The Referee's comments below are in italics, our answer in plain font

In this manuscript, Goosse and co-authors examine the possibility of using modelled glacier length changes to investigate modelled past climatic conditions. To model past glacier changes, they use a newly developed state of-the-art glacier model (OGGM) (Maussion et al., 2018), while for the climatic aspect of the work they rely on variables derived from the PMIP3 and the CMIP5 protocols. A total of 71 glaciers in the European Alps are modelled throughout the previous millennium, and from this it is clear that the modelled length fluctuations are far more sensitive to differences in modelled climate (which is used as an input to drive OGGM), than to the OGGM model parameters. This suggests that modelled glacier length changes can be used as an interesting alternative/complimentary approach to other widely used palaeo-archives (e.g. trees, pollen,...) that are typically used to evaluate model performance.

I greatly enjoyed reading this manuscript. The results are generally well presented and the text is compact, well written and clear. Although not being an expert in past climate variability, I was able to follow the main points and the conclusions drawn from this. Some passages in the text were not entirely clear to me, and sometimes a few additional references would be welcome. I think that with a bit of reworking, this manuscript will be of great value to this journal and its readership. My list of comments may seem long in first instance, but most are suggestions that should be easily implemented.

We would like to thank the Referee for their very careful evaluation and for the suggestions that will greatly help us to improve the quality of our manuscript.

General comments

• *Some of the glaciological aspects of the manuscript were only very shortly treated, and in some cases a bit more explanation / additional framing would be welcome. I understand that most of the OGGM model description is summarized in Maussion et al. (2018), but it would be nice if the manuscript could be read without having to refer to the other one. My (sometimes rather specific) comments on this are detailed below.*

We will expand in the revised version the description of the glaciological aspects of the manuscript, as suggested. The proposed modifications will be described below in the answers to the specific comments.

• *Not all figures were straightforward to interpret and for some of them I had to look several times before I got the main message. A few suggestions on how to possibly improve these figures are formulated at the end of this review.*

We will modify the figures following the suggestions of the Referee in order to make their interpretation more straightforward.

1. Introduction

• *p.2, l.12-14: the emphasize is on the fact that glacier changes are mainly driven by changes in summer temperatures and winter precipitation. But what about insolation changes and other changes related to incoming radiation (e.g. aerosols, volcanism,..)? Aren't these important, especially on longer time scales and before the industrial revolution, when the strong anthropogenic atmospheric temperature signal was not yet present. It would be good if a few words would be spent on this. Furthermore, the effect of precipitation is in fact barely described throughout the manuscript and the focus is almost solely on (summer) temperatures. Some other authors pone that winter precipitation is very important and may for instance have been one of the main mechanisms behind the 1800-1850 glacier advance (e.g. Vincent et al., 2005).*

This sentence in the introduction is very general and we do not expect to go into the details at this stage but we will include in the revised manuscript the potential impact of incoming radiation.

The potential role of precipitation will be expanded at the beginning of section 3.

'Enhanced winter precipitation has been suggested to be an important contributor to some past changes in glacier length in the Alps (e.g., Vincent et al., 2005; Steiner et al. 2005; Steiner et al. 2008). Nevertheless, summer temperature is generally considered as the major driver of European glacier fluctuations at centennial timescales (Oerlemans, 2001; Steiner et al. 2005; Huss et al., 2008; Steiner et al., 2008; Leclerq and Oerlemans, 2012; Zekollari et al., 2014). It is thus instructive to compare first the simulated temperatures with reconstructions before analysing the glacier themselves.'

Additionally, we have repeated the equivalent of Figure 6 for winter precipitation (Figure R1). The link with glacier length changes is much weaker than for summer temperature. It is also less straightforward to interpret. Temperature and precipitation trends are positively correlated in some experiments, with a potential influence on the correlation between glacier length changes and precipitation. This might explain why, for instance, larger precipitation rates are generally associated with a retreat of the glaciers on the panel b of Figure R1. It is also possible that models underestimate the contribution of precipitation changes on glacier length because of a too low decadal variability of precipitation over the Alps, as suggested for some other regions (e.g., PAGES Hydro2k Consortium, 2017; Seftigen et al. 2017), and because of their coarse resolution that limits their ability to reproduce adequately precipitation changes in mountainous areas. Nevertheless, investigating those processes is out of the scope of the present study. Consequently, we have chosen to not include this additional figure in the manuscript.

