ESA's Ice Sheets CCI: validation and inter-comparison of surface elevation changes derived from laser and radar altimetry over Jakobshavn Isbræ, Greenland – Round Robin results

J. F. Levinsen, K. Khvorostovsky, F. Ticconi, A. Shepherd, R. Forsberg, L. S. Sørensen, A. Muir, N. Pie, D. Felikson, T. Flament, R. Hurkmans, G. Moholdt, B. Gunter, R. C. Lindenbergh, and M. Kleinherenbrink.

We appreciate the constructive comments submitted by the two anonymous reviewers. We have carefully considered each comment and revised the manuscript accordingly so we believe it meets the high quality standards for publication in *The Cryosphere*.

This document provides comment-by-comment responses to the reviewers, thus summarizing the changes made to the manuscript.

Response to Reviewers:

In the following, reviewers statements are written in bold font and author response in normal font.

Anonymous Referee #1 (RC C2975)
Received and published: 10 January 2014

General comments:

Over the last few decades, satellite altimetry has become an important tool to monitor the state of the ice sheets. Several research groups have developed their own methods to process the data, and different sensors are being used. This makes that published results are not always directly comparable to one another. Although earlier studies have compared methods and sensors (e.g., Thomas 2008), a thorough inter-comparison as is the purpose of this manuscript is still missing (to my knowledge). This study therefore has the potential to make a relevant contribution to the field. For example, the finding that radar altimetry is able to capture the surface elevation changes in the marginal regions of the Greenland ice sheet is quite important. However, I believe that in its present form, the manuscript shows some methodological (and other) shortcomings which should resolved before publishing the article.

Specific comments:

*** As I mentioned above, the finding that radar altimetry is able to capture the surface elevation changes in the marginal regions of the Greenland ice sheet is novel and would imply that the ERS/Envisat observations have more potential than hitherto assumed. Unfortunately, no details about the processing and, in particular, how it differs to the processing of radar altimetry used so far are given. I assume that the authors will publish this in a separate article, but a short (∼ one paragraph) description should be given, in my opinion.**

The discovery of radar altimetry's potential for change detection even in margin regions of Greenland is indeed novel, and the reviewer is correct that very little information concerning the processing specifics was given originally. This has been changed as we have added both tables describing the processing details of each submission (Tables 3 and 4) as well as short descriptions of how they agree/ differ from one another (Section 2.2). This makes the specifics behind each result more easily understandable and thus allows for comparisons across submissions as well as with previous work. However, as assumed, a paper regarding the specifics of the final surface elevation change production is currently in preparation. Therefore, although more descriptions have been given on each method, the authors find it that the full details e.g. concerning SEC-1's contribution and the resulting final elevation change technique cannot be given here. We hope that the reviewer understands our reasoning.

*** The manuscript focuses on the differences between different methods (crossovers vs. repeat track) and sensors (laser vs. radar) but a discussion of how the contributions of the different research groups given a method and sensor differ from one another (e.g. how do the ICESat repeat track (SEC-2 to 5) compare?) seems equally important to me, for example to give an idea of the spread and methodological uncertainties.**
As mentioned above, both tables and text regarding the specifics of each contribution have been included through a separate methodology (Section 2.2, Tables 3 and 4). Furthermore, the discussion regarding error derivation, data rejection criteria, etc. has been expanded (Section 4), just as well as Table 6 provides the details of the validation. All of the above aid the interpretation of the achieved results and thus provides the reader with a better idea of how the submissions vary from one-another.

