Response to RC1 on "*tc-2021-352*"

We would like to thank Prof. David Parkes for reviewing our paper again. Major changes are summarized here followed by point-to-point responses to each comment. Reviewer's comments are in black while responses are in blue.

General comments:

The revised paper offers considerable improvements over the previous submission and addresses the most significant concerns I had at that stage. In particular, the treatment of LIA substages is considerably enhanced by the addition of a technical description of how they are identified. The improved clarity of the figures is also appreciated.

I still have some questions over how well the moraine ages can be confidently associated with LIA substages identified in the OGGM output; my understanding is that the moraine ages seem to show that a series of substage maxima were hit for a number of glaciers in each of the LIA substages, but not how these cases can be linked to the four regional substages in the output. However, this is a relationship that would probably need to be examined on a per-glacier basis which is beyond the scope of this paper and potentially beyond the capabilities of OGGM runs calibrated regionally rather than to individual glaciers. It is probably not worth adding material on this topic to the paper at this point beyond perhaps some discussion of how future work might tie the observations and the modelled glaciers together more closely.

I would recommend the paper as it stands for publication given the minor changes detailed below. Whether the above point on how regional and individual scale variations can better be tied together should be left to the authors' discretion.

Please note that I have had limited time to perform this review and I have therefore been less selective about language than I would have liked, with requests for changes mostly restricted to cases in which I felt the meaning was obscured rather than technical errors where the meaning was still clear.

Reply: Thanks for your encouragement and suggestions! We agree that the moraine ages used in our study just show a series of substage maxima were hit for a number of glaciers in each of the LIA substages. Observations (dated moraines; Fig. 4b) indicate the occurrence time of glaciers' *maximum peak GLR* is different across different individual glaciers, which supports the dispersion of the simulation results (Fig. 6a). In addition, we can also infer from the observations that more glaciers advanced to their maximum extents at the early LIA rather than the late LIA from a probability point of view.

Actually, except for the glaciers only with one LIA substage (Fig. 5), several moraines should be found for a glacier. However, in most field investigations, some moraines are destroyed by the post glacial advances, fluvial process, and frost weathering. If all moraines of a glacier could be well-preserved and accurately dated, we can directly compare the dated moraine ages with the simulated substages ages. Under this circumstance, the influence of glacier properties on the number of LIA substages will be further explored.

Specific comments/technical corrections:

1. Figure 8: The readability of this figure and the information that is being provided appear negatively impacted by the choice of reference year: what I understand here is that in the 1950s, September and October were particularly cold and August was particularly dry, and these have considerable skewed the visuals away from a focus on the LIA substage bands that are the intended focus. Could this graph instead use the 20th century average as a reference, to try to highlight the features of the dataset which are of greater interest here.

Reply: The original figures (using 1950s as a reference) and the new figures (using the 20th century average as a reference) were shown below. We found only slightly differences between them, indicating the warmer September and October and wetter August during the LIA is not caused by the abnormal climate in 1950s. Instead, as we have written in the original paper, we found the abnormal temperature changes in September and October and precipitation change in August are more likely polluted by the GISS climate dataset (Fig. R3). Therefore, as our paper mainly focuses on the glacier length changes relative to 1950, we keep using the original figures in the revised paper.

