Reviewer #1

Willeit and colleagues present a large ensemble of CLIMBER-X simulations with various combinations of continental ice sheet configurations and atmospheric CO2 concentrations. This unprecedented ensemble allows them to analyse the physical conditions that determine the forcing range in which CLIMBER-X produces DO-like, millennial-scale climate variability. They find that this "sweet spot" is controlled by the sign of the surface buoyancy flux north of 55N. Millennial-scale transitions between week and strong AMOC states occur when the buoyancy flux north of 55N is about to switch sign. A strong/present day like AMOC occurs when the buoyancy flux is negative and deep water formation takes place in the Labador and Nordic Seas. When the buoyancy flux switches sign, this modern-like deep water formation pattern becomes unsustainable. The conditions under which this sign switch occurs are controlled by the boundary conditions. LGM-like ice sheets tend to enhance buoyancy loss, while low CO2 concentrations tend to decrease it. The balance of the two effects seems to be well captured by CLIMBER-X as the strongest DO-like variability occurs at realistic MIS3-like boundary conditions.

Some of the conclusions are not exactly new, e.g. the cancellation of the effects of ice sheet size and CO2 concentration. However the range and combination of covered boundary conditions is unprecedented, and the results are very relevant for the DO- and wider CP community and thus definitely worthy of publication. It is also very much appreciated that the authors define a metric that could be used to compare the physical conditions that control the "sweet spot" across models. Before publication, I would ask the authors to provide more context in some parts and to address a few issues as outlined in my comments below.

We thank the reviewer for the positive appraisal of our work and the constructive comments.

Major Comments

1. Introduction/Discussion: Please provide more context on what has already been suggested in terms of physical control on the sweet spot. At least Galbraith & de Lavergne (2019) and Klockmann et al (2018) provided some suggestions, e.g. the overall volume of Antarctic Bottom Water (AABW) present in the deep ocean, the density difference between AABW and North Atlantic Deep Water, presence of deep water formation in the Nordic Seas. Also spell out more directly how the additional CLIMBER-X simulations can help in pin-pointing the physical control across models. Because the physical control might also be model dependent.

We will expand the introduction with a more extensive discussion of what controls AMOC strength in general, and what has been previously suggested in terms of control of the sweet spot in particular. We will also elaborate a bit more on the relevance of our buoyancy criterium and potential limitations that might arise when applying it to other models.

2. I agree that the buoyancy flux analysis in this paper and the one in Klockmann et al (2018) cannot be compared directly one to one but at least a qualitative comparison should be possible and would actually strengthen the authors arguments even further. This could e.g. take place in the Discussion section.

Overall, the mode transitions in the experiments with PI ice sheets in Klockman et al also occur when net buoyancy flux over their NAtl&LabSea region changes from buoyancy loss to buoyancy gain (Klockmann et al use density instead of buoyancy, so the sign is flipped). In *their Nordic Seas region, the buoyancy flux is close to zero for the CO2 range where the transition takes place, so the Nordic Seas would not change the sign. This qualitative agreement makes the suggested metric M in the discussion of this manuscript even stronger.*

Calculating buoyancy over the deep water formation area(s), as in Klockman et al. (2018), is of limited use because this flux will be strongly negative as long as deep water formation continues. This is because most of the heat is released in this area, but only a small fraction of the freshwater flux enters the surface through this area (e.g. river runoff along continental margins). Therefore, this flux does not necessarily provide information about the stability of the AMOC. On the contrary, the sign of the buoyancy flux integrated over the entire Atlantic/Arctic ocean domain north of 55N, as shown in our paper, provides useful information about the (convective) stability of the AMOC and explains its instability under glacial conditions. In the revised paper we will add some further discussion on the rationale behind using integrated versus local buoyancy flux in diagnosing AMOC instability.

One interesting difference can be seen in the effect of ice sheets on the thermal component. In Klockmann et al, the stronger net buoyancy loss with glacial ice sheets is due to increased heat loss over the deep convection sites, while in the present study, it is due to the reduced freshwater input. I do not have an immediate hypothesis where this difference might arise from. Perhaps it is simply due to the different areas of integration.

