Yang et al review

General comments

This paper exhibits a fairly well-established use case of OGGM that doesn't offer much that's novel over other studies using this model, but I don't consider that particularly necessary – repeated use of models in similar configurations and comparison of results with additional observational datasets is useful in establishing model robustness. The focus on a specific – and important – region and the use of morraine and surface exposure ages adds value to the model runs, and the sensitivity experiments are effective and provide context for glacier responses over the study period.

As written, the identification of sub-stages in the LIA does not seem compelling to me, and lacks sufficient cohesion between models to suggest that each is the result of the same set of real changes in climate across models where they are identified. Ideally I would want to see the material enhanced with an explicit metric by which the sub-stages can be identified both within and between the modelled timeseries. I like the idea of identifying smaller periods of pre-industrial variation within a smaller area, as an enhancement to identification of global- and regional-scale long-term variation, but it needs more work.

Assuming a comprehensive revision of the LIA substages material is conducted, addressing the terminology, methodology, and results, I think the paper will be of clear interest and value to the glaciology community.

Specific comments

Figure 1: The colours for lakes and glaciers in (b) overlap with the colours used in the elevation map they are overlaid on, which hurts readability. Is it necessary to represent modern glaciers with two colours, and to represent lakes at all? I would recommend simplifying to a single modern glacier colour, a single LIA glacier extent colour, and the background image.

Including markers for major volcanic events on the graphs of temperature and glacier change is necessary where the LIA substages are identified, as this is one of the major drivers of the negative temperature anomalies resulting in glacier advance.

The paper uses the term 'LIA substages' to mean apparently two different things: one is 4 periods of time within the LIA, across all of the BH area, and the other is a variable number (sometimes more than 4) of stages of greater glacier extent identified per-glacier. The latter is not covered in sufficient detail despite showing results in fig 3.

Section 2.3 needs considerably more depth on the methods used. This should outline the process of the identification of the LIA substages a glacier has experienced (presumably from multiple morraine locations dated to different ages) as well as details of how the glacier area is calculated from glacier length (is this an empirical scaling approach or a reconstruction of glacier geometry? Does it relate to the way OGGM calculates glacier area of larger-than-present-day glaciers?).

I have refrained from giving a run-down of language errors as providing a comprehensive list would dominate this review and the time and energy is better spent assessing the scientific content. I recognise the biases inherent in a scientific establishment that demands publication in English

from native and non-native speakers alike, and I do not think that most of the errors have a major negative impact on the ability of a reader to determine the content of the study. Nevertheless, if at all possible I would recommend having an expert in technical written English help with the revision process after changes to the content are made. I can provide more detailed notes on subsequent revisions, but I don't want to make a host of minor corrections to sections that are likely to see top-down rewrites.

Technical corrections

Figure 2 – all panels have Y-axis scales on the left except (g), which should also have one, unless I am misreading the reason the bars are stacked.

Figure 2 – specific explicitly the values to which the panels showing percentage change are normalised, and the year to which temperature anomalies are relative.

Figure 3 – does panel (a) really need breaks on the Y axis? All it seems to do is make the graph harder to read and make the lines of best fit less clear. At best it's messy, and at worst actively misleading because the squashing of the data points for 0 substages makes the relationships appear more linear (even though the lines of best fit indicate that they aren't).

Figure 3 (caption) – panel (b) should be described as showing the number of glaciers with each count of identified substages, not the number of substages for each. In general, I find most of the figure captions in the paper could be more descriptive. I don't know how much of this is personal taste, but I find captions that describe the relationships exhibited by the data shown are better for readability than captions that just describe what the data is.

Line 255 – please explain what baseline the sensitivity experiment increases/reductions are relative to.