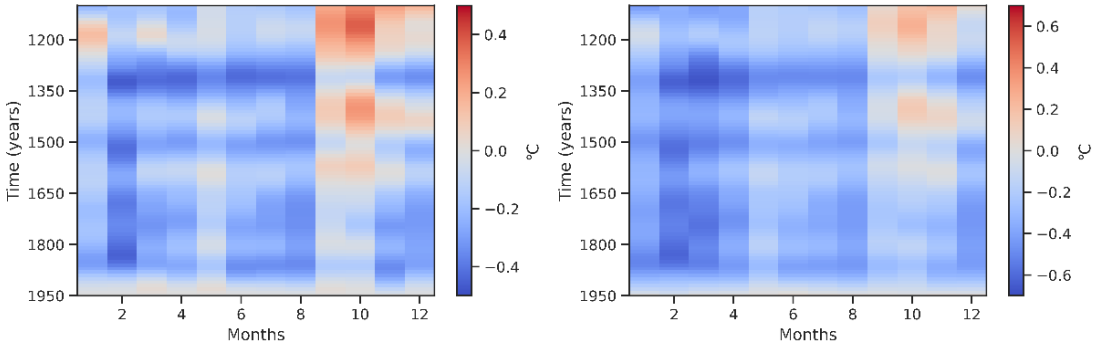

Fig. R1 The monthly temperature changes relative to 1950s (left) and the 20th century average (right). Please notice the different color bars between two figures.

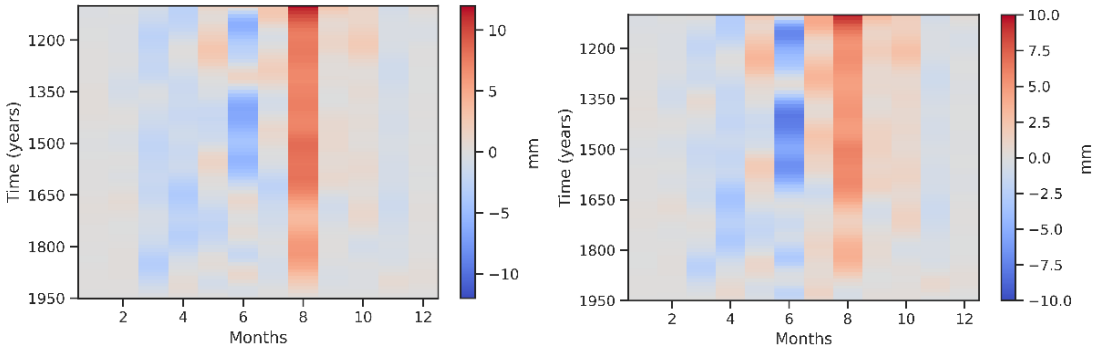

Fig. R2 The monthly precipitation changes relative to 1950s (left) and the 20th century average (right).

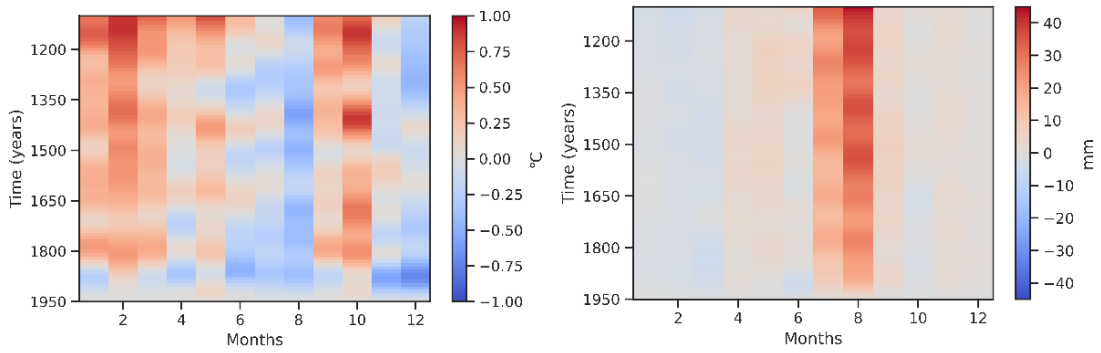

Fig. R3 The monthly temperature (left) and precipitation (right) changes relative to 1950s in the GISS experiment.

2. L23-24: "… has generated increasing interest in exploring the mechanisms behind this."

Reply: We have revised.

3. L31: "post-glacial"

Reply: We have revised.

4. L42: "… understanding of the pattern"

Reply: We have revised.

5. L81: "Specifically"

Reply: We have revised.

