The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC1
https://doi.org/10.5194/tc-2021-176-RC1, 2021

Review of 'A leading-edge based method for correction of slope-induced errors in ice-sheet heights derived from radar altimetry', by W. Li et al.

Anonymous Referee #1

Referee comment on "A leading-edge-based method for correction of slope-induced errors in ice-sheet heights derived from radar altimetry" by Weiran Li et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2021-176-RC1, 2021

This manuscript describes a new approach for relocating radar altimetry measurements acquired over ice sheets; one of the most important processing steps for retrieving reliable surface elevation measurements. The authors outline the method, together with a proof-of-concept study whereby the approach is applied to one year's worth of CryoSat-2 LRM measurements over the interior of Greenland. They perform validation relative to ICESat-2 measurements and an independent DEM, alongside a sensitivity analysis to explore some of the inherent assumptions within their approach.

I found the manuscript very interesting; the proposed methodology is novel and definitely has the potential to improve upon current approaches documented in the scientific literature and implemented within ESA's ground segment. I therefore believe that it will be of interest to the subsection of The Cryosphere's readership that have an interest in radar altimetry processing techniques over ice sheets, ice caps and glacier surfaces. That being said, I believe that there is still some additional work required to (1) convincingly demonstrate the superior performance of the method relative to existing approaches, and (2) to provide the necessary level of methodological detail required to adequately document this promising new method. Without this, I am left feeling that I have a glimpse of an exciting new approach, but have many unanswered questions that prevent me from being fully convinced that it delivers the improvements that the authors claim. I hope that, by addressing these points, the authors will be able to provide a more compelling demonstration for The Cryosphere's readership. I have detailed these major comments below and would like to see each of them addressed in the revisions. Following these comments I have also listed a number of more minor points, which I hope will help to improve the clarity of the manuscript. Finally, I would recommend that the manuscript undergoes a thorough check for grammatical errors, as there were a considerable number throughout.

Major comments

- **Performance of LEPTA relative to other approaches.**

The authors compare LEPTA to the ESA L2I product, and their own in house versions of the slope correction method and the Roemer et al. (2007) relocation method. Whilst the statistics show the superior performance of LEPTA, I am left with several important questions relating to the implementation of other approaches, which make it difficult to determine whether they have been implemented optimally; i.e. whether a better implementation could have yielded improved results more closely matching the performance of LEPTA. Specific points that I would like to see addressed are as follows:

- For ESA L2I – have any of the quality flags included within the product been applied? More L2I data are available than for the in house methods and this makes me wonder whether stricter quality control has been applied in the latter, e.g. the waveform filtering mentioned on line 289. In other words, that some of the improvement of LEPTA relative to L2I is not due to the method used slope correction, but simply down to the quality control applied.
- For the authors' in house 'slope correction' method – the results, e.g. as shown in Fig 4, indicate far worse performance than the ESA L2I implementation, and make me concerned that their slope correction method has been implemented sub-optimally. This, combined with point 1 above, means that I do not think that a convincing case has been made to justify the level of improved performance of LEPTA relative to the slope correction approach. This is not to say that LEPTA is not an improvement, but just that I feel that more work is required to justify this convincingly. Specifically, if the authors really believe that the difference between L2I and their in-house implementation relates to the Doppler slope correction, then I would like to see further analysis to demonstrate (1) that this really is the case (i.e. that the Doppler slope correction can be responsible for a difference of this magnitude), and (2) why it does not affect LEPTA in the same way (and should not be incorporated into the LETPA L2 processing). I would also like the authors to state the DEM resolution used for the slope correction (I couldn't seem to find it anywhere), and if it is 900 m or less, to justify why

this is an appropriate choice. From my perspective, the 'resolution' should be comparable to the beam limited footprint (i.e. 10's of km), not the pulse limited footprint, because it is preferable to relocate using the large scale slope across the illuminated area. If you use the '900 m' slope at nadir, then there is the risk that the slope you use will not be representative of the average slope across the illuminated area. Indeed I think you could be seeing this effect in Figure 7, where performance improves up to a resolution of 900 m, and raises the question as to whether you would see further improvements if the resolution was increased any more. As such, I would like the authors to either provide a justification to counter the above concerns, or to test this by computing the slope over a larger length scale (comparable to the beam limited footprint) and re-evaluating the performance of their slope-based method.

