I agree with the previous reviewers, that the questions that are being addressed in the manuscript are very interesting and fit to the Cryosphere. Significant improvements have been made based to the manuscript based on the previous reviews. Unfortunately the improvements are in my opinion insufficient to recommend publication in its current state. I like the revised introduction and most of it is easy to read, minor comments you will find below. I have however trouble understanding the experimental set-up that was used. The methods need to be clarified, as there are a couple of things that seem inconsistent and/or confusing to me. This is being specified below, together with some other detailed comments on the section. I will keep the remaining part of this review very brief, because based on the method description, I am not certain about what the model set-up is and therefore on how to value the results.

Without a version with marked changes of the manuscript it was very hard to track what improvements have been maybe. However I found some comments that neither have been addressed in the rebuttal letters nor implemented in the revised manuscript. I would recommend the authors to go over the reviews again and check if they addressed all comments. Though this maybe sounds all very negative, I think the changes that are required might be easy to implement and I see potential in the paper.

2. Data and model description

2.1 As you're analyzing changes in glacier length, I would recommend to spend a few words on describing the dynamic part of the model in this section. (At it is current state the reader will only find out it is a flowline model in the "Limitation and Extensions" section.)

Line 88: *"OGGM is an open source model of glacier evolution, which couples a surface mass balance (SMB) model based on solid precipitation and temperature with a model of glacier dynamics."*

This sentence makes me wonder, did you force your simulations with solid precipitation? If so why, did you do so? (OGGM is calibrated with total precipitation and computes the solid fraction based on that.) Please clarify the type of precipitation you used, either here or in line 111 for instance.

Line 94: "*using CRU 3.21 temperature and precipitation data for the 20th century (Jones and Harris, 2013)*"

OGGM uses CRU TS v4.01 (Harris et al., 2014) anomalies with respect to the CRU CL (New et al., 2002) in its calibration procedure (see Maussion et al., 2019 Appendix A for the details). In this context I am wondering, to which CRU data are you referring in the section Line 111-115. Is this the by OGGM preprocessed CRU climate data? If so this needs to be clearly indicated. If not, why do you use a different climate than was used in the calibration procedure, when applying the anomaly method to the GCM data?

Line 109: Again be more specific. What type of temperature (reference height temperature?) and precipitation (total/ solid) did you use?

Line 115-116: "*The processing is done within OGGM, as its default setting.*" and Line 119, "*using OGGM's default preprocessing at level 3"* contradict what is being stated in line 108-110. "*The GCMs provide gridded monthly records of temperature and precipitation, which OGGM uses to determine temperature and precipitation at each specific glacier location using locally-calculated lapse rates for each.*"

OGGM uses by default a lapse rate of 6.5 C/km. This makes me wonder, did you use a different lapse rate in your forward runs than was used in the calibration? Additionally I wonder, where did you get the "locally calculated lapse rate" from? (To my knowledge this is not an implemented option in the OGGM function that process the GCM data.)

Line 118-123: It is odd to start the "Glacier observations" section with the model set-up. The remaining part of the this sections is in contrast to the first part of the method section written less messy and more easy to follow.

Line 123: Maybe you could mention the inversion method somewhere.

Fig. 2: Consider adding the names of the regions in the legend.

Fig. 2: I think the word "number" is missing before "of glaciers".

3. Experimental description

Section 2 and 3: The order in which the methods are described, was inconvenient for me to read, mostly because parts of the experimental set-up is described spread throughout the "data and model description" section and the other parts are in the "experimental description". I recommend to make section 2 strictly a description of the model and data, or merge the two sections. Either way I would just describe the model set-up that you used in one section, instead of spread over multiple sections with other descriptions in between.

Line 188: Do the spin-up runs with the different forcing results in very different glacier lengths? Why did you decide to start from the different simulations from a different initial state?

Line 190: "*For all runs, a 300-year spinup using annual climate data selected randomly from a 51-year window of 875-925 CE from the same model is performed prior to the run, to allow the glacier to develop from the preprocessed glacier geometry (based on RGI data and therefore representative of the year of the observation used) to a more realistic geometry for the climate near the start of the model run."*

I find the statement quite strong. What is it based on?

4-7. Results – Conclusion

- How sensitive are the results to the year that was picked for normalization (line 148-149)?

- Line 215: Are there glaciers in the Leclercq dataset that are expected to have been retreating throughout the last millennium?

- Fig 3&4. Ref table 1. It is not clear from the table which glaciers are included in the plot. Are the glaciers that have failed during one of the the simulations included in figure 3 and 4?

- I'm wondering, how do your results for Europe differ from those in Goose et al. (2018)?

- Line 347-365: Are you aware that there is a function in OGGM which can merge the glacier flowlines? (https://docs.oggm.org/en/latest/generated/oggm.workflow.merge_glacier_tasks.html) Have you tested the claim that you make in the last part of this section?

- Line 376-378: Consider rewording.

- I noticed that very little references to the literature are being made in the *Discussion* and *Limitations and Extensions* section.

Minor details:
- The comment on line 25 (previously line 21), by reviewer 3, hasn't been addressed nor included: *most relevant study from Oerlemans to supportthis statement is Oerlemans (2005)*
- I got lost in the sentence that starts in line 24 and end in line 28. Where does *",however (Zemp et al., 2015; Cogley, 2009),"* refer too?
- Line 33: I think the reference to the website in redundant.
- Line 36: Update the reference to the CRU version that is being used by OGGM v1.1.
- Line 37: I think it would be nice to name the version of the RGI that you used some where. (I know this can be looked up with the reference. However I think it is common practice to name the version some where.)
- Line 46-47: I think the it would be appropriate to use a reference for the timing of the LIA and MWP.
- I would like to repeat the suggestion of reviewer 3, to try to avoid multiple repetitions of brackets. (e.g. line 32 & 118). In addition please remover double brackets (e.g. Line 69).
- e.g. line 208: I would like to suggest to use the names of the regions and not their numbers.