We thank the reviewer for their time and effort reviewing our manuscript, and for the constructive comments which have helped to substantially improve our text. We address the reviewer's suggested improvements by completely rewriting our description and discussion of the processes that occur in the model, adding results to the figures in the main text and adding new figures to the SI, and referencing these figures more thoroughly in the text. Specifically, we make the changes outlined in our replies to the detailed comments below.
Below are our detailed point-by-point replies and suggested manuscript improvements (blue) for each comment (black).

Detailed comments are as follows:

P2L15-18: Freshwater input might be positive feedback to AMOC weakening as well. please refer to Barker et al 2015 and rephrase the sentences accordingly here as well as in L23-24.

We add the suggested reference and mention the possibility for freshwater feedbacks. Specifically, we add the following sentence to our introduction:

"Lags between the appearance of ice-rafted debris and the reconstructed cooling, however, suggest that freshwater fluxes could have instead acted as a positive feedback to AMOC weakening rather than triggering it (Barker et al., 2015)."

P2L25-29: other key relevant paper should be cited, for instance, Zhang et al., 2014, 2017.

We add references to Zhang et al., 2014 and 2017 and Vettoretti, 2022.

P2L33: also consider citing Zhang et al 2021, Vettoretti et al 2022 here.

We add the suggested references.

P2L45-47: Please add relevant papers after the first sentence (e.g. Knorr and Lohmann 2007, Zhang et al 2017, Galbraith and de Lavergne 2018, etc.)

We add the suggested references.

P4 L5: one predominant feature of glacial cycle is the development and demise of northern hemisphere ice sheet, involving both area and height, of which impacts on climate system are not the same. The former, as discussed in this study, via its albedo feedback is a thermal impact, while the latter, via its impacts on winds, is a kinetic impact (Zhang et al., 2014). In addition, there is no change in Bering Strait considered as well (Hu et al., 2011) (P5L4, a typo there). I was wondering how far these additional setups can alter the key messages of the thermal thresholds in this study. As seeing in my following comments, at least a comprehensive discussion around this is required.

We add more discussion of other factors for AMOC stability, see also our answers to further comments on this topic below. We will add the following text to section 3.3:

"Previous studies suggested that pre-industrial or intermediate glacial ice sheet configurations are required to even produce a thermal AMOC threshold in the range of glacial-interglacial CO_2 concentrations in a full GCM and that the presence of a full glacial Laurentide ice sheet prevents such a threshold (e.g. Klockmann et al., 2018). In addition, changes in the interconnection of marine basins, specifically the Bering Strait, also affects AMOC stability (Hu et al. 2012). The values of the thermal thresholds in our experiments are thus likely sensitive to the model design and initiation. Pöppelmeier et al. (2021) showed that the sensitivity of Bern3D to freshwater hosing increases when additional LGM boundary conditions are prescribed (changed wind fields, closed Bering Strait, tidal mixing differences due to sea level changes). The different wind fields and tidal mixing strengthened AMOC and increased the salt and heat transport into the subpolar North Atlantic. This could mean that stronger cooling is required to stabilise the water column in the Irminger Sea and reach the first thermal threshold, when the full range of glacial boundary conditions are applied. Closure of the Bering Strait increased the salt advection feedback, which stabilises the weak circulation state without deepwater formation in the subpolar North Atlantic. Further investigations are needed to determine how changes in strength and location of the wind stress due to the ice sheet's orography, sea level and Bering strait closure would affect sea ice formation in the northern North Atlantic and the AMOC thresholds in our simulations quantitatively. "

P6 3.1: it would be good to present the 800kyr long transient simulation results. In Figure 1, it is of great help to add the radiative forcing curves to enable a comparison with B.slow experiment.

We add the according figure to the SI.

P7L15-17: As alluded, lacking feedback from topo changes might overestimate the LGM cooling caused by radiative forcing decrease because higher NHIS can cause a stronger AMOC which promotes heat release from the ocean and hence North Atlantic warming. This might stimulate some discussion perhaps in data-model comparison or model limitation sections.

We add this to our discussion as suggested (see the previously shown addition to section 3.3.)

P8. Fig3: given the North Atlantic and Nordic Sea are the key regions for AMOC state shift, it would be better to provide a zoom-in plot for this region, especially for the sea ice fraction plot. Please also revise the color scheme for "sea ice cover fraction" to highlight change in the low values (<0.5) or just provide anomalous field as delta Density. Please also include lat-lon info in the plots. In addition, as you are discussing AMOC states, AMOC plots are highly recommended in this figure.

We change the figure (or panel?) to show sea ice anomalies in the North Atlantic and add coordinates. We also add vector plots showing AMOC circulation to the SI.

P8L18-20: in the state (II), deep water formation is enhanced in west and south of Greenland. In general, it is more reddish in State (II) than in State (I), but why the AMOC is weakened in the former. Is this due to that convection in the western North Atlantic is not the key to the strength of the AMOC?

