Response to Referee 2 on tc-2021-176

First, we would like to thank the referee for reviewing and commenting the manuscript. Please find the item-by-item reply below, with the original comments in *italics* and the responses in blue. The suggested changes will be implemented in the revised text.

The authors present a novel method of correcting slope-induced biases in radar satellite altimetry. A major challenge in assessing elevation and elevation changes of ice sheets. First, I have to applause the authors for revisiting this challenge, which has been a considerable error source in radar altimetry since the early work by (Brenner et al., 1983). Novel strategies for dealing with this issue are of interest to the radar community, but I will let it be up to the editor to decide if the topic is within the scope of "The Cryosphere".

I share many of the same general concerns as the first reviewer, and the following review will mainly supply additional comments. However, first I would like to highlight a couple of common issues also raised by the first reviewer. 1) Impact of radar penetration. Operating in the LRM area of Greenland, one would expect a considerable difference between the raw elevation measurements derived by leading-edge retracking at >10-20% and a validation dataset of real surface elevation observations. Hence, before venture into assessing the biases, this needs to be addressed. I cannot see any mentioning of surface penetration in the paper. 2) Performance of LEPTA relative to other approaches. The limited description of the implementation of both the LEPTA and reference methods leaves the readers with whether the observed differences are due to the method or implementation. 3) The figures are not of publication quality:

We thank the referee for the cronstructive review and suggestions. Based on some of the comments, we found that some important (technical) details regarding the methods and the setup of the experiments were missing. We will elaborate on this below and clarify that with relevant details in the revised manuscript. We regret that our explanations were not always clear, and will tackle that in the revised version with specific attention to the three points raised:

1) We will improve the discussion of the role of penetration in the revised manuscript (see also detailed response to Referee 1).
2) The description of the methods will be extended to better show that the results can be attributed to the methods and not their implementation (see also detailed response to Referee 1).
3) Figures will be updates based on the suggestions.

- *The figure is hard to follow from the caption. Besides the equations, the main text only offers the figure to be "methods are briefly illustrated". What is "briefly illustrated" I suggest adding an extensive description of the model flow, both in the main text and caption.*

We understand the concern of both referees. Figure 1 and the caption will be revised.

- *This figure could be one image, I guess h_ice should be h_ice2, what is d_min. Text inside the figure should be avoided as much as possible. What is the geographical location of the plot?*

Indeed, h_ice should have been h_ICE2. d_min is the distance to the nearest ICESat-2 point. We will change figure 2 in line with the suggestions of the referees.

- *This is a fine figure, however, see the following comments about grid sizes.*
- *Same as above*

Please see our later comment about grid size.

- *Could have been one figure, with a double y-axis. The curves show steps, which suggest the tested delta r values to be too coarsely spaced.*

Both panels will be integrated in one figure. Following the suggestion of Referee 1, new thresholds are added (0.5–2 m) which have changed the scale of the figure. We do not agree with the referee that the delta-r values are too coarsely spaced, as one cannot expect a continuous line when the number and locations of the points change for every choice of delta-r.

- *Work is needed to better resolve the signals in this illustration; the flat lines do not offer much information.*

We agree that the flat lines do not add a lot of information, but this is exactly what we would like to show: the slope method and point-based method are not affected by a bias in the DEM, whereas LEPTA is.

This leaves me with the following suggestions for improvements.

- *When assessing the performance of retrackers an informative measure is "slope vs. elevation bias". A better-performing retracker will have a flat response to an increase in the surface slope. Assessing this response would be beneficial for the paper.*

We agree with the 'better-performing retracker will have a flatter response' statement and this is actually also what we observe in Figure 4 with an increase of median absolute deviation from the interior of the ice sheets towards the edges, where the slopes generally are steeper. This increase towards the edges is especially strong for the ESA L2I and slope method and to a lesser extent for the point-based and LEPTA methods. In this comparison, the LEPTA method again performs best. This as such shows that LEPTA has flatter response vs. slope. We prefer to keep the analysis as it is (i.e., based on the maps), while simultaneously adding an additional description of better performance vs. slope based on the maps instead of adding an additional figure. The reason for this is that computing the plot asked for by the referee leaves with the ambiguity on how to define the surface slope (i.e., over which resolution it should be computed) and in which point (i.e., the nadir point or the impact point).