*** I was very surprised to read that 'the method for finding the remaining [error] estimates [of the repeat track SECs] is unknown'. As you mention, 'the error estimates provide important information on the accuracy of the observations and methods and are therefore included', but without any background documentation on how they were computed, they are relatively meaningless. For example, SEC-2 and SEC-5 both are based on ICESat data and use a repeat-track method. Comparing their results in fig. 1, SEC-5 appears to be slightly more noisy than SEC-2. Yet, the SEC error estimates of SEC-5 are an order of magnitude lower than those of SEC-2, which is puzzling. Furthermore, on page 5441, the fact that the error estimates are at the sub-meter level is used to confirm the high accuracy of the results. Without a proper description of the error estimates, it's hard for the reader to tell if such a claim is justified. According to the 'author contributions' section at the end of the manuscript, the Round Robin participants are co-author of this paper, so I would expect that it should be relatively easy to obtain this information and include this in the manuscript, which allows the reader the results correctly. Furthermore, I strongly recommend to ask the author of SEC-10 for his/her error estimates too. In science, every number should come with an error estimate to allow to draw valid conclusions. The fact that SEC-10 is being used in the RR inter-comparison in section 3.2 makes this even more important.**
The reviewer is correct that without proper descriptions of how the error estimates are derived, the different Round Robin contributions cannot be directly compared. Therefore, such descriptions have been included (Section 2.2 and Tables 3 and 4), just as well as SEC-10's errors have been assessed and added. An updated version of Fig. 2 is therefore available in the revised version of the paper.

*** In the inter-comparison section, the motivation for choosing the experiments to be compared is not always clear. Take, for example, the Laser repeat-track vs. cross-over comparison, where you compare SEC-3 (repeat-track) to SEC-7 (XO). Another option would have been to compare SEC-2 (which covers the full study domain, not only the drainage basin) to SEC-8 (which scores better than SEC-7 in the comparison with the LiDAR data), or SEC-4 to SEC-6, or ... This can have a substantial impact on the result of the inter-comparison and following discussion: figs 1 and 3 show that the SEC-7 dH/dt are all nearly-zero, so in essence, the mean and RMSE in table 3 are basically the mean and RMS of the SEC-3 dH/dt. It also explains the low slope of 0.01. Replacing SEC-7 by SEC-6 or SEC-8, which both show a much better (visual) agreement with SEC-3 in fig.**

1, may lead to very different statistics and a different discussion. (The same applies for the other inter-comparisons, where other combinations are possible as well). The choice of experiments should be justified properly, and a more detailed discussion of differences in the results of experiments using the same method and sensor seems required (see my point above).
Initially, elaborate descriptions of the motivation behind the selection of Round Robin submissions to inter-compare were not included. This has been done, just as well as the consequences of this selection on the results have been described.
Finally, it has been clarified that all inter-comparisons are carried out for observations lying within similar spatial domains. Only overlapping and neighboring comparisons falling within a specific search radius are used, thereby making the analyses more reliable and explaining why comparisons with e.g. SEC-7, whose results are near-zero, are carried out.
The search radius is based on the spatial resolution of the given submissions, and whereas this radius is mentioned in Table 5, the use of it was not previously described in the text. This has now been changed.

As also described in the revised manuscript, the reasoning behind the inter-comparisons is as follows:

- SEC-1 to -10:
  - Both based on radar altimetry,
  - Almost similarly sized grid cells (5x5 km vs. 10x10 km) whereas SEC-9 uses 0.5°lat x 1.0°lon.
- SEC-3 to -7:
  - To specifically demonstrate the influence of the size of the area over which observations are used in the derivation of elevation change trends. These submissions have used 1x1 km and 0.5°lat x 1.0°lon, respectively, and the resulting large disagreements thus show the importance of using a high spatial resolution.
- SEC-8 to -10:
  - The submissions have the most similar spatial resolutions,
  - SEC-6 vs. SEC-10 could be another possibility. However, as SEC-6 uses a slightly different approach (any combination of campaign shots is used rather than those from only ascending and descending orbits), this idea was discarded.
- SEC-3 to -1:
  - Best spatial resolution overlap, considering that SEC-5, which has the same spatial resolution as SEC-3, has larger disagreement with the lidar data.

*** The fact that SEC-3 only computes the dH/dt for the drainage basin may affect the statistics in table 4. You mention on page 5444, the differences between the LiDAR and altimetry observations are large north of the glacier basin. These points are not included in SEC-3, which likely leads to a lower mean and std of the LiDAR-altimetry differences. In order to get a fair comparison, the statistics should be computed for identical domains.**
The reviewer is correct that similar geographical areas must be considered in order to accurately compare SEC-3 with the remaining submissions. Therefore, throughout the paper, we have stressed that all comparisons are made for spatially similar areas, just as well as we have pointed out that SEC-3's submission is limited to the drainage basin, and that this may affect the statistical outcome. The latter is mentioned in the beginning of Section 3.