Yes, the mixed layer depth that we diagnosed from the annually-averaged model output is not a good metric to understand changes in AMOC strength. We'll explain the density changes more explicitly in the new manuscript version, and include that some changes in the locations of downwelling occur without changing AMOC strength. We'll remove the plots of mixed layer depth and instead show the absolute vertical density gradients in each state. We will add the following text to section 3.1:

“Initially, the whole Atlantic surface ocean cools and freshens, leaving the temperature and salinity differences between the Irminger and Caribbean Seas almost unchanged (Fig 5e). However, NADW becomes less salty and colder as a consequence of the changes in the surface ocean (not shown) and the vertical density profiles in the subpolar North Atlantic steepen due to the temperature and salinity changes (Fig. SI.7-8). After about 6 kyr, the changes in the North Atlantic density profile shift the location of NADW formation. NADW formation moves south as vertical density profiles in the subpolar east North Atlantic stabilise under a freshening of the surface and density profiles further south steepen due to surface cooling combined with subsurface warming (Fig. SI.7-9). These changes do not cause a step-change in AMOC strength, but freshwater and heat advection into the North Atlantic is reduced, sea ice expansion increases in the eastern North Atlantic and AMOC variance (calculated over a moving 50-year window) is increased (Fig. 5). Transport of heat and salinity into the North Atlantic decreases (Fig. 5f, g) and North Atlantic SST and SSS decrease (Fig. 5e). Reduced influx of subtropical surface waters also causes sudden cooling and freshening in the Irminger Sea (Fig. SI.8).”

P8L25: “south-flowing fresh Arctic waters further stratify …”. This is a key process to stabilize the glacial AMOC state, but in this version, there is not direct evidence to support it. Note that freshwater convergence in Fig5e cannot provide such support to this statement because it is a sum of freshwater flux across both 40N and 70N in the North Atlantic.

We show the freshwater fluxes across each latitude separately in the updated Fig. 5. We also clarify that the spread of cold, fresh surface water in the North Atlantic stabilises the circulation state. We cannot actually determine whether the water comes from the Arctic or just turns more Artic-like due to reduced influx of southern surface water.

P9L10: what is Kolmogrov-Smirnov test? Add details and reference.

We provide some more information and a reference:

“but none are statistically significant in the two-sided Smirnov test, which determines the likelihood that two distributions are the same (Berger and Zhou, 2014), even at the 50% confidence level”

P10 Fig 5: Panel e, it would be good to interpret meanings of positive/negative values of freshwater convergence to help readers understand this plot (e.g. positive values indicate freshwater import and hence a stable AMOC). In addition, the definition of freshwater convergence should be added to the Method section. It is worth noting that this AMOC stability indicator (Liu et al., 2014 Clim Dyn) predict a mono-stable AMOC regime in B.slow., in contrast to the hysteresis feature shown in Fig 4b. In addition, comparing the panel a) with Fig 4b, it appears that B.slow.b is initialized from a AMOC state that is bistable with respect to radiative forcing. If so, why the AMOC recovers to its initial strong mode after removing the freshwater input? Typo in y-axis labels of panel c). it is also good to add radiative forcing panel on the top of it, with a vertical shaded bar to highlight periods when AMOC is bistable.

We amend Figure 5 as suggested. The hysteresis shown in Fig 4b is not the result of a traditional perturbation experiment with freshwater hosing but is the transient response to the applied radiative forcing. We are not sure if stability with regard to a freshwater perturbation is the same as stability in the face of changing boundary conditions, as caused by the radiative forcing. We are therefore careful with the interpretation of freshwater transport as stability indicators in our study. We add the following text to section 3.1:

“The difference between freshwater transport into the South Atlantic at 32°S and into the Arctic at 62.5°N in Fig. 5f can be used as a measure for the basin-wide salinity feedback (Rahmstorf, 1996, de Vries and Weber, 2005). In our simulation, changes in this metric are predominantly caused by changes in the transport across the northern edge, since transport into the South Atlantic remains almost unchanged throughout the cooling phase of B.slow. North Atlantic salinity is instead governed by changing transport from the subtropics into the North Atlantic and between the North Atlantic and Arctic. As such, the processes involved in the sudden AMOC strength changes, namely density changes in the upper water column, and those that stabilise new circulation states (salinity and heat redistributions, sea ice expansion) mostly operate in the North Atlantic region.”

P11L31: how do you identify the reduced heat convergence “off the British Isles” based on the time series in Fig5?

The geographic information is not derived from the time series but provided as extra information to contextualise the time series. We add a figure to SI showing the discussed spatial pattern (new Fig. SI.11).

P11L33: It is also not logically clear why this is the cause to the northward spread of AABW. In Fig5, the northward intrusion of AABW is starting from the beginning of the experiment, not lagging the reduction of heat convergence in North Atlantic.

We agree, and change the sentence to point out coincidence rather than causality. The experiment is initialised with no AABW tracers in the North Atlantic. Initially, changes in the concentration of AABW tracers in the North Atlantic are small. Their amount only begins to rise substantially at ~15 kyr and shows the biggest jump at ~27 kyr when the heat convergence also declines.

P11L35: why “heat advection to >55N stops entirely”? could the authors present the evidence?

We apologise, this should have been ‘heat convergence’. We correct this and add a figure of the spatial pattern of heat convergence changes to the SI (new Fig. SI.11).

P11L37-39: Again, no direct lines of evidence to support this statement. Does the contemporary sea ice expansion and its seasonality contribute to the freshening in the eastern Nordic Sea? As well as in P11L42-43. Please clarify. Is there a bipolar thermal seesaw during abrupt AMOC reduction in B.slow? The results appear to show that bipolar sea ice change out of phase with AMOC/NADW change – sea ice expansion with NADW weakening. The subdued thermal seesaw in B.slow indicates the dominant role of decreasing radiative forcing in controlling bipolar change.