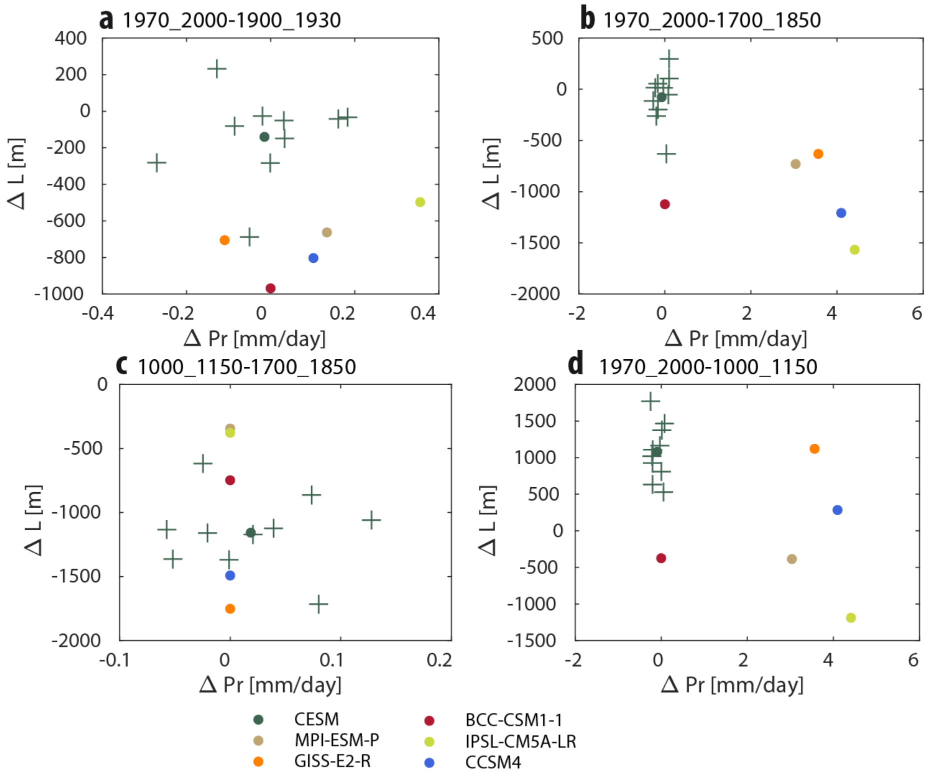

Figure R1. Glacier length changes as a function of winter precipitation (DJF) changes for the differences between a) 1970-2000 and 1900-1930; b) 1970-2000 and 1700-1850; c) 1000-1150 and 1700-1850; d) 1970-2000 and 1000-1150. The crosses represent the individual CESM ensemble members, the ensemble mean being represented by a dot of the same color.

p.2, l.14-15: "Furthermore, glaciers integrate forcing over timescales ranging from a few years to several decades or even centuries". This is correct, and is related to their response time. I would *suggest to explicitly state this here, as this may not be clear to a non-glaciologist, and add some references to some of the classic works on this (e.g. Johannesson et al., 1989; Leysinger Vieli and Gudmundsson, 2004). It would also be in better harmony with your next sentence, in which you describe other records as having ”a much faster response”.*

As suggested, we will add 'because of their long response time' and the references in the revised version.

• *p.2, l.17: would also refer to a recent study of Roe et al. (2017) here, in which glacier retreat is described as an evidence for changing climatic conditions, and in which it is also explained that these changes can not directly be compared to changes in climatic conditions and related proxies (which they circumvent by utilizing a signal-to-noise analysis).*

The reference will be added in the revised version.

• *p.2 l.19-21: many references, but only to recent studies, which gives the impression that this is a recent study field. This is in fact not the case, as simple glacier models have been used since a long time to better understand past climatic conditions. Here a reference to some pioneering studies, such as for instance to Allison and Kruss (1977) and Oerlemans (1986) would be justified. Notice that these older studies used simplified ice flow models that are in fact very similar to the flow model used in the OGGM model. The main difference resides of course in the fact that OGGM can be applied at a much larger scale (cf. 'Global' in its nomenclature), while in the earlier studies typically only one glacier was modelled to better understand past conditions. The authors could also decide on mentioning this in the text. Furthermore a reference could also be added to the recent study by Doughty et al. (2017).*

The suggested references will be added.

• *p.2, l.27-29: cf. earlier comment. What about insolation changes?*

Temperature and precipitation provide just an example. This will be specified in the revised version.

