wording retained. This is to make a transition between "accuracy" (the mean) and "precision" (the standard deviation around the mean).

L23-24: Why welcome? I don't think this the words "tools" and "welcome" fit to the context

Text was proof-read by a native speaker (Harvey Harder). We think "welcome" is OK. Editor?

P4852

L5-9: Sentence too long and therefore complicated

We think this sentence is OK.

L15-22: Sentence way too long and also too complicated also because of parentheses. There are five parentheses in one sentence which is hard to read.

Sentences split into two sentences and most parentheses removed.

L22-24: What gap do you mean since Pléiades DEMs can be extracted at high resolution?

A data gap. Aerial surveys are not feasible everywhere, whereas coarse resolution DEM are not precise enough.

P4853

L1-4: Pléiades-data by ISIS of CNES was available after the launch and not immediately for all European researchers (particularly those that are not affiliated to ORFEO member states)

Clarified now.

L19: "... launched on..."

corrected.

L20-23: Again, too many parentheses with long text make this sentence hard to read. Try to avoid such parentheses and include their information as part of the sentence

Split into two sentences.

L25 ()- L4: Too long parentheses, hard to read. Form new sentences...

Split into two sentences.

I think 12 bits should be clear and must not be explained in particular

Was added this info upon request of the editor and believe it is indeed useful.

P4854

L6-9: You are quite sure about this statement, based on the higher radiometric quality of Pléiades. But still, can you state it in this way?

See Fig. R4 and R5 (shows in our response to reviewer#1 and added in the revised MS).

L10-11: Don't use the expression "thanks to".

Include "along-track" and "pitch" as part of the text if possible... just in general, I am not against parentheses, but there are just a lot of them your manuscript.

"Thanks to" replaced by "due to".

L11-12: Please make it more clearly since you probably only mean the data of your study. Since Pléiades triplet-stereo images are also available for other parts of the world...

Modified to clarify that this statement is not generic but for our study sites.

L12-16: Again the parentheses issue... and I would not provide such a long URL at this place due to readability. Maybe it would be of interest as additional information somewhere else?

URL removed.

L16-L20: Particularly for the second sentence, can you provide a reference for this?

Reference added to (Toutin, 2008).

L21-L8: Much information is here provided about Table 1, and in the caption of table 1 there is also much information given . Try to fuse both information and make it more precise. Remove or omit unnecessary information that is not essential and that can be easily extracted out from the Table. Moreover, some of this information might be maybe better placed elsewhere in your manuscript.

As stated earlier we prefer keeping a self-understanding Table. We added a new column in Table 1 with the precision of the GNSS measurements. We think that all the information provided is needed.

L22-25: Try to include the parenthesis as part of the sentence

Sentence reorganized.

P4855

L9-10: I have problems to correctly understand this sentence. I understand that all elevation differences are errors in the Pléiades DEM. This would mean that there should be no elevation differences at all, what is right on stable terrain, but not on glacierized areas.

Exactly, you understood correctly. This is why we provide an upper limit to the errors in the Pléiades DEMs, in others words a conservative estimate of these errors.

L8-19: This section concerns DEM uncertainty estimation and should be placed in a separate chapter, maybe elsewhere.

The section was moved to the new section "3.2 Pléiades DEM evaluation".

L10-14: Make multiple sentences out of this single and complicated sentence.

A parenthesis removed but the sentence seems clear and short enough to us

L14: I do not clearly understand what upper bound does mean

"Bound" replaced by "limit".

P4856

L1: To what extent was the result not improved? The DEM should be as double as fine as with 4m I think...

Not necessary. DEM pixel size does not mean resolution. Anyway, the sentence was removed following suggestion of reviewer#1.

L6: Cross-reference to subsequent text makes it hard to read. Is it possible to re-arrange your chapters to make reading more fluent?

Text deeply re-organized but we did not find a way to avoid this reference to subsequent text.

L11: PCI can generate an additional score channel image when extracting a DEM to assess the correlation coefficient for each extracted DEM pixel. This can be another metric to describe the DEM quality. Why have you not considered this option?

See our reply to the related general comment.

L22: "Statistics after horizontal co-registration... ". Why did you not employ both horizontal and vertical co-registration to calculate the statistics afterwards?

Because we believe it is interesting to report on the remaining vertical bias after simple horizontal co-registration.

L23: Why did you not used the methodology of Nuth and Kääb (2011) for horizontal co-registration?

See our earlier answer to a similar comment from the same referee.

L28-L2: What does detectable horizontal shift mean? How have you conducted this verification? Visually? Of what magnitude were these shifts, particularly when you mention "small shift" on L2?

Visually. We have clarified our text: "*When GCPs were used to compute the DEMs, we always verified visually that no horizontal shift of more than one to two pixel (0.5 to 1 m) remained between the Pléiades ortho-images and the GNSS tracks acquired along roads and trails*".

