Reply to Referee Comments

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

General Comments

In this study, the authors estimate the ice thickness of Northern Icefield and Kersten Glacier on Mt. Kilimanjaro in 2000 and 2011 using the ice thickness approach by Fürst et al (2017). Three different "experiments" are conducted to estimate the ice thickness, which either improves or are within the estimates from previous studies. The study makes good use of the few available observations, and the method and results are generally sound and interesting. I know this is a brief communication, but there are some key pieces of information that are missing within the data and model descriptions which I think are necessary to understand the manuscript, and the results would benefit from a discussion of uncertainties. In addition, the conclusion needs to be rewritten, as it does not seem to fit with the rest of the paper. In general, the manuscript would also benefit from an increase in specificity and clarity, as I often had trouble following the text. I hope the technical comments are useful for improving this.

Specific Comments

L40: Location of the AWS is not on Figure 1. Also, for how long a period has the AWS measured and what components are measures (and with what uncertainties)?

> The AWS collected data from February 2005 to September 2013. It is located on Kersten Glacier at 5873 m.a.s.l. and the measurements include incoming and outgoing radiative fluxes (longwave and shortwave) with an accuracy of ±10%, air temperature (±0.2°C), relative humidity (±2% units), wind speed and wind direction (±0.3 ms^{-1} and ±5°), air pressure (±0.2hPa) and the distance to the surface (±0.4%) (Section 3 c. in Mölg et al. 2009a).
>
> Due to the limited space in a "brief communication" we added a reference to the corresponding article by Mölg et al.

L48-49: This is not quite correct. The thickness estimate was created from the GPR by doing kringing interpolation, and it is an estimate from the whole area, not just the flat central part. Later in the manuscript (L 169), you seem to use the estimate that Bohleber created from the DEM, so I would mention that result here too. E.g. "In addition, ground penetrating radar (GPR) profiles from September 2015 (Fig. 1) were created by Bohleber et al. (2017). Using a kringing interpolation and the KILISoSDEM, the authors estimated the mean thickness to be between 21.2 ± 1 m and 27 ± 2 m."

> Implemented your suggestion into the manuscript.

L55-56: The Bohleber estimate is from 2015 and the consensus estimate is from 2000, so that should also contribute to the differences.

Added a sentence concerning the different years for clarity.

L39-56: For most of the observations you do not provide uncertainty estimates

We deliberately decided to not account for the input uncertainties in this case study as the focus is to exploit multi-temporal satellite information to better constrain a thickness reconstruction. Uncertainty consideration are covered in the methodological study by Fürst et al. (2017) comprising a spectrum of leave-out and sensitivity experiments. In light of the short communication format, it appears distracting to expand on the propagation of the input uncertainties into the final result. Moreover, input fields (SMB, DEM, outlines, thickness measurements) are often not necessarily provided with a robust error estimate. We therefore decided to refer the interested reader to Fürst et al. (2017) concerning the associated uncertainties.

L61: You have a point measurement in one location. How to you get distributed mass balance maps from one point on one glacier? And do you use the mean SMB from 2005-2013 for the 2000 and 2011 estimation? If you do, you should mention this as a possible source of uncertainty (and if you don't, how do you find the 2000 SMB?).

The surface mass balance model (Mölg et al. 2008, 2009a) creates the distributed mass balance based on a DEM (we used the SRTM), which creates the lower boundary conditions for the model and the meteorological data from the AWS, which are used as the model driver (Mölg et al. 2009a). We use the mean SMB for both, the 2000 and 2011 glacier states.

L61-65: Again, what do you use as forcing for NIF if the AWS is on KG? And how do you use the sonic ranger to test refreezing on NIF if it is mounted on KG? In addition, you should mention the sonic ranger in the data section and not only in the methods.

We use the meteorological data gathered by the AWS on KG as forcing on NIF as well. There is an WS installed on NIF from which we use a plotted time series of the sonic ranger measurements to which we compare our modelled accumulated surface height change. As the ice thickness reconstruction use the mean annual surface mass balance as input, we mainly used the total accumulated surface height change over the time period/ at the end of the modelling period (2013 September) to compare our results to. So the climatic variables (T, RH, ..) are the same for NIF and KG, but the topographic/elevation data differs, as this is directly calculated from the digital elevation model SRTM.

L87: What method do you use for interpolating?

The method used for interpolation is Natural Neighbor/Sibsonian Interpolation.

We mentioned the method now in the manuscript.

L109-122: A table with the different main thicknesses estimates would be useful and make comparison easier for the reader. I would e.g. include the mean thickness for each experiment, the mean thickness in the consensus estimate and Bohleber et al, and perhaps the thickness at the borehole locations. I know this is a brief communication and you are not allowed more figures, but maybe as a supplement.

We added a table containing the mean thickness estimates for NIF and KG and the thickness at the Thompson et al. (2002) borehole locations to the supplement (Supplementary Table 1).