- For the authors' point-based approach, I find the magnitude of the bias surprising, e.g. as shown in Figure 4, and that there is a general lack of detail or discussion required to assess whether this is due to the implementation of the approach. In particular, I cannot find any information relating to the search area that the authors have used; i.e. the illuminated area on the ground where they assume the leading edge reflection could have come from. It would be reasonable to base this upon the 3 dB beamwidth of the instrument, but it is not clear to me what the authors have used. As such, my concern is that an inappropriate choice could lead to a bias in the 'point-based' solution; for example if the criteria used is too strict, and does not allow for the POCA to be sufficiently far away from nadir. I would therefore like to see the authors (1) state what criteria is used, (2) justify why it is appropriate and not impacting the accuracy of the results, and (3) dependent upon these points, consider whether the performance of their point-based approach should be re-evaluated with a refinement to the allowed relocation distance.

- **Choice of delta-r.**

The choice of delta-r seems rather arbitrary, yet central to the LEPTA approach, and so I would like to see some more discussion relating to this point within the manuscript:

- From a theoretical perspective, clearly it would make sense to let delta-r vary according to the width of the leading edge of each waveform. I assume the authors have practical considerations for why they chose not to implement this approach, and I think it would be helpful for readers if they could therefore expand on this within the manuscript, to explain why such an approach was not selected.

- I appreciate this is extra work, and therefore I would not insist upon it, but given the central role that the leading edge plays in the LEPTA approach, I think it would be really valuable for the authors to provide some quantitative measures relating to the characteristics of the CryoSat-2 LRM leading edge over Greenland. For example, can you provide statistics relating to the mean and standard deviation of the range spanned by the leading edge? This would provide really helpful context for judging the validity of the range of delta-r considered.
- Without point 2 being addressed, it's not clear to me why delta-r of 2 metres is a reasonable lower bound. I would therefore like to see the sensitivity analysis expanded below 2 metres, or a justification for why this is not appropriate; as, in theory, choosing a lower threshold would seem a sensible approach to ensuring that you always identify terrain corresponding to the leading edge.
- I also suspect that the optimal choice of delta-r might vary significantly spatially; yet this is impossible to assess based upon the median statistics presented. For example, that a delta-r of 2 m or lower might perform much better over simple topography. Given the central role of delta r in terms of the LEPTA approach, I think it would be interesting to produce spatial maps of the type shown in Figure 4 for a LEPTA-delta-r of 1 m and 2 m, to see the extent to which this can improve upon the 3.5 m case already plotted.

- **Impact of penetration**

Throughout the manuscript, the issue of penetration into the snowpack is never mentioned. I do not think it requires further analysis, but I do think it would be helpful to include some discussion related to this phenomenon, and whether or not it has any implications for the LEPTA method; given that LEPTA uses range information from the leading edge, and the leading edge of LRM measurements can be modified by subsurface scattering.

Manuscript minor comments

Line 2: anomalies in what – mass change, physical properties?

Line 4-5: Perhaps I'm misunderstanding, but I think the 'slope' method and 'point-based' (I assume Roemer?) are correcting for undulating topography within the *beam-limited* footprint rather than the pulse limited footprint?

Line 9: Begin -> beginning.

Line 13: 'slope corrected' – I assume this relates to those using LEPTA? This should be made clear.

Line 13: 'Almost negligible' is rather vague – please be quantitative, especially as you quote statistics for the other methods and ESA L2. E.g. is it better than ESA L2 at 0.01 m difference?

Line 14: Which methods exactly do 0.22 m and 0.69 m refer to? Currently it is not clear. Same applies on line 16.

Line 14: Median absolute deviation from what?

Line 31: surface *slope* parameters.

Line 31: Beam limited footprint?

Line 32: ‘full height information’ is not particularly clear for readers not familiar with the subject – perhaps something like ‘uses a topographic model…’ would be clearer?

