Authors response to the comments of referee #3

We thank the reviewer for the thorough, fast and helpful evaluation of the manuscript. His suggestions, including also the minor comments, helped substantially to improve clarity. We would like to mention that the correction of a small inconsistency in our variance propagation of the GNSS profiles and the variable ICESat campaign precisions (recommended by reviewer #2) slightly changed the results for the campaign biases. However, this does not change anything in the general messages of the manuscript.

[RC] *1. Presentation: A major component of this paper is the derivation of inter-campaign biases in the ICESat data. The authors present a series of histograms showing the residuals between the ICESat-derived heights and those derived by kinematic GNSS in Figure 7b. It would be useful to also have a third panel in the figure showing residuals between ICESat and kinematic GNSS prior to the inter-campaign bias correction, so the reader can see the effect of the correction in improving the height residuals.*

We decided not to include the suggested histograms as they would look very similar to those for the corrected data. In the time invariant treatment of the data (Fig. 7b), the crossover statistics of the uncorrected data will mainly differ in its mean values. This is a result of the mean value of the biases. The standard deviations will not be significantly larger (e.g. 11.1 cm for the uncorrected, 10.7 cm for the corrected data over Lake Vostok) as most of the corrections are in the order of only a few centimetres. The main effect of correcting the biases is to remove the spurious trend before the calculation of temporal changes from the data. We added a sentence towards this comparison in the discussion of the results.

[RC] *Figure 7a needs to be presented with error bars.*

We followed this suggestion and modified Figure 7a accordingly.

[RC] *All histograms shown in this paper should be presented with the number of measurements that go into it (this could be added to all the legends along with the mean and standard deviations).*

OK, this was included for the crossover histogram statistics. For the validation of the DEMs, such a number would not be very meaningful. Here elevations were interpolated and differences calculated for each kinematic GNSS position. Hence, this would just reflect the number of GNSS elevations in a specific zone but say nothing about their spacial coverage of a sufficient area of this zone.

[RC] *2. Stability of reference surface: There can be issues with height changes in the reference surface between the time of acquisition of the GNSS data and the time of the altimetry-derived height measurement (i.e. the reference surface is not always stable). Toward this, the authors should include a time series of GNSS data over Lake Vostok to demonstrate the variability in the observed height over the reference surface and provide statistics on the time difference between GNSS and altimetry datasets at crossovers.*

This point is discussed in section 2.5. For radar altimetry, the errors of the slope correction (up to 10 m and more in the margins) exceed the possible elevation change rate (maximum around 10 cm/yr) in this region by at least an order of

magnitude. As elevation changes are important when determining the ICESat campaign biases, we included $\dot{h}$ in Eq. (4).
However, we agree that a time series helps to get a deeper view into the temporal variations and their origins. Therefore we added Fig. 7c.

[RC] *3. Precision of GNSS estimates: The mean baseline differences presented here are only one potential source of error in the GNSS estimate. In this paper, however, there is no mention of the precision of GPS measurements. To estimate this, the authors could look at GPS data collected by a tractor/trailer in a single place for an extended period (hours) and present the noise in the determined height (see Borsa et al, 2007 Modeling long-period noise in kinematic GPS applications). This is important if the authors include Table 3 in the final version, since this can be a major source of uncertainty. If this cannot be included, the authors should add a note about this in the discussion.*

We use the terms precision and accuracy now in a more precise way. We have already done the suggested analysis for measurements acquired during the stops of the vehicles. The elevations vary within 1–2 cm. However, we did not include this in the paper, as it provides only information about white noise. But there are further possible sources of errors as, for example, the ambiguity resolution, which are not detectable in this way. We think, solving the same epoch as a differential solution using multiple baselines is a rigorous estimate for the repeatability of the epoch solution.
In contrast, the RMS of the final surface elevation is determined from crossovers between different profiles of one season. These profiles are considered independent and thus, this RMS is a realistic estimate for accuracy. In addition, now we include also an independent dataset (ICEBridge) which confirms our accuracy estimate. However, the small amount of crossover points here does not provide sufficient statistical significance.

[RC] *4. Residuals through time: The authors show residuals through time between ICESat and GNSS (Figure 7a). It is not clear if the trend in the residuals is unique to the ICESat period. Since the authors already have the data for Envisat and CryoSat-2, I suggest that the authors also plot residuals between each those missions and the GNSS, with time on the x-axis. One way of doing this would be to bin the GNSS-altimeter residuals into yearly (or other) intervals and plot them over the whole time period.*

There might be a misunderstanding here. Fig. 7a does not show residuals over time between the two datasets, it shows the result of the bias estimation using Eq. (4). The trend is not a true elevation change, it is the apparent trend introduced by these biases. The results of Eq. (4) suggest that $\dot{h}$ is 0.0 $\pm$ 0.2 cm/yr.

