Dear Editor,

We want to thank you and the two anonymous referees for the critical and constructive comments on our manuscript. We have carefully considered the reviewers' and your comments. Our point-by-point response to the reviewers' comments follows below. Our replies/actions are indented and given in blue font.

Reply to Reviewer RC1

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.

<u>Minor Comments</u>

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.

Reply to Reviewer RC2

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.

Technical Comments

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.”

Brief communication: Glacier thickness reconstruction on Mt. Kilimanjaro

Catrin Stadelmann[1], Johannes Jakob Fürst[1], Thomas Mölg[1] and Matthias Braun[1]

[1]Institute of Geography, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, 91058, Germany

Correspondence to: Catrin Stadelmann (catrin.stadelmann@fau.de)

Abstract. ~~Glaciers on Kilimanjaro are unique indicators for climatic changes in the tropical mid-troposphere of Africa. The history of severe glacier area loss raises concerns about an imminent future disappearance. Yet, the remaining ice volume is not well known. We reconstruct thickness maps for 2000 and 2011 for the Northern Icefield (NIF) and Kersten Glacier (KG) that are informed by ground-truth thickness measurements and multi-temporal satellite information. For 2011, we find mean thickness values of 26.6 and 9.3 m, respectively. The existing consensus estimate for global glacier ice thickness shows unrealistically thick values for KG in areas that are meanwhile ice-free.~~

Glaciers on Kilimanjaro are unique indicators for climatic change in the tropical mid-troposphere of Africa, but their disappearance seems imminent. A key unknown is their present ice thickness. Here, we present thickness maps for the Northern Icefield (NIF) and Kersten Glacier (KG) with mean values of 26.6 m and 9.3 m respectively in 2011. In absence of direct measurements on KG, multi-temporal satellite information was exploited to infer past thickness values in areas that have become ice-free and that allow glacier-specific calibration. In these areas, KG is unrealistically thick in the existing consensus estimate of global glacier ice thickness.

1 Introduction

~~The importance of t~~Tropical glaciers at high elevations a~~re~~s unique climate indicator~~s~~ for the tropical mid-troposphere ~~has been previously highlighted~~ (e.g. Kaser 2001, ~~Kaser et al. 2004,~~ Mölg et al. 2009~~a~~). As one of few ~~remaining~~ tropical locations with still existing glaciers, Mt. Kilimanjaro, a stratovolcano with an elevation of 5895 m.a.s.l., is located in East Africa close to the Tanzania-Kenya border (3°04' S / 37°21' E) (Fig. 1, overview). In addition to the very high elevation, the free-standing nature of the mountain causes the glacier on top of the summit to be directly exposed to tropospheric flows at higher altitudes, minimizing the forcing of local climate on the glacier and creating a unique opportunity to study the mid-troposphere climate. The modern glacier recession on Kilimanjaro has been well documented and mapping approaches have shown that from an estimated ice extent of 11.4 km² in 1912, only 1.76 km² remained in 2011, constituting ~~to~~ a severe 85% reduction in glacier area (Cullen et al. 2013). While glaciological research on Kilimanjaro has ~~been~~ focused on mapping glacier area and ~~glacier~~ retreat (~~Kaser et al. 2004,~~ e.g. Cullen et al. 2013), as well as quantifying the mass and energy balance ~~studies~~ (Mölg et al. 2003, 2008, 2009), the research on the ice thickness of different glaciers on Kilimanjaro has been comparably sparse (Bohleber et al.

2017). However, in light of severe glacier recession, an assessment of~~n~~ current glacier thickness is important to better determine future recession. A recent effort was made to reconstruct the distributed ice thickness for all glaciers outside ~~of~~ Antarctica and Greenland ~~from~~ using a consensus of up to 5 models (Farinotti et al. 2019).~~,~~ ~~provides the possibility to compare the global estimate, which generated ice thicknesses from an ensemble of 2 (Northern Icefield, NIF) and 3 (Kersten Glacier, KG) models (Fig. 2 C), to results of this study.~~ This estimate generated ice thicknesses for Northern Icefield (NIF) and Kersten Glacier (KG) using ensembles of 2 and 3 models, respectively. The consensus estimate produces a similar mean ice thickness of 21.5 m for NIF, which is in fair agreement with the observations by Bohleber et al. (2017), considering that~~U ncertainties of the global consensus estimate, such as the separation of NIF into three separate glacier entities in the most recent version of the Randolph Glacier Inventory (RGI6.0), as well as the consensus estimate~~ the consensus was not ~~being~~ informed ~~of~~ by local thickness observations~~, have already been recognized as deficiency of the global estimate~~ (Farinotti et al. 2019). ~~and t~~Moreover, the recently observed separation of KG into two fragments (e.g. Landsat 5 scene 2011-08-22; Image courtesy of the U.S. Geological Survey) ~~casts further doubts on~~ is not in agreement with the high thickness values ~~there~~illustrated in the consensus estimate.

Here, we present the first well constrained thickness maps for KG and NIF using a mass conserving reconstruction approach introduced in Fürst et al. (2017) that readily assimilated thickness measurements (Section 3.4) ~~for the first time (Fürst et al., 2017)~~. In two different experiments we test the influence of varying input of ice thickness ~~input~~ observations for the glacier state of 2000, where we rely on surface mass balance (SMB) data from a physically-based model developed by Mölg et al. (2008, 2009; Section 3.1) and ~~-~~ digital elevation data with global coverage (Shuttle Radar Topography Mission; SRTM; USGS), pursuing a new calibration strategy that uses multi-temporal satellite information on geometric changes in absence of observational ice thickness data on KG. These resulting thickness estimates are then compared to the consensus estimate (Farinotti et al. 2019). In a third experiment, we combine the very high resolution digital elevation model KILISoSDEM (0.5 m ground resolution; Sirguey et al. 2014) with ~~results~~ the calibration strategy from the previous 2000 experiments to ~~improve the reconstruction of the ice cliffs at NIF~~produce a best estimate for the 2011 glacier state.