Yes, we thank the reviewer for pointing this out. There is a small bipolar seesaw effect. We mention this explicitly in the revised manuscript. We add a plot of with spatial patterns of changes in B.slow to the SI and add the following text to section 3.1:

“The biggest AMOC weakening at ~27 kyr is also accompanied by a weak bipolar seesaw effect, which causes a temporary decline in sea ice coverage in the Atlantic sector of the Southern Ocean (Fig. 5). It is, however, too small to reduce the radiation-driven sea ice increase in the longer term.”

P12L27: what’s the statement “… increased heat advection into the North Atlantic” based on?

This statement was erroneous and we delete it. AMOC strength is not constant at the beginning of the experiment but weakens slowly, while the spatial pattern of deep convection and heat *convergence in the North Atlantic change. We rewrite our description of the processes responsible for AMOC changes and use more references to figures. We add the spatial patterns of heat convergence changes to the new Fig. SI. SI.*

P12L29: weakened north ward transport of what? Upper cell of the AMOC?

We clarify that we speak of transport of salt and heat.

P13L1: please show the weakened the meridional salinity gradient in the North Atlantic.

We apologise, this was meant to say weakened meridional salinity transport, i.e. an increased salinity gradient. We correct the statement in the text and add the temporal evolution of salinity in the Irminger and Caribbean Seas to Fig. 5 to show the increased meridional salinity gradient.

P13L3-5: how does the increased surface density promote SST decrease? This is not clear at all here.

We agree with the reviewer that this formulation was misleading. We meant to express that SST changes have a direct effect on water density and an indirect one via influencing evaporation, and that temperature-driven evaporation changes counteract the buoyancy forcing caused by the temperature change. We will rewrite this paragraph as follows:

"The deep water formation regions are sensitive to heat and salt flux changes, because any reduction in sea surface temperatures (SST) increases surface density but simultaneously reduces evaporation in ice-free areas, thus effectively creating a small freshwater forcing and a negative feedback to the buoyancy changes caused by the initial SST decrease. Sea ice covering the downwelling areas stabilise the water column by preventing surface ocean cooling and evaporation. The progressive influx of AABW into the North Atlantic is a further process stabilising new circulation states by stratifying the water column from below (Buizert and Schmittner, 2015)."

P13L5-7: the authors proposed that sea ice expansion over convection sites acts as negative feedback in response to SST cooling, which is not convincing. This process, as demonstrated in this sentence, can avoid further cooling of sea surface, which in turn reduced sea surface heat loss to increase surface density, and thus stratifying the water column. This seems to exert rather positive feedback to stabilizing the cooling-induced AMOC slowdown. Please clarify. In general, positive/negative feedback discussed in this paragraph is hard to follow. Please clarify with more direct evidence/references.

We agree with the reviewer that the discussion of feedbacks in this paragraph is unclear and we will rewrite this. Importantly, the current version suggests that sea ice cover is a negative feedback on SST changes, which is not correct. Sea ice cover prevents evaporation and heat loss to the atmosphere, stabilising the water column. This would be better described as removing the positive feedback buoyancy changes. We will rewrite the paragraph focussed on stabilising mechanisms rather than feedbacks as follows:

"The resulting redistributions of heat and salinity cause sudden shifts in the vertical density profiles and sea ice expansion which consolidate the new circulation state. The downwelling zones are sensitive to heat and salt flux changes, because any reduction in sea surface temperatures (SST) increases surface density but simultaneously reduces evaporation in ice-free areas, thus effectively creating a small freshwater forcing and a negative feedback to the buoyancy changes caused by the initial SST decrease. Sea ice covering the downwelling areas stabilise the water column by preventing surface ocean cooling and evaporation. The progressive influx of AABW into the North Atlantic is a further process stabilising new circulation states by stratifying the water column from below (Buizert and Schmittner, 2015).. The difference between freshwater transport into the South Atlantic at 32°S and into the Arctic at 62.5°N in Fig. 5f can be used as a measure for the basin-wide salinity feedback (Rahmstorf, 1996, de Vries and Weber, 2005). In our simulation, changes in this metric are predominantly caused by changes in the transport across the northern edge, since transport into the South Atlantic remains almost unchanged throughout the cooling phase of B.slow. North Atlantic salinity is instead governed by changing transport from the subtropics into the North Atlantic and between the North Atlantic and Arctic. As such, the processes involved in the sudden AMOC strength changes, namely density changes in the upper water column, and those that stabilise new circulation states (salinity and heat redistributions, sea ice expansion) mostly operate in the North Atlantic region."

P13L11-13: As mentioned in previous comments, providing supportive evidence is of crucial importance since this is important positive feedback to the AMOC slow-down.

We make more references to figures in the re-written paragraph.

P13L15: please clarity and specify the positive and negative feedback mentioned here.

We agree that this paragraph is unclear, and we will rewrite it by using the clearer term 'stabilising process' (see above).