• *p.3, l.7-9: from the sentence and the references, it looks like Farinotti et al. (2017) is also a glacier modelling study. This is not the case, as this study treats ice thickness / bedrock elevation modelling, which is in its turn an important input of glacier models. I would suggest removing the reference to Farinotti et al. (2017) here and instead mention it later: ”...to the glacier model rather than to the climate model” ! ”...to the glacier model and its input/boundary conditions (e.g. the ice thickness (Farinotti et al., 2017)) rather than to the climate model”.*

The reference will be moved in the revised version, as suggested.

• *p.3, l.10-13: when reading this, I found it to be a bit strangely placed here and obstructing the flow of the introduction with little added value. Consider omitting or moving this to the discussion section?*

Internal variability is potentially an important cause of disagreement between simulated and observed glacier length. We consider that it is necessary to cite the key references in the introduction but the last sentence of the paragraph will be suppressed in the revised version to facilitate the flow.

• *p.3, l.17-18: did not understand this sentence in first instance, as I was confused with the 'modelled as well as reconstructed temperature changes'. It only became clear to me what this meant when I read section 3 and got that first the temperatures (modelled and reconstructed) are compared and subsequently they are linked to the glacier fluctuations. Maybe slightly reformulate this sentence to make this point clearer.*

This sentence will be reformulated in the revised version. As testing the compatibility between simulated and reconstructed temperatures is not the main goal of our study, this part of the sentence will be suppressed to avoid confusion.

• p.3, l.21-22. Maybe strange that I suggest this, as the authors of this particular study are co-authors on the study presented here, but think it would also be nice to refer to the very recent study by Marzeion et al. (2018).

The reference will be added.

• p.3, l.23-24: initial focus is on the European Alps, because here the records are long enough for analysis. To my knowledge, there are also several glaciers in Scandinavia with a very long record (e.g. Leclercq et al., 2014). Why where these glaciers not used/excluded? It seems that all necessary data available needed for this study (i.e. data that was used for European glaciers) would also be available for those glaciers.

There are 14 Scandinavian glaciers in the compilation of Leclercq et al. (2014) that meets our criteria (data covering at least the 20th century). Some of those glaciers are strongly influenced by the inflow of marine air, while some others have a more continental climate. This means that the number of samples that corresponds to similar climatic conditions is relatively limited. Consequently, biases in simulated length for one or two specific glaciers, which are expected for a model like OGGM, could have a large impact on the average. With 71 glaciers in relatively similar climatic conditions, the sensitivity of the average to biases in a few glaciers is much smaller in the Alps. We thus expect our results to be more robust in this area. This is the reason why we have chosen to focus on the Alps first. In the near future, we expect to repeat our analysis globally to check how the conclusions differ for different areas.

2.1 Climate model results

• p.4, l.4-5: reference to (Otto-Bliesner et al., 2009). This is a reference to a workshop, related to PMIP2 (vs.PMIP3 in text). Any better/newer reference available?

In the revised version, we will cite for PMIP3 the specific paper describing the forcing that has to be applied for the past millennium and a general paper analyzing PMIP3 simulation results.

2.2 The Open Global Glacier Model

• There is one main, rather crucial part in the model, that I am missing in this section. This relates to modelling/calculations for the regions that are not ice covered today. I understand that it is not the goal here to go into too many details regarding OGGM, but I think this part should be elaborated. It is very difficult to model the dynamics/evolution of the prefrontal areas that are not ice covered today with a simple ice flow model as several issues arise: e.g. how will the shape of the transects look? In which direction/where does the ice flow? And this part is rather crucial in this study, as most of the changes modelled here occur in this pre-frontal area (compared to the present-day period). For studies focusing on the future evolution of glaciers, this is less important, as they are assumed to shrink in the future and the 'action' occurs in regions where the treatment of for instance the cross sectional shape is more straightforward. As the authors (nicely!) show later, the choice of OGGM model parameters is not crucial for their modelled evolution over the past millennium, but despite this I think it would still be important that the 'prefrontal action/modelling' is explained in more detail here.

A central information needed in OGGM to compute the flow is the shape of the glacier bed and the slope along the flowline, which can be computed for areas that are not ice covered today (with actually less uncertainties than for some of the glaciated areas), so we expect the glacier forefront to be well simulated. Specifically, as explained in Maussion et al. (2018), for the prefrontal areas, we compute the direction and path of ice flow by computing the route from the glacier tongue toward the end of the domain that is the least costly in terms of positive altitudinal change. The flowline therefore follows the valley as a river would do. Along this flowline, we estimate the shape of the bed by fitting a parabola to the intersection points between the actual topography and the normal to the flowline.

