Point by point response to the reviewers

We would like to thank you for your constructive feedback on our manuscript. We have done our best to address the points raised by the reviewers and hereby submit a new version of the manuscript together with a point by point reply to each of reviewers' comments. After answering questions that were mentioned from both reviewers, we provide a point by point answer to the reviewers' comments. ("RC" stands for "reviewer comment", "AR" for "authors response")

Synthetic experiments

After going trough the reviews, we realized that most points raised by both reviewers lead back to the design of the synthetic experiments (on which our paper is based) and a better explanation of their benefits is required to make the the relevance of our study more clear.

Reconstructing past glacier states is a complex inverse problem, and the results will depend on (i) the uncertainties involved in the boundary conditions (climate, glacier bed, etc.), (ii) the uncertainties in the glacier model itself (as pointed by reviewer #2), and (iii) a theoretical lower bound (termed "reconstructability" in our manuscript and associated to the concept of "response time" by reviewer #1) tied to the characteristics of the glacier itself (slope, size, the past climate, etc.) as well as how much we know about today's glacier and how far in time we attempt to do our reconstruction. In our opinion, addressing all three issues at once is too much for a single study.

The main point of the synthetic experiments is to separate these issues from one another, and to focus on point (iii) only. Thanks to the synthetic experiments, we are able to isolate and understand the limitations and errors of our method itself, as opposed to uncertainties that derive from unknown boundary conditions and model parameters. They also allowed us to detect what kind of observations are necessary to reduce the uncertainties of our reconstruction (for example, that length observations are less suitable than area, or better: topography).

That being said, we agree with the points raised by both reviewers and the editor, and modified our manuscript by explaining our objectives more clearly and adding a new results section and more discussions. Importantly, we now better show that although theoretical, our study also has real-world applications by providing insights about the glacier themselves. We learn from our study what makes a glacier "reconstructable", and discuss how this knowledge can be transferred to the real world. Our synthetic experiments are designed to have a certain similarity with the real world, so that we can expect to learn something meaningful from them: a point which we didn't explain clearly in our first manuscript and we hope to have improved now.

Minor changes to the method

During the review process, we realized that our mass balance model did not perform as good as it could have. We used OGGM default values for mass balance specific parameters, which were calibrated with a global climate dataset (CRU). We now use a new parameter set which was determined by a cross-validation for the Alps only (41 reference glaciers) using HISTALP data. We updated this parameter set to: precipitation scaling factor $p_f = 1.75$ (instead of 2.5), melt temperature $T_{Melt} = -1.75$°C (instead of -1), liquid precipitation temperature $T_{Liquid} = 2.0$°C (unchanged) and temperature lapse rate $\Gamma = -6.5$ K km^{-1} (unchanged). The new mass balance calibration lead to some changes, as the model is sensitive to this choice. First, the glaciers' response times increased a little bit. We therefore had to extend the 400 years random climate runs to 600 years, to ensure that for every case t_{stag} can be properly determined. Also, the number of experiments that fulfill the area threshold of 0.01km^2 is enlarged to 2660 glaciers thanks to this change. We updated all figures and numbers in the manuscript without any significant change to our conclusions.

1 Reply to Reviewer #1

"RC: *In this manuscript, Eis et al. use a numerical approach to assess how well past glacier geometries can be reconstructed by relying on the present-day glacier geometry. For this purpose, they utilize the Open Global Glacier Model (OGGM), which is a state-of-the art glacier evolution model that has the capacity to model a large ensemble of glaciers (Maussion et al., 2019). The OGGM is used to simulate the transient evolution of > 2000 glaciers in the European Alps between 1850 and present-day, after which they are compared to observed geometries (or rather synthetic geometries close to these). In many cases, different initial geometries lead to an almost identical present-day state (i.e. various initial states lead to unique present-day geometry). The authors also show that when using the entire information about the present-day geometry, the uncertainty on past conditions reduces compared to more simplified approaches in which only the present-day length is considered. I must say that I was very enthusiastic when starting to read this manuscript, but that in the end I have several questions - some more substantial than others. On the hand, I think the idea is very interesting, the model is the correct one to tackle this particular problem, and the presentation of the results is neat: the text is generally easy to follow, and so are the figures. On the other hand, I have some reservations concerning the experimental setup (with a largely theoretical focus, but no incorporation of real data/observations) and the conclusions drawn from this. I have detailed on this in the next section, and hope that the authors will be able to address (some of) the issues raised. There may be some elements/passages that I may have misunderstood, and on which I would gladly be corrected, but in that case I am afraid they may also be problematic to understand for some other readers.*"

AR: We would like to thank you for your detailed feedback and the many suggestions to improve our study. We hope that the following answers and changes in the manuscript sufficiently address your reservations concerning the experimental setup.

General comments

"RC: *When going through this manuscript, the first thing that popped up in my mind is: 'these experiments are all about glacier response time'. Also when reading the entire manuscript, this idea persisted: this is a response time story! I was therefore rather surprised to not see any discussion on this, or not even having it mentioned anywhere. In the end - to me - it boils down to: you can say something about the past glacier geometry (when considering the present-day geometry) over time periods that are* *close to or shorter than the glacier response time (depending on which definition is used for the response time). Or formulated differently: the present-day geometry does not depend on the past glacier geometry when considering time periods that exceed the glacier response time. As the response time of Alpine glaciers is typically in the order of years to decades (e.g. Haeberli and Hoelzle, 1995; Oerlemans, 2007; Zekollari and Huybrechts,2015) it is difficult to picture how the present-day geometry (or a simulated geometry resembling this) can be used to say something about the glacier geometry in 1850.* "

AR: Indeed, we discussed the response time topic during the development of our method and manuscript. And yes, we agree that our study indirectly deals with response times. The main issue we see with this concept is that it is not well-defined and often causes misunderstandings. The most popular quantifiable value is the response time as defined by the e-folding time response to a step function (e.g. Oerlemans, 1997). This value depends on both the glacier state and the step change in climate, and will change over time.

In order to quantitatively address your hypothesis (*"the present-day geometry does not depend on the past glacier geometry when considering time periods that exceed the glacier response time."*), we defined a new non-dimensional measure (termed "reconstructability") which assesses how easy or hard it is to find the correct past geometry. We now quantify and discuss which parameters influence this reconstructability, and among other things discuss the response time as requested. However, note that the present-day geometry is not the only information entering the reconstruction, we also provide the model with information on the past climate evolution, which helps to constrain the reconstructions.

"RC: *Continuing on the above point, do you think that what you derive as the past geometry in 1850 is realistic and that for cases where a non-unique answer (i.e. a non-unique glacier geometry) arises for the present-day: that the 'best' 1850 geometry that you obtain is really an indication of the past geometry?* "

AR: We believe that there may be a slight misunderstanding here. If by "realistic" you mean "in the real world", we would like to refer to our explanation of how we designed the synthetic experiments below. If you mean, "realistic with respect to the true 1850 state as defined in our synthetic experiments": yes we do! Obviously, we can trust the 1850 glacier geometry of a unique case more than the 'best' candidate in a non-unique case. But with the help of the error analysis in Sect. 4.1, we can prove that our 'best' candidate performs well despite of being slightly different than the true state. Especially when we run this state some years forward, the error converges to a negligible range. So in a case with non-unique solutions in 1850, we always should keep in mind that all acceptable states result in the present-day geometry and it might be better to "spin-up" the 'best' candidate some years for a better confidence.

"RC: *Are there not other model uncertainties that play a bigger role?* "

AR: For real world applications, definitely, but not in our experimental set-up. This is precisely what motivated our choice for a synthetic environment in the first place: since the system is known perfectly, and our possible glacier evolution scenarios depend only on their unknown initial state (and not on model or boundary condition uncertainties), we can address the research questions formulated above and in the introduction of the manuscript.

For real-world applications, model uncertainties will have to be accounted for, and will have to be compared to (and added to) the theoretical lower bound defined in this study. We addressed this question by adding this point to the synthetic experiment section and to the conclusion.

"RC: *You mention that you cannot perform tests with real present-day geometries. When reading the manuscript, it does not entirely become clear to me why that is. Could you elaborate on this?* "

AR: Present-day real-world geometries are the result of a series of processes and past-climate that are largely uncertain. In our paper, we remove these uncertainties and ask the question: even if we know *everything* about the system , how constrained are the past geometries?

This said, it is of course our intention to make the reconstruction method applicable to the real world. We tried to clarify our intentions and remove the statement that the method cannot be applied to real-world glaciers in the revised manuscript.

"RC: *It would have made it really interesting if you could have worked with real present-day geometries and performed your simulations based on this, which I was in fact what I was expecting... e.g. (1) reproduce geometries at the end of the LIA and compare these to real geometries at that time or (2) for instance compare the length changes modelled between 1850 and the present-day with observed length changes over this time period (e.g. from Leclercq et al., 2014). Such a validation would really have been of great value here, and would probably be the best way to increase our confidence in the applicability of the method you propose.* "

AR: We apologize for misleading your expectations, and hope to have clarified the main purpose of our study with this revision. A comparison with real-world data does not make sense in the context of the synthetic experiments, and we believe that it should be part of a dedicated study that addresses the reconstruction uncertainties that are caused by uncertainties in the boundary conditions and by the model/parameter uncertainties. At the time of writing, we are confident that our method can eventually be applied to any land-terminating glacier, and we do evaluate the method against observational data as you suggest. However, as described above, this comes with a whole new set of problems, which need to be addressed in a separate manuscript.

"RC: *1. Why not invert the setup? Could potentially be very interesting: for every individual glacier the question could be asked 'how far can you can back in time to have an initial geometry that still determines the present-day glacier geometry?': this is a kind of response time for every glacier. This information could be used to infer something about the response time of Alpine glaciers, even in the rather theoretical framework that you are using so far (with little to no observational data).. Again, I understand that this would be a considerable amount of work at this stage, so I am not saying that this has to be included, but I think it would be a nice addition, which would be more valuable for the reader (vs. the so-far used rather technical setup..)* "

AR: Thank you for this idea! We tested it (see below) and found that while this inverted setup is computationally very expensive, unfortunately it doesn't lead to improved results: we applied our method to different starting times (1850,1855,...1965) and based on this, one can see how far one can go back in time to get a good initial state for this glacier. See Fig. 1 (of this response) for three different examples. For each tested starting year, we determined the median state and conducted an uncer-

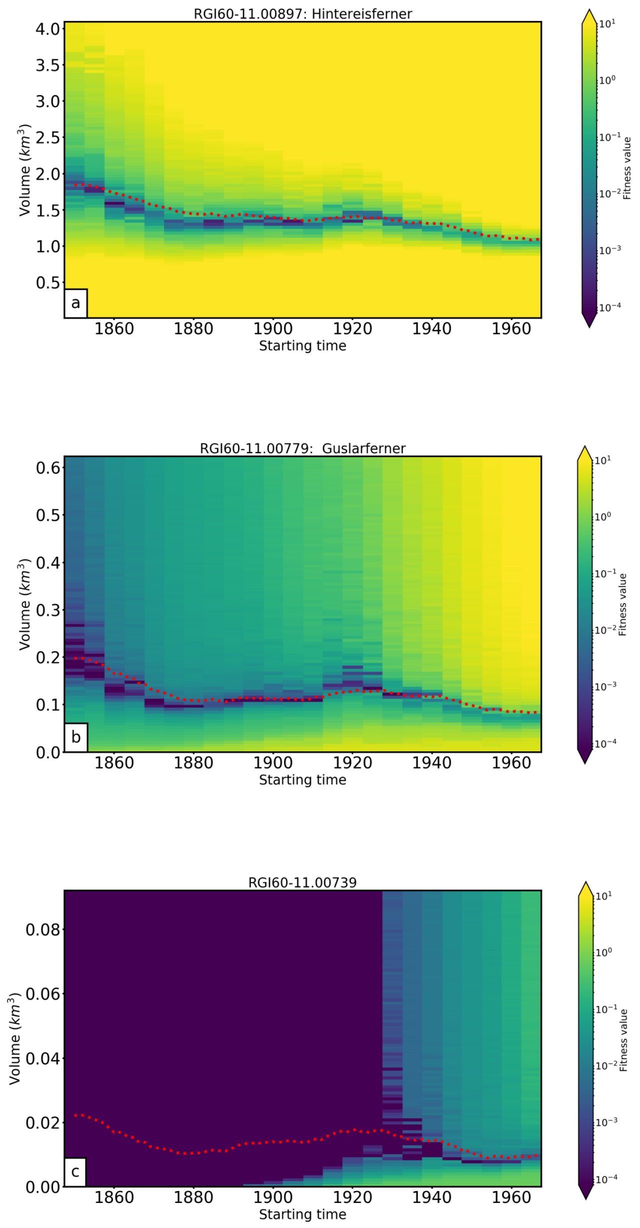

Figure 1. Reconstructability for different starting times. Colors indicate the fitness value of a simulation initialized with a glacier volume indicated by the vertical axis at a time indicated by the horizontal axis. Red dotted line shows the synthetic experiment. Upper panel: example for a glacier with high reconstructive power; the “observed” glacier state in 2000 constrains the past evolution well into the 19 th century, and the reconstruction is close to the goal. Middle panel: example for a glacier with low reconstructability before approx. 1920 and high reconstructability afterwards. Lower panel: example for a glacier with very low reconstructive power; the ”observed” glacier state does not constrain the past glacier evolution before approx. 1930.

tainty analysis (similar to the one in the manuscript). We find that the uncertainties of the median states at the different starting points are higher than doing the initialization for the year 1850 (only) and running this state forward in time. While this is counter-intuitive, the main reason is that by starting in 1850 even with a very large number and range of candidates, the very unrealistic ones are quickly forced to converge by the boundary conditions (i.e., by climate), effectively reducing the number of potential candidates for a later date. In other terms, we make use of our knowledge about past climate to reduce the number of candidates at each later stage. In real-world application, results might be different since uncertainties in past climate are large. We will explore this further, but because of the computational cost, it is hard to imagine an eventual applicability on the global or even large regional scale.

"RC: *In whichever way the manuscript/experiments are reorganised: a discussion on the role of the response time seems crucial!* "
AR: We agree. We added this discussion to the new section about reconstructability and to the conclusions as well.

"RC: *Would really be rewarding to have the whole story a bit less theoretical and more applied to the real world. You model the glaciers in the European Alps, for which there is an amazing dataset on past changes* "
AR: We hope that our revised formulation of the research questions better addresses how we can learn about the real world from our setup, even though our experiments are synthetic (see Sec. 4.3 in the manuscript)

Specific and technical comments

Abstract

"RC: *p.1, l.7: 'alpine'. Is the term 'alpine' not referring to the type of glacier (i.e. a mountain glacier) instead of the region ('Alpine')? See e.g. nsidc.org/cryosphere/glossary/term/alpine-glacier . From my understanding, here, and throughout the manuscript, you want to refer to 'Alpine' glaciers?* "
AR: We agree. We changed the term throughout the manuscript.

"RC: *p.1, l.13-15: the problem does indeed turn out to be 'non-unique' in many cases. As said earlier, even for the cases where a small difference exists in the modelled present-day glacier, I am not really convinced if this can really be seen as an indicator for how the glacier was in 1850... Could be valuable when considering shorter time periods between model initialization and present-day (e.g. multidecadal timescale), but not convinced over time periods > 100 years... See also suggestion 1 at the end of the 'General comments' section.* "
AR: As explained above, we don't want to imply that our reconstructions are valid for the real world. However, in the context of response time, we would argue that in particular reconstructions exceeding the response time can be expected to work better, because in this case, the climate forcing will have stronger a impact on the evolution of the glacier geometry than the initial geometry. Or, to put it differently: A glacier with a short response time will have "forgotten" the initial conditions quickly, and

the reconstructed evolution will be a result of climate over a large extent of the reconstruction; a glacier with a long response time will be influenced by the initial conditions for a longer time period, thus reducing the fraction of time where it is governed by the known climate.

Introduction

"RC: *p.1, l.18-19 and/or p.2, l.2-3: could consider adding some new references here: Zemp et al. (2019) and Wouters et al. (2019)* "
AR: We added the two references.

"RC: *p.2, l.12-13: from this: sounds like in most numerical experiments the starting point is an observed geometry, which is typically not the case (e.g. due to problems related to model drift,...etc) (e.g. Goelzer et al., 2018, which you mention later). Could consider reformulating this.* "
AR: We agree. We have reformulated this to make clear, that this is valid for models that derive the initial surface hypsometry from DEM's and outlines.

"RC: *p.2, l.21-23: do not know the details about this study, but I am surprised that unique glacier area in the past can be used to have correct glacier area at present over such long time periods. Is this an artefact of using the V-A scaling method? Can imagine that in reality, can build up same glacier when starting from two very different states in 1850 (e.g. from zero ice thickness in 1850 and when using the present-day geometry as the 1850 state). Could you comment on this here?* "
AR: On p.4, l.31 we wrote that "the backwards reconstruction is impeded by the non-linear interaction between glacier geometry, ice flow and mass balance." The non-uniqueness in our case basically stem from the SIA equation, on which OGGM is based. This equation can be transformed into a diffusion equation, for which the backwards problem is ill-posed (e.g. because of the non-uniqueness). So yes: The V-A scaling method is not based on a diffusion equation. Thus, it does not lead to non-unique reconstructed glacier areas in the past (although this has not been proven unequivocally in Marzeion et al., 2012).

"RC: *p.2, l.30-31: starting in 1990 for all simulations is indeed somewhat arbitrary. Your setup could potentially be used to suggest a good starting point for every individual glacier, which could be a few decades in the past (suggestion 1 in 'General Comments' section). But again, not convinced you can go back to > 100 years when it comes to deriving past glacier information from the present.* "
AR: Yes, this would be possible. We add this to the discussion.

"RC: *p.2, l.32: 'most work focuses on estimating the present-day state of ice sheets'. Is a bit strangely formulated, as you make it sound like the main goal of ice sheet modellers is to correctly represent the present-day geometry of an ice sheet. Would rather say that their work is focused on making accurate projections of future ice sheet change, for which accurate reconstructions of the present-day state are crucial.* "

AR: We agree, we rephrased the sentence.

"RC: *p.2-3, l.34-35 - l.1: odd sentence. Consider splitting up / reformulating?* "

AR: We split the sentence into two.

The Open Global Glacier Model

"RC: *p.3, l.25: resolution varying from 10 to 200 meters. Which criterion is used to decide which horizontal resolution to use for a given glacier?* "

AR: We here use OGGM's default: $dx = \sqrt{aS}$ with $a = 14$ and S the area of the glacier in km^2. We added this information in the text.