P13L22-23: given the gradual decreasing radiative forcing, it is not clear whether it is the self oscillation or just an increased variability (small magnitude, 0.5Sv) as the system approaches the threshold. It appears that AMOC variance is of comparable or even larger magnitude during 6-11kyr (Fig 5b). Is this also corresponding to self-oscillation?

We are now more cautious with our statement and only write that variability is increased. The large variability at 6-11 kyr is related to density changes in the Irminger Sea. We discuss this in the new manuscript version in section 3.1, but we didn't see indications of oscillations:

"After about 6 kyr, the changes in the North Atlantic density profile result in shifts in the spatial pattern of NADW formation. NADW formation moves south as vertical density profiles in the subpolar east North Atlantic stabilise under a freshening of the surface and density profiles further south steepen due to surface cooling combined with subsurface warming (Fig. SI.7-9). Apart from temporary volatility, the mean AMOC strength is not affected by these changes, but freshwater and heat advection into the North Atlantic is reduced, sea ice expansion increases in the eastern North Atlantic and AMOC variance (calculated over a moving 50-year window) is increased (Fig. 5)."

P13P26-33: as discussed, results from B.slow.b seem not to support the hysteresis behavior with respect to radiative forcing change. What about stability/sensitivity of the AMOC at ~6kyr in B.slow?

We referred to hysteresis behaviour here because the radiative forcing that is required to cause the abrupt weakening of AMOC is not the same as the forcing required for strengthening it again. We are clearer in the revised manuscript. We did not test AMOC stability explicitly at 6 kyr because the high variability seems to cease once the density field has re-adjusted.

P13L36: Orbital configuration consists of three orbital parameters. Their combinations in the chosen time slices are different but this does not mean the associated climatic impacts are significantly distinct, for instance, 21ka versus 0ka. It is thus better to show values of obliquity, precession, eccentricity and boreal summer insolation for the chosen time slices here, which would be helpful to clarify whether orbital forcing matters the transient behavior of the AMOC. A better approach to test roles of orbital configurations is to re-conduct such transient experiments based on orbital sensitivity, for example, high versus low obliquity experiments (e.g. experiments in Extended Data Table 1 of Zhang et al 2021).

We provide the orbital parameter values for each experiment in the SI. We were mostly interested here to see that changing the orbital configuration does not substantially alter the simulation results. It would be interesting to investigate the role of orbital changes for thermal thresholds in more detail in the future.

P14L5-7: not a full list of key relevant papers. Please add Knorr & Lohmann 2007, Banderas et al 2012 and Zhang et al 2017. Re multiple stable AMOC states, the difference in the strength of the AMOC is significantly different with a magnitude of >5Sv. In this context, it appears that the metastable AMOC states proposed here are perhaps sub-states of the interglacial/glacial AMOC state. Given the low AMOC variability in Bern3D, I assume this might not be reproducible by full GCMs nor perhaps in proxies.

We thank the reviewer for these additional relevant references, which we add to the paragraph. We also add further SI figures that show that each of the four persistent AMOC strengths is associated with different Greenland temperatures and North Atlantic sea ice extents, suggesting that they correspond to different climate states. Further, we agree, that it would be interesting to see this tested with a full GCM in future studies

P14L12-16: what's the exact role of 'heat advection" in AMOC mode transition? A positive feedback, a trigger or else? It would be good to have a clearer description here to specify the importance of heat advection.

Yes, we clarify this. Changes in heat convergence only seem to stabilise density profiles in the North Atlantic rather than causing AMOC shifts. We add these two paragraphs to section 3.1:

"After about 6 kyr, the changes in the North Atlantic density profile result in shifts in the spatial pattern of NADW formation. NADW formation moves south as vertical density profiles in the subpolar east North Atlantic stabilise under a freshening of the surface and density profiles further south steepen due to surface cooling combined with subsurface warming (Fig. SI.7-9). Apart from temporary volatility, the mean AMOC strength is not affected by these changes, but freshwater and heat advection into the North Atlantic is reduced, sea ice expansion increases in the eastern North Atlantic and AMOC variance (calculated over a moving 50-year window) is increased (Fig. 5). Transport of heat and salinity into the North Atlantic decreases (Fig. 5f, g) and North Atlantic SST and SSS decrease (Fig. 5e). Reduced influx of Atlantic waters also causes sudden cooling and freshening in the Irminger Sea (Fig. SI.8)."

"The simulated step changes in AMOC strength in our simulations are thus the response to gradual surface cooling and freshening, and occur when NADW formation shifts southwards. The resulting redistribution of heat and salinity cause sudden shifts in the vertical density profiles and sea ice expansion which consolidate the new circulation state."

P14L26-27: it is not true. For instance, Zhang et al 2017 applying a fully coupled AOGCM proposes that atmospheric CO2 levels are of control for glacial AMOC bi-stability.