2 Data

To apply the distributed surface mass balance (SMB) model (Sect. 3.1; Mölg et al., 2008, 2009) and the thickness reconstruction (Sect. 3.4), the following input data was used:~~The two glaciers share the following input for the distributed surface mass balance (SMB) model and the ice thickness reconstruction approach:~~ climate data measured by the automatic weather station (AWS) located on KG (~~Fig. 1,~~ Section 3 c in Mölg et al. 2009), digital elevation information from the SRTM digital elevation model (DEM) from 2000 and the KILISoSDEM from 2012 (Sirguey et al. 2014), the RGI6.0 glacier outlines from 21 February 2000 (RGI Consortium 2017), as well as digitized outlines based on a Landsat 5 image from 22 August 2011. Surface height change was generated by differencing ~~from~~ a merge of two TanDEM-X radar images from 2011 (28 January 2011, 4 April 2011) and the SRTM DEM.

The central plateau area of NIF drains westward into two glaciers, Drygalski Glacier in the south and Credner Glacier in the North. In anticipation of a future separation of NIF, we redefine Credner Glacier (CG) to comprise the northern part of NIF (Fig. 1). Ice thickness measurements on Kilimanjaro are limited to NIF, where apart from three ice cores were drilled to bedrock in 2000, with lengths from 49.0 m (C1) to 52 50.9 m (C2, C3) and 50.8 m (C3) (Thompson et al. 2002; Fig. 2A 1 for borehole locations),. In addition, ground penetrating radar (GPR) profiles from September 2015 (Fig. 1) were collected by Bohleber et al. (2017). Using a kriging interpolation and the KILISoSDEM, the authors estimated the mean thickness to be between 21.2 ± 1 m and 27 ± 2 mGPR profiles from September 2015 show a mean (maximum) ice thickness of 21.2 ± 1.0 m (53.5 ± 1.0 m) for NIF's central flat area (Bohleber et al. 2017). For the anticipated reconstruction in 2000 and 2011, the GPR thickness measurements for NIF are adjusted by linearly interpolating scaling of the above-mentioned elevation change information to the elapsed time between the DEM date and acquisition date of the thickness measurements. To account for different availability of thickness measurements, Fürst et al. (2017) conducted experiments withholding 1% - 99% of the available point measurements on several test geometries on Svalbard. Aggregate errors typically exceed 10-20% of the mean glacier ice thickness when most measurements are withheld but error values quickly reduce as measurements become available. Between the two end-member experiments (1% and 99%), volumes of the test geometries differ by at most 10%. Considering input uncertainties from the DEM and the SMB fields, sensitivity tests revealed that ice-volume differences remain below 5% (only shown for the ice-cap geometry). For more details on associated uncertainties and input sensitivities, we refer the interested reader to Fürst et al. (2017). Here, the focus is rather on assessing the utility of multi-temporal satellite information in a glacier specific calibration.

For KG no thickness measurements are available, but the consensus estimate (Farinotti et al. 2019) provides approximate information for the year 2000 where KG showed a mean ice thickness of 27 m. A similar value was foundfor NIF, which is in fair agreement with the Bohleber et al. (2017) observations considering that the consensus was not informed by the available thickness measurements.

3 Methods

3.1 Mass balance modelling

The mean annual climatic surface mass balance fields were generated using version 2.4 of the distributed, physically-based mass balance (MB) model by Mölg et al. (2008, 2009a), being driven by using meteorological input from the aforementioned AWS (Suppl. Fig. 1). As tThe full MB model has only already been tested on and verifiedcalibrated and validated for KG before., For the application on NIF, surface meltwater is not expected to run off but rather refreeze over the very flat plateau areas (Mölg & Hardy, 2004). Because of the low slope angles, meltwater cannot run off from the surface of NIF's planar top before refreezing sets in (Mölg and Hardy 2004).

To properly reproduce these conditions on NIF, we ~~slightly altered~~ revised the model code so that ~~it allowed~~ refreezing of meltwater can occur ~~on~~ on a bare ice surface with a slope angle below 5 degrees (not captured before). ~~In this way~~ With these changes, the model is capable of reproducing the observed surface height changes observed by a Sonic Ranger mounted to the AWS.

3.2 TanDEM-X~~DX~~ processing

First a DEM with 30 m resolution was generated from the 2011 TanDEM-X imagery (for details refer to Braun et al., 2019). ~~The detailed processing steps to create a DEM, as well as a surface height change layer, from the TanDEM-X image can be found in e.g. Braun et al. (2019).~~ Surface elevation changes between 2000 and 2011 were inferred from DEM differencing with respect to the SRTM DEM (Suppl. Fig. 2). ~~Positive values, which indicate a height gain in the TanDEM-X layer, were removed as a height gain outside the 2011 glacier extent implies an increase in glacier thickness from 2000 to 2011, which is unlikely. In total we removed 92 of 602 grid cells with a mean height gain of 0.19 m/a for NIF and 14 of 254 grid cells with a mean height gain of 0.25 m/a for KG.~~

3.3 Margin thickness generation

For KG, no in-situ thickness measurements are available. Therefore, multi-temporal DEM and glacier outline information is used to infer past ice thickness. First, glacier retreat is delineated from outline information in 2000 and 2011 (Fig. 1 hatched area). In the ~~nowadays~~ currently ice-free area, contemporaneous elevation changes (2000-2011) then give information on past ice thickness. Positive values, which indicate a height gain in the TanDEM-X layer, were removed as a height gain outside the 2011 glacier extent implies an increase in glacier thickness from 2000 to 2011, which is unlikely. In total we removed 92 of 602 grid cells with a mean height gain of 0.19 m/a for NIF and 14 of 254 grid cells with a mean height gain of 0.25 m/a for KG.