"RC: *p.4, l.9: 'realistic climate settings': what do you mean with 'realistic' here?* "

AR: We mean that the forcing of the climate setting is realistic (because it comes from real climate data, but the order is changed). We rephrased the sentence.

Problem description

"RC: *p.5, l.1-2: 'such that the forward modelled state is as close as possible to the observation': OK. Here I get a bit lost however. If I understand it correctly from reading the text further, in the end you try to model a present-day state that is as close as possible to the observation, but for this purpose, as a reference/observed state, you do not use the observed present-day glacier, but a state resembling this ('observed state' of the synthetic experiments). Correct? If so, can you explain why that is? I seem to have missed this piece of information, and find it quite confusing. If not, I am still afraid that it will not be easy to understand for other reader. Potentially consider reworking this.* "

AR: We here mean an observation in general (either the present state or the observed synthetic experiment). But we agree that this can be confusing for the reader. We have reworked the synthetic experiments section to avoid confusion.

"RC: *p.5, l.11-12: 'non-unique': of course $\rightarrow$ related to the fact that the time period (1850 - present-day) considered exceeds the response time of this individual glacier. By here, would have expected to have the response time mentioned somewhere...* "

AR: The non-uniqueness stem from the ill-posedness of the backwards SIA/(diffusion problem) or in other words: because the reconstruction is impeded by the non-linear interaction between glacier geometry, ice flow and mass-balance. Please also note (as mentioned above) that reconstructions exceeding response time are potentially possible through the availability of information on past climate.

Synthetic experiments

"RC: *I get somewhat lost here (in general in section 2.3 and 2.4).. See first comment on previous section: is the real present-day geometry used as the 'observed state': why (not)?* "

AR: We clarified these terms in the manuscript. In general one could say that the method is described for an observation in general (either a real present-day geometry or the synthetic experiment). The method is also implemented in that way. The synthetic experiment in 2000 is then treated as an observation. But we will make this more clear to the reader.

"RC: *p.6, l.1-2: temperature bias of -1 K: seems rather arbitrary. For which other values was it tested? What is influence if other value would have been chosen?* "

AR: Setting the temperature bias to -1K is not completely arbitrary. To justify our choice, we tested different temperature biases for the generation of the synthetic experiments. The results are summarized in Fig. 2:

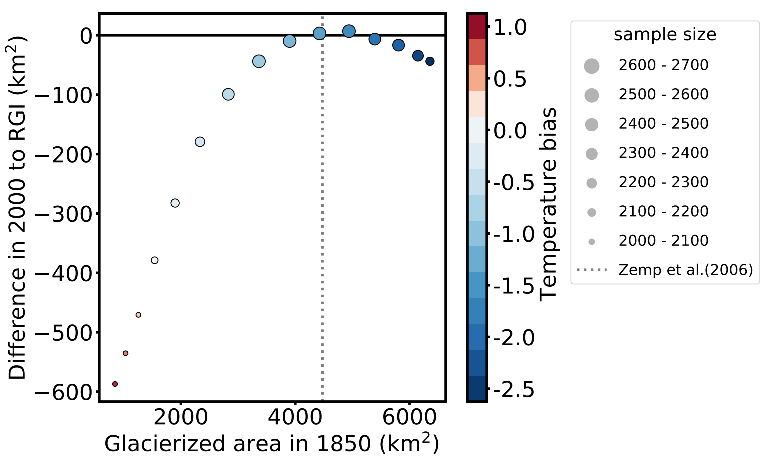

Figure 2. Difference between the total area in 2000 to the total area from the RGI plotted as a function of total area in 1850. Colors mark the applied temperature bias to create the synthetic experiments, and the size of the points mark the sample size (number of glaciers with an area larger than 0,01 km^2 in 2000). The dashed grey line marks the estimated total area of all Alpine glaciers in 1850 from (Zemp et al., 2006)

This figure shows that applying positive or small negative temperature biases to the synthetic experiments results in large area differences to the RGI in 2000, and the total glacierized area in 1850 is also too small. The sample size is reduced, because less glaciers fulfill the area threshold criteria of 0.01 km^2. Negative temperature biases that are too large also reduce the sample size, because some runs fail ("Glacier exceeds boundary", this means the glacier would get larger than the underlying grid). The experiments with a temperature bias of -1K, -1.25K or -1.5K perform best regarding the area difference to the RGI in 2000. But only the experiment with the temperature bias of -1K performs good regarding the estimation in 1850 of Zemp et al.

(2006), whereby it needs to be taken into account that the dots only represent a subset (the small glaciers in 2000 are missing) of the glaciers considered in (Zemp et al., 2006). We added this results to Appendix A.

"RC:*p.6, l.4: Guslarferner. Please state where this glacier is located. Why are these two Austrian glaciers (together with Hintereisferner) considered and not others? As no data is used for evaluation/calibration, it seems that any glacier could have chosen (i.e. also glaciers that are not monitored / glaciers that are less well known).* "

AR: We added information, where the two example glaciers are located and why we picked them as examples. And of course, different glaciers could also be chosen. For the uncertainty analysis, we run each of the 2660 glaciers and we produce the same plots for all. But for a more representative set of glaciers, we added several new examples (15) as new supplementary material.

Reconstruction of initial glacier states

"RC: *Figure 2: from panel a, the temperature bias seems to be varying between -3.0 and 1.95 K. But in the legends a range from -2.65 to 2.95 K is mentioned. Should this be the same or am I misunderstanding this?* "

AR: Yes, they should be the same. During the development of the manuscript, we changed the example glacier in this Figure and we forgot to change the numbers in the description. We corrected them.

"RC: *p.7, l.11: 'where all trajectories have reached this stagnation period': could you be more specific? How do you define stagnation?* "

AR: During the stagnation period the glacier volume does not increase or decrease strongly in comparison to the total volume since the beginning of the simulation. We added this information to the text.

"RC: *p.7, l.11-12: linking the 'stagnation period' (somehow related to glacier response time) to glacier size → questionable. Not always the case that 'longer/larger glacier has got a longer response time' (e.g. Leysinger Vieli and Gudmundsson, 2004; Bahr et al., 1998; Pfeffer et al., 1998; Raper and Braithwaite, 2009; Oerlemans, 2012). It is OK to take the largest glacier to determine the stagnation period, but would be careful with the statement linking the response time to the glacier size/thickness (Jóhannesson et al., 1989)* "

AR: We agree. We removed the link to Jóhannesson.

"RC: *p.8, l.1: smoothing. Over which time period?* "

AR: 10 years, we added this information to the text.

"RC: *p.8, l.3-4: 'equilibrium with the climate around 1850': questionable, as many Alpine glacier were still advancing and reached a maximum extent later, while others were already retreating by then (Leclercq et al., 2014). Not a major issue, but would be good if could shortly discuss this assumption.* "

AR: We agree. To a certain extent, the synthetic experiments simulate this situation since the target states are created from a randomly fluctuating climate, and therefore some of the glaciers are advancing while others are retreating. We cannot really assess if these advances are comparable to those observed in the real world after 1850. We added a mention to advancing glaciers in the synthetic experiments description section.

"RC: *p.8, l.10: 'same model parameters set': could you provide a bit more details here. For instance, how is the ice rheology described (deformation/rate factor) for every glacier. How is this determined/tuned? Is this the same for every glacier? And what is the effect of this on your results?* "
AR: We added more information about the dynamical parameters used in this study (which are the same for every glacier) at the end of Sect. 2.1, but we refer at the same time to (Maussion et al., 2019). There you can also find a discussion about the model sensitivity to these parameters (for the Hintereisferner and on a global scale). As the model is sensitive to these parameter settings, a change consequently would also lead to different ice volumes and response times in this study.

"RC: *p.8, l.15: 'sorted by their fitness value': what do you mean by this?* "
AR: For each glacier we return a DataFrame (similar to a table), containing information about each glacier candidate that was evaluated. This DataFrame is sorted by the fitness values of the candidates. (The ones with the lowest fitness values come first). We reformulate this sentence.

Test site and Input Data

"RC:*p8, l.20: 90 m resolution DEM: this seems to be relatively low, especially given the fact that you consider glaciers as small as 0.01 km^2 (1 grid cell at this resolution).* "
AR: We agree that in these cases the resolution is low, but this is the standard procedure of OGGM and similar models, and is driven by data availability. The 90 m resolution DEM is interpolated to higher resolution grids (down to $dx = 10$m for small glaciers), leading to a smooth field. To make sure that the glacier outline touches more than one grid cell of the DEM, OGGM test if the minimum and maximum value are equal. If they are and error will be raised and this glacier can not be modelled.

"RC: *p.8,l.22: RGI date:2003. Is the case for most glaciers in RGI6.0 (those derived by Paul et al.2011), but not for all*
AR: We changed this.

"RC: *p.8, l.24: '5 minutes resolution'. What is this approx. in meters here?* "
AR: This correspond to a resolution of approx. 9,3 km. We added this information to the manuscript.

"RC: *p.8, l.27: 'threshold for the RGI': for this region? Not sure, but thought this was region-dependent?* "
AR: The threshold is the same for all regions. You can find this information in the RGI 6.0 Technical Report (e.g. page 18, Section 3.4 "Quality Control", step 2)

Results

"RC:*p.9, l.7: non-uniqueness for Guslarferner: as expected, given the fact that response time of this glacier is likely much shorter than the period between 1850 and present day... Do not think you can really say anything about glacier state in 1850 from these experiments, as you also point out. But also for Hintereisferner (p.9, l.25-...): not convinced that the narrower set of 1850 sets having a lower 'fitness value' are indicative for how the glacier was in 1850.Intuitively, expect that you can say something about past Hintereisferner state (from present-day state) for max. a few decades back in time (1950 maybe?)* "
AR: Here, we disagree. First, we think that the concept of "response time" cannot be used to discard any climate signal longer that a few decades: Hintereisferner certainly needed more than 50 years to grow as big as it is today (i.e. today's state does contain information about past states, at minimum by discarding obvious impossible candidates). More importantly, the narrower set of 1850 sets having a lower fitness value is an indication for how the glacier was in 1850, at least in our synthetic experiments - for which we know the "truth", and are able to assess the accuracy of the method. In the case of the Guslarferner, the response time is shorter than in the case of the Hintereisferner. All acceptable states need then to be considered as possible solutions, and the median state is only a good representative of this set. The uncertainties definitely need to be taken into account for these cases, but running the median state some years forward reduces the uncertainties rapidly (as shown in Fig. 7-10). However, we agree that (as explained above), uncertainties in past climate conditions as well as uncertainties of the model and its parameters will increase the reconstruction uncertainty in a real-world application, such that the good reconstructabilities obtained here cannot be expected to be carried over to real-world conditions.

"RC: *To be more convincing, would really be rewarding to compare these 'best' estimates for the 1850 state with observations. And this ideally for a large set of glaciers. Think this could work, also for the length changes between 1850 and present-day, even when considering the fact both states (1850 and present-day) are synthetic → are the length changes close to observations for your 'best' results (i.e. closer when considering the 1850 states that are not being considered as 'good')?* "
AR: The synthetic experiment length changes should not be compared to observations. The 1850 glacier geometry, depends strongly on the observation (here the synthetic experiment) as the candidates are evaluated by the difference to the observation. As the synthetic experiment in 2000 differ to the real-observation, a different trajectory is declared as the best candidate. This trajectory can also differ in gradient. Such a comparison only makes sense, if the candidates are evaluated based on real-world data. It will be done in a follow up study, where we address the uncertainties caused by the model and the boundary conditions.

"RC: *p.9, l.10: '...200 candidates has a...' → '...have a...'* "
AR: Done.

"RC: *p.9, l.16-17: 'in close proximity to the synthetic experiments': OK. Because do not compare to the real observation. Is it not possible to work with real observation? To get match here, would probably have to tune some model parameters. Is* *this the problematic aspect of working with observed present-day geometry? (seems to be the case when reading the 'Discussion and conclusions' section)* "

AR: Yes, the model calibration (mostly for the sliding parameter, the creep parameter and the consideration of lateral drag) also play important roles. This will have to be addressed in a real-world application, but does not influence the results of the synthetic experiments presented here.

"RC: *p.9, l.18: 'the observation': which is also modelled, right?* "

AR: Yes, we clarified this in the text.

"RC: *p.9, l.25: results are different: quite trivial. Would reformulate this or simply omit this* "

AR: We rephrased this.

"RC:*p.9, l.26: only few with small fitness value: see first comment on this section* "

AR: See answer to first comment on this section.

"RC:*p.9, l.27-28: 'need more time to adapt' $\rightarrow$ response time!* "

AR: See answer to your first comment (General comments)

"RC: *p.10, l.1: '11.7 to 12.4 km': from this: deduce a retreat of ca. 5 km. Is this the case? Would be really interesting to compare. For this glacier and for many others. Would make story much stronger.* "

AR: Please refer to the comments to the real-world applications before.

"RC: *p.10, l.6: 'we were able to show': personally not really convinced at this point unfortunately. See suggestion about incorporating 'real' data (observed past states and observed (length) changes)* "

AR: We wrote in the text: "we were able to show, that our method is able to determine the state in $t_0 = 1850$ of the synthetic experiment." and we still think that we could show this very well for the synthetic experiment case.

"RC: *p.10, l.7-8: combining with climatic information. How big is the role of the past climatic information on your results? Does the choice of past conditions (the conditions that you impose) affect the 'best' past states?* "

AR: This is an interesting question. The conditions that we impose are not completely arbitrary, as they are taken from HISTALP. But it is a valid question to ask if different past climate conditions (e.g. a climate which is not getting warmer but colder) would affect the accuracy of the method. This question is a perfect example for the usefulness of synthetic experiments: they would allow to address this question (as well as other factors, such as the influence of parameter uncertainty), and could be the subject of a future study. We added a sentence discussing this question in the conclusions.

"RC: *Figure 9+10: number of glaciers mentioned in figure title = 2619 vs. 2621 before. How come?* "

AR: In Figure 10, we show the relative error. To calculate the error, we need to divide by the volume of the experiment glacier. Here it happens for two glaciers, that the volume of the experiment was close to zero at some time t. That's why these glaciers are not included in this plot. After updating the mass balance related parameters (see General comment2), this issue does not occur any more and the numbers now coincide with each other.

"RC: *p.15, l.1: 'This holds also true': strange formulation. Consider reformulating* "

AR: Done.

Hardware requirements and performance

"RC:*p.15, l.12: 'influence strongly' → 'strongly influences'* "

AR: Done.

Discussion and conclusions

"RC: *p.16, l.3-4: 'identify errors...introduced by model approximation': OK, but this is not very satisfying for the reader and makes the paper almost purely theoretical...* "

AR: The paper presents a new method and thanks to the synthetic experiment world, we are able to determine uncertainties from the method itself. We agree that an actual reconstruction is desirable and plan to do so, but as pointed out above, there are more issues to overcome than fit into one manuscript.

"RC: *p.8-10: synthetic glacier state in 2000. Due to this cannot say something about real glacier state in 1850. Is it not possible to say something about the (length) change during this period and compare this to observations (for the 'best solutions')?* "

AR: No, a comparison is not useful. See also comments to real-world applications before.

"RC: *p.13: 'which we don't address here' → 'which we do not address here'. A pity...I understand that this is difficult, as explained by the authors (p.16, l.4-7), but question may arise at this point how relevant this story is for the community? Seems to be a good starting point for further research by the group and users of OGGM (p.17, l.26-27: 'framework will be useful'), but more limited for outsiders. Feels more like a 'technical' note/paper now. Here again, I am convinced that by incorporating some 'real data' and/or rethinking of some experiments (see e.g. also suggestion 1 in 'general comments' section): the story would become far more relevant for the glacier (modelling) community.* "

AR: We hope that our study is useful for the community (see several comments above). The model and our method are open source and documented.

"RC: *p.16, l.15: 'observed state': really the observed state or a synthetic one?* "

AR: We here mean the observed state in general (it can be a synthetic derived one or a real one)

"RC: *p.16, l.16: 'depends on the situation' → 'depends on the specific glacier setting'?* "

AR: Done.

"RC: *p.17, l.8: 'performs equally good' → 'performs equally well'?* "

AR: Done.

"RC: *p.17, l.9: 'In Sect 4.2...': find it strange to have this formulated in such a manner in the conclusion* "

AR: We reworded this sentence.

"RC: *p.17, l.12: 'could differ strongly in volume and area': would expect the differences to be relatively small with a trapezium/parabola cross-section, no* "

AR: Glaciers with the same length have the same number of grid points with non-zero ice-thickness. At each of these grid points a cross-section (with a parabolic, rectangular or trapezoidal shape) perpendicular to the flowline defines the relation between the ice-thickness and the widths of the glacier at this grid point. Each of these cross-sections can differ, whereas the number of cross-sections is the same. Thus, glaciers with the same length can differ strongly in volume and in area, because they can differ in width (and therefore in ice-thickness).

"RC: *p.17, l.13: 'lead more variability' → 'lead to more variability'* "

AR: Done.

"RC: *p.17, l.18: 'for some glaciers': indeed, for the ones with a short response time* "

AR: Yes, the ones with a low reconstructability measure (indirectly connected to the response time).

2 Reply to Reviewer #2

"RC: *The authors present an initialisation strategy for the Open Global Glacier Model (OGGM). The aim of this strategy is twofold. First, the initialisation should produce a best estimate for the glacier extent and geometry at the end of the little ice age (LIA). Second, the spin-up into present day should appropriately reproduce the observed geometry. The latter aim is formulated as an optimisation of an inverse problem. For testing and validation of the initialisation, the authors suggest synthetic experiments for which the past and present geometry is perfectly known. The main target parameter for the optimisation is a so-called 'temperature bias' . The authors show that their strategy allows to constrain the possible parameter space significantly, certainly if the optimisation accounts for geometric information beyond the glacier length. The manuscript is of great interest to the community as it aims at formulating a standard procedure that should provide the initialisation basis for regional or global ice-dynamic simulations of glacier evolution. In this sense, the authors are trying to solve a pressing issue. Moreover, the manuscript is well written and structured. However, I have major concerns on the pursued parameter sampling strategy in light of other a-priori parameter choices concerning ice-dynamics and surface mass balance. Consequently, I fear that the single parameter problem is oversimplifying any real-world application. To address my comments, the authors will certainly have to expand the manuscript to better justify and motivate their decisions. Consequently, I recommend a major revision of the manuscript.*"

AR: Thank you for your constructive feedback on our manuscript. We tried to implement most of your comments. However, it is a misunderstanding that "the initialization should produce a best estimate for the glacier extent and geometry at the end of the little ice age (LIA)", since we here focus on the methodological aspect of the reconstruction, and therefore do not present an estimate of the end-of-LIA ice volume in the Alps (among other things). We are aware that we probably did not explain this well enough and tried to improve the manuscript. This choice of experimental setup also has direct implications for the issue of uncertainties from model parameters that the reviewer raises. It is correct that they play an important role for real-world applications and, therefore, we agree that we are oversimplifying the problem (see also our general answer above). By using only one parameter (temperature bias) to create a test case, we are able to separate the uncertainty caused by the method itself from the uncertainty caused by model error, by parameter uncertainty, and by uncertainty of the boundary conditions. We hope we were able to sufficiently clarify our reasoning below and in the manuscript.

