Response to Reviewer:

Firstly we would like to thank the anonymous reviewer for the time they have taken to read through the manuscript and make their suggestions. However we do have to refute several of their suggestions as they unfortunately seem to have mistaken the goals of our investigation, suggesting we may need to clarify these goals more explicitly in the manuscript in addition to adding additional content around caveats for the methods employed. An itemized set of responses to their comments is below, beginning with the more substantial points and technical points following.

As highlighted by the authors, the length of their simulation is very short, which strongly hamper the interpretations of the results from those simulations for paleoclimate timescales, which are usually two order magnitude longer, as illustrated in Fig. S1 from the paper
...the first caveat is discussed appropriately in the paper..

We agree that one limitation of our study is the short duration of the simulations in that there were still non-zero trends in some relevant climate metrics (i.e. AMOC). That is why we raise this point in the paper in a way that the reviewer describes as "appropriately discussed". We think it's also important to draw attention to the fact that the duration of the simulations is long enough for the surface transports of freshwater, which is the focus of our work, to have stabilized in most of the runs. Nevertheless, since receiving this review, we have resubmitted the runs, they have been extended by a few years (5-10 years run depending). Updated figures with these additional data and any updated discussion where required are included in the revised text. There have been no major changes to the discussion because of this extension as the relative ordering of the freshening has remained the same, excepting that for the North Atlantic DWF site the OBS Mackenzie River injection has become comparable to the Fennoscandia injection. A main point of our paper is that the two order magnitude longer typical paleoclimate simulations the reviewer is referring to are also difficult to interpret given their own sources of model uncertainty such as limited resolution, uncertainties and errors in boundary conditions/forcings, etc.

The updates with former work is quite far from substantial, and the models used is very close to the one used in e.g. Condron & Winsor (2012). After almost a decade, computing power have strongly increased, so that these simulations now cannot be really considered state-of-the-art anymore, since far higher-resolution ocean-only simulations now exist, and do show that having even stronger resolution play a crucial role for the mean state of the AMOC (cf. Hirschi et al. 2020). As such, I am surprised that the authors still consider such short simulations (cf. point 1).

We would argue that this study represents a significant improvement on previous work. As described in the introduction, there are three common issues in the design of experiments that implement freshwater in paleo contexts: 1) the freshwater is deposited directly over the sites of deepwater formation to compensate for inadequate horizontal resolution, 2) the amounts of freshwater used are unrealistically large, and 3) inconsistent/unrealistic ocean gateways. We also note that previous studies address aspects, but not all, of each of these issues. For example, Roche et. al., (2009) explored the impact of varying geographic regions, used appropriate gateways, and to a lesser extent used reasonable freshwater volumes but lacked the horizontal resolution to capture key transport features (e.g. boundary currents). Condron and Windsor (2012) and Hill and Condron (2014) addressed the horizontal resolution (3) and partially the geographic location issue (2) but used unrealistic volumes (an order of magnitude larger) and had gateways and bathymetric features inconsistent with reconstructions (e.g. Barents-Kara being glaciated, assumption of eustatic sea level adjustment, among other limitations). They did not consider Fennoscandian and GoM freshwater sourcing nor did they consider the impact of open/close Barents Strait. We address 1) via releasing freshwater at coastal locations consistent with glacial reconstructions and by using a model well able to represent small scale features known to be important in the transport of coastally released freshwater. Issue 2) is addressed by bounding our fluxes by the upper limits of a self-consistent glacial reconstruction. Finally, 3) is addressed by using the relative sea level component of the self-consistent glacial reconstruction to configure our land-sea boundary and bathymetry, as well we address a limitation of the reconstruction by examining the impact a key gateway (the Bering Strait) has on our results for the most proximal injection location (the Mackenzie River). In summary, ours is the first such study to address these three common issues simultaneously and as such represents a significant improvement on previous work. These points are made clearer in the manuscript.

We point out the main focus of this work, the representation of surface transports and features generally regarded as subgrid scale, would not benefit from existing updates to the model as the features of interest are already adequately represented in the version we use. Updates to the model appear to largely center around bug-fixes and documentation updates (https://github.com/MITgcm/MITgcm/releases) without substantial effect on the representation of surface transports and eddies. As well, with regards to increasing the resolution of ocean-only simulations, we do note there are some entries in Hirschi et al. (2020) (which for the benefit of those unfamiliar with the work, is a review paper examining the representation of AMOC under present-day conditions from multiple sub 1 degree resolution model simulations extracted from 23 different publications) which are higher resolution. However, only one is a global ocean-only simulation which is above our grid resolution (Moat et al. (2016) which used 1/12 degree). Thus, we contend that the model configuration used in this study is of comparable complexity and resolution to the multi-model ensemble of simulations presented in Hirschi et al (2020). We make this point in the revised submission.

