**Anonymous Referee #2 (RC C3054)**
**Received and published: 20 January 2014**

**As satellite altimetry missions continue to monitor elevation changes of the cryosphere, the assessment of elevation change algorithms is timely and important. This study compares the performance of difference altimetry methods (lidar vs. radar) and processing algorithms (along-track vs. across-track) by comparing reconstructed elevation changes with ground-truth obtained by airborne laser altimetry within the rapidly changing drainage basin of the Jakobshavn Isbræ in West Greenland. The study uses a wide variety of algorithms and relies on a careful examination of the results. However, there appears to be several flaws in the analysis that should be addressed to produce a robust comparison.**

**** Selection of suitable cross-over methods for ICESat change detection.*** **The appr. 70 m footprint of ICESat lidar measurements allows the reconstruction of elevation change time-series within small regions, measuring 1-2 km2. Several participants of the study (SEC-6, SEC-7) applied crossover change detection methods adapted from radar altimetry measurements that use much larger grid cells. These solutions, especially SEC-7, dramatically reduce the resolution of the results, compared with methods developed for ICESat change detection (see for example SEC-8). Therefore, I recommend removing the SEC-7 result from the Round Robin experiment as it does not adequately represent the results obtained by ICESat specific cross-over change detection methods.**
The reviewer is correct that particularly SEC-7 uses a grid-cell size much larger than what is necessary for extracting an elevation change signal from the ICESat measurements. In the inter-comparison, however, this analysis is included to illustrate exactly that: That when using such a large spatial resolution considering ICESat's footprint size, a large amount of the otherwise observable signal is lost. This being the key conclusion from the SEC-3 vs. SEC-7 analysis, we have chosen to include the submission in the Round Robin dataset. In doing so, we have therefore stressed the importance of using a smaller grid-spacing in the case of laser altimetry, a conclusion, which also stands out in the following section regarding the validation.

**** Present and examine the spatial distribution of airborne laser altimetry data used for validating the trends derived from satellite observations.*** **While the mean differences between the satellite derived elevation changes and the validation data sets are small, they have large standard deviation. It is hard to interpret these results without examining the spatial distribution of the elevation change errors.**
In order to facilitate an easier interpretation of the Round Robin results, the assessment of those as well as the validation has now been carried out for observations above and below 2000 m, respectively. This is clear in Sections 3.1 and 3.3 as well as Table 6.
As we received surface elevations from only a number of the participants, the GIMP DEM by Howat et al., (2014) was used as a reference surface. The outcome of this elevation division revealed the expected result that most large errors are found at low elevations, explaining why different change estimates also occur here, as well as that a few submissions have large errors in both regions.

**** Derive separate statistics of elevation errors for fast flowing, crevassed and rugged glaciers and for smooth ice sheet surfaces.*** **According to the manuscript, most elevation change errors are in the range of 0-1 m/yr over the smooth, higher elevation part of the ice sheet. Thus, the overall standard deviation of 1.3-5.5 m/yr of the elevation change errors implies very large errors within the steep, rugged coastal zone and over fast flowing outlet glaciers. A separate presentation of the errors for the smooth, higher elevation part and the rugged coastal region of the ice sheet would**

**allow a better assessment of the different methods used for determining elevation change rates.**
In the new manuscript, the validation of the Round Robin results has been analyzed in a manner similar to the errors, namely relative to a division of observations into areas above and below 2000 m altitude, respectively. Cf. the largest errors being found in the margin region, this is generally where the largest disagreements with validation data occur as well. This allows for analyzing the methods' performance in both rugged, dynamic regions as well as over smooth surfaces.

*** The very large differences between the crossover and along-track elevation change estimates derived from the same sensors and shown in Fig. 3 are especially disturbing. It is likely that the large differences between the laser altimetry derived XO and RT elevation change rates are caused by the extremely large cell size selected for deriving the XO solution (see 1). However, RT and XO solutions were derived using similar cell sizes for radar altimetry. The manuscript states "that is spite of a relatively high R 2 the different methods do not resolve the same signal". This observation indicates a non- linear evolution of elevation change in time. Taken together with the rapid increase of elevation change rates toward the coastal regions and over the fast flowing Jakobshavn Isbræ, the altimetry record indicates a complex spatio-temporal evolution of elevation changes in the study region. Therefore, simple statistics, such as mean and standard deviation of the differences over the whole region might not be suitable for evaluating and comparing the performance of the different methods.**
The reviewer is correct that the large difference between the laser XO and RT results (SEC-3 vs. SEC-7) can be attributed to the different grid-sizes used by the participants. When instead considering the radar results (SEC-1 vs. SEC-10) a disagreement is indeed observed. We believe it arises from a combination of differently sized grid-cells as well as backscatter effects, which Khvorostovky (2012) has proven are necessary to correct for in order to carry out a reliable change detection. Yet another contribution may arise from the observations not completely overlapping, and when considering the margin region, where the differences between SEC-1 and -10 occur, large topographic changes over small distances can change the actual signal a significant amount. Thus, if two measurements are compared, which are physically separated by just a few km, differences may occur, which naturally translate into the results. This has now been stressed in the analysis.
We therefore conclude that considering the mean and standard deviations of the elevation change differences is suitable for comparing the methods, particularly when specifics behind the methodologies are also assessed.

*** Finally, I recommend providing detailed information about the participants of the Round Robin experiments, including the name/affiliation of the research groups, participants and relevant publications, of course without establishing a connection between this information and the sensors and methods listed in Tables 1-4.**
As part of the encouragement for contributing to the Round Robin exercise was for the participants to remain anonymous, we do not believe that more information regarding the specific methods can be revealed without tying each submission with the corresponding group.
We have already listed the Round Robin participants including affiliations in the author list, and from the acknowledgements it becomes clear exactly who are participants and who form the ESA team behind the survey. Furthermore, the updated version of the paper includes two tables in which details regrading data processing and error estimation have been presented. If we also add the relevant publications supporting each submission, it will be straight-forward to compare the results with the list of participants, and this is something we wish to avoid. We hope the reviewer understands.