Temperature bias

"RC: *For the synthetic experiment, you prescribe a 'random climate scenario'(p. 5,line 26). I think that this term refers to a random permutation of the climatic forcing around the year 1850 (within a 30 year period). Is that right?*"

AR: Yes, the random climate scenario is explained on p.4, line 6-9. We shuffle the years (using a 31 year period around 1850) infinitely.

"RC: *I assume that the climatic forcing is taken from HISTALP. You then add a temperature bias of -1K to this climatic forcing. This bias is not well motivated. Why is it necessary?*"

AR: The main motivation is to create a synthetic environment which is realistic, i.e. close to what we would expect from the real word. When created with such a climate, our synthetic 1850 states are such that when evolving under the 1850-2000 climate they reach a state which (on average) is that from the Alps in 2000. To justify the choice of β=-1 K, we also test different values. The results are shown in Fig 3 (see p. 18). This figure shows that applying positive or small negative temperature biases to the synthetic experiments results in large area differences to the RGI in 2000, and the total glacierized area in 1850 is also too small. The sample size is reduced, because less glaciers fulfill the area threshold criteria of 0.01 km^2. Negative temperature biases that are too large also reduce the sample size, because some runs fail ("Glacier exceeds boundary", this means the glacier would get larger than the underlying grid). The experiments with a temperature bias of -1K, -1.25K or -1.5K perform best regarding the area difference to the RGI in 2000. But only the experiment with the temperature bias of -1K performs well regarding the estimation in 1850 of Zemp et al. (2006), whereby it needs to be taken into account that the dots only represent a subset (the small glaciers in 2000 are missing) of the glaciers considered in (Zemp et al., 2006). We added this results to Appendix A.

Additionally,we illustrate the process of the generation of the synthetic experiment again with the example of the Hintereisferner (see Fig. 3). Under climate conditions similar to the climate around 1850, today's glacier would still lose mass. If we do not apply a negative temperature bias to the random climate run, we would create a synthetic experiment with a glacier in 1850 that is smaller than today's glacier.

"RC: *I thought that HISTALP provide you with perfect climate conditions*"

AR: Yes, they do. Note that the temperature bias is only applied to create the true initial states (in which case the bias is fixed but not used by the reconstruction method since it should be unknown), and then by the reconstruction method to generate the potential candidates (in which case many biases are used and tested for). For the past climate runs from 1850 to 2000, we do not change the temperature bias, as HISTALP provide us "perfect" climate conditions for this time period. We add in the manuscript the information that no temperature bias is applied for the past climate runs (see Sect. 2.4.3 Evaluation) to avoid further confusion.

"RC: *Please provide a firm motivation for this bias, why it varies from glacier to glacier and consequently has to be inferred*"

AR: The starting point of our method is always the present state, as this is the only state we know. Provided with the assumption that past glaciers were different than today, we need a way to create physically consistent candidates. Glaciers respond differently to changes in climate and thus the required temperature biases to create these candidate are different for each glacier, too. As we do not know the size of the glaciers in 1850, we also do not know what temperature bias is necessary to create a glacier with a given size in 1850. This is the reason why we create a large set of physically plausible glacier candidates, which

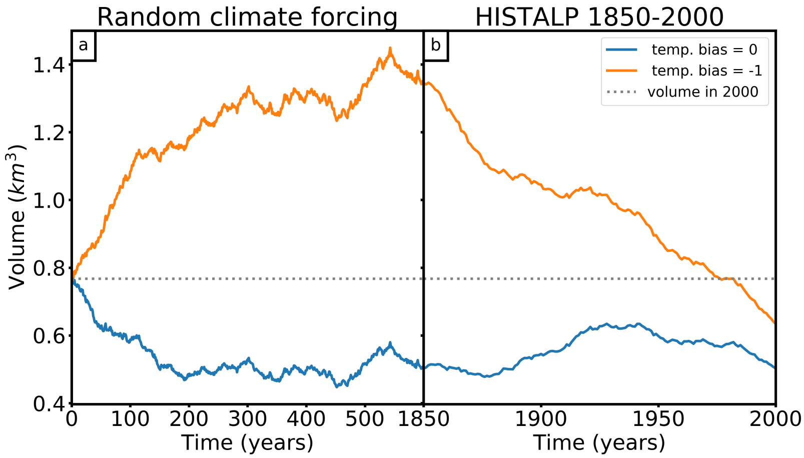

Figure 3. Generation of the synthetic experiment for the Hintereisferner. The dashed grey line marks the volume of today's glacier, the blue line the evolution without a temperature bias during the random climate run and the orange line using a temperature bias of -1 K. **a:** Generation of the synthetic experiment in 1850. These runs are forced with a random climate scenario (31 years around 1850). The glacier state at the end of this run (t=600) is the synthetic experiment in 1850.**b:** This state serves as initial condition for the past climate run (forced by HISTALP, no temperature bias) to create the synthetic experiment in 2000.

we need to evaluate with the forward run from 1850-2000. The range of biases we test in our study is not entirely random, as we assume that the glaciers are likely to be as big (or bigger) than the present-day geometry. But the method does not depend on this range: it can be increased at the cost of more computations (as we would have more candidates to test for). In the text, we motivated the use and the necessity for the temperature bias in Sec. 2.4.1 "Generation of glacier states".

A-priori OGGM calibration vs. parameter sampling

"RC: *Initially, the first question which came into my mind was why you did not enlarge the ensemble by including other uncertain parameters, as for instance the rate factor, basal friction or parameters linked to the SMB model. So ultimately this question links to the comment above. In other words, how do you ensure that the other parameters are well constrained.*"

AR: Including other uncertain parameters is of course possible. Adding e.g. a bias for the precipitation during the random climate run would be an option, too. We think that the variations we could obtain by varying "only" the temperature bias, the permutation and the time point (during the selection of the candidates) are sufficient and we could well demonstrate the functionality of our method. For the moment we ensure that other uncertain parameters are well constrained, because we used a synthetic environment: there is no parameter uncertainty influencing the results presented here. We added more information about the choice of these parameters in the text and explain how other parameter choices might influence the uncertainty in real-world applications in the discussion.

"RC: *I understand that you present a synthetic setup but with regard to any real-world application this issue is very important. Even if there was an a-priori calibration of a combination of the rate factor and the melting parameters, other combinations might also produce plausible glacier geometries at present. Yet the exact choice will affect the past glacier geometry.*"

AR: Yes, we agree. The exact choice of these parameters will affect real-world applications. Parameters linked to the SMB model are validated trough a cross-validation. The most important uncertain parameters of OGGM that will influence the past geometry as well as the response time of the model are the sliding parameter f_s, the creep parameter A, and the consideration of lateral drag. Maussion et al. (2019) analyzed the model sensitivity to these parameters. But again, as long as we consider synthetic experiments, this does not affect the results. For real-world application, the calibration of these parameters and the quantification of their influence on our results has to be part of the reconstruction procedure. The complexity of this issue is why we decided (and believe that it is necessary) to first realise this synthetic study.

"RC: *In many other glaciological applications, basal friction or the rate factor are the central unknowns that are calibrated during the model initialisation. I therefore think that it is inevitable that you include a section on the OGGM procedure for the calibration/choice of these other parameters. On this basis, you should motivate your parameter choice for your ensemble generation. Dependent on how the other OGGM parameters are calibrated, it might well be necessary to include further parameters in the ensemble generation*"

AR: We agree, and are well aware that parameter uncertainty will play an important role in real-world applications. The purpose of this study, however, is to explicitly remove parameter uncertainty and still demonstrate that, even in this controlled setting, the problem of model initialisation is non-trivial and non-unique. We hope that the changes in the introduction and method section of our revised manuscript now make a better case at explaining this point.

Climatic forcing

"RC: *Throughout the manuscript, I sensed some redundancy because I did not find any discussion of the 'acceptable ensemble states' in terms of the initial parameter choice. To put it bluntly, were you able to infer the -1K temperature bias prescribed in the synthetic experiment from your data assimilation? Otherwise, the specific climatic permutation might have had a significant influence?*"

AR: The permutation is necessary to create the random climate time series that are 600 years long (see also the point after the next point). We were able to reproduce the -1 K temperature bias from the synthetic experiment. Figure 4 shows a histogram of the temperature bias of all median states (best candidates). The bias of the majority lies around -1 K. Values that differ more from -1 K, can be explained by the uncertainty of the median state. Especially in cases where most tested candidates are acceptable, the median is only a representative of the large set of acceptable glacier states. The aim of this study is to determine a past glacier geometry, not a temperature bias. The temperature bias is (only) the tool, which creates the desired geometry candidates. For this reason we decided to do the uncertainty analysis based on the glacier geometry and not based on the temperature bias.

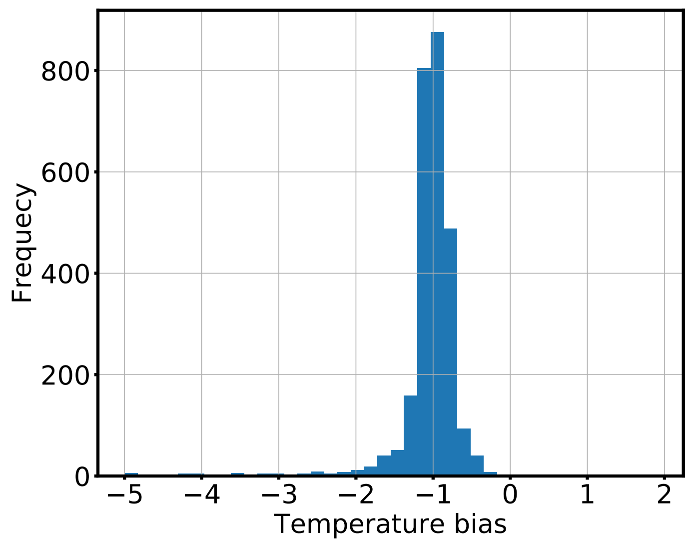

Figure 4. Histogram of the temperature bias from median states.

"RC: *In short, please justify your choice to exclusively focus on climatic quantities in the optimisation. From my experience, you should include other parameters in this optimisation.*"

AR: Our method searches the climate conditions (similar to the ones around 1850) under which the today's glaciers would grow to their "true" size in 1850. Dynamical parameters of the model would have to be calibrated before the initialization method (see also comments to parameter calibration).

"RC: *I wonder about the necessity to permute the initial climate time series during the initialisation. Do you really attain distinctly different glacier geometries in 1850 that you could not generate by only changing β. Please re-assess the dual sampling of climatic variables/input in the ensemble generation*"

AR: The infinite shuffling does not only create more variations, but is also necessary to create a time series that is long enough to reach the stagnation period. Without the random shuffling, we would not be able to create the 600 year long time series which we use for the generation. Of course, we could use the same permutation for all runs and only varying β, but there is no reason to prefer only one out of the huge set of possible permutations. For this reason we decided to use for each different β, which we tested, a different way to shuffle the 31 years around 1850. We added this argument in the manuscript (see Sec. 2.4.1)

"RC: *The past-to-present volume evolution of Guslaferner of all ensemble members is shown in Fig. 3. It surprised me that all ensemble members readily converge to a very similar present-day value. Certainly if you consider the large range of β-values used for the initialisation.*"

AR: The rapid convergence was caused by the wrong parameter set used for the SMB calibration (see also comment to your question about the calibration of the model). We repeated all runs with the correctly calibrated mass-balance model. The convergence is delayed now.

Detailed comments

"RC: *on page 8 lines 10- 11, my attention was then drawn to the fact that all forward runs are conducted with 'exactly the same climate time series' and use the 'same parameter set'. Does this include β?*"
AR: Yes, this include the temperature bias β=0. The past climate runs (from 1850-2000) use the HISTALP time series, which provide us perfect climate conditions (as mentioned earlier). There is no need to change the temperature bias for these runs. The climate, as well as all other parameters are the same for the forwards runs starting in 1850. Only the initial condition (surface elevation and widths along the glacier flowline) is different.

"RC: *If β was set to zero after the initialisation, I would be highly concerned about the abrupt climatic shift you introduce when switching from the initialisation to the forward experiment. A reason for the latter case is the overall quick convergence during the forward simulations*"
AR: The past climate runs are performed with temperature bias $\beta = 0$, but the generation of the candidates is distinct from the initialization process and thus we do not see a major issue that arises due to an abrupt climatic shift. From the selected states we only take the surface elevation and the widths along the flowline and we start a completely new run.

"RC: *What climatic forcing is used for the a-priori OGGM calibration. Is it HISTALP. Is the temperature bias β included? Please clarify.*"
AR: The mass balance model calibration is done with HISTALP[1]. We added this information in the manuscript in Sect. 2.1. The temperature bias (which is an artificial offset that we use to generate glacier candidates) is set to zero for the calibration.

"RC: *p.1, line 20 - Sentence is difficult to understand. Reformulate.*"
AR: We rephrased the sentence.

"RC: *p. 2, line 17 - 'Despite of the importance ...' –> 'Despite the importance ...'*"
AR: Done.

"RC: *p. 2, line 28 - Introduce a comma (,) before which. Please check throughout the document.*"
AR: Done.

"RC: *p. 4 ,line 1 - '... glacier's length ...' –> '... glacier length ...'. Please avoid the 's genitive throughout the document.*"

[1] See our cross-validation monitoring website if you are interested in the results: https://cluster.klima.uni-bremen.de/~github/crossval

AR: Done.

"RC: *p. 5 ,line 5 - Please omit the bedrock difference in Equation (4). As the bedrock does not change during your initialisation this term is zero*"

AR: Done.

"RC: *p. 5 ,line 22 - Remove comma.*"

AR: Done.

"RC: *p. 5 line 28 - Delete 'this' at the beginning of the line.*"

AR: We corrected the sentence.

"RC: *p.6 line11-p.7 line 5 - The initialisation ensemble is formed by the variation of two parameter: β and a permutation of the 30-yr climate forcing time series. For both quantities it remains unclear how many ensemble members are created. As you reduced the ensemble later on to 200 members based on an equal spacing argument, I would assume a sufficiently dense sampling. Anyhow, please provide some numbers.*"

AR: The number of ensemble members differs from glacier to glacier. This depends on t_stag and on the number of successfully completed random climate runs n_r (number of grey lines shown in Fig. 2). We stored yearly glacier volumes and thus the ensemble size is $n_r(600 - t_{stag})$. In the case of the Guslarferner $t_{stag} = 282$ and $n_r = 140$ result in $\sim$ 44500 members. The sampling is sufficiently dense in all cases. We add further information in the manuscript in Sect. 2.4.2 Identification of candidates.

"RC: *p.6 line 13 - It remains vague how the permutation is done. You permute the climatic forcing per year, month, etc.*"

AR: We shuffle the years, but use monthly climate data. The order of the month are not changed. We made this more clear in the text.

"RC: *p.9 line 6 - Please provide values for glacier area and estimates for mean ice thickness for Guslarferner and Hintereisferner at present. Hintereisferner covers a much larger area and is probably much thicker. These values are informative and they are difficult to infer from Figs. 3 and 5.*"

AR: We add this information to the captions of Fig. 3 and 5 in the manuscript. Additionally we improved the labelling of the x-axis of these figures. Please keep in mind, that the values of the synthetic experiment can differ to the reality.

"RC: *p. 9, line 10 - 'has' –> 'have'*"

AR: Done.

"RC: *p. 10, line 2 - ‘not’ –> ‘no’*"

AR: Done.

"RC: *Figure 2 There is a discrepancy between the range of temperature biases given in the caption [-2.65, 2.95] and shown in panel (a) [-3.0, 1.95]. Moreover, I would try to remove some redundant information. For panel (b), limit the graph to the stagnation period. For panel (c), do show the initialisation period with the coloured points but extend the figure by the forward simulation 1850-2000 as in Fig3c.*"

AR: We corrected the values for the temperature range in the figure and limit the graph in Figure 2b and 2c to the stagnation period only. We decided against the extension of the simulation 1850-2000, because the Figure becomes too confusing otherwise.

References

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, Geoscientific Model Development, 12, 909–931, https://doi.org/10.5194/gmd-12-909-2019, https://www.geosci-model-dev.net/12/909/2019/, 2019.

Oerlemans, J.: Climate Sensitivity of Franz Josef Glacier, New Zealand, as Revealed by Numerical Modeling, Arctic and Alpine Research, 29, 233 – 239, https://doi.org/10.80/00040851.1997.12003238, http://www.jstor.org/stable/1552052, 1997.

Zemp, M., Haeberli, W., Hoelzle, M., and Paul, F.: Alpine glaciers to disappear within decades?, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL026319, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL026319, 2006.

Initialization of a global glacier model based on present-day glacier geometry and past climate information: an ensemble approach

Julia Eis[1], Fabien Maussion[2], and Ben Marzeion[1,3]

[1]Institute of Geography, University of Bremen, Bremen, Germany
[2]Department of Atmospheric and Cryospheric Sciences, University of Innsbruck, Innsbruck, Austria
[3]MARUM - Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Correspondence: J. Eis (jeis@uni-bremen.de)

Abstract. To provide estimates of past glacier mass changes over the course of the 20^{th} century, an adequate initial state is required. However, empirical evidence about past glacier states at regional or global scale is largely incomplete, both spatially and temporally, calling for the use of automated numerical methods. This study presents a new way to initialize the Open Global Glacier Model from past climate information and present-day glacier states~~. We~~ , since it is the only data available for almost all glaciers globally. We use synthetic experiments to show that even with these perfectly known but incomplete boundary conditions, the problem of model initialization is an ill-posed inverse problem leading to non-unique solutions, and propose an ensemble approach as a way forward. The method works as follows: we generate a large set of physically plausible glacier candidates for a given year in the past (e.g. 1850 in the Alps)~~. All these candidates are~~ , all of which are then modelled forward to the date of the observed glacier outline ~~,~~and evaluated by comparing the results of the forward runs to the present-day states. We ~~apply~~ test the approach on ~~2621 alpine~~ 2660 Alpine glaciers and determine error estimates of the method from the synthetic experiments. The results show that the solution is often non-unique, as many of the reconstructed initial states converge towards the observed state in the year of observation. We find that the median state of the best 5 percent of all acceptable states is a reasonable best estimate. The accuracy of the method depends on the type of the considered observation for the evaluation (glacier length, area, or geometry). Trying to find past states from only present-day length instead of the full geometry leads to a sharp increase in uncertainty. Our ~~method~~ study thus also provides quantitative information on how well the reconstructed initial glacier states are constrained through the ~~combination of present-day outlines with past climate conditions. We show that even with perfectly known but incomplete boundary conditions the problem of model initialization is non-trivial and non-unique~~limited information available to us. We analyse which glacier characteristics influence the reconstructability of a glacier, and discuss ways to develop the method further for real-world applications.