3.4 Ice thickness reconstruction

A detailed description of the two-step ice-thickness reconstruction, of which we only used the first model step as surface velocities were not available, can be found in Fürst et al. (2017). The reconstruction approach is based on the principle of mass conservation and computes a glacier-wide flux field from the difference between the surface mass balance (Section 3.2) and contemporaneous elevation changes. The flux solution is converted into thickness values using the Shallow Ice Approximation (SIA; Hutter, 1983). This conversion involves the ice-viscosity parameter *B*, which is a-priori unknown. This parameter stems from assuming a Glen-type flow law, linking deviatoric stresses to strain rate components $\dot{\varepsilon}_{ij}$ via the effective viscosity $\eta = 0.5\, B^{-1/n3} / \dot{\varepsilon}^{(1-n)/n}$. Here, $\dot{\varepsilon}$ is the second invariant of the strain rate tensor (for further information on the equation see Pattyn 2003) and n=3 ~~is Glen's law exponent~~. After the flux solution is obtained, *B* ~~can be quantified~~ is calibrated at locations where thickness measurements are available. This point-information is then expanded to the entire glacier domain using a Natural Neighbor Sibsonian Interpolation, resulting in a spatially variable field. Before interpolating the *B* values from each measurement location to the entire glacier basin~~at each measurement sites~~, ~~the~~ an average value is prescribed along the glacier outline to avoid spurious extrapolation effects. For the reconstruction in 2000, a nominal mesh resolution of 25 m was chosen. With the higher DEM quality in 2011, the resolution was iteratively increased from 25, via 10 and 5, to 2 m~~to up to 2 m~~. The processing was conducted separately for NIF and KG. To smooth the surface slope during reconstruction, we use a coupling length parameter (introduced in Fürst et al. 2017), which is defined as a multiple of the local ice thickness. In this way, flux streamlines become less erratic and their alignment increases. ~~The coupling length parameter is defined as a multiple of the local ice thickness serves to smooth the surface slope during the reconstruction.~~ For KG, the parameter is set to 1, a typical value for valley glaciers (Kamb & Echelmeyer, 1986). For NIF, it had to be reduced ~~because otherwise the step in the elevation profile over ice cliffs was not imprinted~~ so that the steep elevation increase at the vertical ice cliffs is depicted in the thickness field. A compromise value of 0.3 was chosen to still guarantee sufficient smoothing of the flux streamlines.

3.5 Experimental Setup

The general strategy is to reconstruct a thickness field for KG and NIF by combining SMB, elevation changes and glacier geometry with in-situ measurements of ice thickness for two points in time, namely 2000 (Experiment 1 and 2) and 2011 (Experiment 3). In Experiment 1, we reconstructed the glacier state for 2000 with the generated margin thickness data (Section 3.3) for both NIF and KG. ~~At these thickness measurements, an appropriate viscosity value is inferred.~~ As KG is rather small, we expect a homogeneous ice viscosity. In Experiment 2, we therefore decided to simply average the point information on ice viscosity and use a constant viscosity value over the entire glacier basin~~average all viscosity values and impose a constant value for KG~~. In this way, ~~generated~~ lateral thickness values are no longer reproduced but spurious spatial viscosity variations stemming from the generic margin data are suppressed. For NIF in Experiment 2, we chose to use the thickness measurements from Bohleber et al. (2017) as input, to check how observational data influences the glacier~~-~~wide ice thickness in comparison to only using margin thickness information. ~~With~~ In Experiment 3, the aim is to benefit from the 2011 KILISoSDEM showing very high resolution. For NIF, the reconstruction can still be calibrated by GPR measurements from Bohleber et al. (2017) acquired in central areas. For KG, the retreat information falls outside the ice-covered domain in 2011. Therefore, we use the mean viscosity information as inferred for the reconstruction in year 2000 (Experiment 2). The KILISoSDEM is further exploited to investigate the resolution influence. ~~we reconstructed the glacier state in 2011, using again the mean ice viscosity of 2000 as input for KG and the observations from Bohleber et al. as input for NIF, as these thickness inputs produced the best results for the respective glaciers in the 2000 reconstruction. We decided to do a reconstruction of the 2011 glacier state, to test the influence of a higher resolution DEM, as well as a higher model resolution on the reconstruction result.~~ Table 1 summarizes the three different experimental setups (Tab. 1).

4 Results

~~In the following section we discuss the results of the three experimental setups (Section 3.5).~~ Results show generally larger ice thickness for NIF than for KG in all three experiments. For Experiment 1 (Fig. 2 A), KG shows thickest~~er~~ ice of up to 15 m at the flat ~~top~~ plateau parts of KG. For the central areas on the mountain flank thickness values show a mean of 6.2 m and locally reach up to 7.5 m, ~~and thickness up to 7.5 m at the central area of the steep part,~~ with patches of thinner ice towards the glacier margins ~~and a mean ice thickness of 6.2 m~~. NIF is up to 40 m thick in its center, decreasing towards the glacier margins and towards CG and has a mean ice thickness of 13.7 m. At the borehole locations C1, C2 and C3 ice thicknesses of Experiment 1 are 19.9 m, 23.9 m and 36.6 m thinner, respectively (see Suppl. Table 1).