P4858

L5: "...prominent features such as large boulders..."?

corrected.

L21: "The last column of this table...". Make it a little more clearer that this and the previous text is still referring to Table 3

References to Table 3 added more clearly.

P4859

L11-L24: Your approach with tiles is good, but why did you not calculate trend surfaces to evaluate the spatial pattern of these varying errors? It would be interesting of what polynomial degree this trend surface is, should be linear in case of satellite attitude

recordings, isn't it? This section is quite detailed and long, and again explained in the caption of Figure 3, you should shorten it I think.

We think it is important to clearly explain the procedure because it has not been used previously. Following the reviewer comment, we fit a polynomial surface to the 2003-2012 elevation difference for the Mont Blanc area where we want to compute the most accurate geodetic mass balance (but we found a very small influence of this correction). Here in the case of the Icelandic study site, we do not want to get rid of the spatial bias but want to illustrate it and we think that the tile strategy is a clear way to do so. See also our response to the General comment "Vertical co-registration" of reviewer#2.

P4860

L3-8: Try to reduce these three parenthesis

We note that one of the three parentheses is for a reference so cannot be skipped. One of the two others parenthesis has been removed.

L16-21: Precision of Pléiades DEMs: In your study you have a very good reference surface and the resolution of your DEM is probably well adapted to the resolution of this DEM reference. So I would not expect considerable curvature effects and dispersion as result of different DEM resolutions in your study which is well proved by your low NMAD. You argue that precision is more influenced by the landscape than by the DEM processing what is surely right. I might be wrong, but what about the precision in regard of my statements for the other studies that you mention?

"Very good reference surface": only for the Icelandic study sites the Pléiades DEM is evaluated against another DEM. Otherwise, it is compared to GNSS point-wise measurements. If this is what the reviewer asked, we do not want to speculate about the influence of the DEM resolution on the precision of the DEMs derived by others (Stumpf et al., 2014; Poli et al., 2014) for different types of landscape. Sorry but in fact we do not really understand what the reviewer meant here.

P4861

L5-25: Maybe I misunderstood, but you generally have employed correction of spatially-varying elevation errors to correct for mean vertical biases? Make this maybe more clear

In our procedure to evaluate the DEM (see section 3.2 of the revised MS) we never corrected for spatially-varying elevation errors. As stated already in this letter, such a correction was not performed for our Icelandic study site because we wanted to illustrate whether spatially-varying elevation error exist or not. Following reviewer#1 and #2 comments, a correction of a tilt in the 2003-2012 Mont-Blanc elevation difference map has been introduced. But with only a minor influence on the region-wide elevation change (and thus mass balance).

L26: Write out the abbreviation for TP or explain

TP was already defined but to avoid any headache to the reader we defined it again in the title of the subsection.

P4862

L6-11: Why has a nadir/back-pair stronger distortions? I would expect this for backward/forward views, probably you meant these views, since in L11 you again mention nadir/back.

Thanks for spotting this error. This is indeed the Front/Back pairs for which distortions are the strongest.

L11-15: The way you combined both DEMs is good. But why did you not use the DEM pixel that obtained the higher score? You can use such a setting in PCI (Score channel). Then you would not use the mean elevation, but the elevation value with the higher quality.

The procedure described by the referee makes senses. It was indeed evaluated (for the 2013 tri-stereo in the Mont-Blanc area) but did not lead to improved results when the DEM is compared to GNSS measurements on glaciers. The suggested procedure, based on the Score Channel, is also specific to a software (Geomatica) so we prefer to propose a simple but more generic merging of the different DEMs from a tri-stereo. For example, our simple merging could be used with DEM generating tools that do not output a score channel.

L21-L23: What do you mean with homogenous? Did you observed that vertical biases showed less "noise" in regard of their spatial distribution, or was it less systematically and trend-like?

"Homogeneous" was maybe ambiguous. Replaced by "similar".

P4863:

L12-16: How good is the agreement? Try to include the parenthesis in your sentence

Agreement now quantified. Parenthesis removed by splitting the sentence into two.

P4864

L9-14: You observed thickening in the accumulation zones. I would expect that there are DEM elevations in such snow covered areas which are of poor quality because of high saturation, despite the high radiometric resolution of Pléiades imagery. Terrain extraction might be hampered in such areas and when regarding your difference image in Figure 5 I wonder if obviously high noise of difference elevation values in these accumulation zones are not an indicator of such worse DEM pixels.

See our statistics for the altitudes above 4000 m a.s.l. in the Mont Blanc area and see also Figure R4 & R5. Both show the excellent quality of the Pléiades DEMs in the accumulation areas and in Antarctica.