Furthermore, we would argue that the existence of higher-complexity or higher-resolution simulations for present-day phenomena does not negate the value of a study focussing on past oceanic phenomena using a model with slightly less complexity and lower resolution. The study here and those previous studies by Condron, Windsor, and Hill are still the highest-resolution, ocean-only simulations to date using bathymetry and boundary conditions that are not either pre-industrial or present-day (though the AWI group in Bremerhaven have conducted some very interesting paleo work with their unstructured high-resolution FESOM configuration). With regards to computing power having strongly increased, indeed some features of computing power have increased substantially but unfortunately model wall time does not decrease as per Moore's law as one might hope and enterprise computing focuses on parallel compute performance with a focus on stability, not single-thread performance, which does not translate into performance gains nearly as effectively.

Line 90: such ocean-only model are simulations are not that costly within present-day computing time standard (e.g. Penduff et al. 2018 who considered 50 members of multi-decadal high resolution simulations…). Improvements as compared to former work with Condron as co-author, dating than almost a decade is not clear at all, while the main message remains also quite similar with this former work.

It appears to us that the reviewer is making two separate claims in these comments. Firstly, they would like us to have run a larger ensemble of longer simulations or use a higher spatial resolution in the ensemble we did produce on the basis that such has been done in a previous study examining a different scientific question altogether. Secondly, they argue that the updates with former work (in the reviewer's words, "the use of glacial boundary conditions, and a more systematic analysis of the different potential outlet locations as well as the consideration of smaller rate of freshwater release, more in line with recent reconstructions.") are not substantial.

In regards to the first claim, we would argue that setting the bar for minimal requirements in an experiment to equal the most resource-intensive project published to date is illogical as doing so would rule out almost all researchers except those with the greatest access to resources. The simulations we have conducted represent an advancement over previous studies and are more than sufficient to provide important insight into the surface transport of continental runoff given we explicitly address 3 significant weaknesses in previous studies. Expecting us to greatly expand our simulation numbers and durations just because other multi-institutional projects have done so in other unrelated contexts is not reasonable nor necessary. As it is, these simulations occupied a substantial component of our computational allocation budget for the years during which they were run, the cost of which was O(10,000-100,000+CAD/year). Conducting over 100 years of simulation has consumed sizable compute resources unavailable to many researchers and required a Compute-Canada Resource Allocation Competition grant on the Niagara national system for both the storage (several hundred TB of data) and the compute time. Furthermore, we remind the reviewer how resolution, time-stepping, and compute-cost of a numerical model scales with resolution (generally cubic or higher (if the number of vertical levels is increased), such that a halving of horizontal resolution requires about 8x or more flops), Penduff (2018) having used a model roughly 33% coarser would be be able to execute their goals with more moderate compute resources (potentially even more so given no grid topology was provided for their experiments whereas we used a cubed-sphere topology which results in a generally uniform horizontal grid spacing of ~18km globally). Our experiments, when taken as a whole, are comparable to those presented in Hirschi et al. (2020).

As to the second claim, given we have conducted a study which explicitly addresses the primary drawbacks of the previous relevant works through "a more systematic analysis of the different potential outlet locations as well as the consideration of smaller rate of freshwater release, more in line with recent reconstructions." it would seem the reviewer contradicts their own claim of insufficiency. Three key limitations of previous studies, as stated above, severely limited what conclusions could be drawn from them. We have addressed those limitations.

Improvements in our understanding of the impact of ocean resolution from models of oceanic circulation need to be more appropriately discussed (cf. Hirschi et al. 2020, Le Corre et al. 2020)

A valid point, we included additional text to address this in the experimental design section.