~~In comparison, r~~Results from Experiment 2 show a similar thickness pattern on KG (Fig. 2B). For NIF, the magnitude differs significantly ~~(Fig. 2 B)~~. Now one large part of NIF's flat area and two smaller parts of CG exceed 40 m. Moreover, the ice thickness in the steeper western areas of NIF and CG has increased by a factor of 2. The mean ice thickness also increases to 23.4 m. At the borehole locations C1, C2 and C3 ice thicknesses of Experiment 2 are 4.4 m, 8.3 m and 26.1 m thinner, respectively (see Suppl. Table 1). Concerning the GPR surveys from Bohleber et al. (2017), the thickness map of NIF (Fig. 2D) largely reproduces these measurements. Turning to the consensus estimate map (Farinotti et al., 2019), larger discrepancies prevail (Fig. 2E), especially towards the eastern part. KG shows a similar thickness, ~~but a distribution is smoother~~ but the ice body on the mountain flank becomes thicker in the central parts. As before, the thickest ice patch remains on the plateau~~as compared to Experiment 1 and the flat top part of KG is still thicker than the slope part~~. The mean ice thickness with 6.9 m is very similar to Experiment 1.

For Experiment 3 (Fig. 1), KG is now split into two parts and shows an ice thickness of up to 10 m at the flat top part and most of the slope being between 5 and ~~-~~7.5 m thick. KG's mean ice thickness is 9.3 m. NIF's thickness distribution is similar to Experiment 2, with the thickest areas of over 40 m at its flat part on the plateau. For NIF, the decrease in thickness is less noticeable than the lateral retreat and decrease of glacier area. The mean ice thickness of NIF in Experiment 3 is 26.6 m. At the three ice-core locations, the thickness mismatch remains comparable to Experiment 2 (see Suppl. Table 1), with a mean relative absolute difference of 26%. This value is rather large and exceeds inferred error estimates for the majority of glacier on Svalbard (Fürst et al., 2017). Here, we want to use it as a rough orientation for the overall uncertainty of the 2011 reconstruction.

5 Discussion

We will first discuss the reconstructions for the year 2000 (Fig. 2): Generally, our experiments produce results with a higher difference in thickness magnitude between KG and NIF as compared to the consensus thickness map (Fig. 2C; results from Farinotti et al. 2019). ~~shows less difference in thickness magnitude between KG and NIF as compared to our reconstruction.~~ For KG, no ice thickness measurements are available, and it is uncertain to what extent the generated thicknesses along the glacier margin (Section 3.3) are useful to inform the reconstruction. The We find that the margin values result in a spatially varying viscosity field, which is transmitted into the ice thickness field (Experiment 1; Fig. 2 A). As no strong viscosity variations are expected for the small KG, a second run was conducted with constant viscosity (Experiment 2; Fig. 2 B). Results indicate a thick central flow unit, as one might expect for a steep glacier. In this case, the result provides, as well as a smoother ice thickness distribution, with higher thickness in the center of the glacier and thinning towards the margins. In the absence of ground truth data, it is unclear, which thickness field is more plausible.Whether this smoothed result, or the less smooth reconstruction generated from margin observations, is closer to the reality is unclear. Considering the input of thickness observations was only available on the glacier margin, calibrating the model, which is based on ice flow, with data where ice flow is generally small, puts the reconstruction to its limits However, as the thickness of most glaciers on Earth is unsurveyed, 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 (Huss and Farinotti, 2012). The consensus map shows a similar thickness pattern as Experiment 2. The most notable difference is found for the thickness magnitude of KG. For the consensus estimates, thickness values exceed 35 m both for the flat top part and the central steep slope part. The consensus mean thickness of 27.1 m, is more than twice as large as in our reconstruction. Since there are no actual thickness observations for KG, it is not certain that the ice was only up to 15 m thick in the year 2000. However, KG split in two parts by 2011. The separation line follows a contour line just below the plateau. Mean elevation changes between 2000 and 2011 of -0.64 m/yr suggest that not more than 7 m of ice was present in 2000. With 35m ice in this area, the consensus estimate seems too large. NIF's peculiar geometry poses a challenge and it is difficult to reconstruct the ice thickness distribution using generic thickness observations around the margin (Fig. 2A). The ice is much too thin in the interior (Fig. 2 A), which underestimates the ice core lengths from Thompson et al. 2002 by 48, 52 and 71% for the core locations C1, C2 and C3 respectively (Fig 2. A). When the interior GPR measurements are assimilated used as model input (Experiment 2; Fig. 2 B), differences decrease to 10, 17 and 53%. The consensus estimate for NIF also underestimates the ice core lengths by 34, 38 and 72%, respectively. The consensus map is a weighted mean of two model results on NIF. The better model shows a difference of 10, 12 and 76% for C1, C2 and C3, respectively. C3 is located near a separation line between the three RGI glacier units for NIF. This might be a reason why the consensus estimate has problems to reproduce a larger value. The upper value of the error margin has to be assessed cautiously, as the location of the drill site of one of the cores coincides with the margin of the three separate glacier entities on NIF, where model 1 and consequently the composite show a notably larger difference to the measurements than for the other two locations. While this location is the one where our reconstructions show the largest difference to the measurements as well, the proximity to an internal separation line should not influence our reconstruction. If we only consider the other two core locations (C2, C3), our reconstructions underestimate the measured thickness by 48 and 52% and 10 and 17% for Experiment 1 and 2, respectively. The consensus result (model 01; model 02) underestimates the measurements by 34 and 38% (10-12%; 65-72%). This shows that one of the models is strongly influenced by the separation of NIF into three separate entities and in turn worsens the composite result. NIF's mean ice thickness of 21.5 m is similar to our results, especially to the mean of Experiment 1 and 2 (21.6 m). Increased mismatch values, especially for borehole C3, might as well be explained by the very flat plateau. Therefore, ice motion is expected to be rather slow. Stagnant and flat areas are challenging for a reconstruction based on ice-flow and Fürst et al. (2017) show that uncertainties in the reconstructed thickness values significantly increase towards the ice divide of an ice cap. They further show that measurements along divides and ridge areas are most valuable to constrain the reconstruction approach used here. Although GPR measurements are available on the NIF plateau, we expect that uncertainties increase quickly away from these measurements. This can partly explain the mismatch with C3. Turning to the consensus estimate the mismatch is significantly larger with relative underestimations of 34%, 38% and 72% for boreholes C1, C2 and C3. Despite that no GPR measurements were considered on Mt. Kilimanjaro, the complex topography posed a similar challenge for the models participating in the consensus. This is also reflected in the similar mean ice thicknesses, which are 27.1 m for the consensus estimate (Farinotti et al. 2019) and 23.4 m for Experiment 2.