6. L119-121: 'initial condition' in this section is used to mean two things:
    1. The glacier state as determined by directly from the RGI input, and
    2. The glacier state at the start of the model run with historical climate (post- spinup)

    It is necessary to refer to these differently. Perhaps 1should be the 'pre-run condition' and 2 should be 'post-spinup condition' (and then clarify the post-spinup condition is what is being selected as the initial conditions for the historical run)

    E.g line 120 "The β will regulate the post-spinup condition (i.e. the length of glaciers at the start of the historical run)"

Reply: Thank you for your suggestions! We have used the words 'pre-run condition' and 'post-spinup condition' instead of 'initial condition' in the revised paper for clarity. In addition, we have added a sentence to state the post-spinup condition is what is being selected as the initial conditions for the historical run. "We spinup the model to avoid the influence of the pre-run condition and tuned the parameter, temperature bias (β) in Eq. 1, to obtain a better post-spinup condition. Note that post-spinup condition would be used as the initial condition for the historical run. The β directly regulates the post-spinup condition and largely impacts the *GLR* during early LIA (e.g., LIA4). We alter β from -1 to 1 °C with an increment of 0.1 °C during the spinup period to select the best initial condition for the historical run. For all experiments, a 5000-year spin-up forced by the climate data selected randomly from a 51-year window of 875 - 925 CE is conducted prior to the historical run."

7. L123: It is ambiguous what is meant by the altering of beta. Is it being modified within the

course of a single run (and if so, why?)

Or are you running several versions of the spinup with different values of beta to see what is best? (it looks like the latter from context but it should be clear from this first mention)

Reply: In this study, we run several versions of the spin-up with different values of beta to see what is best. We have rephrased the sentence for clarity. "The β directly regulates the post-spinup condition and largely impacts the *GLR* during early LIA (e.g., LIA4). We alter β from -1 to 1 ℃ with an increment of 0.1 ℃ during the spinup period to select the best initial condition for the historical run."

8. L296: The intention behind this sentence is unclear. The fact that regional glacier evolution is an integral of individual glacier changes is true by definition rather than established in the results. Perhaps just simplify and say "The climatic mechanisms behind the LIA substages described here are still unclear."

Reply: We have rephrased this sentence as "The above discussions explained why there are four glacial substages in BH, but the climatic mechanisms behind these substages described here are still unclear.".

9. L335: Increase

Reply: We have corrected.

10. L356: Overall I think this paragraph could use some rewriting. I understand most of what is being discussed but I have a feeling it won't be clear to anyone not familiar with sensitivity experiments for glacier models. Specifics below. It should also be noted somewhere in this paragraph that the maximal sensitivity at unmodified temperature/precipitation is the expected case due to the negative feedback mechanism of changing ELA as glacier length changes.

Reply: Following your suggestions here and below, we have rewritten this paragraph and added the sentence "The maximal sensitivity at unmodified temperature/precipitation is the expected case due to the negative feedback mechanism of changing ELA as glacier length changes" in the revised paper.

11. L358-359: "The sensitivity of glaciers to temperature/precipitation changes – in the form of the rate of change for GLR per ℃/% respectively – is highest for unmodified temperature/precipitation and decreases as they are varied further from the values given in the historical climate runs"

Reply: We have revised this sentence according to your suggestion.

12. L365-367: Is this a measure of observed relative changes in precipitation and temperature, a measure of how much precipitation changes in response to temperature changes (this is suggested at present by the 'will'), or a measure of how much the precipitation would have to change for a given temperature change to maintain the same SMB?

Reply: The understanding that it is a measure of how much precipitation changes in response to temperature changes is what we want to express. We have rephrased the sentence into "we introduced an index $k = \frac{\Delta p}{\Delta T}$, which is a measure of how much precipitation changes in response to temperature changes at present (Jeevanjee and Romps, 2018)" for clarity.