1 Introduction

Glaciers contributed significantly to past sea-level rise ~~(SLR; e.g. Gregory et al., 2013; Slangen et al., 2017a; Cazenave and et al., 2018)~~ (S they will continue to be a major contributor in the 21$^{\text{th}}$ century ~~(e.g. Church et al., 2013; Slangen et al., 2017b)~~(e.g. Church et al., 2013; Slan A large fraction of this contribution will be caused by the ongoing adjustment ~~to the past climate~~ of glaciers to previous climate change (Marzeion et al., 2014, 2018). Reconstructions of past glacier mass change are therefore not only necessary to determine the budget of past sea-level change (Gregory et al., 2013) and to increase the confidence in projections (by allowing to quantify the agreement with observations, Marzeion et al., 2015), they also enable us to quantify the pattern of the ongoing adjustment of glaciers to present-day climate. Estimates of global glacier mass change are based on in-situ measurements in mass and length changes (e.g. Zemp et al., 2015; Leclercq et al., 2011), on remote sensing techniques ~~(e.g. Gardner et al., 2013; Jacob et al., 2012; Bamber et al., 2018)~~(e.g. Gardner et al., 2013; Jacob et al., 2012; Bamber et al., 2018; Woute or on mass balance modelling driven by climate observations (Marzeion et al., 2012, 2015). Since observations of temperature, and to a smaller degree, precipitation, are more ubiquitous (e.g. Harris et al., 2014) than glacier observations (WGMS, 2018), reconstructions of glacier change produced by forcing a glacier model with climate observations have the potential to increase the understanding of past glacier behaviour. Finally, reconstructing glacier change based on climate model output allows to test the skill of climate models (Goosse et al., 2018).

A number of global glacier models were developed in the past (e.g., Radić and Hock, 2011, 2014; Giesen and Oerlemans, 2012, 2013; Marzeion et al., 2012, 2014; Huss and Hock, 2015; Maussion et al., 2019). The more recent and complex of these models (e.g. Huss and Hock, 2015; Maussion et al., 2019) require Digital Elevation Models (DEMs) and outlines from the Randolph Glacier Inventory ~~(RGI; RGI Consortium, 2017)~~ (RGI; Pfeffer et al., 2014) to derive the initial surface hypsometry. ~~The~~ Hence, their starting date of a glacier evolution simulation ~~thus~~ depends on the recording date of the DEM and the outline, which typically do not coincide with one another, nor with the required starting date of a projection. The model of Huss and Hock (2015) indicates a high sensitivity to the initial ice volume. Similarly, Maussion et al. (2019) remark that great uncertainties, especially on local and regional scales, derive from unknown initial conditions.

Despite ~~of~~ the importance of glacier contribution to past sea-level rise, so far only one model was able to provide estimates of glacier mass changes over the course of the entire 20$^{\text{th}}$ century on the global scale (Marzeion et al., 2012). All other global modelling studies limit their application to the recent past and future projections. The reconstruction by Marzeion et al. (2012) was possible because of the highly parameterized representation of ice dynamics and glacier geometry change, applying a Volume-Area-Time scaling to translate mass change into surface area and elevation range changes. Based on this approach, it was sufficient to iteratively optimize one variable (glacier size in the year of interest, e.g., 1850) such that when run forward to the year of the observed glacier outline, the modelled glacier area agreed with the observed glacier area.

An increase of model complexity impedes the process as more and more variables are required for initialization. Flowline models require input data along the coordinates of the flowline (e.g. bed topography, surface elevations and/or widths) and thus more complex initialization methods are needed. For example, van Pelt et al. (2013) developed an iterative inverse method to reconstruct distributed bedrock topography and simultaneously initialize an ice flow model. Zekollari et al. (2019) added an ice flow model to Huss and Hock (2015), which required an automated initialization for glaciers in 1990 (prior to the glacier inventory date) to avoid spin-up issues and so that the reconstructed initial states fit the glacier geometry at the inventory date after being modelled forward. By choosing a decade long initialization, they avoid problems of non-uniqueness (as we discuss below), but raise the question of how arbitrarily this date can be chosen. Similar approaches exists for the initialization of ice sheet models, where most work focuses on estimating the present-day state of ice sheets in order to make accurate projection of future ice sheet change (e.g. Heimbach and Bugnion, 2009; Lee et al., 2015; Mosbeux et al., 2016). Goelzer et al. (2018) divide the existing initialization approaches into 3 methods: spin-up, assimilation of velocity, and assimilation of surface elevation. Spin-up procedures are typically used for long-term and palaeoclimate simulations, the required spin-up time is unknown ~~,~~ and can be relative long~~, and~~ . Additionally, the reconstruction cannot be expected to represent effects from internal climate variability correctly. The data assimilation approaches typically determine model parameters (e.g. basal parameters like basal friction or bedrock topography) that reduce the mismatch between observed and modelled velocities or surface elevations.

In this study, we aim to ~~bring elements of response to the following questions:~~ identify fundamental limitations that narrow the reconstruction of past glacier states from present day geometries, under the assumption of perfectly known boundary conditions and a perfectly correct glacier model. I.e., more specifically:

- ~~to~~ To which degree does the past evolution of a glacier constrain its present day geometry?
- ~~how~~ How much information does the present day glacier geometry contain about its past states?
- ~~is~~ Is it possible to reconstruct past glacier states from the partial information available to us?
- How far can we go back in time to have an initial geometry that still determines the present-day glacier geometry?
- Which glacier attributes influence the answers to the questions above?

To this aim, we present a new method estimating past glacier states and apply it to synthetic numerical experiments, and we show the obstacles that need to be overcome before applying our method to real-world problems. After introducing the relevant features of the Open Global Glacier Model (OGGM; Maussion et al., 2019) in Sect. 2.1, we describe the design of the synthetic experiments in Sect. 2.3. The synthetic framework serves to test the skill of our approach in a surrogate model world where everything is known, and allows to apply data denial experiments to address the questions listed above. The initialization method is presented in Sect. 2.4. The developed method consists of 3 steps: generation of plausible glacier states, identification of glacier state candidates, and their evaluation based on the misfit between the modelled and the observed ~~surface elevation~~ geometry at the year of the observation. We applied our approach to ~~2621 alpine~~ 2660 Alpine glaciers and present the results for the reconstructed initial states in the year 1850 in Sect. 4.1. The influence of the considered type of observation (e.g. glacier length, area or geometry) is shown in Sect. ~~??~~4.2 and a statistical analysis of glacier attributes that influence the reconstructability of a glacier is presented in Sect. 4.3. Finally, we summarize the results and discuss the ~~method's limitations~~ limitations of the method and its applicability to real-case studies, as well as needed and possible future developments in Sect. 6.

2 Methods

2.1 The Open Global Glacier Model

The Open Global Glacier Model (OGGM; Maussion et al., 2019) is an open source numerical framework that allows the modelling of past and future changes of any glacier in the world. Starting with a glacier outline, provided by the Randolph Glacier Inventory (RGIv6.0; Pfeffer et al., 2014), a suitable surface DEM is automatically downloaded and interpolated to a local grid. The size of the local grid is given by a border parameter, which is the number of grid points outside the glacier boundaries. We choose a border value of 200 grid points to ensure that also large glacier states can be generated. The resolution of the map topography dx depends on the size of the glacier ($dx = a\sqrt{S}$, with $a = 14\mathrm{m\ km}^{-1}$ and S the area of the glacier in km^2) and is constrained to ~~$10\ m \leq dx \leq 200\ m$~~$10\ \mathrm{m} \leq \mathrm{dx} \leq 200\ \mathrm{m}$. After the preprocessing, glacier centerlines are computed using a geometrical routing algorithm (adapted from Kienholz et al., 2014). They are then considered as glacier flowlines, and grid points are generated using a fixed, equidistant grid spacing, which is twice that of the underlying 2D map topography. Surface elevations along the flowline coordinates are then obtained from the underlying topography file and glacier section widths are computed by intersecting the ~~flowline's normal~~ normal of the flowline to the boundaries of the glacier. By making assumptions about the shape of the bed (parabolic, rectangular or a mix of both), OGGM estimates the ice thickness with a mass-conservation approach (Farinotti et al., 2009, 2017, 2019). Information on bed topography at each grid point results from the calculated ice thickness and the surface elevation. From this, the glacier ~~'s~~ length, area, and volume can be determined. These values depend strongly on the surface topography and are based on the (often wrong) assumption that the recording date of DEM and that of the outline coincide. The ~~model then computes the glacier mass balance~~ dynamical flowline model of OGGM can then be used to determine the evolution of the glacier under any given climate forcing by solving the shallow ice approximation along the flowlines.

The mass balance is computed at each grid point using climate data (monthly temperature and precipitation). Climate data can be used from different sources, including gridded observations or reanalyses for past climate, projections for future climate, or randomized climate time series~~for more specialized applications. To force~~ . The purpose of forcing the mass balance model with ~~a randomized climate time series, a window size h and~~ randomized climate is to easily produce a great number of realistic climate forcings representative of a given time period, characterized by a center year y_0 ~~need to be set first. All~~ and a window size h (typically 31 years). All climate years $\in [y_0 - \frac{h-1}{2}, y_0 + \frac{h-1}{2}]$ are then shuffled infinitely in the next step. Additionally, it is possible to set a temperature bias β, which ~~shift~~ shifts all values of the temperature series ~~. The objective of forcing the mass balance model with randomized climate is to easily produce a great number of realistic climate settings, representative of a given time period. The dynamical flowline model can then be used to determine the evolution of the glacier under certain climate forcings by solving the shallow ice approximation.~~ towards warmer or colder climates.

Identically to the study of Maussion et al. (2019), we only calibrate the mass-balance model while the creep parameter A and the sliding parameter f_s are the same for each glacier and set to their default values ($A = 2.4 \times 10^{-24}\mathrm{s}^{-1}\mathrm{Pa}^{-3}$, $\mathrm{f_s} = 0$, no lateral drag). The following mass balance related parameter values were used in this study: $p_f = 1.75$, $T_{Melt} = -1.5°\mathrm{C}$, $T_{Liquid} = 2.0°\mathrm{C}$ and $\Gamma = -6.5\mathrm{Kkm}^{-1}$. This parameter set was determined with a cross-validation done with the HISTALP data set and tested for the 41 Alpine glaciers with more than 5 years of mass-balance observation. For more details concerning the glacier model ~~,~~ (e.g. the mass-balance calibration or sensitivities to the dynamical parameters of the model) please refer to Maussion et al. (2019) and http://docs.oggm.org.

2.2 Problem description

Here, we define a *glacier state* (hereinafter referred as *state*) as follows:

Definition 1. *Let $m \in \mathbb{N}$ be the total number of grid points of all flowlines of a glacier. Then $s_t = (z_t, w_t, b)$ is a* ***glacier state*** *at time t, with surface elevation $z_t \in \mathbb{R}^m_+$, widths $w_t \in \mathbb{R}^m_+$, and bed topography $b \in \mathbb{R}^m_+$. The set $\mathcal{S}_{t_i} = \{s_t | t = t_i\}$ contains all physically plausible glacier states at time t_i.*

The construction of an initial state is an inverse problem and can be defined in opposition to the direct problem. The *direct problem* corresponds to a forward model run: ~~Given~~ given an initial state $s_{t_0} \in \mathcal{S}_{t_0}$ at time t_0, the state $s_t \in \mathcal{S}_t$ at time $t > t_0$ can be computed by:

$$s_t = G_{\text{past}}(s_{t_0}) \tag{1}$$

where $G_{\text{past}} : \mathcal{S}_{t_0} \rightarrow \mathcal{S}_t$ is an operator representing the equations of OGGM, using known climate time series as boundary condition.

For *inverse problems*, the solution is known by direct observations: $s_{t_e} = s^{obs}_{t_e}$, whereas the desired initial state s_{t_0} is unknown. The inverse problem consists of finding the initial state $s_{t_0} \in \mathcal{S}_{t_0}$, such that the forward modelled solution at time t_e fits the observations from the same year t_e:

$$s_{t_0} = G^{-1}_{\text{past}}(s^{obs}_{t_e}) \tag{2}$$

Unfortunately, we do not have an explicit formulation for G^{-1}_{past} in our case. A backwards reconstruction is impeded by the non-linear interaction between glacier geometry, ice flow and mass balance. Optimization methods can be used to solve inverse problems. To this end, we introduce a minimization problem such that the forward modelled state is as close as possible to the observation:

$$\min_{s_{t_0} \in \mathcal{S}_{t_0}} j(s_{t_0}) \tag{3}$$

with

$$j(s_{t_0}) := \frac{1}{m} \parallel s^{obs}_{t_e} - \underbrace{G_{\text{past}}(s_{t_0})}_{s_{t_e}} \parallel^2_2 = \frac{1}{m}\left(\sum_{i=0}^{m} \left((z^{obs}_{t_e})_i - (z_{t_e})_i\right)^2 + \left((w^{obs}_{t_e})_i - (w_{t_e})_i\right)^2 + \underbrace{\left((b^{obs})_i - (b)_i\right)^2}_{=0} \right) \tag{4}$$

This function calculates the averaged difference in surface elevation and width between the observed and forward modelled glacier state. Differences in bed topography can be neglected, as we assume the bed topography to remain the same over the inspected time period.

In many cases, however, OGGM ~~'s~~ forward integrations of different initial states result in very similar states at time t_e. This implies that there exist many local minima of the function $j(s_{t_0})$. As uncertainties of the model can safely be assumed to be larger than the differences between those states at time t_e, it is impossible to identify the global minimum of $j(s_{t_0})$. I.e., the solution of our inverse problem is non-unique.

The objective of our approach is therefore to identify the set $\mathcal{S}_{t_0}^{\epsilon}$ of all states, which correspond to the observed state $s_{t_e}^{obs}$ within a given uncertainty ϵ after being modelled forward. We call this condition *acceptance criterion*:

$$J(s_{t_0}) := \frac{j(s_{t_0})}{\epsilon} < 1 \tag{5}$$

The function $J(s_{t_0})$ is called in in the following *fitness function*. Assuming a vertical error of 5 m in x and an horizontal error of 10 m in w, we propose to set $\epsilon = (5\mathrm{m})^2 + (10\mathrm{m})^2 = 125\mathrm{m}^2$. These numbers can be changed easily, and in a real-world application should be based on the vertical uncertainty of the ~~used DEM~~ reconstructed ice thickness and the horizontal uncertainty of the used outline. All states $s_{t_0} \in \mathcal{S}_{t_0}^{125}$ that have a fitness value smaller than one are called *acceptable states*. The first expectation would be that the glacier candidate with the smallest fitness value ~~, is be~~ is also the best solution. However, due to uncertainties that derive from the model integration itself, this is not always the case. As an alternative, we determine the 5th quantile of all states in $\mathcal{S}_{t_0}^{\epsilon}$. This set contains the best solutions of all acceptable states referred to ~~the fitness value~~their fitness values. We choose the median state ~~,~~ as a representative of this set and compare the state with the minimal fitness value and the median state in Sect 4.1.

2.3 Synthetic experiments

~~In order to be able to evaluate the results of the~~

2.3.1 Design

We create a time series of glacier states which range from the target year of initialization t_0 (e.g. 1850) to the present date t_e (see Fig. 1 for an example). These are the glacier states that we aim to reconstruct with our initialization method, ~~we use synthetic experiments. To this end, we use OGGM to initialize a model run. We~~ using only partial information (here the "observed" state at present day). This type of experiment is sometimes called "inverse crime" in the inverse problems literature (e.g. Colton and Kress, 1992; Henderson and Subbarao, 2017), and we explain their rationale below. To generate them, we apply a random climate scenario (window size $h = 31$ years ~~,~~ and center year ~~$y_0 = t_0$~~ $y_0 = t_0 = 1850$) and run the model ~~400 years forward~~ 600 years forward (see Fig. 1c). The temperature bias is set to $\beta = -1K$ to ensure that a relatively ~~big~~ large 1850 glacier state is created ~~, because this the case~~ (as expected for most real glaciers ~~in 1850.~~ at the end of the Little Ice Age). The resulting state is defined to be the synthetic experiment state in year t_0 (see Fig. 1a). We model this state forward, applying the past climate time series from t_0 until t_e (here: 2000) (see Fig. 1d) and obtain the *observed* state of the synthetic experiment ~~. Through this procedure, a time series of glacier states is created, which in the following we will try to recover applying the initialization method described below, but only using information about the observed state from the synthetic experiment in~~

~~t_e~~(see Fig. 1b). Thanks to the initial temperature bias of $\beta = -1K$, these synthetic states in t_e are very close to the real observed states in 2000 on average (total area difference for the Alps of about 1%, but individual glaciers can vary) ~~. In the following we call the state~~ and the total synthetic glacierized area in 1850 fits well to an estimate of Zemp et al. (2006) (see appendix A for more details). We call the states derived from the synthetic experiment s_t^{exp}. ~~Figure 1 shows the surface elevations in 1850 and~~

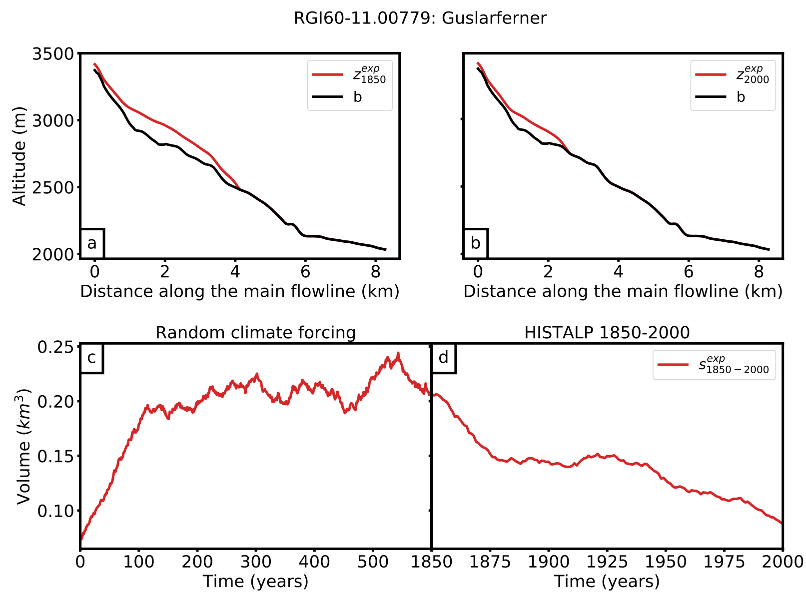

Figure 1. ~~Synthetic~~ Illustration of the generation of the synthetic experiment ~~applied to~~ with the example of Guslarferner (~~Alps~~Oetztal, Austria). ~~Cross-Sections~~ **a:** glacier thickness along the main flowline ~~in **a:**~~ at $t_0 = 1850$ and **b:** $t_e = 2000$. ~~Black~~ The black line indicates the bed rock and the ~~dashed~~ red line the ice surface of the synthetic experiment. **c:** generation of s_{1850}^{exp}, which is the state at t=600 (the end of the trajectory) and **d:** the volume of the glacier states s_t^{exp} from 1850 to 2000. Note that the synthetic year 2000 glacier does not necessarily correspond to the "true" year 2000 glacier.