- *The differences between the different methods should be judged in terms of statistical significance.*

In general, we agree with this comment. Doing so to assess the significance of the differences in the median requires, however, the differences to be normally distributed. From the top panels in figures 3 and 4 we can see that this is not the case. Hence, doing a t-test does not make sense. Therefore, we prefer to show the actual distributions so the reader can see the differences instead of relying on a t-test that violates the normal distribution assumption.

- *How does the gridding of 50x50 km tiles influence the results? Why is the point-based method the only method missing data in the trunk of Jakobshavn isbrae? Why is there no data for all the methods east of the line from 79fjord to Helheim of figure 3? The data coverage seems different in figure 4.*

The difference in data coverage between figures 3 and 4 originates from a minor issue in the plotting which will be solved in the revised manuscript. The fact that we lack data in the trunk of the Jakobshavn isbrae is caused by the fact that outliers are removed (please see attached figure).

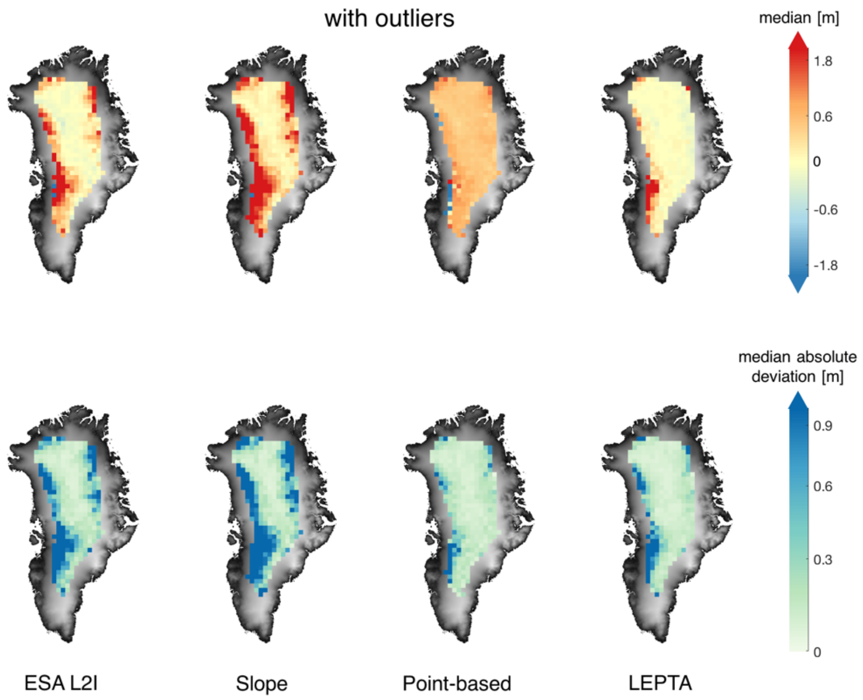

Figure 1. Median (upper panels) and median absolute deviation (lower panels) with all points (both with and without outliers) included.

Regarding our choice to use 50x50 km, this was a compromise between visualization and robustness of the statistics. Switching to 25x25 km tiles (please see map below) results in more empty tiles (because we lack data points) and enhanced pepper and salt effect, but the main conclusions remain the same.