Our intention here was to understand why Oka et al. (2021), specifically, required a stronger forcing in the southern hemisphere for thermal thresholds to arise, while we see thermal thresholds in our model under a globally uniform forcing. We clarify this and also refer to Zhang et al. (2017):

"In their examination of thermal forcing of both hemispheres in an ocean-only model, Oka et al. (2021) found that thermal AMOC thresholds only exist in the ocean model COCO if the Southern Hemisphere is cooled more than the Northern Hemisphere. In contrast, Zhang et al. (2017) found sudden AMOC changes also due to changes in well-mixed greenhouse gases. In our simulations with Bern3D, we also find thermal thresholds with similar cooling rates in both hemispheres, but only after salinity redistributions and changing meteoric freshwater fluxes in response to about six thousand years of global cooling. In accordance with Lohmann et al. (2023), we found that the observed shifts in AMOC strengths were thus preceded by cascades of density and circulation field changes, the number and sequence of which depend on the strength of the forcing. It would be interesting to test whether models seemingly without thermal thresholds under North Hemispheric cooling can reach such thresholds on the timescale of tens of thousand years. It is also possible that changing meteoric freshwater fluxes are essential for the existence of such a thermal threshold, which does not therefore appear in an ocean model without a thermally responsive atmosphere with a climate-driven freshwater balance."

P14L28-30: this may be true if comparing with other EMICs or simple models but not for GCM. Please clarify.

This was a wrong conception, we remove this statement from the manuscript.

P14L35: please provide modeling results or relevant literatures to support this statement especially regarding poleward moisture transport. It appears to me Fig 5e would be the right panel to refer to given the different trends between Atlantic and North Atlantic freshwater convergence. Sentences in P11L13-14 seem already touch this point, but it requires future clarification to link them to moisture transport and so on.

The statement on moisture transport was specific to our model. The wind field is constant but the water holding capacity of air decreases with the temperature decline. Hence, less moisture is transported polewards by the large-scale atmospheric circulation. We clarify this in the text. We also add more specific metrics of the changing water balance to the figures. We add SSS and SST timeseries for the Caribbean and Irminger Sea, as well as marine freshwater fluxes across latitudes 37.5°N and 62.5°N to Fig. 5, and spatial changes of P-E in the North Atlantic to the SI.

The additional text in section 3.3 is:

"In a model with a climate-sensitive meteoric freshwater balance, climate cooling reduces evaporation and the water-holding capacity. In such a model, cooling can then affect seawater density directly via changing temperatures, and indirectly via changing the meteoric freshwater balance and seawater salinity. In this context, it is also important to consider spatial changes in atmospheric dynamics, which are kept constant in our simulations, i.e. the moisture content of air changes with climate but not the direction or strength of winds which disperse it. In our model, a decrease in the water-holding capacity of air therefore directly leads to a reduction of the large scale atmospheric moisture transport from low to high latitudes. Several studies found that wind stress changes in the North Atlantic and Southern ocean are important controls on AMOC stability (e.g. Arzel et al., 2008, Yang et al., 2016) and thermal thresholds (Oka et al., 2012). The specific effects of atmospheric dynamics on meteoric freshwater forcing on AMOC would be an additional relevant topic for future studies."

P15: it is good to see the discussion about potential impacts of other parameters, especially ice sheet topography and associated wind, on the simulated AMOC change in different transient runs. In a glacial cycle, both changes in radiative forcing (e.g. CO2) and wind circulation/gateway caused by ice volume changes play a role in the strength/stability of the AMOC (Hu et al., 2011, Zhang et al 2014, 2017, 2021). Of most relevance here is their opposite impacts on the strength of the AMOC through glacial cycles in comparison to the thermal forcing (Barker and Knorr 2021). In this study, the authors investigated the roles of changes in radiative forcing in AMOC stability, which is the half story of AMOC multi-equilibria in glacial cycles. How do changes in those key parameters influence the results of A experiments? I would be happy to see more comprehensive discussion around this here as well as in Section 3.4 and 3.5. Perhaps, Section 3.3-3.5 can be integrated to one section to highlight and discuss the current understanding of AMOC stability, impacts of current model limitation on the current results and data-model comparison, and their implications and future perspectives.

We follow the reviewer's advice and add more discussion of AMOC stability and model limitations by combining sections 3.3 and 3.5. The new paragraph reads:

"In our transient simulations covering the past 788 kyr, the AMOC strength decreases during glacial phases solely due to changes in the hydrological cycle and sea ice that are induced by orbital, greenhouse gas and dust-driven temperature changes. The existence of multiple stable AMOC modes under varying thermal or radiative forcings has been found in various GCMs (e.g. Knorr and Lohmann, 2007, Oka et al., 2012, Banderas et al., 2012, Brown and Galbraith, 2016, Zhang et al., 2017, Klockmann et al., 2018). In agreement with previous studies, we found multiple (meta)stable AMOC circulation modes with distinct AMOC strengths for radiative forcing levels between full glacial and interglacial climate states. Moreover, we find that the transitions between these states occur abruptly, some within as little as 100 years. In accordance with Lohmann et al. (2023), we found that these shifts in AMOC strengths were thus preceded by cascades of density and circulation field changes, the number and sequence of which depend on the strength of the forcing. Similar to the findings from Oka et al. (2021), these AMOC transitions arise primarily from salt redistribution in the ocean and sea ice expansion into deep convection zones. In our simulations, each transition in AMOC strength is associated with a shift in the convergence of heat and salt fluxes and a southward expansion of sea ice into the North Atlantic which increasingly decouples the surface ocean buoyancy from the atmosphere. In the meta-stable modes, the density gradients in the main North Atlantic deep convection zones are strongly dependent on surface buoyancy fluxes. In these modes, small changes in buoyancy or sea ice cover can cause resumption or cessation of convection, which makes the AMOC sensitive to small perturbations. The AMOC is only pushed into its weakest mode when the net heat advection into the North Atlantic is strongly reduced and all former convection sites in the subpolar North Atlantic are sea ice-covered. In their examination of thermal forcing of both hemispheres in an ocean-only model, Oka et al. (2021) found that thermal AMOC thresholds only exist in the ocean model COCO if the Southern Hemisphere is cooled more than the Northern Hemisphere. In contrast, Zhang et al. (2017) found sudden AMOC changes also due to greenhouse gas changes without a special focus on the Southern Hemisphere. In our simulations with Bern3D, we also found thermal thresholds with similar cooling rates in both hemispheres, but only after salinity redistributions and changing meteoric freshwater fluxes in response to about six thousand years of global cooling.