Response to RC2 on "*tc-2021-352*"

We are very grateful for your constructive comments and suggestions after careful reading. As your main concerns are the model's ability to simulate present-day glaciers and how the chosen value of β influences the simulation results, we have added a paragraph to discuss the model applicability on modern glacier simulations and some sentences to describe the influences of β on the number and timing of glacial substages. The revised parts are marked in red in the paper. Major changes are summarized here followed by point-to-point responses to each comment. Reviewer's comments are in black and our responses are in blue.

Major comments:

While it is an interesting problem, I do have concerns about how grounded in reality the glacier simulations are. The only glacier observations presented are their RGI outlines, and some LIA lengths. Given this paucity of glacier-specific data, one would expect that careful attention would be paid to ensuring that these glaciers are adequately simulated in the present-day, to establish a baseline for comparison to the last millennium.

Unfortunately, the model (as set up) seems unable to simulate many of the present-day glaciers, with a maximum of about ~600 glaciers of the 803 in the RGI have length > 0 for any tested value of β (Figure 2c). It might well be that these are small glaciers which rely on local topography, preferential deposition and redistribution of snow, or avalanching for their existence, and the degree-day model is unable to simulate these. However, in the manuscript there is no explanation provided. How do the modelled volumes at the year 2000 compare with global estimates of ice volume in this area? While this is a regional study, and the reluctance to focus on individual glaciers is understandable, the calibration of the model is rather general, and its not clear whether the model set-up performs adequately for individual glaciers, a necessary prerequisite before analysing the collective glacier response.

For the last millennium it is not reasonable to say that the simulated length changes have been validated by geomorphological maps (L51), the comparison is not that specific. The comparison is largely done through the comparison of 'maximum peak GLR'. Further, only two of the dated moraine ages comes from within the study area, a limitation which is not clear at first.

Many models spin-up simulations are run with the temperature bias β adjusted, and the value $\beta = -0.4$ is taken to best represent the glacier regime during the spinup period, to ensure that not too many glaciers over-run their LIA domains. From Figure 2(b) it would seem that the value of β chosen has a large influence on the early part of the simulation.

In further revising this paper, I would like to seem more consideration given to ensuring that the model system adequately simulates present-day glaciers, and how the chosen value of β influences the results, in terms of the number and timing of glacial substages.

Reply: Thank you for your suggestions. We have added some discussion on the applicability of OGGM to modern glacier simulations and the influences of β on the number and timing of glacial substages in the revised paper.

For some small glaciers that cannot be captured by OGGM, we have added an explanation that "Note that the model is unable to capture some small glaciers, which rely on local topography,

preferential deposition and redistribution of snow, or avalanching for their existence."

For the model performance on the modern glaciers, we have added a paragraph that "The modern ice volume is estimated by the ice inversion module in OGGM (Maussion et al., 2019). This module is designed to diagnose the glacier thickness distribution under the constrains of modern glacier extents (such as RGI outlines) and climate scenario (such as CRU dataset), which can provide the best estimation of glacier volume (Maussion et al., 2019; Farinotti et al., 2019). The simulated BH ice volume at 2000 increases with decreased β, results from the reduction of available glacier numbers. Compared with best estimation, the simulated regional average ice volume has a small bias ranging from -0.006 to 0.010 km^3, especially for a zero bias when $\beta = -0.4$. This confirms the ability of OGGM to simulate the glaciers at regional scale and $\beta = -0.4$ is the best choice for our study."

For the influences of β on the number and timing of glacial substages, we have added some sentences. For example:

"Only two substages have been detected with $\beta = -0.8$ (LIA2 and LIA1) and $\beta = -1.0$ (LIA3 and LIA1) while only the latest substage could be probed with $\beta = -0.9$."

"In addition, the occurrence time of LIA-4, LIA-3, and LIA-2 becomes earlier with a smaller β, but the occurrence time of LIA-1 is stable with various β."

Minor comments:

1. One of the tuning constrains is that the regional average GLR is greater at LIA4 than LIA1 (L128); but this is presented as a result in the discussion (L225, and again at L288).

