Reply to Referee Comments

First of all, we want to thank the referee for the critical and constructive comments on our manuscript. We considered all comments. Our replies/actions are indented and given in blue font.

Summary

The manuscript presents estimates for the ice thickness distribution for the glaciers on Mount Kilimanjaro. The estimates refer to the years 2000 and 2011, and are based on a combination of in-situ observations, past ice thickness reconstructions derived for areas that are now ice free and a numerical, ice-flux based approach. The paper seems to have two main points. For one, the available global-scale ice thickness estimates seem to have overestimated the ice thickness for one of the two investigated glaciers. For another, the idea of using a combination of past and present digital elevation models (DEMs) to derive ice thickness observations that can be passed to ice-flux estimation approaches is suggested to hold promise for future applications. The paper has a general good quality, and the findings are certainly worth conveying to the larger audience. Slight improvements seem necessary in the way that individual details are presented. The discussion section could benefit from a somewhat more substantial revision.

Major Comments

- Somehow, I was left in doubt on how the available Ground Penetrating Radar (GPR) measurement enter the game. They are briefly mentioned in the Data section (L. 48), do not show up in the Methods, and re-appear in the Discussion (L. 145). In particular, clarification is required for what the mentioned "assimilation" (L. 33 and 145) actually entails. As the manuscript stands now, no information is provided, and that should be rectified.

    We agree that the article format required us to shorten many technical details. The old document did however specify that viscosity values are computed at the location where thickness values are available (L86-87). In section 3.5, we now further expanded on the details of how GPR measurements are used. We hope that these extra sentences provide the necessary clarification.

- I had some reasonably hard time in following the discussion. I found it particularly hard to keep track of the many comparisons done for the two glaciers, targeting at the three Experiments performed in the work itself, the two (or three?) models used in the consensus estimate, and the two available sources of in-situ observations (boreholes and GPR data). To me, it would seem natural to show a figure depicting the various model results along the available GPR transects. Since both surface DEM and thickness are available for any of the various results, all information required to generate such a plot seems available. Most likely, this would help the readers to better grasp the main outcome of the discussion which, as far as I understand, rather focuses on the performance of the consensus estimate than on the results of the manuscript itself?

    The consensus estimate shows a mean of three (Kersten Glacier) and two (Northern Icefield) separate models. We have decided to omit discussing the models in the consensus estimate separately to avoid misunderstandings.
    As you mentioned in the annotations directly in the manuscript, Farinotti et al.'s consensus estimate has a twice as high ice thickness for Kersten Glacier, while it underestimates the thickness at the boreholes, which are located on the Northern Icefield.
    We reworded parts of the discussion to make it easier to understand by clarifying whether the discussion is about the Northern Icefield or Kersten Glacier.
    We chose to not change the figure as showing the thickness along the GPR transects would not adequately depict the ice thickness distribution across the whole Northern Icefield. Thickness

surveys were only available for NIF. These were directly assimilated by our method and are reproduced. The only thing such a profile graph would show that other approaches deviate. We therefore decided to extract the thickness values at the unconsidered ice core locations and added them into Supplementary Table 1.

- An important point of discussion that seems to have been missed is that ice thickness estimation approaches as used in this study require the investigated glaciers to have some ice flux. Otherwise, the main idea behind the approaches somewhat breaks down. This point is skimmed in the Conclusions & Outlook section (L. 183) but would probably deserve some space in the Discussion section as well. May it help to explain some of the discrepancies noted between model results and observations?

As Kersten Glacier is located on the steep flank of Mt. Kilimanjaro, we expect some glacier deformation with a clear directional preference even if rates remain small. For NIF, we agree with the reviewer that the situation is more complex. Its central areas are characterized by flat plateaus and the abrupt step changes in the topography over the cliff features. As we suspect little deformation, we can only alleviate this concern by pointing to the error assessment in Fürst et al. (2017). The approach has there been applied to an ice-cap geometry on Svalbard. There it is shown that error estimates associated to the thickness reconstruction increase substantially towards the flat interior where no thickness measurements are available. The reason is that the associated error estimates are inversely proportionate to the ice flux. For NIF, we are however in the favorable position that thickness values were measured over the flat plateau area giving some confidence in the results. We inserted a brief discussion of this issue into the discussion section.