Analyzing the three separate models of the consensus, none of them creates a large bias, which could influence the composite result, but the KG reconstruction is too thick in areas where the glacier has disappeared in 2011. Thus, we can confidently say that the consensus estimate is too thick for KG, but we cannot confirm how well our reconstruction results are able to depict reality.

Experiment 3 repeats the reconstruction for the year 2011 at a very high resolution. The overall general distribution of ice thickness is barely affected by the increase in resolution. This stability under resolution increase is assuring and illustrates the effects of inherent smoothing via the coupling length parameter a desired effect and stems from the coupling length that scales with the thickness. For NIF however, resolution is key, and the cliff geometries are much better imprinted in the final thickness map. Further experiments with 10 and 5 m model resolution (not shown) showed barely any difference in thickness distribution, verifying this effect. The mean ice thickness for KG and NIF have increased in comparison to Experiment 2 to 9.3 m and 26.6 m, respectively. This increase likely stems from the difference in DEMs that were used for the reconstruction. While an increase for the period 2000-2011 in mean glacier-wide ice thickness is unlikely, results for NIF now match the thickness estimates from Bohleber et al. (2017), where the very high resolution KILISoSDEM is used as input as well. While Bohleber et al. (2017) have reconstructed parts of NIF with a linear extrapolation of the bedrock, the thickness distribution in the areas located on top of the summit plateau matches the result of our reconstruction. However, the slope area is generally thinner, which is caused by the influence of several ice-free areas that could not be identified from the coarse 2011 Landsat image. From this we can deduce that the increase in ice thickness could be caused by a difference in DEM and not increased model resolution. As observed elevation changes do not support an increase, remaining explanations comprise model resolution, outline differences and DEM quality. Resolution can be excluded from a 25 m reconstruction in 2011 (not shown). Concerning the 2011 outlines, some internal ice-free areas (on both NIF and KG), present in the RGI, could not be confirmed from the coarse Landsat imagery, resulting in thicker ice. The quality difference between SRTM and KILISoSDEM is certainly also a contributing factor explaining part of the larger thickness values.

Finally, we want to briefly discuss the reconstruction approach, used here, with respect to other strategies for inferring distributed thickness information. The Ice Thickness Models Intercomparison eXperiments (IMTIX; Farinotti et al. 2017) concluded that as long as no thickness measurements are available, no single strategy generally outperforms the others. In this case, an ensemble result from multiple models is preferable. Yet here, observations are either available or are inferred from multi-temporal satellite imagery. Measurement availability was used in the global consensus estimate to infer performance scores for the participating models (Farinotti et al., 2019) and the approach by Fürst et al. (2017) was attributed the highest value. Yet, with regard to applications on individual geometries as for Kilimanjaro, comparable results, as presented here, might well be attainable with various approaches. Regarding input requirements, approaches based on the perfect plasticity assumption are least exigent, only requiring information on the ice geometry (e.g. Frey et al., 2014).

6 Conclusion & Outlook

This study has a multi-disciplinary character as we apply modelling approaches for glacier surface mass balance, infer remotely sensed elevation changes and utilize available information in a data assimilation. The aim of ~~this study was first and foremost~~ the assimilation is to accurately determine the ~~volume~~ thickness and distribution of ice for NIF and KG on Mt. Kilimanjaro. ~~For this purpose~~As ice thickness observations were not available for KG, ~~thickness observations~~the reconstruction approach was calibrated with past thickness values inferred ~~are generated~~ from multi-temporal satellite information ~~on the glacier geometry. These generic values can be inferred~~ in areas that became ice-free in the last decade ~~or were ice-covered in the past~~. Our reconstructions for 2011 show mean ice thicknesses of 9.3 m for KG and 26.6 m for NIF. A comparison of modelled thickness to the ice core lengths (Thompson et al. 2002) results in a mean relative absolute error of 26%.