It is possible that changing meteoric freshwater fluxes are essential for the existence of such a thermal threshold, which does not therefore appear in an ocean model without a thermally responsive atmosphere with a climate-driven freshwater balance. In a model with a dynamic energy moisture balance component, atmospheric cooling reduces evaporation, the water-holding capacity of the atmosphere and the atmospheric poleward transport of moisture. In such a model, cooling can then affect seawater density directly via changing temperatures, and indirectly via changing the meteoric freshwater balance and surface salinities. These changes would induce additional kinematic changes (i.e., in the wind fields) in fully dynamic atmosphere models, but are kept constant in our simulations, i.e. the moisture content of air changes with climate but not the direction or strength of winds which disperse it. Accordingly, wind stress fields are also kept constant here. Changes in wind stress have been documented to exert important controls on AMOC stability (e.g. Arzel et al., 2008, Yang et al., 2016) and thermal thresholds (Oka et al., 2012). These effects have been investigated in detail with the Bern3D model by Pöppelmeier et al. (2021) focusing on LGM boundary conditions.

The primary importance of salinity and heat redistributions as well as sea ice extent in the North Atlantic for the simulated AMOC shifts resembles the findings from Ando and Oka (2021)'s hosing experiments under LGM conditions and Zhang et al. (2017)'s simulations of AMOC shifts in response to CO2 changes in intermediate-glacial conditions. While our experiments are run with pre-industrial boundary conditions, the initial location of convection sites between Greenland and the British Islands (areas with lowest density differences over upper 1000m in Fig. SI.8) resembles the LGM and intermediate-glacial circulation states in Ando and Oka (2021) and Zhang et al. (2017). Ganopolski and Rahmstorf (2001) found that the possibility of a southward shift of deep convection depends on the latitude of prior deep convection and the density field further south, and Oka et al. (2012) showed that the location of deep convection and its distance from the winter sea ice edge define thermal thresholds in AMOC strength. Several controls on the location and strength of deep convection in the North Atlantic, that would have affected AMOC stability over glacial cycles, have been established. The location of deep convection is dependent on wind fields, climate and sea level/bathymetry (Ganopolski and Rahmstorf, 2001, Oka et al., 2012, Zhang et al., 2017), and thus the thermal AMOC thresholds are model and forcing dependent (Oka et al., 2012). Our simulations capture the albedo effect of varying terrestrial ice sheet extent, but we do not consider their orography or sea level effects, including impacts on the atmospheric circulation, which were shown to affect AMOC (Li and Born, 2019, Pöppelmeier et al., 2021). Previous studies suggested that pre-industrial or intermediate ice sheet configurations are required to even produce a thermal AMOC threshold in the range of glacial-interglacial CO_2 concentrations and that the presence of a full glacial Laurentide ice sheet prevents such a threshold (e.g. Klockmann et al., 2018). In addition, changes in the interconnection of marine basins, specifically the Bering Strait, also affect AMOC stability (Hu et al. 2012). The values of the thermal thresholds in our experiments are thus likely sensitive to the model design and initiation. Pöppelmeier et al. (2021) showed that the sensitivity of Bern3D to freshwater hosing increases when additional LGM boundary conditions are prescribed (changed wind fields, closed Bering Strait, tidal mixing differences due to sea level changes). The different wind fields and tidal mixing strengthened AMOC and increased the salt and heat transport into the subpolar North Atlantic. This could mean that stronger cooling is required to stabilise the water column in the Irminger Sea and reach the first thermal threshold, when the full range of glacial boundary conditions are applied. Closure of the Bering Strait increased the salt advection feedback, which stabilises the weak circulation state without deep water formation in the subpolar North Atlantic.

Further investigations are needed to determine how changes in strength and location of the wind stress due to the ice sheet's orography, sea level and Bering strait closure would affect sea ice formation in the northern North Atlantic and the AMOC thresholds in our simulations quantitatively. Since we chose to focus on radiation driven AMOC changes in our experiments, we would not expect a close model-data match with reconstructed AMOC changes from paleo-records. Our simulations show that the reconstructed temperature changes had the potential to alter the density field in the North Atlantic by redistributing heat and salt, and that some of these changes might have resulted in abrupt changes of AMOC strength. By testing a wide range of glacial-interglacial temperature changes, our experiments demonstrate that the cooling during glacial periods likely contributed to a weakened AMOC. The strength and timing of the weakening depends on the actual temperature change in the North Atlantic which would have been modulated by changes in winds and ice shields."

P15 L14-15: Please add relevant reference to "different representations of processes affecting AABW density changes".