This difference could indeed be at least partly due to the different areas of integration. Since most of the heat transported northward by the AMOC will be released over the convection areas, the much stronger AMOC simulated with LGM ice sheets compared to present-day ice sheets (for a given CO2 concentration) will result in a strong increase in the surface sensible heat loss and consequent increase in buoyancy loss over the convection regions, which is well captured by the integration areas in Klockmann et al. 2018. However, the integration areas in Klockmann et al. 2018 capture only part of the changes in hydrological cycle and the resulting changes in the net surface freshwater fluxes.
That said, it is also clear from Fig. 15 that CLIMBER-X tends to show a larger haline buoyancy response between pre-industrial and LGM compared to PMIP models. This could be attributable to a substantial CLIMBER-X AMOC weakening at LGM, which results in a cooler northern North Atlantic and therefore a decrease in precipitation, while most PMIP models, and also the MPI-ESM used in Klockmann et al. 2018, show a strengthening of the AMOC at LGM.

3. What is the role of sea ice in the buoyancy flux? Is the effect of freezing/brine release and melting included in the freshwater and heat budgets? Sea ice typically plays a big role in feedback loops regarding convection patterns. Even though it can be difficult to determine whether sea-ice is driving the change in the convection patterns or responding to it, it is still worth to be included more explicitly in the analysis.

The effect of sea ice formation and melt is explicitly included in the computation of the surface buoyancy flux. This will be made more explicit in the revised paper. As long as sea ice is formed and melted inside the area of integration of M (north of 55°N), the net contribution of sea ice to the integrated buoyancy flux will be small and only due to the non-linear equation of state if sea ice is formed and melted in regions which differ in their sea surface temperature.

4. It might be insightful to show the buoyancy flux also for the equilibrium simulations, e.g. in a similar style as Fig. 3 with buoyancy flux as the colour coding. That would help in linking the results from the transient and equilibrium simulations.

Thanks for the suggestion, we will consider including the suggested figure in the revised paper.

Minor Comments

l.3 "latitudinal reach" or "northward extent" instead of "latitude reach"?

We will change it from 'latitude reach' to 'northward extent'.

l.38-40: see major comment 1

As outlined in the response to the major comment above, we will expand this section.

l.45: what is the climate-only setup? Are there other setups?

CLIMBER-X also includes a global carbon cycle model and an ice sheet model, which are not used in the present study as CO2 and ice sheets are prescribed as constant over time. However, considering that this statement might confuse the readers, we will remove '*in a climate-only setup*'.

l.75: How sensitive is the model to the area where the noise is applied? Why is it applied only locally and not globally?

Initial sensitivity tests (not shown in the paper) indicated that the results are not very sensitive to the details of where noise is applied, as long as it covers the areas in the North Atlantic where deep water forms. Noise is introduced in the model to mimic synoptic-scale and interannual climate variability, and applying the same noise globally would be unrealistic as it would assume that this climate variability is globally uniform, which is definitely not the case. A more realistic, global, application of noise would require some kind of weather generator, which is beyond the scope of this study and would very likely have no impact on the results presented in the paper.
In the revised paper we will add the sentence: '*Sensitivity tests indicated that the model results are not very sensitive to the details of where the noise is applied, as long as it covers the areas in the North Atlantic where deep water forms*'.

l.115: please also state the temperature changes over Greenland in the simulations and in the reconstructions. What does it mean if Greenland change is not capture well but the Iberian margin yes?

In the revised paper we will explicitly add the modelled and reconstructed temperature changes in Greenland.
The deficiency in the simulated Greenland temperature response in the model is somewhat expected as the atmosphere in CLIMBER-X works best over relatively flat terrain, while the Greenland ice sheet is characterized by large slopes and the circulation over steep slopes is not properly resolved by the model. DO events are expected to affect mainly winter temperature in the northern North Atlantic, primarily as a response to the retreat in sea ice.

This temperature changes are going to be largest in a relatively thin layer close to the surface and since in the atmosphere model the transport of heat is mostly horizontal, the warming over the ocean is not very efficiently transported to the summit of the Greenland ice sheet. Also other models, including many GCMs, tend to underestimate the DO warming over Greenland (e.g. Menviel et al., 2020; Li et al., 2010; Kuniyoshi et al., 2022).

l.123-124: "The heat transport [...]" What do you base this sentence on? Is it based on previous studies (if yes, please cite)? Or do you infer it from your results (if yes, please elaborate shortly)?