~~2000 of the synthetic experiment for the Guslarferner as an example.~~

~~We will discuss the limitations of the usage of the synthetic experiments in the context of real-world applications in Sect. 6~~

2.3.2 Rationale

These synthetic states therefore provide a realistic setting with a strong advantage over actual observations: they are perfectly known. As stated in the introduction, reconstructing past glacier states is a complex inverse problem, which accuracy will depend on (i) the uncertainties in the boundary conditions (climate, glacier bed, etc.), (ii) the uncertainties in the glacier model itself, and (iii) a theoretical lower bound (termed "reconstructability" in this study) tied to the characteristics of the glacier itself (slope, size, the past climate, etc.). The main objective of the synthetic experiments is to separate these issues from one another, and to focus on point (iii) only. This allows us to isolate and understand the limitations and errors of the developed method itself, as opposed to uncertainties that derive from unknown boundary conditions and model parameters. They also allow us to realize data denial experiments and detect which kind of observations are necessary to reduce the uncertainties of our reconstruction (Sect. 4.2), and to determine which glacier characteristics affect the reconstructability of a glacier (Sect. 4.3).

2.4 Reconstruction of initial glacier states

Our initialization method consists of three main steps: generation of a set of physically plausible glacier states $\mathcal{S}_{t_0}$, identification of glacier candidates $s_{t_0} \in \mathcal{S}_{t_0}$, and their evaluation based on the fitness function $J(s_{t_0})$ (see Sec. 2.2).

2.4.1 Generation of potential glacier ~~states~~candidates

In a first step, we generate a set of different, physically plausible states from which we will pool our "candidates" (Fig. 2a). For this purpose, we utilize a random mass balance model with a window size of $h = 31$ years and the center year $y_0 = t_0$ to create different climate conditions. Obviously, we do not use the same permutation as for the creation of the synthetic experiments (~~see~~ Sect. 2.3). This procedure ~~has the advantage that a realistic~~ generates a climate representative of a given time period ~~can be created, while interannual variability is~~ with an interannual variability uncorrelated to that of the ~~period. Hence,~~ original period. For each random climate a different way of permutation is used. This ensures that all generated climate ~~models~~ time series differ from each other, but at the same time all represent the climate conditions around t_0 ~~at the same time . Additionally, we~~ (and an associated temperature bias β). The infinite permutation is necessary to obtain a time series that is long enough for the glaciers to reach an equilibrium (while maintaining the impact of interannual climate variability) with the forcing climate (here 600 years). To create a large set of states, we additionally vary the temperature bias β~~to create further variations~~. Glaciers respond differently to changes in climate and thus the required temperature biases vary from glacier to glacier and have to be inferred. We start with temperature biases $\beta \in [-2, 2]$ K. If $\beta = 2$ K is not large enough to result in a present day glacier with zero ice thickness, higher values will be used. If $\beta = -2$ K is not small enough to result in a glacier that reaches the boundary of the local grid (200 grid points outside of the glacier ~~'s boundary~~outline), smaller values will be used. ~~We use these climate conditions to force a 400-year glacier simulation. 400 years turned out to be a sufficient time to generate a large number of different states. This model run is initialized with the only known state, which is the actual observed glacier topography, taken from the RGI (see Fig. 2a).~~

2.4.2 ~~Identification~~ Selection of candidates

Figure 2a shows the evolution of the volume of the generated glacier states over time. In the first years, the time series clearly diverge (mostly caused by the temperature bias β), but after a certain time all time series begin to fluctuate around an equilibrium value. We refer to the period of fluctuations around the assumed equilibrium as the *stagnation period*. During the stagnation period the glacier volume does not increase or decrease strongly in comparison to the total volume change since the

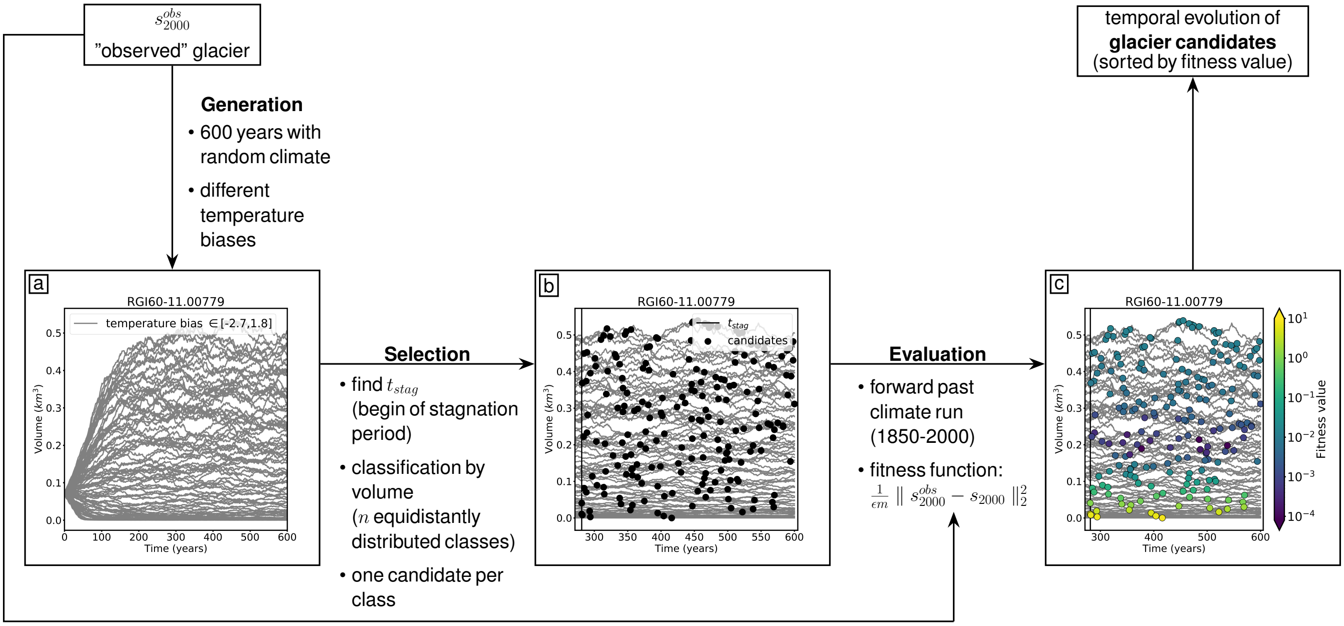

Figure 2. Workflow of the ~~initialization~~ candidates generation, selection and ranking method, using Guslarferner (~~Alps~~Oetztal, Austria) as an example. **a:** ~~The~~ generation of ~~different~~ potential glacier candidates. The grey lines indicate the glacier volume evolution for a set of different random climate scenarios over ~~400~~ 600 years each. The temperature biases vary between ~~-2.65~~ -2.7 and ~~2.95~~ 1.8 K. **b:** selection of candidates. The black vertical line indicates t_{stag} and the black points show 200 candidates. **c:** ~~The~~ glacier candidates colored by their fitness value. Violet marks candidates with a small ~~missfit~~misfit, whereas yellow marks states that don't meet the acceptance criterion (Eq. 5).

beginning of the simulation. We define t_{stag} as the point in time where all trajectories have reached this stagnation period ~~. The required time to reach a stagnation period is longer for thicker glaciers (Jóhannesson et al., 1989). Glacier volume increases with decreasing temperature biases. Thus, we~~ and choose the upper ten volume trajectories, corresponding to the lowest temperature biases, to determine t_{stag}. To this end, we smooth each of the ten curves with a 10 years rolling-window and calculate the time point of their first maximum. t_{stag} is defined as the latest of all previously determined time points (see Fig. 2b). Defining t_{stag} is necessary, because we determine initial glacier states at $t_0 = 1850$ and the searched glaciers are assumed to be in equilibrium with the climate around 1850. Hence, each state that fluctuate around an equilibrium value is a potential glacier state candidate (in the following referred as candidate). This holds true for all states s_t with $t > t_{stag}$. Depending on t_{stag} and the number of successfully completed random climate runs n_r (number of grey lines in Fig. 2) the sample size is $n_r(600 - t_{stag})$ (glacier states are stored yearly). The sample size is sufficiently large for all cases, e.g. in the case of the Guslarferner (Fig. 2) the sample contains approx. 44.500 members. In order to avoid testing very similar states, we classify all states by their volume and select one candidate per class. We choose n equidistantly and approximately uniform distributed classes, where n (default: $n = 200$) is the number of candidates to evaluate in step three.

2.4.3 Evaluation

The last step evaluates all previously selected candidates. Each candidate is used as initial condition for a forward run, using observed past climate time series, e.g. from $t_0 = 1850$ until $t_e = 2000$. All runs use the same model parameter set, except for the initial condition and exactly the same climate time series (e.g. no temperature bias β is applied for the past climate runs). Afterwards, we compare the resulting modelled states s_{t_e} with ~~the~~ an observed state $s_{t_e}^{obs}$ (here taken from the synthetic experiment) by applying the fitness function $J(s_{t_0})$ (Eq. 5). This function calculates the averaged difference between the glacier geometries at the grid points, more specifically between the surface elevations z_{t_e} and the widths w_{t_e}, of the observed and the modelled glacier. In Fig. 2c the candidates are colored by their fitness value. ~~The resulting output of this method is the temporal evolution of all candidates from t_0 until t_e sorted by their fitness value.~~

3 Test site and Input Data

We tested our approach on ~~alpine~~ Alpine glaciers. The glacier outlines are ~~derived~~ taken from the Randolph Glacier Inventory (RGI v6.0, region 11; Pfeffer et al., 2014). ~~For each of the glacier outlines a transverse Mercator projection, which is centered on the glacier, is defined in order to preserve map projection properties (e.g. area, distances, angles). Next, the topographical data is automatically downloaded by OGGM. The source of the DEM is~~ We use topographical data from the Shuttle Radar Topography Mission (SRTM) 90m Digital Elevation Database v4.1 (Jarvis et al., 2008)~~as the study area is located in the $[60°S; 60°N]$ range (Maussion et al., 2019) and the~~. The SRTM aquisition date (2000) matches well that of the RGI (2003 for most glaciers). The climate dataset we use for this approach is the HISTALP database (Auer et al., 2007, http://www.zamg.ac.at/histalp). The temperature time series covers the period 1780 to 2014 and the precipitation time series 1801 to 2014. Both data sets are available on a regular grid of 5 minutes resolution (approx. 9.3 km in the Alps).

We generate synthetic experiments (see Sect. 2.3) for all glaciers in the Alps, and determine their glacier states in 1850 if the area of the observed synthetic state s_{2000}^{exp} is larger than ~~$0.01 km^2$~~ 0.01 km^2. This value is consistent with the minimum-area threshold of the RGI. The condition is satisfied for ~~2621~~ 2660 synthetic experiments of the 3927 glaciers included in the Randolph Glacier Inventory in Central Europe (region 11).

4 Results

Here we show ~~(i)~~ the results for two example glaciers in 1850 ~~,~~ as well as an error analysis for all tested glaciers ~~, and (ii)~~(Sect. 4.1), the influence of the choice of the fitness function on the quality of our results (Sect. 4.2), and a statistical analysis of the reconstructability of a glacier (Sect. 4.3).

4.1 Initial glacier states in 1850

Following the method described in Sect. 2.4, we determine reconstructed initial glacier states in $t_0 = 1850$. Figures 3 and 4 show the results of the Guslarferner, as an example glacier with a large set of accepted candidates. A second case with a more narrow set of acceptable states, the Hintereisferner, is shown in Fig. 5 and 6. More examples can be found in the supplementary material.

Especially the result of the Guslarferner shows clearly that the determination of past states is not unique (see Fig. 3). Multiple initial states (violet and blue colored) merge to the observed state in the year of observation. The fitness values, which means the averaged difference between the forward modelled states and the observation at $t_e = 2000$, are extremely small for most candidates. The fitness values of all candidates range from ~~1.6×10^{-5} to 1.26. Only two~~ 1.08×10^{-6} to 7.98. Only 16 of the 200 candidates ~~has~~ have a fitness value higher than one and thus do not fulfill the acceptance criterion (Eq. 5); for these states, the glacier in 1850 is too small to reach the volume of the observed glacier within 150 years. Also Fig.4 illustrates the diversity of the different acceptable solutions (grey, shadowed area). The length of all states in S_{1850}^{125} varies between ~~0.5 km and 8.3~~ 0.98 km and 8.1 km. The acceptance criterion in this example is not strong enough to provide any information about the searched state in $t_0 = 1850$, as any of the candidates would lead to an acceptable result. Figure 4 also shows the 5th percentile of all acceptable states (blue, shadowed area). This set contains the 5 % best solutions, based on the fitness value. All candidates of the 5th percentile are in close proximity to the synthetic experiment. The range of fitness values of all candidates of the 5th percentile is ~~$[1.6 \times 10^{-5}, 1.84 \times 10^{-4}]$~~ $[1.08 \times 10^{-6}, 7.95 \times 10^{-5}]$ and the length of the states in 1850 only varies from about ~~3 km to 4.1~~ 3.6 km to 5.3 km. All these candidates match the ~~observation~~ synthetic experiment in $t_e = 2000$ very well and converge to the synthetic experiment by ~~1880~~ 1900 at the latest, which can be seen in Fig. 4c. As a representative of this set, we choose the median state of the 5th percentile of S_t^{125} (in the following referred as s_t^{med}). Figure 4a shows that the surface elevation of s_{1850}^{med} in 1850 corresponds very well to the synthetic experiment, whereas the state with the minimum fitness value (in the following referred as s_t^{min}) ~~slightly~~ mismatches the synthetic experiment at the ~~tail~~ tongue of the glacier. Regarding the volumes, ~~neither~~ s_t^{med} ~~nor s_t^{min}~~ match exactly the volume of the synthetic experiment in 1850, ~~but the differences are small (0.007 km^3 for s_t^{med} and 0.008 km^3 for s_t^{min})~~ whereas the volume of s_t^{min} differs by 0.4 km^3.

~~The results of Hintereisferner are different to the results from Guslarferner~~ In the case of the Hintereisferner (see Fig. 5) ~~. The~~ the fitness values of most candidates are large compared to the ones of the Guslarferner. Only a few candidates have extremely small fitness values ~~. In the case of the Hintereisferner,~~ and the past state is thus much more narrowly confined. The different states need more time to adapt to the climate conditions and therefore they do not converge as quickly to one state. As a result, the differences between the forward modelled states and the observed one in 2000 are larger. The fitness values of all candidates range between ~~1.2×10^{-3} and 47.~~ 2.8×10^{-5} and 43.

~~68~~ 36 candidates fulfill the acceptance criterion (Eq. 5). Figure 6 shows that the acceptance criterion in this case confines the ~~method~~ result better than in the case of the Guslarferner. The length of all glaciers in S_{1850}^{125} range from ~~8.5 km to 15.7~~ 8.4 km to 12.3 km. In this case the 5th quantile of S_t^{125} is again in close proximity to the synthetic experiment and all candidates of the 5th quantile have extremely small fitness values (between ~~1.2×10^{-4} and 2.16×10^{-3}~~ 2.8×10^{-5} and 5.4×10^{-4}). The length of the candidates of the 5th quantile in 1850 only varies from ~~11.7 km to 12.4~~ 9.1 km to 10.3 km and is thus more precise than in the Guslarferner example. ~~Also in~~ In this example, all candidates of the 5th percentile converge ~~not later than 1880~~ no later than 1920 to the state of the synthetic experiment. ~~s_t^{med}~~ Here s_t^{min} matches the surface elevation of the synthetic experiment in 1850, as well as the volume trajectory over time, slightly better than ~~s_t^{min}~~ s_t^{med}, but the volume differences to the synthetic ex-

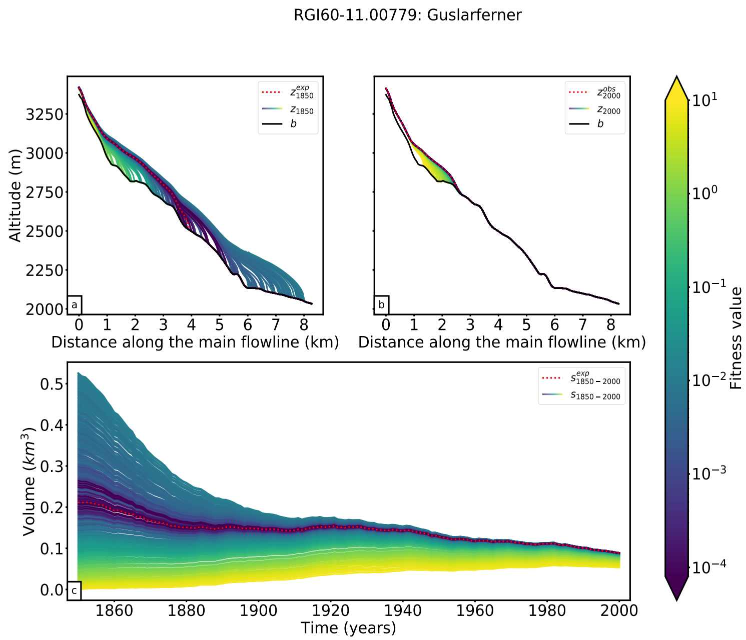

Figure 3. Results for the Guslarferner (Oetztal, Austria). Top: Cross-sections along the main flowline in **a:** 1850 and **b:** 2000. Black line indicates the bed rock, the red, dotted line the surface elevation from the synthetic experiment, and the remaining lines the modelled ice surfaces of all candidates, colored by their fitness value. The synthetic experiment state in 2000 has a length of 2.7 km, an area of 1.71 km^2, a volume of 0.09 km^3 and a mean thickness of 62.8 m. Bottom: Volume changes from 1850-2000, colored by their fitness values.

periments in 1850 are also very small in this example (~~0.007~~ 0.004 km^3 for ~~s_t^{med} and 0.03~~ s_t^{min} and -0.08 km^3 for ~~s_t^{min}~~ s_t^{med}).