Line 35: Not clear whether you are referring to pulse limited or beam limited footprint. As a more general point, I would recommend that you make sure that through-out the manuscript that it is unambiguous which you are referring to.

Line 54: Please state which CS2 product baseline was used.

Line 56: Is the data used inclusive of these end months? Important for future reproducibility.

Line 66: use -> used.

Line 66: What is 25% more realistic than? Do you have any supporting evidence for this statement?

Line 68: Please explain what you mean by ‘a distinguishable noise’ and what criteria exactly were used to identify the waveforms that failed this and the ‘beginning of leading edge’ tests; i.e. so that the reader has sufficient information to be able to reproduce your method, should they wish.

Line 70: ‘as *a* benchmark’.

Line 71: ‘which has a resolution’?

Line 84: Please also mention which ATL- product was used.

Figure 1. I find this figure pretty hard to interpret and I think it would benefit from some more attention:

- Why is the low resolution DEM only given in the slope method panel?
- I don't think 'apply the satellite-terrain range' really makes sense.
- I think 'block mean averaged' could do with more explanation in the caption – presumably you mean the average range over either a square, rectangular or circular search window? What is the radius used in these graphs, or is it just a cartoon drawing to illustrate the concept?
- Might be also worth annotating with the true POCA as well?

Line 98: 'is the central angle between the satellite and Ps'. I don't think this is very clear. I'm not sure what the 'central angle' means, and also that it is correct. Doesn't it depend upon the instrument boresight, which might not necessarily be pointing at nadir?

Figure 2: This figure feels somewhat rough and not ready for publication:

- Where is this data from? A location map would be helpful.
- Why does the header say natural neighbour but the caption say nearest neighbour?
- What is h_ICE? ICESat-2 elevation? If so, how should the statistics be interpreted given that point-based method is much further from the IS-2 track than LEPTA? For example, has a correction been applied to account for the effect of surface slope between the CS2 and IS2 locations?
- What is d_min?
- How were the ICESat-2 tracks that are plotted selected?
- Visually, I think it would be easier for the reader to interpret if the DEM was displayed

as a contour map; but this is only a recommendation, not essential.

Line 117: It’s not clear to me why you are dividing by R? Also equation is not numbered.

Line 123: I think it would be helpful to expand upon this final sentence slightly, as I think it is important to convey this point, as it’s your main argument relating to the limitation of Roemer. It is not clear what ‘It’ refers to in ‘It also shows’. For example, I don’t think that Roemer uses DEM points other than the POCA within equation 4, rather it is only in identifying the location . This distinction is not clearly articulated with the current wording.

Line 134: Please provide justification for why 8 x 8 km is chosen as the search radius for the intersection points. It seems quite an arbitrary choice, with no justification given. For example, why not use something closer to the 3dB beamwidth, which would seem to have a much better physical justification? Otherwise, how can you be confident that you are not incorrectly locating measurements where POCA is greater than 4x4 km from nadir but still within the 3dB beamwidth, and therefore sensitive to the antenna gain pattern? At the very least, I would like to know how many measurements fail to identify DEM points within the search window?

Line 136: ‘In case no DEM grid points are identified…’. Please provide a clearer explanation of what you are doing here, as it seems important but I cannot understand exactly what you are doing here. Is the interval expanded? Or is it shifted? Is this the same as finding the DEM points that are *closest* to the retracked range, even if they are not within the search interval? If this is the case then I would like to see some more analysis to support this approach; e.g. are these points commonly at the edge of the 8 x 8 search window? Is there a systematic bias in terms of whether the retracked range is normally higher or lower than the DEM range? I think this is required because it seems like this is somewhat at odds with the central tenet of your method which is to only use points within the leading edge interval, so it’s not clear to me why this is justified. It relates to the previous point too – in that the underlying issue might be that in these cases POCA lies beyond the 8 x 8 km search window – and it isn’t clear to me that what you are doing here is an appropriate way to correct for this issue.

Line 137: Do you mean here that P(x,y) is computed as the average of the x and y coordinate values? If so, is this the mean, median or mode? Using this approach, I guess you could get a P(x,y) that is located outside of the LEPTA search area? Can you comment on this; e.g. how often it occurs and what the implications are?