We add a reference here: "(e.g. brine rejection, Bouttes et al., 2011)"

P17 Figure 9: it would be good to flip y-axis of d18O curve upside down, given the tradition of plotting LR04/sea level curves.

We invert the y-axis of the $\delta^{18}O$ panel as suggested.

References

Ando, T. and Oka, A., 2021. Hysteresis of the glacial Atlantic meridional overturning circulation controlled by thermal feedbacks. Geophysical Research Letters, 48(24), p.e2021GL095809.

Arzel, O., England, M.H. and Sijp, W.P., 2008. Reduced stability of the Atlantic meridional overturning circulation due to wind stress feedback during glacial times. Journal of climate, 21(23), pp.6260-6282.

Banderas, R., Álvarez-Solas, J. and Montoya, M., 2012. Role of CO 2 and Southern Ocean winds in glacial abrupt climate change. Climate of the Past, 8(3), pp.1011-1021.

Barker, S., Chen, J., Gong, X., Jonkers, L., Knorr, G. and Thornalley, D., 2015. Icebergs not the trigger for North Atlantic cold events. Nature, 520(7547), pp.333-336.

Berger, V.W. and Zhou, Y., 2014. Kolmogorov–smirnov test: Overview. Wiley statsref: Statistics reference online.

Bouttes, N., Paillard, D., Roche, D.M., Brovkin, V. and Bopp, L., 2011. Last Glacial Maximum CO2 and δ13C successfully reconciled. Geophysical Research Letters, 38(2).

Bozbiyik, A., Steinacher, M., Joos, F., Stocker, T.F. and Menviel, L., 2011. Fingerprints of changes in the terrestrial carbon cycle in response to large reorganizations in ocean circulation. Climate of the Past, 7(1), pp.319-338.

Broecker, W.S., Blanton, S., Smethie Jr, W.M. and Ostlund, G., 1991. Radiocarbon decay and oxygen utilization in the deep Atlantic Ocean. Global Biogeochemical Cycles, 5(1), pp.87-117.

Buizert, C. and Schmittner, A., 2015. Southern Ocean control of glacial AMOC stability and Dansgaard-Oeschger interstadial duration. Paleoceanography, 30(12), pp.1595-1612.

de Vries, P. and Weber, S.L., 2005. The Atlantic freshwater budget as a diagnostic for the existence of a stable shut down of the meridional overturning circulation. Geophysical Research Letters, 32(9).

Galbraith, E. and de Lavergne, C., 2019. Response of a comprehensive climate model to a broad range of external forcings: relevance for deep ocean ventilation and the development of late Cenozoic ice ages. Climate Dynamics, 52, pp.653-679.

Ganopolski, A. and Rahmstorf, S., 2001. Rapid changes of glacial climate simulated in a coupled climate model. Nature, 409(6817), pp.153-158.

Gregory, J.M., Dixon, K.W., Stouffer, R.J., Weaver, A.J., Driesschaert, E., Eby, M., Fichefet, T., Hasumi, H., Hu, A., Jungclaus, J.H. and Kamenkovich, I.V., 2005. A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO2 concentration. Geophysical Research Letters, 32(12).

Hu, A., Meehl, G.A., Han, W., Timmermann, A., Otto-Bliesner, B., Liu, Z., Washington, W.M., Large, W., Abe-Ouchi, A., Kimoto, M. and Lambeck, K., 2012. Role of the Bering Strait on the hysteresis of the ocean conveyor belt circulation and glacial climate stability. Proceedings of the National Academy of Sciences, 109(17), pp.6417-6422.

Jackson, L.C. and Wood, R.A., 2018. Hysteresis and resilience of the AMOC in an eddy-permitting GCM. Geophysical Research Letters, 45(16), pp.8547-8556.

Jackson, L.C., Alastrué de Asenjo, E., Bellomo, K., Danabasoglu, G., Haak, H., Hu, A., Jungclaus, J.H., Lee, W., Meccia, V.L., Saenko, O. and Shao, A., 2023. Understanding AMOC stability: the North Atlantic hosing model intercomparison project. Geoscientific Model Development, 16, pp.1975-1995.

Joos, H., Madonna, E., Witlox, K., Ferrachat, S., Wernli, H. and Lohmann, U., 2017. Effect of anthropogenic aerosol emissions on precipitation in warm conveyor belts in the western North Pacific in winter–a model study with ECHAM6-HAM. Atmospheric chemistry and physics, 17(10), pp.6243-6255.

Klockmann, M., Mikolajewicz, U. and Marotzke, J., 2018. Two AMOC states in response to decreasing greenhouse gas concentrations in the coupled climate model MPI-ESM. Journal of Climate, 31(19), pp.7969-7984.

Knorr, G. and Lohmann, G., 2007. Rapid transitions in the Atlantic thermohaline circulation triggered by global warming and meltwater during the last deglaciation. Geochemistry, Geophysics, Geosystems, 8(12).

Li, C. and Born, A., 2019. Coupled atmosphere-ice-ocean dynamics in Dansgaard-Oeschger events. Quaternary Science Reviews, 203, pp.1-20.

Lohmann, J., Dijkstra, H.A., Jochum, M., Lucarini, V. and Ditlevsen, P.D., 2023. Multistability and Intermediate Tipping of the Atlantic Ocean Circulation. arXiv preprint arXiv:2304.05664.