- The last few sentences of the Conclusions & Outlook (L. 190-198) seem the paper's strongest and most valuable point. Shouldn't these implications be highlighted in the abstract as well?

Due to the abstract being limited to a maximum of 100 words, we were unable to highlight it in the abstract as well.

Minor Comments

1) There are several undefined acronyms, including, amongst other, SRTM at L. 35, MB at L. 61, TDX at L. 66.

We added the definitions for the previously undefined acronyms.

2) I could not follow the logics exposed at L. 61-63. According to the sentence, the surface mass balance model applied in the study was "slightly altered" because (sic) "it was never tested for Kersten Glacier before". I imagine that the model was actually tested by the authors before altering it, and that the matter is only one of wording?

The surface mass balance model has previously only been testes on Kersten Glacier. After applying the model with the exact same parameters and settings on the Northern Icefield, we found that it could not reproduce the observed surface height changes measured by the Sonic Ranger mounted to the Automatic Weather Station on the flat parts of NIF. We did test different ways within the scope of the model that would influence the model output to better fit the measurements. We believe that in this case it is a matter of the wording used in the manuscript and we adjusted it to reflect that.

3) At L. 69-72 the authors state that they removed all positive elevation differences from the analysis because such positive changes are "unlikely" to happen. The issue is that this removal apparently affects

some 15% of the area of the Northern Icefield, which calls for some more detail. For example: What is the spatial distribution of these removed cells? Is it completely scattered, suggesting random noise, or is it clustered, indicating that the signal might be real after all? What is the confidence in the individual DEMs? Etc.

> The referee rightly asks for more clarification here. In this section, we failed to clarify that this selection only concerns the DHDT values that are later used to determine past thickness observations in the nowadays ice-free areas. We adjusted our explanation accordingly. Positive DHDT values cannot be considered in the reconstruction because they imply that the formerly ice-covered area had a lower elevation than the nowadays ice-free part. As we aim for distilling useful information from the retreat these values could only be ignored.

4) I was not able to follow L. 90-94. A “coupling length parameter” is introduced without further explanation (I assume the definition is found in Fuerst et al. 2017, which is ok) and, as far as I understand the wording, is first said to control how the surface DEM is “imprinted in the thickness field” (I’m not entirely sure what this means) and later said to control the “smoothness” on not further specified “flux streamlines”. I don’t want to exclude that the wording makes perfect sense to a reader familiar with the details of Fuerst et al (2017) but I think that some additional words of explanation will help the majority of the readership.

> The coupling length parameter is introduced in Fürst et al. 2017 and controls the horizontal smoothing of the surface slope field with the aim to infer smooth streamlines for the flux computations. We reworded the sentence for clarity.

Line-by-line Comments

A (rather long, I apologise) set of line-by-line comments is found in the annotated document, attached to this review. The comments provided above are contained therein as well.

> In our response below, we only address line-by-line comments that were not addressed above and that do not refer to style, punctuation, grammar, etc.

L. 10: Please state at least a standard deviation.

> We have calculated a mean relative (absolute) error of 26% for the reconstructions at the borehole locations. The value is not small as it exceeds error estimates for the majority of glaciers on Svalbard (Fürst et al., 2017). This value can only be a rough orientation for the uncertainties associated with our reconstruction and we therefore refrain from stating it in the abstract. Yet we included it in the results and the conclusions.

L.11: how is it for NIF?

> We added details for NIF.

L. 27: If the "results of this study" are mentioned, shouldn't they be introduced first? At this stage of the text, the reader doesn't really know yet what the study will be about.

> We reworded the sentence and removed “results of this study”.

L. 31: From the context, this "there" seems to refer to the dataset of Farinotti et al., not to KG glacier itself (which is what the sentence seems to say). Possibly reword slightly?

> We reworded accordingly.

L. 32: I'm not sure to understand the meaning of "for the first time". The sentence seems to say that the approach existed before but that no thickness measurements were assimilated so far. However, this is probably not how the sentence was meant?