Line 139: Should there be a 1/K averaging in equation 5?

Line 141: Slight aside and not essential, but do you have any statistics relating to the size of the LEPTA footprint – i.e. the intersect between the leading edge and DEM – it would be really interesting to see how much the reality diverges from the classical footprint size over a flat smooth surface.

Line 142: ‘but *are* outside’.

Line 151: Please explain what a ‘conceptual assessment’ actually means.

Line 153: It’s not clear to me how meaningful the median statistic is, given the effective timestamp of the ArcticDEM. I.e. isn’t ArcticDEM referenced to ICESat, and in which case surely you need to account for the intervening elevation change of the surface?

Line 160: In the case of nearest neighbour, is a correction applied to account for the effect of surface slope between the CS2 and IS2 locations? If not, why not and what are the implications? Given that ArcticDEM is already integrated into your processing flows, I assume it would be pretty simple to do this.

Line 167: 'h_DEM / h_ICE2' – replace '/' with 'or' to avoid any ambiguity with a division operator.

Line 169: Would it not make sense to also consider sensitivity to how the start of the leading edge is defined? Surely this is relevant too?

Line 186: 'best' relative to what – I assume you mean of all methods, but it could be construed as ArcticDEM vs IS-2, so worth making clear.

Table 1:

- 'Before' and 'after' are not used in the table, so I would recommend not needing to refer to them in the caption.
- Please state whether differences are calculated as CS2 – ref, or ref – CS2.
- I'm not sure how useful it is to list all the percentiles in a table. Have you considered showing these as a cumulative distribution figure instead? I think this would be much easier for the reader to interpret.

Line 203: Suggest ‘most ideal’ -> ‘performs the best’.

Fig 3: Comparing the LEPTA and L2I pdf’s it looks like the main benefit form LEPTA is to reduce positive rather than negative differences. Any thoughts on why this might be? Could the lack of impact on the negative differences be due to the relatively large delta-r leading to DEM elevations beyond the leading edge being included – i.e. a smaller delta-r might deliver improvements here as well? I guess it would be fairly clear by looking at the full pdf in the sensitivity analysis, rather than just the central value?

Line 208: I would recommend using ‘positive’ and ‘negative’ elevation differences, rather than ‘right’ and ‘left’ side of the median.

Figure 4: It seems that LEPTA is much more clearly the best performer when compared to IS-2, rather than ArcticDEM in Figure 3. It is not clear to me why this is the case – is it linked to spatial coverage, differences in the timestamp of ArcticDEM relative to IS-2, or something else? I think it would be helpful for the authors to expand upon this here.

Line 218: Also covered in previous points. I’m not sure it is the timestamp of the optical images that is important – isn’t it the data used to provide the absolute reference?

Fig’s 5-7: Captions should state what it is the median and median absolute deviation of – i.e,. CS2-minus-IS2.

Line 229: More detail needed – is this for the full dataset or a subset? Is this with outliers removed?

Line 230: It's not very clear to me how this choice of 2-5 m actually relates to the properties of the leading edge. I think it would help to justify this choice in the minds of the readers, if the authors could describe the typical width of the leading edge, and show that delta-r is a sensible choice within this context. For example, with the current analysis as it is presented, I am left wondering how common it is for the leading edge to be less the 20% OCOG + 2 m; i.e. to lie outside of the range tested. From a theoretical expensive I could see that a delta-r value of 0.5-1 m could make sense, but there is no analysis to explain why this parameter range was not explored; nor indeed why the actual range of ranges spanned by the leading edge of each waveform was used. Did the authors evaluate what happened when delta-r < 2 metres?

Line 244: I'm interested in why the sensitivity to a bias in the DEM is not symmetrical about zero. Can the authors expand upon this point; i.e. why having a biased-low DEM has very little effect, but biased-high does? Is this somehow connected to a generous choice of delta-r, i.e. that at 3.5 metres, it is actually including a significant buffer beyond the leading edge, such that when you bias the DEM low the true POCA still remains within the delta-r range? I think a slightly more in-depth evaluation and discussion for the observed behaviour would be useful here in terms of understanding the method, rather than a simple 1 paragraph summary of the sensitivity results with minimal interpretation.