The use of glacial boundary conditions apparently lead to a collapse of the AMOC in the ocean-only GCM used. The authors qualified it as a glacial state, but Fig. 6 shows a weakening AMOC index in the control simulation (which is thus not equilibrated at all) towards values of 2-4 Sv that rather correspond to an off state than a weak glacial states, according to e.g. Ganapolski and Rahmtorf (2001). AOGCMs indeed do not produce such weak state in glacial condition (e.g Kageyama et al. 2013, with all AOGCMs showing value

larger than 5 Sv in their mean state). Considering an off state has major implications in term of barotropic circulation, notably in the subpolar gyre, which makes the relevance of those results doubtful for examining freshwater pathways at the beginning of e.g. the Younger Dryas as it is suggested in the paper.
Line 145-146: This claim is not supported by anu figures, and I strongly doubt of this, given the very small value at 26°N. The AMOC is state rather resemble an off-state. Can we see the meridional streamfunction in the last 10 years of the control simulation?
and
Line 152: "glacial mode" sounds very optimistic. The authors might need to discuss more what is known from data and models concerning the mean state of the AMOC during the LGM…
and
Line 153: "reasonable". This might be a bit too much optimistic as well I think. Please discuss appropriately the state of your AMOC, or provide more evidences to support that it can be considered as a glacial state.
and
*The authors qualified it as a glacial state, but Fig. **S**6 shows a weakening AMOC index in the control simulation (which is thus not equilibrated at all)*

AMOC is not the focus of this paper and is only provided for context as it was expected portions of the community might seek information only on this metric despite the simulations here not being of long enough duration to make robust conclusions with regards to its behaviour (as further emphasised by the fact it was among the supplemental figures and not a central figure of the study as is common). Given this comment and others made, we clearly need to emphasise this further in the text. We emphasise this by the relocation of the AMOC discussion to the supplemental section where it will not distract readers but remains readily accessible for those who are interested in the results despite the limitations of this metric for this study.

Furthermore the unclear language "Off-state" can be interpreted in multiple ways. If the reviewer is referencing the Off/Heinrich mode of AMOC operation as well summarised in Rahmstorf (2002) then we argue this is incorrect, as this mode of operation precludes the formation of deep water in the North Atlantic whereas Figure S5 clearly shows a robust mixed layer (note: this figure shows the average calculated over a full year, this reduces the magnitude of the mixed layer depth by comparison to a shorter averaging period like monthly maximum) . Furthermore, we note that 26N was chosen to correspond with the present day RAPID array and to allow for easier comparison to the previous work of Condron and Windsor (2012). This is not the location of the peak value of overturning in the North Atlantic basin as is typically reported in most investigations for whom AMOC is a constructive climate metric and thus the reviewer's comparisons to previous works based upon this value are unfortunately not readily accomplishable. Additional information regarding this value is now provided in the supplement, as there is a roughly -6Sv offset resulting from using 26N rather than the peak (that is, our AMOC maximum averages are ~9.5-10Sv rather than ~3.5-4Sv).

With regards to the effect that a reduced AMOC has on features closer to or at the surface, we find that the surface circulation tends to lead the deeper ocean in studies examining this coupling, not vice versa. One of the potentially most important surface features would be the subpolar gyre, which on glacial timescales modulates the salt transport to deep water formation regions (Klockmann, 2020). Furthermore, this coupling is found to be weaker in higher resolution eddy-permitting models than in coarser resolution models (Meccia, et. al, 2021), further reducing the impact of this feature. The other main surface feature of note which can be strongly affected by a weaker AMOC would be the Gulf Stream (as the reviewer has pointed out in another comment). Caesar, et. al., (2018) indicates that the latitude of the separation point of the Gulf Stream from the coast of North America is modulated by the AMOC, with a weaker AMOC resulting in the Gulf Stream shifting northwards and closer to shore. As raised in the other comment, this is now mentioned in the modified text. However, this does not change the impact of the Gulf Stream on our simulations or conclusions, whereby the Gulf Stream acts as an effective barrier to meridional transport of freshwater.

With respect to the structure of the AMOC, there was not a figure included as AMOC and its structure is not the focus of the paper nor is it relevant for inclusion in the manuscript given our primary interest is surface transport over short durations. Regarding the disequilibrium of the AMOC, it is indeed trending downwards initially but is relatively flat within the annual variability (one standard deviation is 1-1.5Sv as noted in the caption of Figure S6) for the years 10+ in Figure S6. Furthermore, we make no claims on the equilibrium of the deeper ocean whose equilibrium time is well understood to be several thousand years. The focus of this investigation is the very uppermost layers of the ocean (the majority of the anomaly is contained only within the top 30m of the ocean) whose equilibrium time is within the range of our investigation's duration.

Finally, to make clear that we do not consider analyses of the AMOC appropriate on the basis of these simulations, we added a corresponding line to the AMOC discussion section in the supplemental materials,

---

The discussion of the implications of their results for paleoclimate understanding is very weak and deserve to be strengthen. What does those results mean in regard to existing literature that GOM and GSL affect so weakly the convection zones? What does that mean in terms of last deglaciation storylines?

line 267: A proper discussion of the implications in terms of the storyline of AMOC changes over the deglaciation and the link with freshwater release should be provided. As an example, we can assume that those experiments strongly support a major role for freshwater release from Fennoscandia, as suggested in e.g. Toucanne et al. 2009. Please, further elaborate on this topic in light of existing literature.