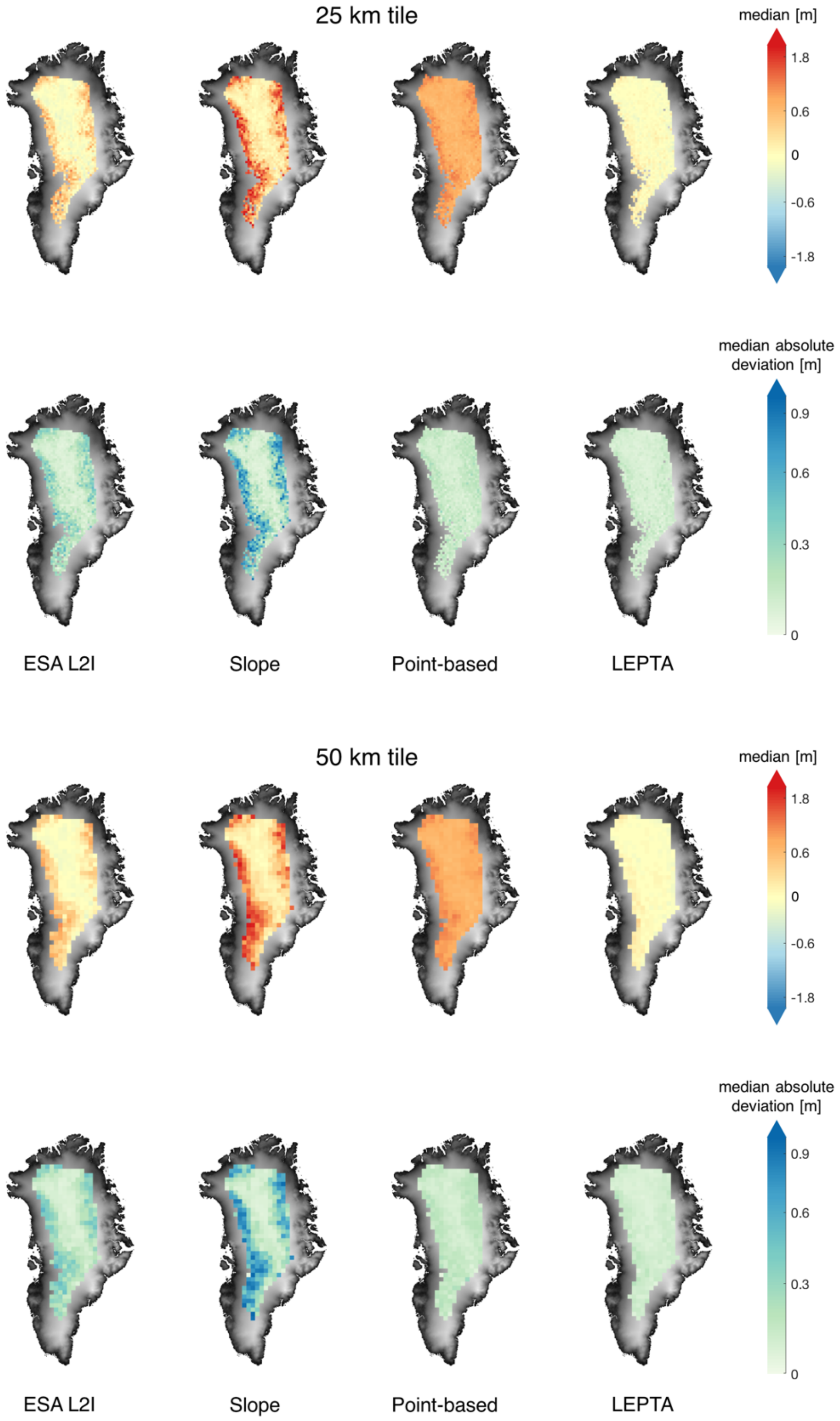

Figure 2. Comparison between using a 25 km tile (upper figure) and 50 km tile (lower figure).

- *Table 1, for the discussion it would be informative to also have the arcticDEM vs. ICESat statistics.*

We agree this can be informative. We can't however add this to Table 1 as it would become too large. We will add a table to the Appendix.

- *Table 2, The surface penetration biases may relate to the retracking threshold chosen. How is the statistics changing between valid choices? (10%-90%)*

Agreed. Note, this was our main motivation for the sensitivity analysis presented in section 4.6. In the revised manuscript, we will use a broader range of thresholds and discuss the issue in more detail.

Minor comments:
L2: What is "assessing snow/ice anomalies"
This statement refers to the 2012 melt observed by Nilsson et al. (2015). It will be rephrased in the revised manuscript.

L14: Is the difference between the 1cm and 0cm bias between LEPTA and ESA significant and therefore needs to be differentiated from the performances of LEPTA. I think that it is the standard deviation that is the important measure.
This sentence will be reformulated. In our view, the results show that the median differences need a careful interpretation. Indeed, penetration plays a role. At the same time, the results suggest that part of the observed bias is due to the applied slope correction method.