For both examples we were able to show ~~,~~ that our method is able to ~~determine~~ recover the state in $t_0 = 1850$ of the synthetic experiment by only using information about the observed state of the synthetic experiment in $t_e = 2000$ and combining it with information about the past climate evolution. s_t^{med}, as well as s_t^{min} match the synthetic experiment in $t_0 = 1850$ extremely well. In the following, we provide an error analysis including all glaciers in the Alps ~~to~~ on which we applied our method ~~.~~ to. For each of the ~~2621~~ 2660 glaciers we have calculated the absolute volume error to the synthetic experiment:

$$e_{abs}^{med/min}(t) = v^{med/min}(t) - v^{exp}(t), \tag{6}$$

where $v^{exp}(t)$ is the volume of the synthetic experiment in year t and $v^{med/min}(t)$ is the volume of s_t^{med} or s_t^{min} in the same year t. Figure 7a shows the absolute volume errors in km^3 for s_t^{med}, as well as for s_t^{min} . ~~Median and the range between the 5th and 95th percentile ($Q_{0.05-0.95}$) of the absolute volume errors e_{abs}^{med} and e_{abs}^{min} over time.~~ Whereas the absolute volume errors in 1850 vary widely from approx. ~~-0.45~~ -1.1 km^3 to ~~8~~ 2.9 km^3, they reduce rapidly within ~~50~~ 60 years. In ~~1900~~1910, the errors range from approx. ~~-0.17~~ -0.25 km^3 to ~~0.3~~ 0.17 km^3. The range of errors in the first ~~50~~ 60 years is largely influenced

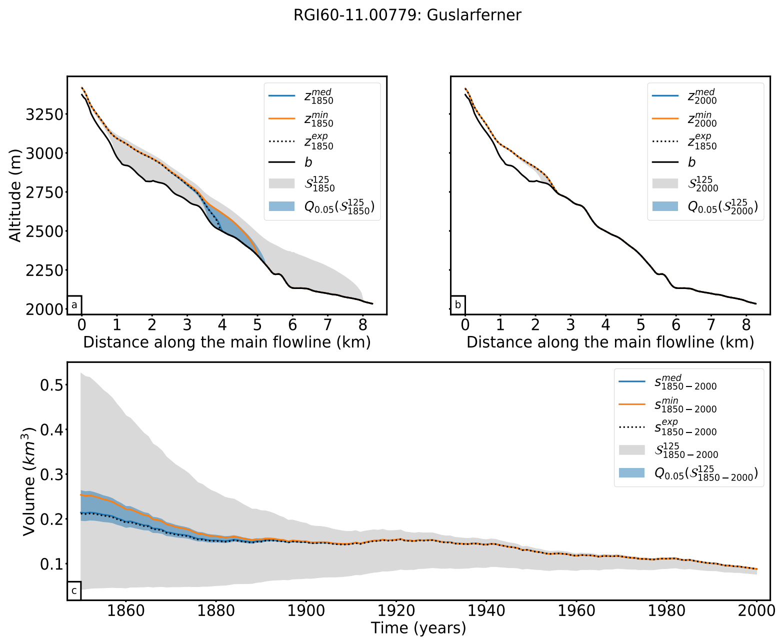

Figure 4. Results for the Guslarferner (Oetztal, Austria). Top: Cross-sections along the main flowline in **a:** 1850 and **b:** 2000. Bottom: Volume changes from 1850-2000. The grey shaded area indicates the range of all solutions with a fitness value smaller than one ($\mathcal{S}_t^{125}$). The blue shaded area shows the range of the 5th quantile of $\mathcal{S}_t^{125}$, the blue line s_t^{med}, and the orange line s_t^{min}.

by a few single outliers. Differences between s_t^{min} and s_t^{med} are small. Figure ~~??~~ 7b shows the median and the range of the 5th-95th percentile of e_{abs}^{med} and e_{abs}^{min} over time, indicating the robustness of our method. The median of e_{abs} of both analyzed states is very small; ~~0.0033~~ 0.00028 km^3 for s_t^{min} and ~~0.0044~~ 0.00076 km^3 for s_t^{med} in 1850. The improvement with time can also be seen here: the median of ~~$e_{abs}^{med}(1900)$~~ $e_{abs}^{med}(1910)$ is of the order of 10^{-5} km^3 and that of ~~$e_{abs}^{min}(1900)$~~ $e_{abs}^{min}(1910)$ of the order of 10^{-6} km^3 . ~~2490 of the 2621 sates with a minimal fitness value s_t^{min} (95 % of the tested glaciers) have in 1850 an absolute volume error smaller than 0.1 km^3. Regarding s_t^{med}, 95 % of the glaciers even have in 1850 error values smaller than 0.08 km^3.~~

As our test site contains large and small sized glaciers, we also evaluate relative errors (in %):

$$e_{rel}^{med/min}(t) = \frac{e_{abs}^{med/min}(t)}{v^{exp}(t)} * 100 \tag{7}$$

Figure 8a shows the histogram of the relative errors in 1850, whereas the ~~the~~ evolution from 1850-2000 of the median and the 5th-95th percentile of the relative errors are shown in Fig. 8b. The median of the relative volume errors in 1850 is ~~14.8~~ -0.97 % for s_t^{med} and ~~10.4~~ -2.69 % for s_t^{min}. The 95th percentile value of ~~e_{rel}^{min} is 286~~ e_{rel}^{med} is 70 %. With ~~248~~ 48% the value of ~~e_{rel}^{med} is slightly~~ e_{rel}^{min} is smaller for the ~~s_t^{med}~~ s_t^{min}.

Whereas ~~s_t^{med}~~ s_t^{min} have in 1850 a slightly smaller 5th-95th percentile range than ~~s_t^{min}~~ s_t^{med}, the median error of ~~s_t^{min}~~ s_t^{med}

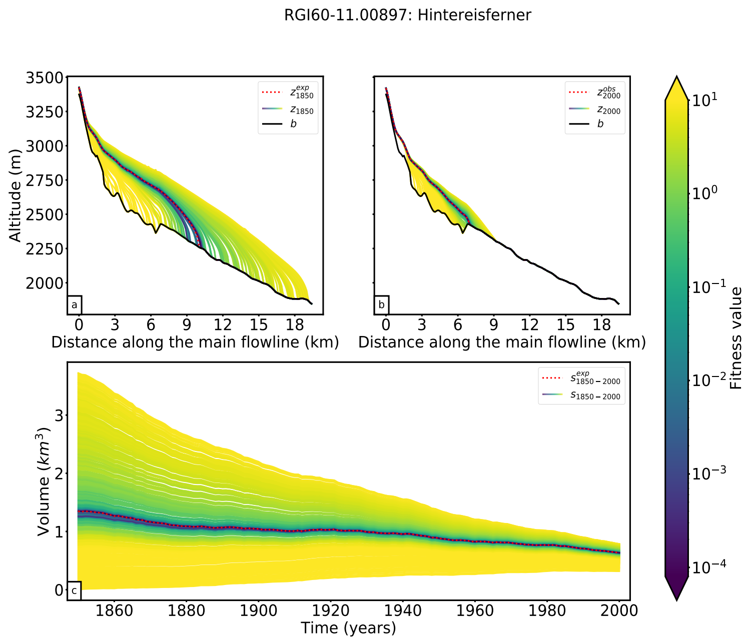

Figure 5. Results for the Hintereisferner (Oetztal, Austria). Top: Cross-sections along the main flowline in **a:** 1850 and **b:** 2000. Black line indicates the bed rock, the red dotted line the surface elevation from the synthetic experiment ,and the remaining lines the modelled ice surfaces of all candidates, colored by their fitness value. The synthetic experiment state in 2000 has a length of 7.3 km, an area of 7.76 km^2, a volume of 0.63 km^3 and a mean thickness of 105.5 m Bottom: Volume changes from 1850-2000, colored by their fitness values. All violet and blue glacier states merge to the observed glacier in 2000.

is slightly smaller than the one of $\cancel{s_t^{med}}$ s_t^{min}.

Both states fit well the synthetic experiment. In many cases, s_t^{med} is equal to s_t^{min}, but for some glaciers either s_t^{min} or s_t^{med} have a clearly better performance. In all cases, the uncertainties quickly reduce after around ~~1880 to 1900. The error analysis~~ 1900 to 1930.

Figure 8a also shows that the error distribution is skewed and our method has a slight tendency to ~~overestimate~~ underestimate the glacier volume~~, and that the error distribution is strongly skewed (see Figure 8a).~~. Although 64% of the relative errors have a negative sign, a few large positive outliers influence the mean error and shift it to a positive value of 16% (in 1850) for the minimum states or 23% (in 1850) in case of the median states.

4.2 Impact of the fitness function

For the evaluation of the glacier candidates we used a fitness function based on differences in the geometry of the glacier (see Eq. 5). In this section we want to test the influence of limited information on glacier geometry on the ~~reconstructabilty~~ reconstructability of past glacier states. Thus, we additionally evaluate the candidates by only using information about the

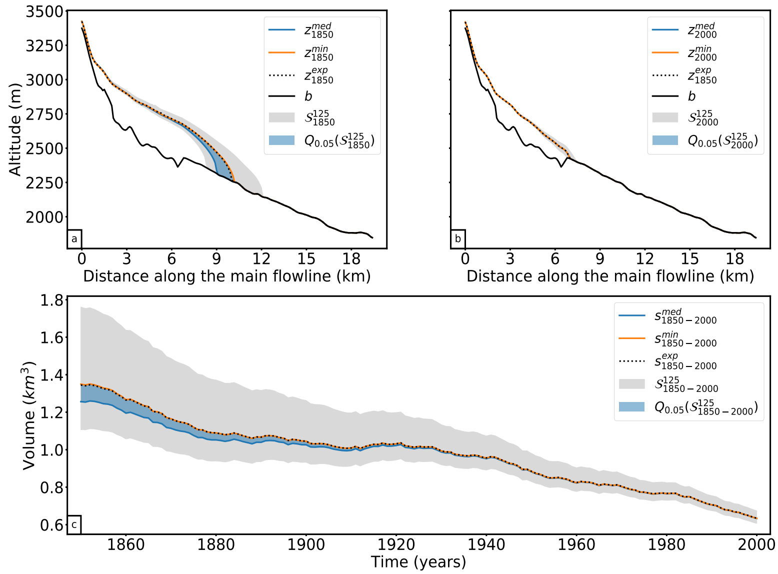

Figure 6. Results for the Hintereisferner (Oetztal, Austria). Top: Cross-sections along the main flowline in **a:** 1850 and **b:** 2000. Bottom: Volume changes from 1850-2000. The grey shaded area indicates the range of all solutions with a fitness value smaller than one ($\mathcal{S}_{1850}^{125}$). The blue shaded area shows the range of the 5th quantile of $\mathcal{S}_t^{125}$, the blue line shows s_t^{med}, and the orange line s_t^{min}.

glacier area or length.

For the glacier area based evaluation, we used the following fitness function:

$$J_A(A_{t_e}) = (A_{t_e}^{obs} - A_{t_e})^2 \tag{8}$$

where A_{t_e} is the glacier area at time t_e. The fitness function that takes only information about the glacier length $l(t_e)$ at time t_e into account is similar:

$$J_l(l_{t_e}) = (l_{t_e}^{obs} - l_{t_e})^2 \tag{9}$$

For each glacier in our test site, we evaluate the 200 candidates also with the fitness functions J_A and J_l. For each evaluation method (geometry, area and length based), we determine the state with the minimal fitness function [1]and calculate the relative volume error to the synthetic experiment.

Figure 9 shows the relative errors of all three evaluation methods. Figure 8a shows the distribution of the relative errors of the 5th-95th percentile in 1850, whereas the the evolution from 1850-2000 of the median and the 5th-95th percentile of the

[1]Instead, it is also possible to choose s_t^{med} for the uncertainty analyses, but this would require acceptance criteria for the fitness functions J_A and J_l, which would have influence on the state. For simplification, we choose the state with the minimal fitness function.

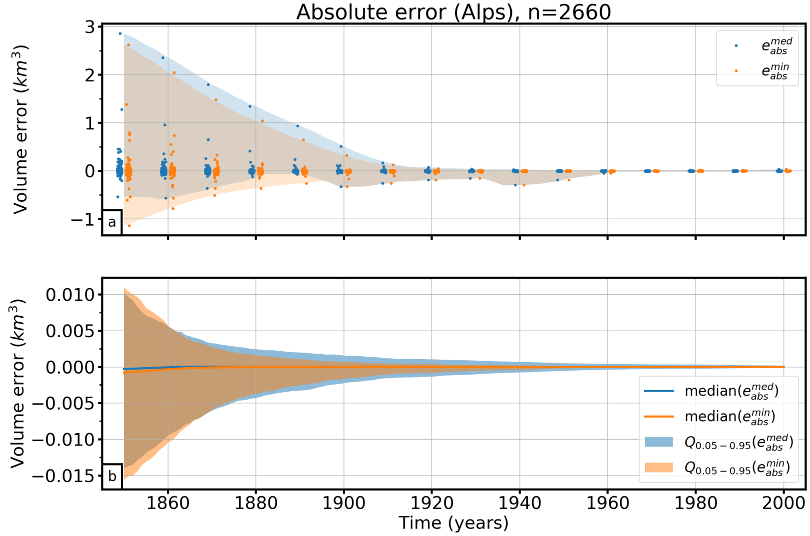

Figure 7. Absolute volume errors in km^3 over time of s_t^{min} and s_t^{med} of all tested glaciers. **a:** The blue points mark ~~the~~ all individual errors e_{abs}^{med} of s_t^{med} and the orange points mark ~~the~~ all individual errors e_{abs}^{min} of s_t^{min}. The blue, shadowed area shows the total range of the errors ~~of the~~ ~~s_t^{med}~~ e_{abs}^{med} and the orange, shadowed area the total range of errors ~~s_t^{min}~~ e_{abs}^{min} over time. **b:** The shadowed areas show the 5th and 95th percentile ($Q_{0.05-0.95}$) of the absolute volume errors e_{abs}^{med} (blue) and e_{abs}^{min} (orange) over time, as well as the median of e_{abs}^{med}(blue line) and e_{abs}^{min} (orange line).

relative errors are shown in Fig. 8b. The more information is taken into account for the evaluation, the smaller are the errors. The greatest uncertainties are associated with using the glacier length based fitness function (Eq. 9), whereas the differences between the area based evaluation (Eq. 8) and the geometry based evaluation (Eq. 5) are small. While the median errors in 1850 of the geometry and the area based evaluation are close (~~10.4~~-2.69% for the geometry and ~~11.5~~ -2.83 % for the area approach), the median error in 1850 of the glacier length evaluation has with ~~137~~107% the worst performance. This ~~holds also true regarding~~ also applies for the values of the 95th precentile; 95% of the tested cases have in 1850 a relative volume error smaller than ~~1008~~1043%, if the length based fitness function is used for the evaluation. In contrast to the other two evaluations, this approach overestimates the volume. For the area based evaluation ~~this value is 329~~95% of the tested glaciers have an error smaller than 90% and for the geometry based fitness function ~~95% of the tested glaciers have an error smaller than 286~~the error is smaller than 49%. This shows that the advantage using the geometry instead of the glacier area to evaluate the candidates is not very high; both evaluations shows a very good performance. Especially if the states are modelled forward (e.g. to 1900), both approaches perform well. However, it is not advisable to use the glacier length based evaluation.

4.3 Reconstructability

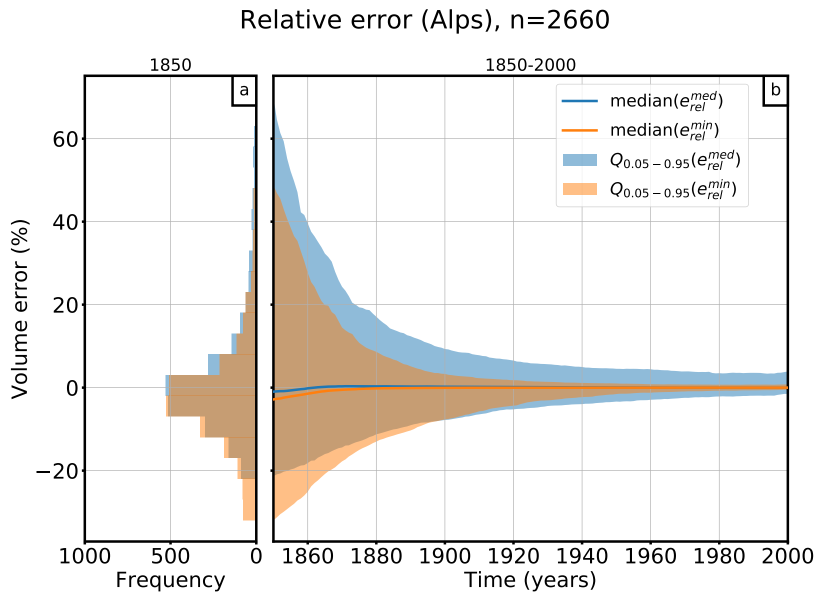

Figure 8. Relative volume errors of s_t^{med} (blue colored) and s_t^{min} (orange colored). Figure **a:** shows a histogram of all errors in the 5th-95th percentile in year 1850 and **b:** the evolution of the relative errors from 1850-2000. The line indicate the median error and the shadowed area the 5th-95th percentile range $Q_{0.05-0.95}$.

The examples from Sect. 4.1 as well as in the supplementary material indicate a high variation of the number of viable reconstructed candidates between glaciers. This number can range from a few viable solutions in a well defined range to many solutions without any constraints (all tested candidates have the same fitness value). In other words, some glaciers can be reconstructed easily, and some cannot.