> For the first time referred to the reconstruction approach being used on Mt. Kilimanjaro for the first time.
> We reworded accordingly.

L. 34: What is the meaning of "thickness input" here?

> Thickness input refers to different data sets of ice thickness observations used as input for the reconstruction approach.
> We reworded accordingly.

L. 35: The wording is slightly confusing: it seems to imply that "satellite information" does not qualify as "observational data". Does the "observational data" only refers to "ground-based observational data" then? And why are the thickness observations called "ground truth" in the abstract then?

> “Observational data” refers to measured ice thickness data, including radar measurements (such as the Bohleber et al. GPR data) as well as the ice core measurements (Thompson et al.), but these observational data sets are not available for Kersten Glacier.
> We reworded accordingly.

L. 35: I'm not sure, which one were the first and the second? Is the first one the one introduced with L.32, or are both first and second referring to what follows in L.32-33?

> We reworded the passage for clarity.

L. 35: Consider providing the resolution explicitly. What is "very high"? 10m, 1m, 10cm?

> Very high resolution in this case means 0.5 m ground resolution. We added the information into the text.

L. 39: I'm not following: Which is "THE distributed SMB model"? There was no SMB model mentioned so far, was there?

> We reworded for clarity and added a cross-reference to Section 3.1 in which the SMB model is described.

L. 43: "from a merge" or "by differencing"? I imagine the latter? Otherwise I'm not sure to understand what is happening.

> Two separate TanDEM-X scenes were merged and then by differencing them from the SRMT DEM, the surface height change was generated.
> Reworded for clarity.

L.46: Please point at a figure where this can be seen. As now, the sentence is pretty abstract.

> Added reference to the corresponding figure (Fig. 1).

L.50: linearly interpolating (I imagine?)

> Adjusted phrase accordingly.

L. 54: I'm not sure: "found" by whom? By the Bohleber et al. study? Or by the Farinotti et al. one?

The consensus estimate provided a similar value.
Rephrased the passage accordingly.

L. 63: Can a rational be given for this slope angle threshold? Is the idea that for steep slopes, the meltwater runs away and therefore does not refreezes in place? I'm not entirely sure I would agree with that.

Yes, the reviewer is correct about the basic idea, but not about the fact that meltwater on steep slopes cannot refreeze in the model. The value 5° is an effective compromise to prevent runoff from the almost horizontal surfaces of the Northern Icefield, since there are virtually no surfaces in this portion of the glacier that would be steeper than 5°. Meltwater, however, can still refreeze in the model on steeper surfaces, which is described in one of the model reference papers (Mölg et al., 2009). However, note that the modification only applies to bare ice (the standard code deals with refreezing only in presence of a snow pack).
We added a sentence to clarify the 5° threshold.

L. 73: I don't understand the meaning of "margin" here. A little wordy, but may "Past ice thickness for areas that have become ice free" be an alternative?

We decided to stick with "margin" as this phrase is used throughout the manuscript and is defined in Section 3.3.

L. 85f: a) Please don't mix the notation $^{-1}$ and $/$. b) I believe this $/$ should not be here at all? (See Pattyn's Equation 11); From the equation above, I understand that this was set to n=3?

We have corrected the notation as suggested.

L.86: I'm somewhat guessing but I imagine that, rather being "quantified", B is "tuned" as to ensure that the flu solution matches the ice thickness.

We changed the wording to "tuned" as suggested.

L. 92: I'm not sure to understand what this means.

The "step in the elevation profile over ice cliffs" refers to the "steep elevation increase at the vertical ice cliffs".
Reworded accordingly.

L.98f: I'm not sure: What was done in Experiment 1 then? I understood that this averaging happened in that experiment already? If not, what viscosity value were used for the locations at which there were no ice thickness observations?

In Experiment 1 the generated margin thicknesses are used as thickness observations. During this experiment, the mean viscosity is generated within the reconstruction approach. This mean viscosity is in turn used in Experiment 2 as thickness input.
Rephrased for clarity.

L. 100: What is the meaning of "generic data" here?

Generic data refers to the margin thickness data.
Reworded for clarity.

L. 109: As far as I'm concerned, this sentence an be removed.

We decided to remove the sentence.