L122: is it possible to calculate an uncertainty on the mean numbers? e.g. by leaving some GPR points out of the simulation and using those points for validation? Or if that would be too much work, you could give an approximation from the core location values (but then only for 2000). You already give it in percentage in the discussion, but here you could use the maximum absolute value.

We added the suggested approximation of absolute values at the core locations. For Experiment 1(2) the ice thickness at the core locations differ by 19.9 m (4.4 m) at C1, 23.9 m (8.3 m) at C2 and 36.6 m (26.1 m) at C3.

L124-174: The discussion would benefit from a short discussion on model uncertainties. For example for the constant viscosity runs, did you conduct a sensitivity analysis? Can you give an approximate uncertainty estimate of the SMB? And are there any uncertainties associated with the use of SIA?

The inferred **viscosity values** not only depend on the structural and temperature properties of the glacier body. They are also affected by the uncertainties of all other input fields and measurements. As the input uncertainty is already analyzed in depth by withholding GPR measurements in Fürst et al. (2017), it seemed redundant to repeat this exercise here. Certainly, in light of the short article format.

We cannot give an approximation on the **SMB uncertainty**, but as previously shown in Fürst et al. (2017) its influence on the ice thickness reconstruction is only minor. It was shown that, by changing the SMB input drastically, the mean ice thickness is reduced by 5% and the estimated ice volume by 4%. These values were found when the least amount of direct thickness measurements was assimilated. Moreover, this influence was estimated for various glacier geometries, including an ice cap, on Svalbard. They also noted that, where ice thickness data is available, the influence of SMB input is compensated by direct observations (Fürst et al. 2017).

First of all, the **SIA** is a key component of this type of reconstruction. An expansion to include the solution of more complete forms of the force balance would require fundamental adjustments in the method. Though more complete, the problem might become even less well-posed and the computing requirements would increase unproportionate. Some of the uncertainties associated to the choice of the SIA are covered in Fürst et al. (2017).

L181-183: Why did you use a method which uses the SIA if the glacier is dynamically inactive? Would a plastic approach not be a better choice? Also, I think this section would fit better in the discussion.

We use the SIA in our reconstruction, as it is the method implemented in our reconstruction approach. KG is located on the steep flank of Mt. Kilimanjaro and we expect some ice motion. For NIF, this issue might be more relevant, and we expanded the discussion of this aspect in the revised manuscript. In such setups, the mass-conserving SIA approach is not ideal. We have no model to use a plastic flow assumption, so this approach was also not viable for us.

Concerning a plastic approach, it would certainly be an alternative here. Yet such approaches have often been applied in flowline setups with appropriate spatial averaging of the geometric input. Although one could theoretically apply them in 2D to each grid point, it would require an extra article to assess what the best strategies would be for spatial smoothing of the required input. We are unaware of a precursor study that applies the perfect plasticity concept in 2D (without final spatial interpolation) that is readily transferable to the complex topography of NIF.

L184-190: I was a bit puzzled on how you reach this conclusion. You suddenly mention “mean viscosity” experiments for NIF, although you did not mention this anywhere in the paper (Only for KG, as written in Table 1). For all three experiments, you always generated a viscosity field from observations for NIF (first from the margins, then using Bohleber et al data). You write that “the reconstructions reveal that if there are no thickness observations available, better results can be achieved with a mean viscosity value as input for ice thickness, instead of margin ice thickness generated from DEMs and glacier outline difference” but from what do you reach this conclusion? For KG you wrote the results for the margin method and the viscosity method were almost equal (and you use the margin method to get the mean viscosity in the first place), and for NIF you did not test it. Please clarify. And if you did do the mean viscosity test for NIF too, you should provide it in the paper.

We fear that we have not been careful enough in presenting the experiments which raised this concern. In the case of 'directly using lateral thickness information' and in the case of the 'mean viscosity', the thickness information from the retreat (ice-free area) is used. The difference is only how the reconstruction deals with this data. The two options are that the viscosity of each 'lateral thickness point' is used individually for an interpolation over the domain, resulting in a spatially variable ice viscosity (Experiment 1, NIF and KG). Otherwise, the viscosity point information is simply averaged, and a uniform value is used for the entire glacier (Experiment 2 and 3, KG).

L196-198: Wouldn’t how well the margin method / mean viscosity method works depend on the size of the glacier?

We believe that the size of the glacier would most likely influence the outcome of the margin/mean viscosity method, but we have not tested the approach on glaciers of different sizes so we cannot comment on that further. The Kilimanjaro setup is quite special, and it is difficult to assess the glacier size dependence. Yet, glacier retreat is mostly expressed at low elevations. It is there that we expect to acquire past thickness values from multi-temporal satellite information. As the frontal area represents an increasingly smaller portion of the entire system as glaciers become larger, the size dependence is certainly an interesting question. We can unfortunately not answer this here on the basis of the two very different glacier types on Mt. Kilimanjaro.