We define a new measure of reconstructability r, where we set the volume range of the 5th percentile in relation to the volume range of all acceptable states of the glacier:

$$r = 1 - \frac{\mathrm{range}(Q_{0.05})}{\mathrm{range}(S^{125})} \tag{10}$$

For a glacier with a unique solution, this measure is equal to one. If all accepted candidates have exactly the same fitness value, the measure will be zero (this occurs if all candidates converge to exactly the same state before the year 2000). Thus, a small measure represents a glacier with low reconstructability and a measure close to one imply a higher reconstructability of the glacier. For example, r is equal to 0.857 for Hintereisferner, and 0.879 for Guslarferner. The similarity of the two values can be explained by the similar proportion of the range of the 5th percentile to the range of all acceptable states in both cases (see Fig. 3 and 6). A histogram of the reconstructability values of all 2660 tested glaciers in the Alps is shown in Figure 10a. The distribution is bimodal and slightly skewed towards a high reconstructability. Values in the middle range are rare.

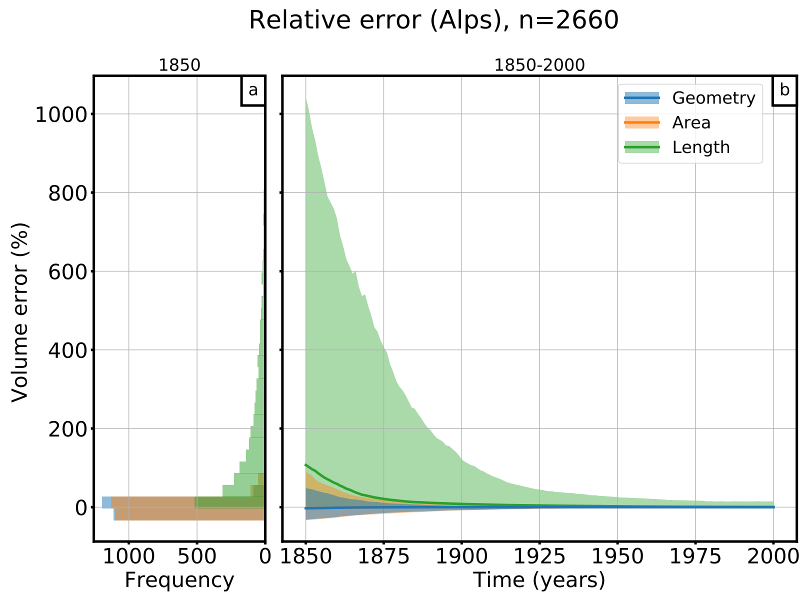

Figure 9. Relative volume errors of s_t^{min} derived from different fitness functions based on the geometry (blue), the glacier area (orange) and glacier length (green). Figure **a:** shows a histogram of all relative errors in the 5th-95th percentile in year 1850 and **b:** the evolution of the relative errors from 1850-2000. The line shows the median error and the shadowed area the 5th-95th percentile range.

What glacier characteristics will influence this reconstructability? The working hypothesis is that it is likely to be associated with the concept of glacier "response time" (here formulated qualitatively). Glaciers with a short response time tend to be less sensitive to initial conditions, and will "forget" their initial state after a short period of time. This will probably lead to low reconstructability values. Inversely, glaciers with a long response time will be easier to reconstruct.

To test this hypothesis, we used the e-folding approach (as defined in Oerlemans, 1997, 2001) and calculated the time response to a step function. To this end, we first run the 1850 state of the synthetic experiment glacier into an equilibrium state by using a constant climate (mean climate of the years 1835-1865, temperature bias = -1 K). We choose the same settings that were used for the generation of the synthetic experiments in order to obtain an equilibrium state s_{eq1} close to our synthetic experiment in 1850. Next, we apply to s_{eq1} a constant climate obtained by the mean climate of the years 1850-1880 using no temperature bias and receive the corresponding equilibrium state s_{eq2} (i.e. a step change of 1 K). We calculate the e-folding time of these two states for each glacier, but exclude the glaciers where the volume of s_{eq2} reaches zero (which was the case for approx. 500 glaciers out of 2660[2]).

The scatter plot in Fig. 10b indicates a relation between the reconstructability measure and response time. The variance of the response time increases for reconstructability values close to one. Dependencies with the reconstructability could also be detected for the position of the equilibrium line altitude (ELA) (Fig. 10c), the mean surface slope in 2000 (Fig. 10d) and the

[2] We also tested a step of 0.5 K, leading to a larger sample size but no significant change to the correlation analysis and our results. Thus, we kept the 1 K step change here.

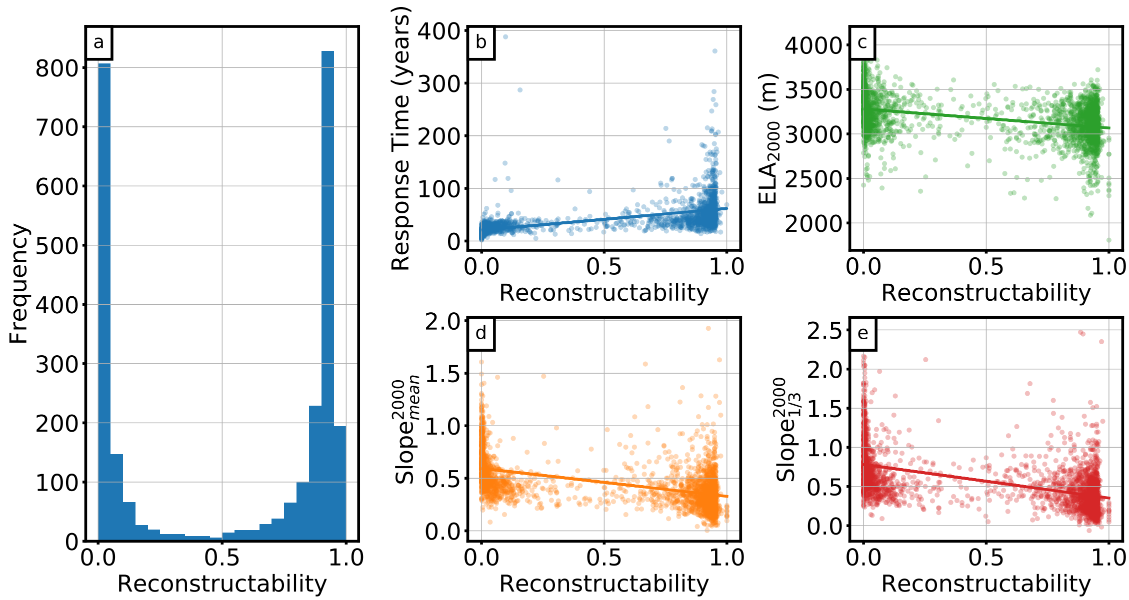

Figure 10. Reconstructability measure. **a:** Histogram of the reconstructability measure of all 2660 glaciers. **b-e:** Scatter plots with linear regression. The x-axis always shows the reconstructability measure. The y-axis shows **b:** the e-folding response time (n=2149), **c:** the equilibrium line atitude in 2000 (n=2660), **d:** the mean surface slope in 2000 (n=2660) and **e:** the mean surface slope of the last third of the glacier in 2000 (n=2660).

mean surface slope in 2000 of the last third of the glacier (Fig. 10e).
Furthermore, we calculated correlations of both reconstructability and response time with the following variables: glacier length (in 2000), area (in 2000), volume (in 2000), equilibrium line altitude (ELA) in 2000, equilibrium line altitude change from 1850-2000, mean surface slope (in 2000), and mean surface slope over the lowest third of the glacier (in 2000) (Fig. 11). The variable explaining reconstructability best is the glacier response time (correlation: 0.54). Both values correlate with the same glacier characteristics. Against a common misunderstanding, glacier length, area and volume do not correlate well with the reconstructability measure nor with the response time. The variable having the main influence is slope: generally, the larger the slope, the lower the reconstructability measure or the response time of a glacier. These findings coincide with results from Lüthi (2009), Zekollari and Huybrechts (2015) and Bach et al. (2018), who concluded that response times depend more on the steepness of the surface than on the glacier size. The correlation of the mean surface slope can be further increased by taking the lowest third of the glacier. Besides that, the position of the ELA in 2000 also influences the reconstructability, whereas the ELA change from 1850 to 2000 only plays a minor role.

Taken alone, these correlation values remain quite low and do not provide enough predictive power to create a statistical

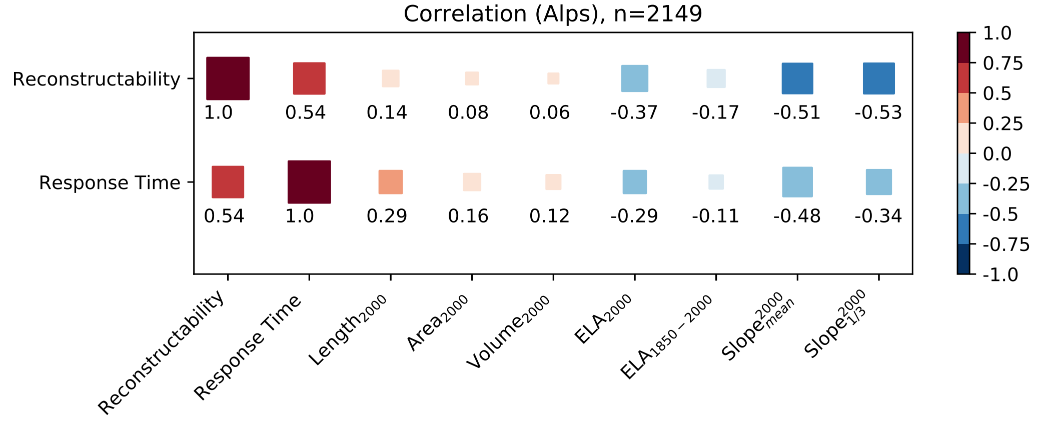

Figure 11. Correlation of the reconstructability measure and the e-folding response time to various glacier characteristics. Correlation values are represented by the square color and size.

model of “reconstructability”. However, they provide a good indication about which factors should be taken into account for future applications.

5 Hardware requirements and performance

For this study we used a small cluster comprising two nodes with two 14-core CPUs each, resulting in 112 parallel threads (two threads per core). Our method requires to run hundreds of dynamical model runs for each single glacier and, as described in Maussion et al. (2019), the dynamical runs are the most expensive computations. The size of the glacier and the required time stepping to ensure a numerical stability ~~influence strongly~~ strongly influences the required computation time. The computation time needed to apply our initialization procedure to one glacier varies from 30 seconds to ~~37~~ 26 minutes. In total~~initializing the 2621~~, initializing the 2660 glaciers in 1850 ~~took 3.5~~ takes about 3.75 days on our small cluster.

6 Discussion and conclusions

In this study, a new method to initialize past glacier states is presented and applied ~~in~~ to synthetic experiments. Assuming a perfectly known world allows us to identify the errors of our method alone and to separate them from uncertainties in observations and errors introduced by model approximations~~. A differentiation between the different error sources in a~~ , a task impossible to realize in real-world ~~application is difficult and thus an evaluation of the skill of the method would be difficult, too. For real world applications, we will have to validate the reconstructions~~ applications. However, the synthetic experiments do not allow external validation, e.g against past outlines derived from moraines, historical maps or remote sensing ~~(e.g. as provided by GLIMS, Raup et al., 2007).~~ (such as provided by GLIMS, Raup et al., 2007). Model uncertainties will have to be accounted for, and will have to be compared to and added to the theoretical lower bound discussed in this study. Similarly, our results do not provide information about actual past glacier mass change. Since in our synthetic experiments glaciers states in 2000 may be different from the real ones, the modelled initial glacier states in 1850 do not correspond to reality either. The past states determined in this study only can serve to verify the functionality of the developed method.

~~The results in Sect. 4.1~~ Our results have shown that the solutions are not unique. Multiple candidates match the observation in $t_e = 2000$, sometimes with a large spread. This raises interesting questions about the use of past glacier change information to reconstruct climate variations, which we don't address here. In the context of model initialization, this non-uniqueness is impeding the reconstruction. We evaluated the candidates with a fitness function based on averaged geometry differences between the forward modelled and the observed state in $t_e = 2000$. The threshold value $\epsilon = 125$ m^2 was derived by assuming a typical error of 5 m in surface elevations and 10 m in glacier width, but how these values should be chosen depends on the ~~situation~~specific glacier setting. Especially in cases where many of the candidate states have extremely small fitness values, a more strict acceptance criterion can help to narrow the results. ~~Otherwise~~On the other hand, an ϵ that is too strict could lead to none of the candidates fulfilling the criterion.

Due to uncertainties that derive from the model ~~itself~~integration, the glacier state with the minimal fitness value is not always close to the synthetic experiment. As a more robust alternative, we propose to use s_t^{med}, the median state of the 5th percentile of all acceptable states as the best estimate. In Sect. 4.1 we compared the errors of both approaches. ~~On the one hand, the~~ The median error of ~~s_t^{min}~~ s_t^{med} is slightly smaller than that of ~~s_t^{med}. On the other hand,~~ s_t^{min} and the total range of absolute errors is smaller for s_t^{med} in 1850, too. Modelling the reconstructed initial states forward in time approximately ~~50~~ 60 years leads to a rapid reduction of the error, and s_t^{min} ~~and s_t^{med} perform equally good.~~ perform a bit better than s_t^{med}. By making use of the knowledge about the past climate, the number of candidates at later stages are through this forward run more constrained than by initializing them directly at a later time (see appendix B for a more detailed description of the inverted approach at different times).

~~In Sect. ?? we compare~~ By comparing different fitness functions for the candidate evaluation~~. We show~~ , we showed that using limited information only (glacier area and glacier length) lead to an increase of the errors in 1850. This indicates what kind of observation is needed to be able to reconstruct past glacier states from today's state. The differences between the geometry based evaluation and the area based evaluation are small, but the differences to the length based evaluation are significant. ~~This reflects that glacier states with the same length could differ strongly in volume and area.~~ But this effect is also influenced by the spatial resolution of the model grid: a higher resolution of the grid would lead to more variability in fitness values and hence to a more precise initialization. At the same time, a higher resolution would increase the computational demands of the initialization method. We strongly recommend to use either the geometry or the area based fitness function for the evaluation. In this study, we only take the observation of the year t_e into account. Multi-temporal outlines ~~would~~ are likely to greatly reduce uncertainties at prior times.

Our results are relevant for future glacier evolution modeling studies, as they indicate that at least for some glaciers the time needed to converge to a similar evolution regardless of the 1850 state is comparatively short. Our study might also be useful to determine a good starting point of a past simulation, e.g. to improve the initialization date in Zekollari et al. (2019). A correlation analysis of the ~~estimated reconstruction spread and error with glacier characteristics (e.g. size, slope, ice thickness distribution per altitude, climatic setting, etc.) may lead to improved understanding of the factors that influence the reconstructability of a glacier's past states, and will be the topic of a separate study~~reconstructability and glacier characteristics showed the position of the ELA, as well as the slope (especially in the lower part of the glacier) influence the reconstructabilty, whereas attributes like the glacier size do not have a strong impact. We could also show that the reconstructabilty measure correlates well with an separately obtained response time of the glacier.

Future work will ~~also~~ include the application of the method on real-world cases, which will come with additional challenges. For example, we will have to consider the merging of neighboring glaciers when growing~~(a work already under way, Dusch, 2018)~~. Importantly, the effect of uncertainties in the boundary conditions (in particular the glacier bed, its outlines and uncertainties in the climate forcing) will have to be quantified. This also includes to test the influence of the choice of climate conditions on the accuracy of our method. Here again, the ~~"surrogate world"~~ synthetic framework will be useful by allowing data denial and data alteration experiments. To ensure the robust reconstruction of real-world glacier states additional changes and model developments are necessary~~, but our study is a first important step in this direction~~. This includes e.g. the development of a glacier-individual calibration method for dynamical parameters (e.g. sliding parameter, creep parameter) as well as of the mass-balance model.

Code availability. The OGGM software together with initialization method are coded in the Python language and licensed under the GPLV3 free software license. The latest version of the OGGM code is available on Github (https://github.com/OGGM/oggm), the documentation is hosted on ReadTheDocs (http://oggm.readthedocs.io), and the project webpage for communication and dissemination can be found at http://oggm.org. The OGGM version used in this study is v1.1. The code for the initialization module is available on Github (https://github.com/OGGM/initialization).

Appendix A: Temperature bias for the synthetic experiments

For the generation of the synthetic experiment state in 1850, we use a temperature bias of -1 K in order to create a relatively big glacier state. To justify the choice of this value, we have tested different temperature biases: the results are summarized in Fig. A1. This figure shows that applying positive or small negative temperature biases to the synthetic experiments results in large area differences to the RGI in 2000, and the total glacierized area in 1850 is also too small. The sample size is reduced, because less glaciers fulfill the area threshold criteria of 0.01 km^2. Negative temperature biases that are too large also reduce the sample size, because some runs fail (the glacier gets larger than the underlying grid). The experiments with a temperature bias of -1K, -1.25K or -1.5K perform best regarding the area difference to the RGI in 2000. But only the experiment with the temperature bias of -1K performs good regarding the estimation in 1850 of Zemp et al. (2006), whereby it needs to be taken into account that the dots only represent a subset (the small glaciers in 2000 are missing) of the glaciers considered in (Zemp et al., 2006).