[RC] *Furthermore, there is an offset (5 cm) in the residuals between ICESat and GNSS. Through the caption over Table 3, the authors imply that the ICESat elevations need to be corrected for that offset. However, there is no evidence that ICESat-derived heights are biased by 5 cm (this would be a major finding if this is real), and this is likely a bias in GNSS-derived heights. The authors need to discuss potential causes of this discrepancy.*

Yes, the mean value of our biases is 5 cm and we believe this is real. This result is not exceptional: Other authors even calculate much bigger mean offsets

(Zwally(2015): -28.8 cm). We know, that the accuracy of our profiles is only some centimeters (see answer on question 3.). Nevertheless, by comparing different profiles of one season and even ICEBridge, we can limit this uncertainty to below 10 cm. Using profiles from 9 different seasons we think that in average we come very close to what is the 'true' surface elevation.
Besides this, the biases are usually used when calculating elevation CHANGES from ICESat data, where any possible mean offset vanishes.

[RC] *5. DEM analysis: The section on DEM validation detracts from the major substance of the paper, i.e., validation of L2 heights derived using various altimeters. The comparison of the ICESat and CryoSat-2 DEM's could be useful in terms of the assessment of their accuracies, but I suggest the authors consider removing the comparisons with the Bamber-DEM and the Bedmap2-DEM.*

We do not think, that there is a conflict between the L2 elevations and the DEMs (or so called L3 grids). They are just a step further to what most end users need. As these grids are created from satellite altimetry data, this nicely matches the scope of this paper. We think there are some very interesting coincidences between the DEMs and the datasets which have been used to create them. Furthermore, especially the Bamber-DEM and the Bedmap2-DEM are widely used products and importand to include in this study.

[RC] *6. Crossover analysis: Section 3.2.1 (Paragraph 4) This paragraph mentions that the crossover method is outlined in Section 2.4, but there is no mention of how altimeter-GNSS crossovers are defined (which is a major aspect of the manuscript). Does the altimeter footprint need to overlap with the GNSS track, or are the authors interpolating the altimeter track crossing the GNSS traverse to obtain a measurement on the traverse? A further technique one could use would be to fit a line to a few altimeter measurements in the along track direction around the crossover point - the prediction of the line fit at the crossover location would be the altimeter-derived height. The technique the authors use should be discussed and justified.*

This is now explained in more detail in the manuscript.

Minor comments:

1. Section 3.1.3 - Include reference for the saturation correction.

Done.

2. Section 3.2.1 (Paragraph 3) This whole paragraph is confusing, with no citations or anything of substance. Maybe the entire paragraph can be rephrased as "some studies that used GNSS data for satellite altimeter calibration/validation use a 2-D gridded reference DEM (cite studies), but we do not adopt this here due to observational limitations". If not, this paragraph can be deleted, since it does not add too much to the discussion.

The whole paragraph has been tightened.

3. Section 3.2.1 (Paragraph 4) I think the along track sampling is around 290 m for CryoSat-2. Check (Wingham 2006) or the CryoSat-2 product handbook for details.

Changed.

4. Section 3.2.2 (Paragraph 1) Consider replacing "Between the campaigns systematic biases exist. If not corrected carefully, these biases corrupt the inference of temporal surface-elevation changes and estimates of height change" with "If not accounted for carefully, any systematic biases between campaigns can corrupt the inference of temporal surface-elevation changes and estimates of height change"

Changed.

5. Section 3.2.2 (Last Sentence) Provide reference for or justify using this value (10 cm).

Modified according to review #2 to use campaign specific values now.

6. Section 3.3.1 (Paragraph 2) Replace "For the GSFC product those errors are even significantly larger." with "For the GSFC product those errors are significantly larger."

Changed.

7. The conventions used are inconsistent (CryoSat vs. CryoSat-2 vs. Cryosat-2; SARIn mode vs. SARIn Mode vs. SARIN; 18.000 vs 18,000.

Changed.

8. There is no punctuation in the caption for Figure 1.

Changed

9. Units are inconsistent (sometimes m sometimes cm).

We are aware that this is not consistent, but e.g. in Fig. 7 we do this for good reason. As the campaign biases are very small, they are easier distinguished in cm. However, for better visual comparability of the histograms, Fig. 7b uses meters. We are confident that this does not lead to confusion as the units are stated clearly everywhere.