A brief description of those processes will be included in the revised version.

• p.4, l.18-19. RGI version 5. Is the latest version of the RGI (RGI 6) not automatically used in OGGM? Not a big deal anyway, since the outlines did not change for Europe between these two versions (I think).

We made the first simulations with RGI version 5, which was the latest version available at that time, and did not update the data set as we checked that there are no differences for the selected European glaciers, as suggested by the reviewer.

• p.4, l.30: melting occurs if the monthly temperature is above -1C. Why was this limit chosen? It would not seem unreasonable to have some melting also in months with a lower mean monthly temperature. If based on a study/observations, would be good if could be mentioned.

This temperature was chosen on the result of a cross-validation procedure similar to the one conducted by Marzeion et al. (2012). The model is not very sensitive to the choice of this value as the calibration procedure of the mass-balance model is very robust and efficient in compensating for unrealistic temperatures as well as approximations inherent to the simple temperature index model (Maussion et al. 2018).

The way this value is obtained will be included in the revised version.

• p.5, l.3: 'based on the shallow-ice approximation'. Is evident was this is for glaciologists (for modellers at least), but here the main public are not glaciologist (or not only limited to them). Would be good if you could explain the shallow-ice approximation in a sentence (horizontal scale much larger than vertical scales considered, ice flow depends on local geometry,..etc.) and add a reference to Hutter (1983).

As suggested by the reviewer, a brief description of the shallow-ice approximation will be added in the revised version "In the shallow ice approximation, the vertical variations of ice flow are neglected and only a depth-integrated ice velocity is computed. This is a common approximation for computationally efficient ice flow models, and it is largely valid as long as the considered horizontal scales are much larger than the vertical scales. (Hutter 1983)."

• p.5, l.5-7: strangely formulated. Here you say that the rate factor mainly changes as a function of ice temperature (which is true!). But in fact all glaciers in the European Alps are (very close to) their melting point, i.e. they are temperate glacier. So by reading this, it does not really make sense that there would be a wide spread in the values of the rate factor. However, due to impurities, ice fabrics, crystal orientation,...etc, the values for the rate factor vary by quite a lot, even for the glaciers in the European Alps (I come back to this point in my comments on section 4). Would be good if you can reformulate this passage.

We agree with the reviewer that the sentence may be confusing as it mixes information on the processes that may affect *A* in the real world, at global scale, and the origin of the uncertainty of *A* in OGGM when applied to the Alps. As our goal is not to get into the details here, we propose to simplify the sentence and mention 'while in reality *A* may change by a factor 10 between glaciers due to a wide range of processes (Cuffey and Paterson, 2010) '.

• *Basal sliding is not treated/neglected it seems. This is an acceptable approach, and has been used in several studies for Alpine glaciers (e.g. Gudmundsson, 1999), as ice motion due to internal deformation and basal sliding is typically occurring in the same places (e.g. Zekollari et al., 2013). This treatment of basal sliding should briefly be mentioned somewhere.*

Basal sliding can be included in OGGM but it is not activated in our simulations. This will be specified in the revised version.

2.3 Glacier length observations

• *p.5, l.23-25: for the historical sources, (old topographical) maps should also be named, as they are widely used to document past changes. Would suggest adding a reference to Purdie et al. (2014) also. Could also update (Nussbaumer and Zumbühl, 2012) to a more recent work of these authors: Zumbühl and Nussbaumer (2018).*

As suggested, we will add in the revised version the old maps as an example of historical source and will cite the suggested references.

• *p.5, l.27: Leroy et al. (2015) ! Le Roy et al. (2015).*

Thanks. This will be corrected.

• *p.5, l.28-30: see my earlier comment regarding glaciers in Scandinavia with a long length record and the question why these are not considered here.*

As discussed above, we have preferred to focus this first study on the European Alps only.

3. Simulated and reconstructed glacier changes

This section starts by comparing the modelled temperatures with observations. Subsequently the observed and modelled length variations are explained. The observed ("reconstructed") changes were already described earlier, and therefore I found the title of this subsection not entirely logical. Also taking into account that first temperatures are described, it would be more logical to name this subsection 'Simulated temperatures and glacier changes' or 'Simulated versus reconstructed glacier changes'. I found this elaborate section well written and liked the short paragraphs and summarizing sentences at the end of various parts (e.g. p.7, l.22-24), which made it easy to follow.