Line 254: Again, I think the manuscript would benefit from critical interpretation here, rather than simply reporting the bare results. For example, can the authors expand on why the point based approach degrades so quickly with increasing resolution – is it due to topographic peaks being smoothed? Wouldn't you expect the point-based approach to tend towards the slope based approach; i.e. with sufficient smoothing then you remove all high frequency topography and are just left with the long wavelength slope?

Line 259: ‘reduces with 30 cm’ – doesn’t make sense.

Line 261: ‘are slightly more off’ – please rephrase this more precisely.

Line 261: ‘choice of retracker *threshold*’ – I don’t think you have compared different retrackers?

Table 2: Does this suggest that LEPTA is more sensitive than the other methods to choice of threshold, for the median absolute deviation; i.e. when 50% is chosen then its performance is comparable to the point based approach? Do you think this ties into delta-r; i.e. if you choose a higher threshold then you are including more terrain at large ranges beyond the leading edge, which might degrade the LEPTA solution in a way that doesn’t happen for the point based approach?

Line 264: Most of this opening paragraph seems simply to be repeating and summarising the results; i.e. not adding new insight, as I would expect in a discussion section.

Line 266: I don’t think the wording ‘in terms of spatial patterns, the LEPTA method outperforms’ is very clear; i.e. what it means for one pattern to outperform another. Please consider rewording.

Line 268: CryoSat-2 *LRM* height estimations'.

Line 268: 'Our results show moreover that the method is not very sensitive to changes in the definition of the end of the leading edge as it shows only millimetre-level uncertainties for the corrected heights when including multi-metre uncertainties on the definition of the end of the leading edge.' I don't think this statement is accurate – If I understand correctly, then I think you see mm-level differences on the median bias averaged over the whole LRM zone in 2019; but this is very different to saying that the height measurements themselves will only vary by mm when you change delta-r by metres – I'm not sure that you have demonstrated this?

Line 271: "the definition of the leading edge should be adjusted accordingly" – please clarify what you mean here. How do you adjust the definition of the leading edge, and what does it mean to be 'adjusted accordingly'?

Line 273: Please explain why you recommend a high resolution DEM when it has little impact on the method. Also what counts as 'high resolution'?

Line 276: 'the importance of *accurately determining the* impact point'?

Line 277: Do you show that your point based approach outperforms the slope based method? There seems to be a pretty large bias in the former, compared to the latter?

Line 286: I am unconvinced by the attribution of differences to the Doppler slope correction – see previous major comment. Please provide more justification that this could be the source of such large differences and, if indeed this is the case, why (1) it does not affect LEPTA, and (2) you don't correct for it in your L2 processing.

Line 289: I think more explanation of how you filter waveforms is required, and how you can be sure that some of the performance improvement you see in LEPTA relative to ESA L2I is not simply due to the fact that you are applying stricter filtering criteria.

Line 290: I'm not sure I understand this – doesn't it contradict line 136 where you say you adjust the interval if no DEM points are within the leading edge range?

Line 293: I don't think the opening sentence makes sense – applying to what? Also needs to make clear this is for non-interferometric data only.

Line 299: Recommend that you do not need to start a new paragraph here.

Line 301: 'by the begin and end of the leading edge' – I don't think this is strictly correct; i.e. you do not limit yourself to the actual end of the leading edge.

Line 304: ‘almost identical’ and ‘good improvement’ – it would be much more helpful to provide quantitative measures here.

Line 305: You don’t evaluate LEPTA at the margins, as your analysis is restricted to the LRM zone only.

Line 306: ‘radar altimetry’.

Line 306: I think somewhere you should flag that you have only assessed performance (1) in the interior, and (2) for Greenland. Therefore it still remains to be shown how the method performs over the more complex ice margin terrain and also over Antarctica. Obviously your thoughts on whether you expect comparable, better or worse performance would be of interest to the reader, and this might fit better within your discussion section.