We have added additional discussion in the paper to address these points from the viewpoint of freshening of deep water formation regions rather than AMOC given the previously discussed de-emphasis of AMOC in our work.

Line 99-100: more should be said concerning the experimental design. Since these are ocean-only simulations, how are considered the boundary conditions? Is there any SSS restoring? How evaporation fluxes are computed,

We are not entirely clear what "more" the reviewer would like described in the experimental design section. However, between addressing the specific questions posed here (with relevant additions to the revised text) and those in Reviewer 2's review, we hope that we have satisfied the reviewer's request. There is no surface restoration, this would defeat the purpose of the experiments conducted. Evaporation is handled internally by the model in the EXF (external forcing package) from provided precipitation, relative humidity, and surface runoff fields.

Line 134: this very zonal Gulf Stream might also be related with the fact that the AMOC is in an off-state, since this can strongly impact Gulf Stream pathway (e.g. Caesar et al. 2018)

As discussed in the manuscript the zonal Gulf Stream is an artefact of the surface forcing, see plots below demonstrating the magnitude of the zonal component of our surface winds relative to a pre-industrial control simulation from CCSM4 (the closest model to what generated our original surface forcing). However a brief discussion of the northward/southward shifting of the separation point of the Gulf Stream from the East Coast of North America as a function of AMOC is now included with appropriate caveats in the supplemental section. Furthermore, we have conducted a sensitivity experiment replacing our glacial winds with that from the ERA40 reconstruction and find that this results in a less zonal Gulf Stream as expected. This additional information is available in the modified supplementary materials.