As suggested, we will modify the title. As the focus is on the glacier, not on temperature, we propose to use 'Simulated glacier changes'.

• *p.7, l.7: "trend" ! "growth trend"?*

We will modify the text as suggested.

• *p.7, l.14-24: would merge these two paragraphs, as the second one is a continuation of the first one.*

The two paragraphs will be merged in the revised version.

• *p.7, l.34: "have a large positive trend": of what? Temperature trend?*

It is a trend in the length of the glacier. This will be specified in the revised version by using 'growth trend'.

• *p.8, l.28: "between Alpine regions or glaciers": found this a bit vague. Maybe just change to "between regions" (clear that is in Alps), or simply drop this part of the sentence.*

In the revised version, we will use 'between regions' because it is indeed clear that it is in the Alps.

• *p.8, l.28: Leroy et al. (2015) ! Le Roy et al. (2015).*

This will be corrected, thanks.

• *p.8, l.33-34: sentence not entirely clear. Consider reformulating*

The whole paragraph will be reformulated: 'The early 15^{th} century corresponds to a minimum for glacier advances in many models (Fig. 5) and a relative minimum in glacier length (Fig. 3). Although the simulated temperatures are generally mild during this period, they are not high and, in particular, are generally lower than in the beginning of the millennium (Fig. 1). This clearly illustrates the impact of the long response timescales of glaciers. The simulated glacier retreats in the early 15^{th} century appears as partly due to the temperatures at that time but also to the recovery from the large advances in the 13^{th} and 14^{th} century.'

• *p.9, l.1-2: "have a long response timescales" ! "have long response timescales"*

We will modify the text as suggested.

Sensitivity of glacier changes to model parameters

• *p.9, l.22-23: the rate factor is doubled as a sensitivity experiment. Why is there no experiment with a lower rate factor (e.g. halving)? In various (flowline) modelling studies in which the rate factor was calibrated (e.g. to reproduce observed surface velocities), lower values than the 2.4E-24 adopted here were obtained.*

Three studies that have for instance found/adopted a (much) lower rate factor:

– 2E-24 $Pa^{-3}a^{-1}$ (Le Meur et al., 2004)

– 6E-25 $Pa^{-3}a^{-1}$ (Stroeven et al., 1989)

– 6E-25 $Pa^{-3}a^{-1}$ (Letréguilly and Reynaud, 1989)

As suggested, we have made an additional experiment with a creep factor divided by two. The difference with the standard solution are also small, not modifying our conclusions. This additional experiment will be included in the revised version. Additionally, to save space, we will only show in the revised version the results for CCSM4 as we consider that the results for one model only are sufficient to illustrate the weak sensitivity of the results to OGGM parameters (the results of two models were shown in the submitted version).

• *p.10, l.9: "similar results have been obtained for other ones" ! "similar results have been obtained for other climate models"?*

This sentence will be modified as suggested.

• *p.10, l.9-10: "The results for CESM ensemble" ! "The results for the CESM ensemble"*

This will be corrected, thanks.

5. Conclusions

• *p.10 last lines - p.11, first lines. Discussion about the end of the LIA and the potential role of black carbon and how a recent study by Sigl et al. (2018) shows that this is unlikely. Lüthi (2014) also showed that the role of black carbon is most likely very limited to non-existent to explain the observed retreat. Also mention this here?*

The reference will be added in the revised version.

• *p.11, l.6-11. Not entirely sure whether your results 100% support this statement. Some simulations based on a particular climate model may result in a correct retreat, but this may in some cases be related to other reasons. By this, I hint in the direction of the role of other important variables for the glacier SMB, such as the winter precipitation and the insolation, which are not really elaborately treated/described in your work. Would be nice if this could be slightly reframed.*

For us, the most important point is that the skill of the climate models can be tested despite the contribution of internal variability and the uncertainty in OGGM parameters or boundary conditions. A disagreement between observed and simulated glacier length for the 'robust diagnostics' indicates thus likely a climate model bias. This bias could be either on simulated temperature or precipitation, which are the two driving climatic variables in OGGM (a disagreement due to insolation changes would have an effect only indirectly as it is accounted in OGGM through the effect of insolation on temperatures). Nevertheless, if an independent comparison between simulated and reconstructed temperature confirms the underestimation of temperature changes, it is compatible with the hypothesis that a bias in temperature is at the origin of the underestimation of glacier length fluctuations. To underline that our hypothesis about a role of temperature changes, as discussed here, does not at all discard a potential contribution of other variables, those sentences will be reframed in the revised version as:

'Nevertheless, some diagnostics appear robust enough for assessing the overall climate models skill in the region studied. In particular, some simulations underestimate the amplitude of the glacier changes between the 18th century and the end of the 20th century. This disagreement may have several origins, such as model biases in temperature or precipitation changes. However, an independent comparison between simulated and reconstructed temperature suggests that those models have a too weak warming over the past two-three centuries, suggesting an important contribution from this variable in the glacier model behavior.'

• *p.11, l.15-16: could you maybe mention how this relates to other regions? (cf. work of Solomina et al., 2015, 2016) i.e. how does in other regions the present-day length compare to minimum glacier lengths over the past 1000 years?*

This is indeed potentially interesting to compare the conclusion from the Alps to other regions but our simulations are only for the Alps and the evolution of glacier length is geographically complex. We would thus be obliged to make strong hypotheses and the resulting discussion would not be very informative. Therefore, we consider that this comparison is out of the scope of our manuscript and prefer to let this point to future studies.

• *p.11, l.18: "simulated and observed values". What does "values" refer to here? Climate or glacier length?*

This is glacier length. This will be specified in the revised version.

• p.11, l.19-21: How can a strong retreat be modelled for this period if there is no real temperature signal? Your story is build up around the dominant effect of (summer) temperatures on glacier MB (and thus length fluctuations). How come then that you model a glacier retreat for this period? Is this related to their response time (i.e. reacting to an earlier T signal) or is there a signal in the precipitation for instance? You make it sounds as if the cause-effect mechanism is unclear, but as you model this, it should be feasible to find this out, no?

We are sorry that this sentence was unclear. This is related to a point raised above in a paragraph that we proposed to rephrase. The difference is simply due to the long response timescales of the glaciers. This will be mentioned explicitly in the revised version.

• p.11, last sentence: consider adding "...of past climate reconstructions"

This will be added in the revised version.

Figures

• Fig.1, 2, 3, 5 & 6: found it very difficult to distinct between CESM and CCSM4, as the lines have almost the same colour. Would be nice if more distinct colours could be used.

The figures will be updated using more distinct colors as suggested.

• Fig.1: nice to have the reconstruction a bit thicker, so that it can be distinguished in a relatively easy manner from the modelled ones. It would be nice if you could also do this (i.e. bolder line for the 'Observation') for the length reconstructions (Fig. 2, 3, 5, 7). Especially in Fig.5, it is difficult to distinct the observed form the modelled values in an intuitive way.

The thickness of the line corresponding to observations will be thicker in the revised version.

• Fig.2, caption: "The reference period is 1901-1930": drop "the years" (cf. caption Fig. 3)

The captions will be modified as suggested.

• Fig.4a,b: had to look several times at the figure before I got it. Find it a bit strange that you present the observation as the last 'column' in each figure, in the same style as the modelled values. As it is a single value, it would be nice if the observations could shown as a dotted horizontal line that crosses the entire box. This would make the figure easier to interpret.

We agree that showing the observations as a horizontal line will make the figure easier to interpret. This will be done for the revised version.

• Fig.6: a comment in the same line as the previous one: it is difficult to detect what the observation is. Given its different nature (observed vs. modelled), a different symbol for the observation, e.g. larger, with different face and edge colour for instance, would be adequate here.

A larger black square will be used for observations in the revised version. Additionally, in the submitted version, we made the link with temperatures averaged over the whole Europe while in the revised version we will show the temperature averaged over the Alps. This does not modify the conclusions but this choice is more logical as it is the local temperature variations that are important for the glaciers.

References

PAGES Hydro2k Consortium, 2017. Comparing proxy and model estimates of hydroclimate variability and change over the Common Era. Clim. Past, 13, 1851-1900, https://doi.org/10.5194/cp-13-1851-2017.

Seftigen K., H. Goosse, F. Klein, and D. Chen, 2017. Hydroclimate variability in Scandinavia over the last millennium - insights from a climate model-proxy data comparison. Climate of the Past 13, 1831–1850, https://doi.org/10.5194/cp-13-1831-2017.