L16: Reformulated: "we recommend the LEPTA method for obtaining... "
This will be changed in the revised manuscript.

L17: What is complex topography? The work is done in the LRM area.
We will rephrase 'especially in regions with complex topography.' to 'especially towards the margins of the LRM area where the surface slopes increase.'

L23: Concerning elevation change, you could add a reference to (Hurkmans, Bamber, and Griggs, 2012)
This will be added.

L51-52: Suggestion to move this to the last part of the introduction.
Agreed.

L79: Why not use the official releases of the downsampled ArcticDEM?
Because in this way we can define the resolution ourselves.

L84: Is it ATL03 or ATL06 being used? From the link it seems to be ALT06, is there a bias of using the downsampled product?
It is indeed ATL06, which has a known geolocation accuracy/bias of less than 10 m (https://nsidc.org/sites/nsidc.org/files/technical-references/ICESat2_ATL06_Known_Issues_v005.pdf, last access Dec. 3, 2021). In the revised manuscript, we will add these details.

L89: This should be moved to the acknowledgment.
This will be changed in the revised manuscript.

L92: This sentence needs to be elaborated.
Agreed. This will be changed in the revised manuscript.

L134: 8x8 km seems small. When looking at the SARIn retracked data from ESA, relocation distances of up to 12 km from the nadir point can easily be found.
Agreed. In line with the point-based method, we will enlarge it to approximately 14x14 km.

L157: The bias can be evaluated in monthly intervals, but at some of 50x50 km tiles closer to the coast a seasonal difference in the bias is expected. How is this seen in your data?
This analysis is part of our ongoing work for future publication where we focus on spatio-temporal variability in penetration depth. We agree that it is interesting, but we consider it outside the scope of this manuscript.

L159: Why use both the nearest and natural neighbor interpolations? You give some reasoning. However, would the two algorithms not converge in your case, and thereby there is no need for adding a user-defined threshold?
We prefer to use a natural neighbour interpolation over a nearest neighbour one. Indeed, doing so allows to compensate for the difference induced by the surface slope as the points never overlap exactly. As pointed out in the manuscript, this requires, however, that the CyroSat-2 data point is surrounded by ICESat-2 points. If this is not the case, we rely on a nearest neighbour interpolation (regarding the use of nearest neighbour, we would like to point to our response to Referee 1). The user-defined threshold the referee points to is required to avoid interpolation over large distances. We will clarify this in the revised manuscript.

L197-198: I guess the eastside is a result of topography? Could you give some insights into the differences on the east and west-side which will be the reasoning for this reported difference.
The highter median values on the eastern side of the ice sheet are indeed likely due to topography. We will study this in some more detail and include our analysis in the revised manuscript.

L219: Having a setup at 50km tiles it would be rather easy to take the time-tagged ArcticDEM tiles into the analyses. This might be a large job to undertake this effort, but one or two tiles would be very informative for the analysis.
We agree that it might be interesting to compare our results with the time-tagged ArcticDEM tiles, but that is indeed an enormous job and difficult as the tiles are not referenced and still include potential offsets and tilts (https://www.pgc.umn.edu/data/arcticdem/, last access Nov. 23, 2021). There is meta-data available to correct them based on ICESat, but then we de-facto replace the time-tag with the ICESat time-tag. Therefore, we think such a comparison with ArcticDEM tiles does not make sense (without enormous pre-processing of the tiles) and we do consider it an unrealistic experiment.

L238: The observed change in bias is an important observation, please elaborate on this.
See also our response to Referee 1. We will elaborate on this in the revised manuscript.

L244: Please clarify the statement: "relative sensitive". Relative is a difficult word as it might be different for you and me.
Agreed. This will be changed to quantative descriptions.

L247: "although not directly visible" please improve the figure.
The figure will be improved.

L284: Any insights into why ESA outperforms the other methods?
Unfortunately, we cannot give a conslusive response.

L290: Please elaborate on this last statement.
This statement was not supposed to be in the manuscript and will be removed in the revised version.