In the reconstructions for 2000 ~~W~~we assessed the utility of ~~lateral~~ this margin thickness information in constraining glacier ~~volume~~thickness by comparing our reconstructions to the recent global consensus estimate (Farinotti et al. 2019). For Kersten Glacier, we report significantly smaller thickness values as compared to the ~~current~~ consensus estimate. The latter reconstruction is shown to be inconsistent with the observed glacier separation between 2000 and 2011. For NIF, our reconstruction (Experiment 2) and the consensus estimate both show a very similar mean ice thickness, which is surprising as the consensus estimate was not informed by any thickness measurements.~~geometry,~~ ~~T~~the lateral glacier retreat information seems less utile as central ice thickness is strongly underestimated. Reasons for this worse performance might be the complex topography and the dynamic inactivity of NIF. We therefore speculate that thickness information from retreat is most useful in areas that have been dynamically more active in the past.~~dynamically more active areas.~~

~~Furthermore, 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. However, if the glacier topography is as complex and peculiar as for NIF, both mean viscosity values and generated margin ice thickness observations underestimate the overall glacier thickness. In the case of NIF, ice thickness based on observational data and ice thickness based on mean viscosity differed by a factor of 2. Based on results obtained for NIF and KG, the mean viscosity can be used in future studies to generate an estimate for the ice thickness of retreating glaciers without actual thickness observations. As ice thickness reconstructions on a global scale can only rely on thickness observations for a small percentage of all glaciers, the approach of using generic margin thickness observations, presented in this study, can~~

provide additional input for glaciers with visible retreat. The first validation of this generic-thickness approach on the glaciers on top of Mt. Kilimanjaro might have been influenced by the complex topography of NIF and the absence of ground-truth data on KG. However, ice thickness reconstructions were still comparable to results by Bohleber et al. (2017) and Farinotti et al. (2019), which supports the use of the approach with mean viscosity values as ice thickness input to adjust the viscosity locally. We imagine that the inference of margin thickness values for retreating glaciers is readily transferrable to many glaciers worldwide and it would provide a mean for glacier-specific calibration of reconstruction approaches on regional or global scales.

The unique glacier settings on Mt. Kilimanjaro are certainly not ideal for this first assessment of utilizing glacier retreat information to allow a glacier-specific calibration of thickness reconstruction approaches. In absence of ground truth data on KG, it remains unclear if the retreat information is best used spatially distributed or as a bulk average. Yet we can state that increased quality and resolution of more recent DEMs are key for capturing sharp transition zones (e.g. cliffs).

To conclude, glacier retreat is palpable all around the planet and it will continue in the future (e.g., Hock et al., 2019). As time progresses, the suggested strategy to infer past ice thickness values from multi-temporal satellite information will produce an increasing wealth of calibration data. Moreover, the approach is readily transferable and provides a means for glacier-specific calibration of reconstruction approaches on regional or even global scales.

Acknowledgement

CS received primary funding from the project BR2105 /14-1 within the DFG Priority Program "Regional Sea Level Change & Society". JJF was funded by the German Research Foundation (DFG) under grant number FU1032/1-1. Results presented in this publication are based on numerical simulations conducted at the high-performance computing center of the Regionales Rechenzentrum Erlangen (RRZE) of the University of Erlangen-Nürnberg. We would also like to thank Nicolas Cullen and Pascal Sirguey (both at University of Otago, NZ) for constructive discussions and for providing the KILISoSDEM.

Author contribution

CS led writing of the manuscript, in which she received support from all authors. The research aims and setup were developed in regular discussion with JJF, TM and MB.

Competing interests

The authors declare no competing interests.

References

Bohleber, P., Sold, L., Hardy, D. R., Schwikowski, M., Klenk, P., Fischer, A., Sirguey, P., Cullen, N. J., Potocki, M., Hoffmann, H. and Mayewski, P.: Ground-penetrating radar reveals ice thickness and undisturbed englacial layers at Kilimanjaro's Northern Ice Field, The Cryosphere, 11, 469–482, doi: 10.5194/tc-11-469-2017, 2017.

Braun, M.H., Malz, P., Sommer, C., Farías-Barahona, D., Sauter, T., Casassa, G., Soruco, A., Skvarca, P. and Seehaus, T.: Constraining glacier elevation and mass changes in South America, Nat. Clim. Change, 9, 130–136, doi: 10.1038/s41558-018-0375-7, 2019.

Cullen, N. J., Sirguey, P., Mölg, T., Kaser, G., Winkler, M. and Fitzsimons, S. J.: A century of ice retreat on Kilimanjaro: the mapping reloaded, The Cryosphere, 7, 419–431, doi: 10.5194/tc-7-419-2013, 2013.

Farinotti, D., Brinkerhoff, D. J., Clarke, G. K. C., Fürst, J. J., Frey, H., Gantayat, P., Gillet-Chaulet, F., Girard, C., Huss, M., Leclercq, P. W., Linsbauer, A., Machguth, H., Martin, C., Maussion, F., Morlighem, M., Mosbeux, C., Pandit, A., Portmann, A., Rabatel, A., Ramsankaran, R., Reerink, T. J., Sanchez, O., Stentoft, P. A., Singh Kumari, S., van Pelt, W. J. J., Anderson, B., Benham, T., Binder, D., Dowdeswell, J. A., Fischer, A., Helfricht, K., Kutuzov, S., Lavrentiev, I., McNabb, R., Gudmundsson, G. H., Li, H., and Andreassen, L. M.: How accurate are estimates of glacier ice thickness? Results from ITMIX, the Ice Thickness Models Intercomparison eXperiment, The Cryosphere, 11, 949–970, https://doi.org/10.5194/tc-11-949-2017, 2017.

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F. and Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth, Nat. Geosci., 12, 168–173, doi: 10.1038/s41561-019-0300-3, 2019.