Reply: According to the references of Murari et al. (2014) and Xu & Yi (2014), we only know the GLR during LIA-4 should be larger than that during LIA-1. However, the glacier changing patterns are worthy to be discussed based on the simulations. Therefore, the purpose of the discussion is to show the occurrence time of each LIA substage, glacier changing patterns and justify the rationality of the results by a comparison between simulation results and dated moraines.

We have also rephrased the sentence for clarity. "Based on our tuning strategy (Murari et al., 2014; Xu & Yi et al., 2014), the *maximum peak GLR* occurred during LIA-4 in MC experiment which is also confirmed by the dated moraine ages in monsoonal influenced Himalaya that the majority of glaciers advanced to their LIA maximum extents at the early LIA rather than the late LIA (Fig. 4b). "

2. L110: it is good to know the method of finding peak lengths, but it would be nice to have an explanation of what 'findpeaks' this actually does, rather than having to read the Matlab documentation to find out (what does the minimum peak prominence mean?)

Reply: We have added some explanations of 'findpeaks' and 'minimum peak prominence' in the revised paper. "… based on the "findpeaks" function embedded in Matlab Software. A local peak is a data sample that is larger than its two neighboring samples. We set the minimum peak prominence to 0.2 to eliminate the peaks that drop smaller than 0.2 on either side."

3. L122: 'all experiments can reach their steady states after a 5000-years spin-up'. However, looking at Figure 2a, it is clear that some of the simulations still have increasing volume at the end of the spin-up period. For the chosen $\beta = -0.4$ the simulations does appear to have reached a steady state, but the general statement here is not correct.

Reply: We have corrected this sentence. "For all experiments, a 5000-year spinup forced by the climate data selected randomly from a 51-year window of 875 - 925 CE is conducted prior to the historical run."

4. L137: the description of the sensitivity experiments is not clear: what is the 51-year window used for? What is the observed SMB? (this is the only mention of observed SMB)

Reply: The 51-year window is used for CRU data. The observed SMB is referred to the World Glacier Monitoring Service (WGMS; WGMS, 2017) datasets. We have recognized this sentence for clarity. "The window size of CRU data is set to 51-year and centered on t^*. t^* is the year when the model best reproduces the observed SMB for glaciers in the World Glacier Monitoring Service (WGMS; WGMS, 2017) datasets (Marzeion et al., 2012; Maussion et al., 2019)."

5. L137: 'We set ε to 0 in Eq.1 in order to maintain the contemporary glacier geometry under the contemporary climate condition.'. This is an important point, how well is the contemporary glacier geometry simulated under contemporary climate conditions? Fig 2c would indicate that many glaciers are simulated with length =0

Reply: The simulated regional ice volume at 2000 is comparable with the best estimations (please refer to the reply to the major comment for detailed information), indicating that the OGGM is capable to reconstruct the glacier change within our study area. In addition, previously studies have also confirmed a good performance of OGGM in simulating alpine glaciers (Farinotti et al., 2017; Pelto et al., 2020) and reproducing the millennial trend of glacial evolution in mountainous regions (Goosse et al., 2018; Parkes & Goosse, 2020).

6. L162: earlier RGI v6.2 is mentioned, here it says v6 is used.

Reply: We have corrected.

7. L183: the best fit RMSE of maximum peak GLR is rather high (> 100%), ad larger than GLR % change through the LIA (Figure4).

Reply: Previous studies indicated that OGGM is capable to reproduce the regional glacier changes but suffered from large bias in simulating the individual glacier changes because the model is calibrated such that it performs well on a regional/global scale (by accepting larger errors for single glaciers; Goosse et al., 2018; Maussion et al., 2019; Parkes & Goosse 2020). If we want to apply OGGM on an individual glacier, it would be a task of us to make sure that OGGM's parameters are well calibrated from the observed SMB of that glacier. Unfortunately, the lack of observations over BH (no glaciers have been continuously observed in BH) makes this task impossible. Therefore, the RMSE is expected to be large. However, the aim of our study mainly focuses on the glacier evolution in Bhutanese Himalaya at regional scale, regardless the detail changes of individual glaciers.