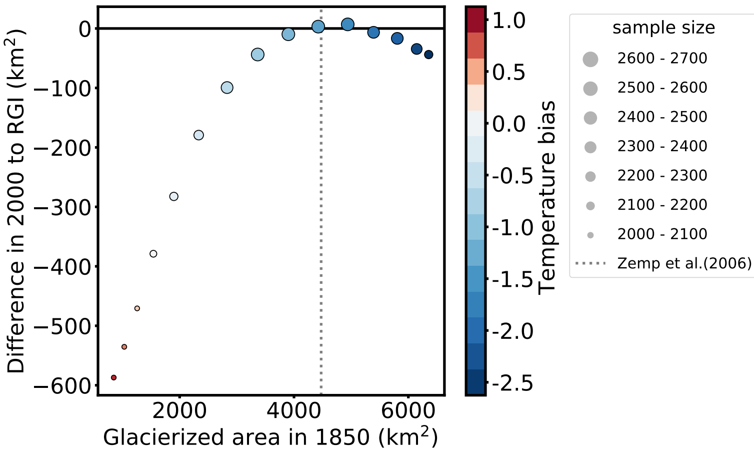

Figure A1. Difference between the total area in 2000 to the total area from the RGI plotted as a function of total area in 1850. Colors mark the applied temperature bias to create the synthetic experiments, and the size of the points mark the sample size (number of glaciers with an area larger than 0,01 km^2 in 2000). The dashed grey line marks the estimated total area of all Alpine glaciers in 1850 from (Zemp et al., 2006)

Appendix B: Initialization at different starting times

We applied our method to different starting times (1850, 1855, ..., 1965) to test how far one can go back in time to get a good initial state for this glacier. While this inverted setup is computationally very expensive, unfortunately it does not lead to improved results. See Fig. B1 for two different examples. For each tested starting year, we determined the median state and conducted an uncertainty analysis (similar to the one in Sect. 4.1). We find that the uncertainties of the median states at the different starting points are higher than doing the initialization for the year 1850 (only) and running this state forward in time. While this is counter-intuitive, the main reason is that by starting in 1850 even with a very large number and range of candidates, the very unrealistic ones are quickly forced to converge by the boundary conditions (i.e., by climate), effectively reducing the number of potential candidates for a later date. In other words, we make use of our knowledge about past climate to reduce the number of candidates at each later stage. In real-world applications, results might be different since uncertainties in past climate are large. While this should be explored further, because of the computational cost it is hard to imagine an eventual applicability on the global or even large regional scale.

Author contributions. JE is the main developer initialization module and wrote most of the paper. BM and FM are the initiators of the OGGM project and helped to conceive this study. FM is the main OGGM developer and participated in the development of the initialization module.

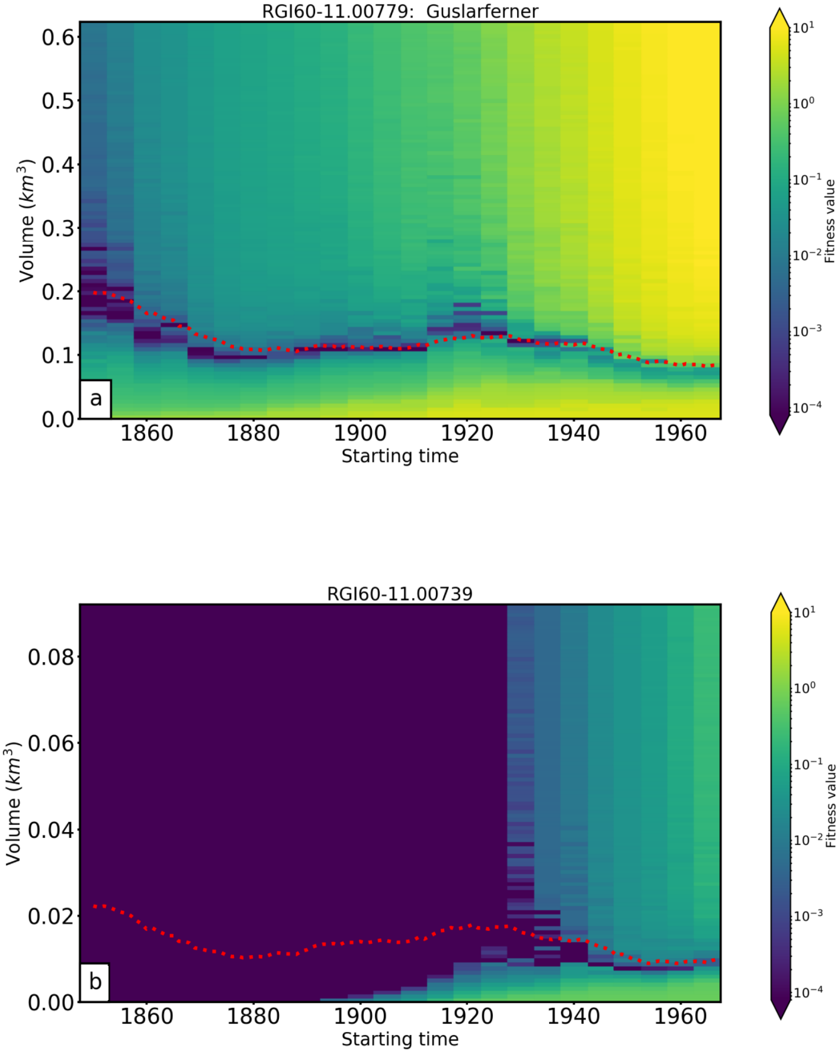

Figure B1. Reconstructability for different starting times. Colors indicate the fitness value of a simulation initialized with a glacier volume indicated by the vertical axis at a time indicated by the horizontal axis. Red dotted line shows the synthetic experiment. Upper panel: example for a glacier with ordinary reconstructive power; the “observed” glacier state in 2000 constrains the past evolution well in the 20th century, and the reconstruction is close to the goal. Lower panel: example for a glacier with very low reconstructive power; the ”observed” glacier state does not constrain the past glacier evolution before approx. 1930.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. BM, and JE were supported by the German Research Foundation, grant MA 6966/1-1.

References

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplova, E.: HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region, International Journal of Climatology, 27, 17–46, https://doi.org/10.1002/joc.1377, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1377, 2007.

Bach, E., Radić, V., and Schoof, C.: How sensitive are mountain glaciers to climate change? Insights from a block model, Journal of Glaciology, 64, 247–258, https://doi.org/10.1017/jog.2018.15, 2018.

Bamber, J., Westaway, R. M., Marzeion, B., and Wouters, B.: The land ice contribution to sea level during the satellite era, Environmental Research Letters, https://doi.org/10.1088/1748-9326/aac2f0, 2018.

Cazenave, A. and et al.: Global sea-level budget 1993–present, Earth System Science Data, 10, 1551–1590, https://doi.org/10.5194/essd-10-1551-2018, 2018.

Church, J., Clark, P., Cazenave, A., Gregory, J., Jevrejeva, S., Levermann, A., Merrifield, M., Milne, G., Nerem, R., Nunn, P., Payne, A., Pfeffer, W., Stammer, D., and Unnikrishnan, A.: Sea Level Change, book section 13, p. 1137–1216, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.026, www.climatechange2013.org, 2013.

Colton, D. and Kress, R.: Inverse acoustic and electromagnetic scattering theory., vol. 93, Berlin: Springer-Verlag, 1992.

Dusch, M.: OGGM feature development: "Allow individual glaciers to merge together", https://github.com/OGGM/oggm/pull/624, [Online; accessed 28-May-2019], 2018.

Farinotti, D., Huss, M., Bauder, A., Funk, M., and Truffer, M.: A method to estimate the ice volume and ice-thickness distribution of alpine glaciers, Journal of Glaciology, 55, 422–430, https://doi.org/10.3189/002214309788816759, 2009.

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, https://www.the-cryosphere.net/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, 12, 168 – 173, https://doi.org/10.1038/s41561-019-0300-3, 2019.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., and Paul, F.: A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009, Science, 340, 852–857, https://doi.org/10.1126/science.1234532, http://science.sciencemag.org/content/340/6134/852, 2013.

Giesen, R. H. and Oerlemans, J.: Calibration of a surface mass balance model for global-scale applications, The Cryosphere, 6, 1463–1481, https://doi.org/10.5194/tc-6-1463-2012, https://www.the-cryosphere.net/6/1463/2012/, 2012.

Giesen, R. H. and Oerlemans, J.: Climate-model induced differences in the 21st century global and regional glacier contributions to sea-level rise, Climate Dynamics, 41, 3283–3300, https://doi.org/10.1007/s00382-013-1743-7, https://doi.org/10.1007/s00382-013-1743-7, 2013.

Goelzer, H., Nowicki, S., Edwards, T., Beckley, M., Abe-Ouchi, A., Aschwanden, A., Calov, R., Gagliardini, O., Gillet-Chaulet, F., Golledge, N. R., Gregory, J., Greve, R., Humbert, A., Huybrechts, P., Kennedy, J. H., Larour, E., Lipscomb, W. H., Le clec'h, S., Lee, V., Morlighem, M., Pattyn, F., Payne, A. J., Rodehacke, C., Rückamp, M., Saito, F., Schlegel, N., Seroussi, H., Shepherd, A., Sun, S., van de Wal, R., and Ziemen, F. A.: Design and results of the ice sheet model initialisation experiments initMIP-Greenland: an ISMIP6 intercomparison, The Cryosphere, 12, 1433–1460, https://doi.org/10.5194/tc-12-1433-2018, https://www.the-cryosphere.net/12/1433/2018/, 2018.

Goosse, H., Barriat, P.-Y., Dalaiden, Q., Klein, F., Marzeion, B., Maussion, F., Pelucchi, P., and Vlug, A.: Testing the consistency between changes in simulated climate and Alpine glacier length over the past millennium, Climate of the Past, 14, 1119–1133, https://doi.org/10.5194/cp-14-1119-2018, https://www.clim-past.net/14/1119/2018/, 2018.

Gregory, J. M., White, N. J., Church, J. A., Bierkens, M. F. P., Box, J. E., van den Broeke, M. R., Cogley, J. G., Fettweis, X., Hanna, E., Huybrechts, P., Konikow, L. F., Leclercq, P. W., Marzeion, B., Oerlemans, J., Tamisiea, M. E., Wada, Y., Wake, L. M., and van de Wal, R. S. W.: Twentieth-Century Global-Mean Sea Level Rise: Is the Whole Greater than the Sum of the Parts?, Journal of Climate, 26, 4476–4499, https://doi.org/10.1175/JCLI-D-12-00319.1, https://doi.org/10.1175/JCLI-D-12-00319.1, 2013.

Harris, I. C., Jones, P. D., Osborn, T., and Lister, D.: Updated hight-resolution grids of montly climatic observations- the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, https://doi.org/10.1002/joc.3771, 2014.

Heimbach, P. and Bugnion, V.: Greenland ice-sheet volume sensitivity to basal, surface and initial conditions derived from an adjoint model, Annals of Glaciology, 50, 67–80, https://doi.org/10.3189/172756409789624256, 2009.

Henderson, L. S. and Subbarao, K.: 'Inverse Crime' and Model Integrity in Lightcurve Inversion applied to unresolved Space Object Identification, The Journal of the Astronautical Sciences, 64, 399–413, https://doi.org/10.1007/s40295-016-0105-1, https://doi.org/10.1007/s40295-016-0105-1, 2017.

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

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Frontiers in Earth Science, 3, 54, https://doi.org/10.3389/feart.2015.00054, https://www.frontiersin.org/article/10.3389/feart.2015.00054, 2015.

Jacob, T., Wahr, J., Pfeffer, W. T., and Swenson, S.: Recent contributions of glaciers and ice caps to sea level rise, Nature, 482, https://doi.org/https://doi.org/10.1038/nature10847, 2012.

Jarvis, A., Guevara, E., Reuter, H., and Nelson, A.: Hole-filled SRTM for the globe : version 4 : data grid, published by CGIAR-CSI on 19 August 2008., 2008.

Jóhannesson, T., Raymond, C. F., and Waddington, E. D.: A Simple Method for Determining the Response Time of Glaciers, in: Glacier Fluctuations and Climatic Change, edited by Oerlemans, J., pp. 343–352, Springer Netherlands, Dordrecht, 1989.

Kienholz, C., Rich, J. L., Arendt, A. A., and Hock, R.: A new method for deriving glacier centerlines applied to glaciers in Alaska and northwest Canada, The Cryosphere, 8, 503–519, https://doi.org/10.5194/tc-8-503-2014, https://www.the-cryosphere.net/8/503/2014/, 2014.

Leclercq, P. W., Oerlemans, J., and Cogley, J. G.: Estimating the Glacier Contribution to Sea-Level Rise for the Period 1800–2005, Surveys in Geophysics, 32, 519, https://doi.org/10.1007/s10712-011-9121-7, https://doi.org/10.1007/s10712-011-9121-7, 2011.

Lee, V., Cornford, S. L., and Payne, A. J.: Initialization of an ice-sheet model for present-day Greenland, Annals of Glaciology, 56, 129–140, https://doi.org/10.3189/2015AoG70A121, 2015.

Lüthi, M. P.: Transient response of idealized glaciers to climate variations, Journal of Glaciology, 55, 918–930, https://doi.org/10.3189/002214309790152519, 2009.

Marzeion, B., Jarosch, A. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, The Cryosphere, 6, 1295 – 1322, https://doi.org/10.5194/tc-6-1295-2012, 2012.

Marzeion, B., Cogley, J. G., Richter, K., and Parkes, D.: Attribution of global glacier mass loss to anthropogenic and natural causes, Science, 345, 919–921, https://doi.org/10.1126/science.1254702, http://science.sciencemag.org/content/345/6199/919, 2014.

Marzeion, B., Leclercq, P. W., Cogley, J. G., and Jarosch, A. H.: Brief Communication: Global reconstructions of glacier mass change during the 20th century are consistent, The Cryosphere, 9, 2399–2404, https://doi.org/10.5194/tc-9-2399-2015, 2015.

Marzeion, B., Kaser, G., Maussion, F., and Champollion, N.: Limited influence of climate change mitigation on short-term glacier mass loss, Nature Climate Change, 8, 305–308, https://doi.org/10.1038/s41558-018-0093-1, 2018.

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T., and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, Geoscientific Model Development, 12, 909–931, https://doi.org/10.5194/gmd-12-909-2019, https://www.geosci-model-dev.net/12/909/2019/, 2019.

Mosbeux, C., Gillet-Chaulet, F., and Gagliardini, O.: Towards a better ice sheet model initialisation and basal knowledge using data assimilation, Geoscientific Model Development Discussions, pp. 1–24, https://doi.org/10.5194/gmd-2016-7, 2016.

Oerlemans, J.: Climate Sensitivity of Franz Josef Glacier, New Zealand, as Revealed by Numerical Modeling, Arctic and Alpine Research, 29, 233 – 239, https://doi.org/10.80/00040851.1997.12003238, http://www.jstor.org/stable/1552052, 1997.

Oerlemans, J.: Glaciers and Climate Change, Swets and Zeitlinger, Lisse, 2001.

Pfeffer, W. T., Arendt, A. A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., Hagen, J. O., Hock, R., G., K., Kienholz, C., Miles, E. S., Moholdt, G., Mölg, N., Paul, F., Radić, V., Rastner, P., Raup, B. H., Rich, J., Sharp, M. J., and the Randolph Consortium: The Randolph Glacier Inventory: a globally complete inventory of glaciers, Journal of Glaciology, 60, 537–551, https://doi.org/10.3189/2014JoG13J176, 2014.

Radić, V. and Hock, R.: Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise, Nature Geoscience, 4, 91–94, https://doi.org/10.1038/ngeo1052, 2011.

Radić, V. and Hock, R.: Glaciers in the Earth's Hydrological Cycle: Assessments of Glacier Mass and Runoff Changes on Global and Regional Scales, Surveys in Geophysics, 35, 813–837, https://doi.org/10.1007/s10712-013-9262-y, 2014.

Raup, B., Racoviteanu, A., Khalsa, S. J. S., Helm, C., Armstrong, R., and Arnaud, Y.: The GLIMS geospatial glacier database: A new tool for studying glacier change, Global and Planetary Change, 56, 101 – 110, https://doi.org/https://doi.org/10.1016/j.gloplacha.2006.07.018, 2007.

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

Slangen, A. B. A., Adloff, F., Jevrejeva, S., Leclercq, P. W., Marzeion, B., Wada, Y., and Winkelmann, R.: A Review of Recent Updates of Sea-Level Projections at Global and Regional Scales, Surveys in Geophysics, 38, 385–406, https://doi.org/10.1007/s10712-016-9374-2, https://doi.org/10.1007/s10712-016-9374-2, 2017a.

Slangen, A. B. A., Meyssignac, B., Agosta, C., Champollion, N., Church, J. A., Fettweis, X., Ligtenberg, S. R. M., Marzeion, B., Melet, A., Palmer, M. D., Richter, K., Roberts, C. D., and Spada, G.: Evaluating Model Simulations of Twentieth-Century Sea Level Rise. Part I: Global Mean Sea Level Change, Journal of Climate, 30, 8539–8563, https://doi.org/10.1175/JCLI-D-17-0110.1, https://doi.org/10.1175/JCLI-D-17-0110.1, 2017b.

van Pelt, W. J. J., Oerlemans, J., Reijmer, C. H., Pettersson, R., Pohjola, V. A., Isaksson, E., and Divine, D.: An iterative inverse method to estimate basal topography and initialize ice flow models, The Cryosphere, 7, 987–1006, https://doi.org/10.5194/tc-7-987-2013, https://www.the-cryosphere.net/7/987/2013/, 2013.

WGMS: Fluctuations of Glaciers Database, World Glacier Monitoring Service, Zurich, Switzerland, https://doi.org/10.5904/wgms-fog-2018-11, 2018.

Wouters, B., Gardner, A. S., and Moholdt, G.: Global Glacier Mass Loss During the GRACE Satellite Mission (2002-2016), Frontiers in Earth Science, 7, 96, https://doi.org/10.3389/feart.2019.00096, https://www.frontiersin.org/article/10.3389/feart.2019.00096, 2019.

Zekollari, H. and Huybrechts, P.: On the climate–geometry imbalance, response time and volume–area scaling of an alpine glacier: insights from a 3-D flow model applied to Vadret da Morteratsch, Switzerland, Annals of Glaciology, 56, 51–62, https://doi.org/10.3189/2015AoG70A921, 2015.

Zekollari, H., Huss, M., and Farinotti, D.: Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble, The Cryosphere, 13, 1125–1146, https://doi.org/10.5194/tc-13-1125-2019, https://www.the-cryosphere.net/13/1125/2019/, 2019.

Zemp, M., Haeberli, W., Hoelzle, M., and Paul, F.: Alpine glaciers to disappear within decades?, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL026319, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL026319, 2006.

Zemp, M., Frey, H., Gärtner-Roer, I., Nussbaumer, S. U., Hoelzle, M., Paul, F., Haeberli, W., Denzinger, F., Ahlstrøm, A. P., Anderson, B., and et al.: Historically unprecedented global glacier decline in the early 21st century, Journal of Glaciology, 61, 745–762, https://doi.org/10.3189/2015JoG15J017, 2015.

Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J., Barandun, M., Machguth, H., Nussbaumer, S. U., Gärtner-Roer, I., Thomson, L., Paul, F., Maussion, F., Kutuzov, S., and Cogley, J. G.: Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016, Nature, 568, 382–386, https://doi.org/10.1038/s41586-019-1071-0, https://doi.org/10.1038/s41586-019-1071-0, 2019.