P4865

For geodetic glacier mass balance determination, have you conducted some statistical analysis to exclude outliers of elevation differences within your glaciated areas? In Figure 6 I remark some noisy areas of elevation differences within your glacier accumulation areas. Also, have you filled gaps within glacier areas particularly in the accumulation zone? To my understanding I would first try to eliminate described outliers and then fill remaining gaps of elevation differences for each glacier. Then I would calculate the mass balance from the gap-filled and cleaned difference image for each glacier. What do you think? You difference a SPOT-5 DEM for your Pléiades DEM. Due to the worse radiometric resolution of SPOT-5, I particularly wonder how good its quality is particularly in such snow covered glacier accumulation zones.

The procedure for deriving mean elevation changes in each altitude band is now explained. A classical 3-sigma filter is used to exclude outliers in each altitude interval.

Treatment of data gaps is also more clearly described and, very conservatively, the error bars for un-surveyed pixels is 5 times higher than for measured pixels.

P4866
L2: His name is Kropacek (I omitted the accents here)

Thanks, this was a typo introduced during type-setting. It was OK in the submitted (and now revised) MS.

Table 1: You mention for the format of the datum is DDMMYYYY, but in the table column "Pléiades date" is written DD Month (not as number) YYYY. There are too many parenthesis in the caption text. You explain B/H-Factor very well, but better do this in the manuscript text. On the other hand, I would be interested what "Stop and Go" GNSS exactly is.

The type-setter changed the format of the dates in the table but did not update the caption. We did not detect this during proof reading, thanks for spotting this. This is now corrected. Most of the parentheses are removed. B/H is never used in the text, so we prefer to define it directly in the caption of Table 1.

Table 2: Parameter settings are particularly the case for PCI software, no other software. You tested various settings, but in regard of terrain type and DEM detail you only tested the most extreme settings which are not the default settings (for terrain type). What about the other terrain types (e.g. hilly) and DEM detail settings (e.g. extra high)? Also this caption text is quite long, explain some details maybe in the manuscript text.

As already explained, we prefer self-contained items (Table and Figures) so the captions are long.

Processing parameter settings: There are a large number of combinations that could be tested in PCI: 4 possibilities for DEM details, 3 for the type of terrain and then different resolutions. We tested nearly all of them but to avoid lengthening the MS and losing the reader, we preferred to present the results for a few very different settings. We are careful to state in the paper that: "*We acknowledge that the differences obtained for different processing parameter settings are not very large and hence that other settings may be more appropriate in some cases*."

Table 3: Which settings were finally used?

The settings are provided in the caption "*The parameter settings used to generate all the DEMs are: DEM detail = Low, Type of terrain = Mountainous, pixel size = 4 m, Data gaps = not filled*".

Table 4: This is obviously the most important table, but I have problems in completely understanding it. You mention that the front/back views were often not applicable, so maybe it is not necessary to mention it here. This table mentions the accuracy / precision measures of your DEMs. Which are the final and valid uncertainty estimates of your work? You provide here different values depending on the number of GCPs and your image combination. What was finally used or can be seen as valid? Why do you provide these values on-glacier and not strictly off-glacier? For some study sites you provide both, but for some sites only for on-glacier surfaces. I would not expect accuracy estimates from on-glacier surfaces... or I do get something here completely wrong...

"front/back views were often not applicable": we do not understand this comment.

"Final uncertainty estimate": we were careful in the entire article to separate the mean bias and the dispersion around this bias. So one cannot really define a "final" uncertainty estimate. Without GCPs and no other reference data to correct the mean bias, one can expect a vertical bias (and thus systematic error) of up to 7 m (maybe more for other Pléiades stereo paris). As stated in the abstract and as a rule of thumb, dispersion (i.e. precision) of 1 m or less is achieved for nearly all our sites. Our goal is not to provide a "valid" final DEM on each site but rather to test the influence of the processing parameter settings and assess the necessity of GCPs and TP.
Our final section "Summary and conclusion" already includes our recommendations: a cost effective solution is to obtain a stereo-pair with a base-to-height ratio of about 0.35 and 0.4 and collect 1 to 2 accurate GCPs (at least, more are welcome for the sake of redundancy) and tie points.
Comparing value on/off glaciers is very interesting and useful if the GNSS survey was performed close in time to the Pléiades acquisitions. This is not available on all sites. As stated in the main text "*This is confirmed by the results for the two study sites (Agua Negra and Tungnafellsjökull) where GNSS data have been collected on and off glaciers (Table 4). The precisions are always higher on glaciers. The improvement is particularly spectacular on the Tungnafellsjökull study site where the standard deviation of the elevation difference is 0.53 m on the ice cap and 1.33 m elsewhere.*"

Figure 2: Your ortho-images are pretty small and do not show much information to my opinion. In some of them it is hard for me to get an idea of the terrain and the environment. You should not use a similar color for your scale bar and for the limits of the Lidar DEM.