8. One first reading this manuscript, I thought that the dates presented in Figure 4b were from within the model domain, i.e. were directly related to the glaciers which are modelled. However, only two dates are actually within the domain (Figure 1, Table S1). There is an assumption, which I do not thick is anywhere mentioned, that the dated moraines outside of the study are also representative of the dates of glacial advances within the study area.

Reply: We have explicitly stated the assumption we made. "The simulated timing and extent of

glacial advances are validated with the ^{10}Be surface exposure ages and ^{14}C ages of the LIA moraines across the monsoonal Himalaya and the mapped LIA glaciers over BH. Here, we assume that the dated moraines outside of the study area also can represent the dates of glacial advances within the study area because the terrain and climatic conditions are similar (Owen & Dortch 2014; Murari et al., 2014). With this assumption, more observations can be included in this study, making them more representative of regional features."

9. As a consequence of the previous point, the statement at L227 'specifically, about 12 of the 30, moraine age shows that the related glaciers reached their maximum peak GLR during LIA-4 compared with only 2 of them during LIA-2', it should be made clear that these 'related glaciers' are not within the study are, but rather in the wider monsoonal Himalaya.

Reply: We have clarified the statement. "Specifically, about 12 of the 30 moraine ages across the monsoonal Himalaya shows that …"

10. Also related to that point, the statement L271 'The occurrence time of each glacial substage also varies from glaciers, supported by the dispersal of moraine ages (Fig. 4b)' also implies that it is that moraine ages are directly related to the occurrence time of each glacial substage, but the comparison is actually more general that that (LIA dates in the monsoonal Himalaya compared to occurrence time in the study area).

Reply: We have clarified the statement. "The occurrence time of each glacial substage also varies from glaciers, supported by the dispersal of moraine ages across the monsoonal Himalaya (Fig. 4b)."

11. The qualitative link makes to volcanic eruptions, perhaps not a compelling as one might expect, given the forcing for PMIP3. Reference should be made to the forcings of PMIP3, of which volcanic aerosols is one of the most important. It seems somewhat round-a-bout, inferring a cooling from volcanic aerosols via a glacier model, when the response of the PMIP3 models to volcanic aerosols would most directly be compared by examining the input and output of the GCMs themselves.

Reply: Thank you for the nice suggestion! Our logic is that four compelling glacier sub-stages detected by OGGM are caused by four cold intervals (from PMIP3 model results) during LIA, and four cold intervals are related to the volcanic eruptions as volcanic eruption will inject abundant aerosols into the upper atmosphere cooling the climate. We have reorganized the paragraph as follows: "The four cold intervals during the LIA in BH are forced by four large stratospheric sulfur-rich explosive eruptions events (sulfate aerosol loadings > 60 Tg; Fig. 7c; Gao et al., 2008) as the volcanic aerosols will inject abundant aerosol into the upper atmosphere, cooling the climate. (Schmidt et al., 2012)."

12. In the abstract (L12) and the conclusion (L379) it states 'OGGM broadly captures the pattern of glacier length change, but underestimates its amplitude'. I'm not sure which result that statement comes from.

Reply: We have corrected these two sentences.

"Compared with geomorphologically-mapped glacial landforms, the model can well capture the patterns of glacier length change."

"Compared with the geomorphological maps and moraine ages, OGGM broadly captures the pattern

of glacier length change."

Minor Edits:

Much of the text is a little hard to understand and needs careful editing. I have made a few comments here, but there are many more minor changes to the text to improve clarity required.

1. L8: evolutions -> evolution

Reply: We have corrected.

2. L20 'glacial history for himalayas...' -> 'glacial history of the Himalaya...'

Reply: We have corrected.