This sentence is based on our results, but from simulations not shown in the paper. We will therefore delete this sentence in the revised paper.

Fig.7: Please correct the caption. The interstadial sea-ice extent is drawn in dark teal and not grey

Will be corrected, thanks.

Fig.8: In the experiment description and in Fig.9 you mention a total of six noise amplitudes. Here you show only four. Why are 0.0625 and 0.125 not shown? Or did you not cover the full CO2 range for these amplitudes? If so please mention this in the experiment description.

We performed the 0.0625 and 0.125 noise amplitude simulations only for a CO2 concentration of 170 ppm, which is why those noise levels are not included in Fig. 8. This will be specified in the revised manuscript.

Fig.9: Which CO2 concentration was used in the respective simulations displayed here?

The simulations in the figure are for a CO2 concentration of 170 ppm. We will clarify this in the caption.

l.152: Please briefly state, how do you define stable here (and elsewhere in the manuscript). Also, how realistic are the deep convection patterns in CLIMBER-X given the very coarse resolution?

Here (and elsewhere), ‘stable’ will be removed as it does not add any relevant information. The coarse model resolution is obviously a limitation of our model. However, the present-day deep convection patterns compare well to ocean reanalysis in the North Atlantic as shown in Fig. 13 in Willeit et al. 2022. There are unfortunately no reconstructions of the mixed layer depth for DO Stadials and Interstadials, but some information on the convection patterns can be derived from sea ice extent reconstructions, which are tightly linked to the locations of deep water formation. As shown in Fig. 7 and discussed in the text, it seems that the CLIMBER-X sea ice extent change between Stadials and Interstadials is in qualitative agreement with reconstructions, providing some support for the simulated deep water formation patterns.

l.154: "two modes" instead of "two stable modes". The "stable" in the latter half of the sentence ("are stable under the same CO2") is sufficient.

Will be fixed, thanks.

l.161/Fig.10: what about the smaller oscillations that occur around 160ppm with interglacial ice sheets and around 240ppm with mid-glacial ice sheets? In these cases, the buoyancy flux does not change sign.

The smaller oscillations for interglacial ice sheets and CO2 around 160 ppm are not reflected in the buoyancy flux because they involve changes in convection pattern that are mostly confined to latitudes south of 55°N, which is therefore not reflected in *M*.
The oscillations at ~240 ppm for mid-glacial ice sheets involve a reorganization of deep water formation inside the domain north of 55°N. This therefore shows a clear imprint on *M*, but does not cause a change of sign of *M*, as convection remains present north of 55°N.
We will add this discussion in the revised paper.

l.161/273: Is the Arctic Ocean included in the integral of the buoyancy flux?

Yes, the Arctic Ocean is included in the integral of the buoyancy flux. We will add this explicitly in the revised paper.

Fig.11: What is the averaging period shown here? What does the grey circle around the North pole indicate?

What is shown is the mixed layer depth at the times corresponding to the CO2 concentrations indicated in the panel titles. There is no averaging in time implied in this figure. The grey circle around the North Pole is an artifact and will be removed.

l.169-188: This part is difficult to read with the many "increases" and "decreases". Try to spell out more specifically whether the listed factors induce a buoyancy loss or gain. It can become difficult to correctly interpret increase and decrease if a property (such as M) can have different signs with small or large absolute values.

Thanks for pointing this out. We agree and will try to make this section better readable by following the reviewer's suggestions.

Fig.13 and related text: This figure is discussed very briefly, approximately with one and a half sentence. It might be worth to spend a few more words on this figure and to also make the connections between the left half and the right half clearer. Especially because the information in the right half has already been shown in Fig. 10 and 12. Also the relation between hosing and noisy freshwater forcing could be explained some more.

Indeed, we agree that it makes sense to discuss the relation between the left and right part of this figure in some more detail and will do so in the revised paper.

l.194-201: Compare diapycnal diffusivity results to previous work, e.g. diapycnal diffusivity seems to have played a key role in generating the DO oscillations under LGM conditions in Peltier&Vettoretti (2014).

We will expand the discussion on the role of diapycnal diffusivity, including some previous work, as also suggested by Reviewer #2.