***p. 5444, line 10: It might be worth to point out that SEC-7 uses a similar grid spacing as SEC-9, and has a similarly high mean(diff_lidar).**
This sentence has been changed.

***p. 5445, line 3: Borsa et al. 2013 discuss the Gaussian vs. Centroid (G-C) offset, which explains only part of the inter-campaign bias. Even after correcting for the G-C offset, intercampaign biases will remain, they conclude.**
That is correct: The GC offset determined by Borsa et al., (2013) does not explain the entire inter-satellite bias contribution, and thus the text in the Discussion has been changed accordingly. It now describes how Borsa et al., (2013) reproduced part of the inter-satellite bias signal, and that more work is still needed in order to gain a full understanding of its implications.

Technical comments:
*** I'm not a big fan of using abbreviations in the title. ESA is probably well known, but CCI should be written in full.**
As the abbreviation CCI is explained in line two of the abstract, and the title is already long, the authors believe that leaving CCI herein does not lead to confusing the reader.

*** p. 5435, line 4, "13 projects either affecting or affected by the concurrent changes": The projects are not affecting/affected by climate changes, but rather study parts of the Earth system affecting/affected by the changes.**
This sentence has been changed.

*** p. 5435, line 12: "most optimal method" -> "optimal method", there is only one optimal method**
This has been changed.

*** p. 5435, line 17-18, "... due to the high accuracy of the former and the high spatial resolution of the latter": this implies that the Envisat measurements have the high accuracy and the ICESat measurements the high spatial resolution, but this probably should be the other way around (as you also mention on line 21).**
The sentence was meant to refer to the high accuracy obtained with the cross-over technique as well as to repeat-track measurements' higher spatial resolution. In order to avoid confusion it has therefore been changed so the correct meaning hopefully comes out clearly.

*** If necessary, the introduction could be shortened. At present, it discusses the ESA CCI in quite some detail, but this is not essential to understand the remainder of the article.**
Some general information about the CCI project has been left out, and the section now provides a more brief description, slowly narrowing it down from the specifics of the CCI project to what is in focus in this work. This should provide a good transition to the essentials of the Round Robin exercise.

*** p. 5436, line 17-18: Wouters et al., NatGeo, 2013 provided a detailed discussion and analysis of the increased mass loss of the ice sheets.**
The authors thank the reviewer for the notification regarding Wouters et al., (2013). The focus of the sentence has been changed to the previous decade's mass loss rather than the acceleration hereof as such acceleration does not exceed the natural variability of the ice sheet's SMB component.

*** p. 5436, line 23: What is the motivation for the dimensions of the grid (5x5km)?**
The motivation for the mentioned grid size is based on a compromise between the spatial data availability, the footprint size obtained with radar altimetry, and the data accuracy. This is described in detail in the Discussion and briefly summarized in the Conclusions. The grid spacing is based on the outputs from the Round Robin exercise as e.g. SEC-7 and -9 have large grid spacings resulting in correspondingly large disagreements with the LiDAR data used for the validation (0.5°lat x 1°lon; see Tables 2

and 6). Further confidence was obtained as SEC-1 demonstrated that radar altimetry data is capable of resolving elevation changes even in margin regions using a 5 x 5 km spatial resolution.

*** p. 5437, line 23: explain what CRYOLIST is, or provide url. Not every reader will be familiar with this mailing list.**
A description of CRYOLIST has been given.

*** p. 5438, line 12-13: since only one of the submissions was rejected, it would make more sense (and be less ambiguous) to simply state which of the two potential rejection criteria applies (i.e., was the submission rejected because it is comparable or independent, or both).**
The sentence has been changed.

*** p. 5448, line 19: add the url where the Envisat SEC prototype can be found.**
An url has been included.

*** p. 5455-5456: To interpret the results in tables 3 and 4 correctly, it would be very helpful to list the number of points used in the computation of the statistics.**
The respective tables have now been updated with the number of points used in the statistics.