3. L23: more and more interest on -> increased interest in

Reply: We have changed this sentence to "… which has generated increasing interest in exploring its mechanisms".

4. L24: mechanisms behind what?

Reply: We have changed this sentence to "… which has generated increasing interest in exploring its mechanisms".

5. L35: modelling complements field work, but this sentence implies that its a substitute for fieldwork.

Reply: We have revised this sentence as "it can be a complement for field-based approach in capturing the glacier evolution on regional scale.".

6. L39: 'study provided...' -> 'study provides...'

Reply: We have corrected.

7. L45: exist -> are

Reply: We have revised.

8. L48: 'The small glaciers...' -> 'Small glaciers...'

Reply: We have revised.

9. Figure 1: This map contains much of the information needed to understand this study and its context. It is however quite small and it's hard to even see where the mapped LIA extents are for all but the largest glaciers.

Reply: This figure mainly gives the overview of the study area and researched glaciers. A very high resolution is needed to show the mapped LIA extents clearly, but the publication requirements limit the figure size. Therefore, we will provide the shapefile of LIA glaciers for clearly displaying the extents of the mapped LIA glaciers.

10. L77: exceeding -> exceeds

Reply: We have corrected.

11. L139: is this temperature bias different from β?

Reply: The temperature bias has the same meaning of β. We have changed the "temperature bias" into "β" for clarity.

12. Figure 2: it's a little confusing having the symbols in (c) the same colors as the color scale for β, but different meaning.

Reply: We have reworked the symbols' colors in (c).

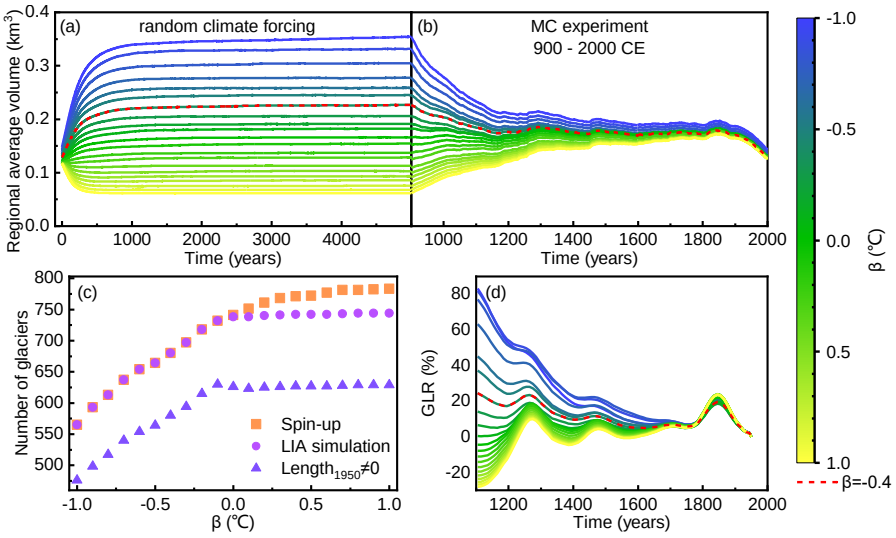

Figure 2. (a) The regional average glacier volume during the 5000-year spinup with various β. (b) The simulated regional average glacier volume from 900 to 2000 CE with different post-spinup condition. (c) The number of available glaciers with various β. (d) The simulated regional average *GLR* from 1100 to 1950 CE.

13. L202: please rearrange this sentence

Reply: We have split this sentence into two sentences for clarity. "The *second peak GLR* occurred during LIA-1 in MC experiment. This founding is the same as the results in the CCSM4, GISS, and MPI experiments but different from the results in BCC (LIA-2), CESM (LIA-2), and IPSL (LIA-3) experiments."

14. L223: extremums -> extrema

Reply: We have corrected.

15. L260: significant -> significance

Reply: We have corrected.

16. L361: is -> are

Reply: We have corrected.