<u>Technical Comments</u>

L10: Add the thickness in 2000 too

We refrained from adding the 2000 thicknesses into the abstract as the word limit did not allow us to explain the difference (thickness increase) between the 2000 and 2011 reconstructions sufficiently and we believe it might cause confusion without the proper explanation.

L11: Write the unrealistically thick value

Changed the manuscript accordingly.

L11: change “meanwhile” to ”have become”

Changed the manuscript accordingly.

L13: change “indicator” to “indicators”

Changed the manuscript accordingly.

L14: delete “As”

We decided to stick with this wording.

L20: delete “to”

Changed the manuscript accordingly.

L24: “assessment on” to “assessment of”

Changed the manuscript accordingly.

L25-28: You haven’t introduced what you will do in this study yet, so a bit odd to talk about comparison already. I would suggest changing to: “A recent study attempted to reconstruct the distributed ice thickness for all glaciers outside of Antarctica using a consensus of up to 5 models (Farinotti et al. 2019). This estimate generated ice thicknesses estimates for Northern Icefield (NIF) and Kersten Glacier (KG) using ensembles of 2 and 3 models, respectively.” Then at the end of line 37 you can add “The resulting thickness estimates are then compared with the consensus estimate” or similar.

Reworded the passage according to the suggestion.

L28-31: I would suggest dividing the sentence in two to make it easier to read: “. . . (Farinotti et al. 2019). In addition, it was recently discovered that KG has separated into two fragments, which is not in agreement with the estimated high thickness values in the study.” I would also add a citation for the separation.

Divided sentence as suggested. Added reference to the Landsat scene used in the study.

L34: I would suggest adding a line describing the model here, e.g. something like L80-83. Currently you mention a SMB model in L 39 without introducing that you even use it first.

Added information on the SMB model in the introduction. As this manuscript is a “brief communication” we refrained from adding further information on the reconstruction approach in the introduction. We added a cross-reference to the corresponding section 3.4.

L39: either delete “the distributed surface mass balance (SMB) model and” or introduce the model in the introduction.

We briefly introduced the model in the introduction.

L41: define DEM the first time you use it

Added definition of DEM.

L41-43: missing reference for SRTM and Landsat 5

We added references to the data sets used.

L43: change "from a merge of" to "by merging"

Reworded to "by differencing from a merge of two …"

L46: reference Fig 1 after describing the redefinition

Added reference to Figure 1.

L46: Future separation? Earlier you wrote it already separated?

We anticipate a future separation of the Northern Icefield. Kersten Glacier has already separated. Reworded for clarification.

L47: delete "apart from" and add "were" before drilled

Deleted words as suggested.

L48: can you add the borehole locations to figure 1 instead? It would be nice to have all the observations in the same figure.

Added borehole locations to Fig. 1 and removed them from Fig. 2.

L48: Definite GPR first time you use it

Defined acronym.

L54: change "showed a mean" to "had a mean"

Changed wording as suggested.

L54: give the value for NIF, "similar value" is too vague

Removed passage from manuscript.

L61-65: You should explain the reason for the model changes first, as it will be easier for the reader to follow. E.g. "The full MB model has only previously been verified for KG. However, because of the low slope angles of NIF, meltwater cannot run off from the surface of its planar top before refreezing sets in (Mölg and Hardy 2004), which was not captured by the model. Therefore we upgraded the model so that refreezing of meltwater is allowed on a bare ice surface with a slope angle below 5 degrees. With these changes, the model is capable of reproducing the observed surface height changes observed by a Sonic Ranger mounted to the AWS."

Rephrased the section for clarity with the suggestions in mind.

L76: change "nowadays" to "currently" or "2011"

Changed "nowadays" to "currently".

L89: change "increase" to "increased"

Changed word as suggested.

L90-91: I suggest changing the structure so the reasoning is before the how, e.g.: "In order to smooth the surface slope during reconstruction we use use the coupling length parameter, which is defined a a multiple of the local ice thickness."

Changed wording as suggested.

L95: add "by" before "combining"

Added the word "by" as suggested.

L98: the values are inferred and then the values are interpolated for the whole area?

We rephrase this passage and hope that it became clearer now.

L117: change "a distribution" to "the distribution"

Reworded the sentence.

L144: reference is missing a year

Added missing year to the reference. The reference is Thompson et al. 2002.

L147: what is "the better model"?

Removed the distinction of the two models that make up the consensus estimate for NIF for easier understanding and reworded the passage.

L149: change the end of the sentence to ".. the consensus estimate underestimates the the thickness at these points."

We rephrased a large part of the discussion for clarity, so the sentence referred to here was completely changed.

L165: mention the 10 and 5 m experiments in methods

Mentioned the 10 and 5 m experiments in the methods section 3.4.

"With the higher DEM quality in 2011, the resolution was iteratively increased from 25, via 10 and 5, to 2 m."

L169: remove "where the very high . . . as well"

We rephrased a large part of the discussion for clarity, so the sentence referred to here was completely changed.

L178: remove "became ice free or"

Rephrased the sentence to "in areas that became ice-free in the last decade."