Frey, H., Machguth, H., Huss, M., Huggel, C., Bajracharya, S., Bolch, T., Kulkarni, A., Linsbauer, A., Salzmann, N. and Stoffel, M.: Estimating the volume of glaciers in the Himalayan–Karakoram region using different methods, The Cryosphere, 8, 2313–2333, doi:10.5194/tc-8-2313-2014, 2014.

Fürst, J. J., Gillet-Chaulet, F., Benham, T. J., Dowdeswell, J. A., Grabiec, M., Navarro, F., Pettersson, R., Moholdt, G., Nuth, C., Sass, B., Aas, K., Fettweis, X., Lang, C., Seehaus, T. and Braun, M.: Application of a two-step approach for mapping ice thickness to various glacier types on Svalbard, The Cryosphere, 11, 2003–2032, doi: 10.5194/tc-11-2003-2017, 2017.

Hock, R., Bliss, A., Marzeion, B., Giesen R., Hirabayashi Y., Huss M., Radić V. and Slangen A.: GlacierMIP – A model intercomparison of global-scale glacier mass-balance models and projections., J. Glaciol., 65, 453-467, doi:10.1017/jog.2019.22, 2019.

Hutter, K.: Theoretical glaciology; material science of ice and themechanics of glaciers and ice sheets, D. Reidel Publishing Company, Tokyo, Terra Scientific Publishing Company, 1983

Huss, M., Farinotti, D.: Distributed ice thickness and volume of all glaciers around the globe, J. of Geophys. Research, 117, doi: 10.1029/2012JF002523, 2012.

Kamb, B. and Echelmeyer, K. A.: Stress-Gradient Coupling in Glacier Flow: 1. Longitudinal Averaging of the Influence of Ice Thickness and Surface Slope, J. Glaciol., 32, 267–284, 1986.

Kaser, G.: Glacier-climate interactions at low latitudes, Journal of Glaciology, 47, 195–204, 2001.

~~Kaser, G., Hardy, D. R., Mölg, T., Bradley, R. S. and Hyera, T. M.: Modern glacier retreat on Kilimanjaro as evidence of climate change: observations and facts, Int. J. Climatol., 24, 329–339, doi: 10.1002/joc.1008, 2004.~~

Mölg, T., Cullen, N. J., Hardy, D. R., Kaser, G. and Klok, L.: Mass balance of a slope glacier on Kilimanjaro and its sensitivity to climate, Int. J. Climatol., 28, 881–892, doi: 10.1002/joc.1589, 2008.

Mölg, T., Cullen, N. J., Hardy, D. R., Winkler, M. and Kaser, G.: Quantifying Climate Change in the Tropical Midtroposphere over East Africa from Glacier Shrinkage on Kilimanjaro, J. Clim., 22, 4162–4181, doi: 10.1175/2009JCLI2954.1, 2009~~a~~.

~~Mölg, T., Cullen, N. J. and Kaser, G.: Solar radiation, cloudiness and longwave radiation over low-latitude glaciers: implications for mass-balance modelling, J. Glaciol., 55, 292–302, 2009b.~~

Mölg, T. and Hardy, D.R.: Ablation and associated energy balance of a horizontal glacier surface on Kilimanjaro, J. Geophys. Res., 109, D16104, doi: 10.1029/2003JD004338, 2004.

Pattyn, F.: A new three-dimensional higher-order thermomechanical ice sheet model: Basic sensitivity, ice stream development, and ice flow across subglacial lakes, J. Geophys. Res., 108, 2382, doi: 10.1029/2002JB002329, 2003.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0, Global Land Ice Measurements from Space, doi: https://doi.org/10.7265/N5-RGI-60, 2017.

Sirguey, P. and Cullen, N. J.: Surveying from Outer Space. KILISoSDEM2012: a very high resolution DEM of Kilimanjaro, Survey Quarterly, 76, 5–7, 2013.

Thompson, L. G., Mosley-Thompson, E., Davis, M. E., Henderson, K. A., Brecher, H. H., Zagorodnov, V. S., Mashiotta, T. A., Lin, P.-N., Mikhalenko, V. N., Hardy, D. R. and Beer, J.: Kilimanjaro ice core records: evidence of holocene climate change in tropical Africa, Science, 298, 589–593, doi: 10.1126/science.1073198, 2002.

USGS: Shuttle Radar Topography Mission Void Filled, doi: /10.5066/F7F76B1X, 2014.

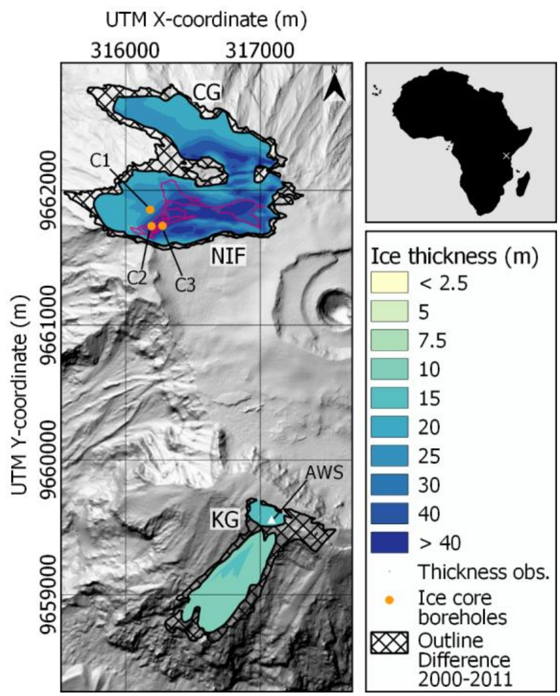

Figure 1: Reconstructed ice thickness (m) for Northern Icefield (NIF) and Kersten Glacier (KG) for the year 2011, based on thickness observations (NIF) and mean viscosity (KG) with a model resolution of 2 m (Experiment 3). The magenta path on NIF represent the GPR ice thickness measurements by Bohleber et al. (2017). Orange dots (C1-C3) indicate the drill locations of the ice cores from Thompson et al. (2002). The AWS location from Mölg et al. (2009) is marked by the white triangle. The background layer is the hillshade ofBackground: KiliSoSDEM hillshade. The overview map depicts the location of Mt. Kilimanjaro (white cross) on the African continent.