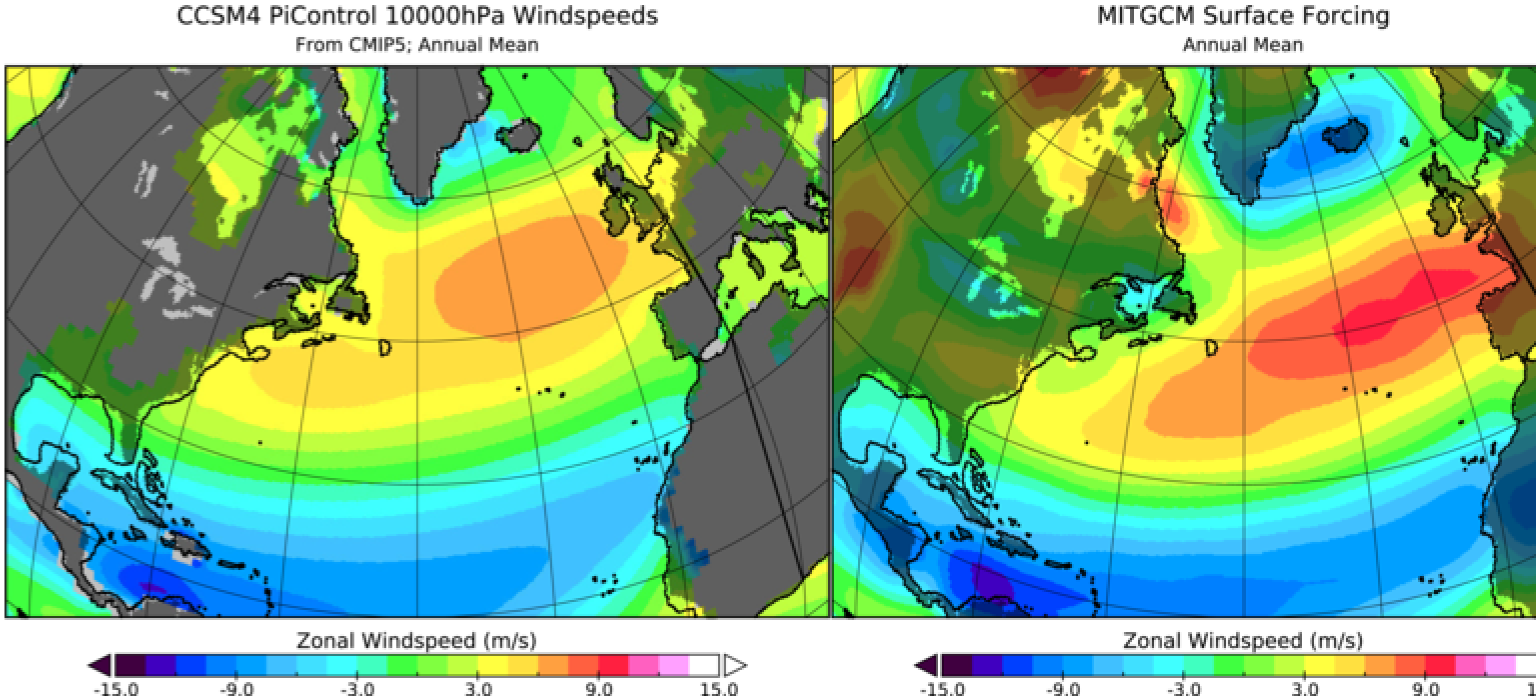

Figure 1: The data are difficult to see during YD due to very strong red. Please consider another colour to allow proper examination of the curves

Accepted, color changed to black and line weight increased for better contrast.

Line 41 and elsewhere: "eg." Should be replaced by "e.g."

Accepted

Line 89-90: How many vertical levels in the model?

The model features 50 vertical levels, this is now noted in the model description section.

Line 141: “Labrador” Sea (not sea)

Fixed.

Line 225: “yr” is not defined.
yr is the CP style guide requested abbreviation for year. Defining this is not requested by the style guide and can be understood from context.

Line 267: An additional caveat is not properly discussed which is the fact that the authors consider here ocean-only model, which prevent from considering any potential coupled ocean-atmosphere feedback, which might play a role.

A useful suggestion, a brief discussion of this point has been added.

Line 268: “under Younger Dryas conditions”: this statement does not really reflect the off state that is simulated in the control simulation.

A weakened AMOC (McManus, et. al, 2004) and stadial surface conditions are very much the expected configuration of a Younger Dryas climate (Carlson A.E., 2013).

Line 276-279: it is quite unclear from where those estimates come from, which is weird to provide in the conclusion, since not shown in the result section. I assume, they are estimated from a similar approach as in line 232-241, which is considering an ocean without any circulation at all. This is quite a strong hypothesis… Thus, I’m not sure those estimates are really useful, especially in the conclusion.

They are indeed estimates using the same simple method as in lines 232-241, this has been made clearer in the conclusions. We chose the simpler of assumptions when making these ‘back of the envelope’ estimates, as the alternative would be to assume some structure of flow under a regime for which we do not have data (one would expect 2dSv of freshwater into a region to affect transport in/out of a region) and could easily scale the results for dramatic impact by making such assumptions. As noted on lines 237-238 this approach is reasonable for a simple estimate, as we find our calculated salinity anomaly for Fennoscandia to be comparable to the simplified hosing estimate within the first year. This localized hosing which we compare to is something which has been done before in previous investigations (see Roche, et. al., 2010) and hence why it was done for comparison.

Line 283: “better ways to mitigate this problem”: this sentence is quite enigmatic. Can you please clarify what is meant here?

This is the subject of upcoming work which is outside the scope of this manuscript. We have modified the text to express that this is the subject of upcoming work.

Line 284-285: it should be stated here that these investigations are done in an off-state for the AMOC, and during only 20 years.

AMOC is not the focus of this paper and is only provided for context (see previous discussion) and is not a relevant discussion point for the conclusions of this paper. The duration of the investigations are already described in the body of the paper and associated figures. We make clear that the AMOC state is reduced glacial model (it is not off).

References:

Carlson A.E. (2013) The Younger Dryas Climate Event. In: Elias S.A. (ed.) The Encyclopedia of Quaternary Science, vol. 3, pp. 126-134. Amsterdam: Elsevier.

Klockmann, M., Mikolajewicz, U., Kleppin, H., & Marotzke, J. (2020). Coupling of the subpolar gyre and the overturning circulation during abrupt glacial climate transitions. Geophysical Research Letters, 47, e2020GL090361. https://doi.org/10.1029/2020GL090361

McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D., & Brown-Leger, S. (2004). Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. nature, 428(6985), 834-837.

Meccia, V.L., Iovino, D. & Bellucci, A. North Atlantic gyre circulation in PRIMAVERA models. Clim Dyn (2021). https://doi.org/10.1007/s00382-021-05686-z

Rahmstorf, S. (2002). Ocean circulation and climate during the past 120,000 years. Nature, 419(6903), 207-214.

Roche, D. M., Wiersma, A. P., & Renssen, H. (2010). A systematic study of the impact of freshwater pulses with respect to different geographical locations. *Climate Dynamics*, *34*(7-8), 997-1013.