L. 124: Please clarify: is this wording referring to the results of Farinotti et al.?

Yes, this phrase refers to results from Farinotti et al.
Added source for clarity.

L. 127ff: I'm not sure to follow, is this discussion still referring to the "consensus thickness map"? Somehow, the focus seems to have shifted without noticing; Now I'm lost: What is this "second run" referring to? Is this Experiment 2? That's what the caption of Fig. 2 suggests. The wording is confusing.

Added Experiment numbers for clarity.

L. 132ff: Please split this sentence in at least two parts. I apologize, but I could not follow.

Split the sentence for easier readability.

L. 145: What is the meaning of "assimilated" here? Was the viscosity tuned again, as it was done for the ice thickness at the margin?

For NIF, the GPR measurements are used as thickness input. This was referred to here.
Rephrased for clarity.

L. 150: The concept of an "error margin" was not introduced, was it? I'm not sure to understand what is meant by that.

Reworded for clarity.

L. 151: I'm again in the need of guessing: are these "separate entities" something defined by the RGI? Fig. 1 doesn't show three entities on NIF, though?

The three different glacier entities are defined by the RGI and are also used in the Farinotti et al. consensus estimate. We have merged the three entities into one for our reconstruction, as the approach by Fürst et al. assigns the glacier margin an ice thickness of 0, which did not appear reasonable for the boundary lines between the different entities on the Northern Icefield.

L. 151ff: Sorry, I'm lost here: what is "model 1"? Is this meant to refer to Experiment 1 perhaps? This would be my first guess, but the next sentence is somewhat at odds with that. Has it something to do with "model 01" mentioned at L. 155?

Model 1 (and later Model 01) refers to one of the models from the consensus estimate.
Removed the single model data for more clarity and focused only on the consensus.

L. 159: Sorry, I'm lost again: didn't L. 155 say that the consensus has two models? Where is the third one coming from now, or why wasn't it mentioned at L. 155?

The consensus estimate is made up of two model for NIF and three models for KG.
Rephrased for clarity.

L. 166ff: Is my understanding correct: For NIF the consensus thickness is thus relatively close to both the GPR measurements and the results presented in this paper? This should probably be said explicitly as well, I imagine?

Yes, the mean ice thicknesses for the consensus estimate, our Experiments 2 and 3, as well as the reconstructions by Bohleber et al. are relatively close to one another. We have added a sentence stating this into the conclusion.

L. 176: Well, why is the volume never mentioned in the text then?

We replaced "volume" with thickness.

L. 181: What is the meaning of "retreat information"?

This refers to the lateral glacier retreat information, which was digitized from Landsat scenes and used in the margin thickness generation (Section 3.5)

L. 184f: Wait, didn't this "mean viscosity" come from the "margin ice thickness generated from DEMs and glacier outline" as well? How can this claim be made then?

The mean viscosity is generated from the margin ice thicknesses, but while local uncertainties from the margin thickness generation can influence the ice thickness distribution over the whole glacier (Fig. 2A KG), the mean viscosity shows a smoother ice thickness distribution, which seems more likely for Kersten Glacier (Fig. 2B KG). But as there are no thickness observations available for KG we cannot verify if the smoothed ice thickness distribution (Fig. 2B) is closer to reality or not.

L. 189f: I might be completely off track, but where would this mean viscosity come from at this stage? And isn't this claim somewhat in contradiction with what said at L. 135-136, i.e. that "the use of margin thickness information, generated from outline differences enabled a local glacier-specific viscosity tuning which might be preferential to an empirical temperature relation" (since, I assume, the latter would result in a mean viscosity as mentioned in the sentence)?

The mean viscosity is generated from within the thickness reconstruction approach. It is generated during the reconstruction using the margin thickness information generated from glacier outline differences. This means that by using glacier outline differences we can generate margin thickness information and then in a second step the mean viscosity. The results from our experiments show that for KG, where no ground/radar thickness observations were available, using the mean viscosity creates a smoother ice thickness distribution. This result might then be used preferential to approaches using empirical temperature relations to assess a glacier ice thickness as it is locally tuned from the direct glacier retreat as seen in satellite data.