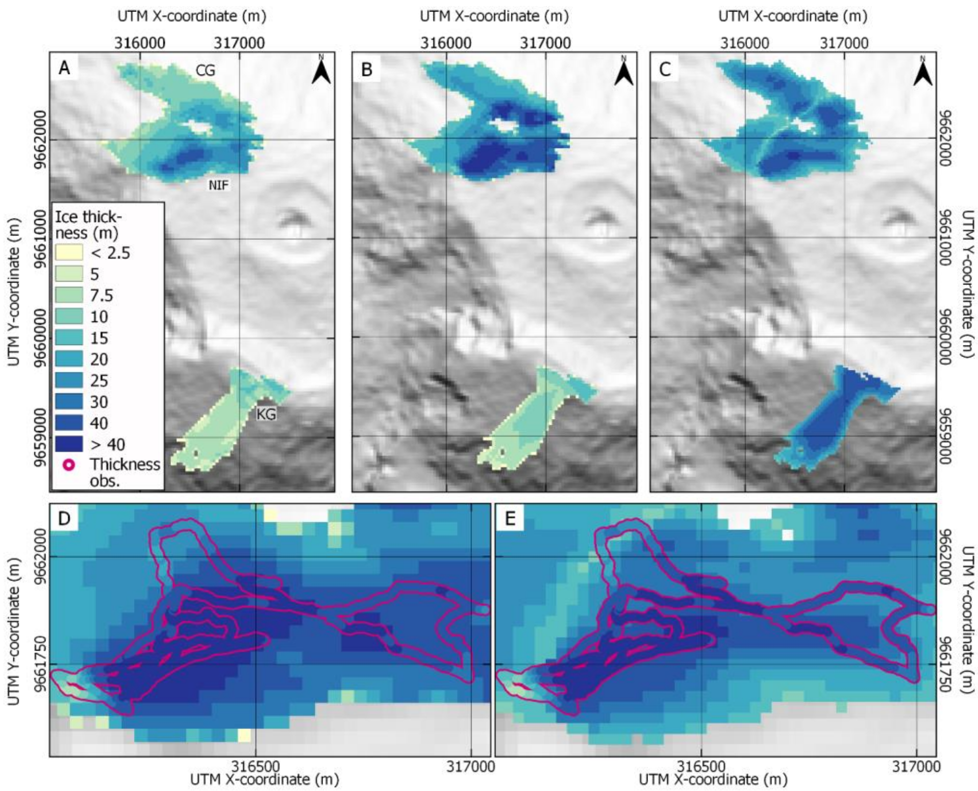

Figure 2: Reconstructed ice thickness (m) for Northern Icefield (NIF) and Kersten Glacier (KG) for the year 2000.~~,~~ Panel (A) shows results for Experiment 1 making exclusive use of past thickness information in nowadays ice-free areas. Panel (B) presents results from Experiment 2, which uses a bulk viscosity inferred from the retreat information on KG, whereas for NIF, the reconstruction is only calibrated by in-situ GPR measurements (see Table 1).~~based on thickness observations (A; Experiment 1) and based on mean viscosity (B; Experiment 2).~~ As comparison panel (C) depicts the composite ice thickness from Farinotti et al 2019. Panel (D) shows a closeup of NIF, overlying the thickness map of Experiment 2 with the GPR thickness measurements (colored dots with magenta outline, showing measured values in the same colourbar) by Bohleber et al. (2017). Panel (E): same as Panel (D) but showing the consensus thickness map (Farinotti et al., 2019; cf. Panel C). ~~Orange dots (C1-C3) in (A) represent the drill locations of the ice cores from Thompson et al. (2002). The background is the hillshade of the~~ Background: SRTM DEM hillshade.

Table 1: ~~Lookup table for the e~~Experimental setup of the ice thickness reconstruction (Section 3.5).

	Representing the glacier state for the year	**Thickness input KG**	**Thickness Input NIF**	**Surface Elevation Information (year of acquisition)**	**Glacier Outlines (acquisition date)**	**Mean ice thickness KG (m)**	**Mean ice thickness NIF (m)**
Experiment 1	2000	Generated margin thicknesses	Generated margin thicknesses	SRTM DEM (2000)	Randolph Glacier Inventory 6.0 (2000/02/21)	6.2	13.7
Experiment 2	2000	Mean viscosity	Observations from Bohleber et al. (2017)	SRTM DEM (2000)	Randolph Glacier Inventory 6.0 (2000/02/21)	6.9	23.4
Experiment 3	2011	Mean viscosity	Observations from Bohleber et al. (2017)	KILISoSDEM (2012)	Digitized from Landsat 5 Image (2011/08/22)	9.3	26.6

Formatierte Tabelle

hat formatiert: Schriftart: Nicht Fett