We will ensure that this figure (R2) is printed at full width. It is very important to show the location of the GNSS reference measurements. A different color is now used for the scale bars.

Figure 3: Nice looking map! I wonder for the most south-eastern tile, this is covered by almost no values of your difference image. From a statistically point of view, how can you be sure that 0.10 is representative for the median when there are only very few values?

Thanks for this comment that help us to detect and correct an error during the production of this figure. When creating this figure, we imported as background a Pléiades ortho-image that had a larger extent than the map of elevation difference. Thus dividing the whole image into 3 by 3 tiles as we did in the submitted MS did not equal to dividing the map of elevation difference in 3 by 3, as done in our script. See the revised figure (R3) where the tiles have been redrawn correctly. Note that the script is correct thus the values for the mean elevation difference off glacier in each tile are unchanged. This was only a visual issue in the figure. The area covered in the south-eastern tile is still relatively small (0.2 km²) but given the high resolution of the differential DEM (4 m), it includes >10 000 values of elevation changes. This is 1 to 2 orders of magnitude less values than in other tiles but still sufficient to provide a useful estimate of centrality of the distribution. This is confirmed by the similarity between the mean and median and the low standard deviation. In this south-eastern tile, the mean elevation difference is 0.12 m, the median is 0.10 m and the standard deviation is 0.37 m.

Figure 4: The scatter plot is comparably small compared to your difference image, particularly in regard of the pretty thick scale bar. You should adapt the colors that you attached to your elevation changes. I think it is not a good idea to use two different colors (red and blue) for only negative values. Blue communicates ice mass gain, what is not the case here. Better use only the red color. Why is the highest loss a much larger class (-7 to -4.5m)? Better mention more as -4.5m loss.

Now Fig. R6.

A color scale with a single color did not show the variability of the elevation changes as well. The color scale was kept unchanged. -7 to -4.5 is prefer to <-4.5 because it shows the most negative value of the elevation changes. Our blue-to-red color scale has the advantage to show in red the warm regions of strong summer compaction and melt (at low elevation) and in blue, cold regions of the ice cap where only limited melt and compaction occurred.

Figure 5: It is generally good, as you also did in your other figures, to provide the entire difference image and not to cut the glaciers. How confident / trustable are the difference values particularly in the glacier accumulation zones? What about outliers in regard of the noise of elevation differences in this region? Please have in this regard a look to my comments that I made for P4864 and P4865. The triangle signatures in this figure are hard to see. I think a hillshade in the background (here instead of SPOT-5) in combination with your difference image at a certain transparency would improve its quality. I remark gaps in this difference image, where are they coming from? Is it because you did not employ void filling during DEM extraction? Have you filled these voids by some way in glacier areas for mass balance calculation? From a cartographic point of view, it is not good to use more as about seven to ten different classes, because it is then hard to correlate the colors from your scale bar to the map. Or use a continuous color scale bar instead when having many classes. Use <-7 and >+7 instead of the maximum value of +-22.

Now Fig. R7.

Accumulation areas: see our previous responses

Size of the dark triangles slightly enlarged

Data voids not filled. The MS has been clarified in Section 4.1: "*With the parameters "Type of relief" set to "Mountainous", "DEM detail" set to "Low" and without filling data voids, the dispersion is just slightly larger and the area covered is greatly improved (nearly 99% versus less than 93%). All DEMs examined in the rest of this study were generated using these parameter settings*"

Treatment of data voids for mass balance calculation is now clarified in the MS.

The number of color classes has been reduced to 9.

We prefer to show the maximum value of elevation changes in the color scale bar.

Figure 6: Please consider similar remarks as for Figure 5 in regard to the scale bar and in regard to outlier detection of poor quality difference values. This is particularly in regard to differencing from a SPOT-5 DEM at worse radiometric quality. Why have you not used a background image at all (I would prefer a hillshade)? The inset figure is too small, the text and numbers in this inset figures are almost not readable. Same for the text boxes in the main figure, they are too small (or the figure is simply not large enough here). I also wonder in regard of the stable (non-glacier) terrain, some of it is quite blueish (too high), particularly to the north and south, and to the east it seems that stable terrain is quite redish (too low).

When regarding your color scale bar, I would expect this at a magnitude of about +-10m. Is there still some spatially dependent offset existing, maybe a tilt? What could be the reason of it?

Now Fig. R8.

No background: this is in purpose to really show where data gaps are (important for mass balance calculation); Visualization of data gaps is less obvious when there is a background.

We will ensure that the figure is printed in full size and will verify carefully his readability.

The number of color classes has been reduced to 9.

Tilt: Now corrected as explain in our previous answers and in the revised MS.