Menary, M.B., Roberts, C.D., Palmer, M.D., Halloran, P.R., Jackson, L., Wood, R.A., Müller, W.A., Matei, D. and Lee, S.K., 2013. Mechanisms of aerosol-forced AMOC variability in a state of the art climate model. Journal of Geophysical Research: Oceans, 118(4), pp.2087-2096.

Menviel, L., Timmermann, A., Mouchet, A. and Timm, O., 2008. Meridional reorganizations of marine and terrestrial productivity during Heinrich events. Paleoceanography, 23(1).

Menviel, L., Joos, F. and Ritz, S.P., 2012. Simulating atmospheric CO2, 13C and the marine carbon cycle during the Last Glacial–Interglacial cycle: possible role for a deepening of the mean remineralization depth and an increase in the oceanic nutrient inventory. Quaternary Science Reviews, 56, pp.46-68.

Mikolajewicz, U., Santer, B.D. and Maier-Reimer, E., 1990. Ocean response to greenhouse warming. Nature, 345(6276), pp.589-593.

Oka, A., Hasumi, H. and Abe-Ouchi, A., 2012. The thermal threshold of the Atlantic meridional overturning circulation and its control by wind stress forcing during glacial climate. Geophysical Research Letters, 39(9).

Oka, A., Abe-Ouchi, A., Sherriff-Tadano, S., Yokoyama, Y., Kawamura, K. and Hasumi, H., 2021. Glacial mode shift of the Atlantic meridional overturning circulation by warming over the Southern Ocean. Communications Earth & Environment, 2(1), p.169.

Okazaki, Y., Timmermann, A., Menviel, L., Harada, N., Abe-Ouchi, A., Chikamoto, M.O., Mouchet, A. and Asahi, H., 2010. Deepwater formation in the North Pacific during the last glacial termination. Science, 329(5988), pp.200-204.

Pöppelmeier, F., Scheen, J., Jeltsch-Thömmes, A. and Stocker, T.F., 2021. Simulated stability of the Atlantic meridional overturning circulation during the Last Glacial Maximum. Climate of the Past, 17(2), pp.615-632.

Praetorius, S.K. and Mix, A.C., 2014. Synchronization of North Pacific and Greenland climates preceded abrupt deglacial warming. Science, 345(6195), pp.444-448.

Rahmstorf, S., 1996. On the freshwater forcing and transport of the Atlantic thermohaline circulation, Clim. Dyn., 12, 799–811.

Sherriff-Tadano, S., Abe-Ouchi, A., Yoshimori, M., Ohgaito, R., Vadsaria, T., Chan, W.L., Hotta, H., Kikuchi, M., Kodama, T., Oka, A. and Suzuki, K., 2023. Southern Ocean surface temperatures and cloud biases in climate models connected to the representation of glacial deep ocean circulation. Journal of Climate, 36(11), pp.3849-3866.

Stommel, H., 1961. Thermohaline convection with two stable regimes of flow. Tellus, 13(2), 224– 230. https://doi.org/10.3402/tellusb.v13i2.12985

Tetard, M., Licari, L. and Beaufort, L., 2017. Oxygen history off Baja California over the last 80 kyr: A new foraminiferal-based record. Paleoceanography, 32(3), pp.246-264.

Vettoretti, G., Ditlevsen, P., Jochum, M. and Rasmussen, S.O., 2022. Atmospheric CO2 control of spontaneous millennial-scale ice age climate oscillations. Nature Geoscience, 15(4), pp.300-306.

Wang, Y.J., Cheng, H., Edwards, R.L., An, Z.S., Wu, J.Y., Shen, C.C. and Dorale, J.A., 2001. A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. Science, 294(5550), pp.2345-2348.

Weijer, W., Cheng, W., Drijfhout, S.S., Fedorov, A.V., Hu, A., Jackson, L.C., Liu, W., McDonagh, E.L., Mecking, J.V. and Zhang, J., 2019. Stability of the Atlantic Meridional Overturning Circulation: A review and synthesis. Journal of Geophysical Research: Oceans, 124(8), pp.5336-5375.

Weijer, W., Cheng, W., Garuba, O.A., Hu, A. and Nadiga, B.T., 2020. CMIP6 models predict significant 21st century decline of the Atlantic meridional overturning circulation. Geophysical Research Letters, 47(12), p.e2019GL086075.

Winckler, G., Anderson, R.F., Fleisher, M.Q., McGee, D. and Mahowald, N., 2008. Covariant glacial-interglacial dust fluxes in the equatorial Pacific and Antarctica. science, 320(5872), pp.93-96.

Yang, H., Wang, K., Dai, H., Wang, Y. and Li, Q., 2016. Wind effect on the Atlantic meridional overturning circulation via sea ice and vertical diffusion. Climate Dynamics, 46, pp.3387-3403.

Zhang, X., Prange, M., Merkel, U. and Schulz, M., 2014. Instability of the Atlantic overturning circulation during Marine Isotope Stage 3. Geophysical Research Letters, 41(12), pp.4285-4293.

Zhang, X., Knorr, G., Lohmann, G. and Barker, S., 2017. Abrupt North Atlantic circulation changes in response to gradual CO2 forcing in a glacial climate state. Nature Geoscience, 10(7), pp.518-523.