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. An itemized set of responses to their comments is below, beginning with the more substantial points and technical points following.

-The main story and motivation of the study is hidden behind all the text. It should be re-written with a clear and standard structure, where general description of the problem, goal of the paper and previous studies are clearly discussed.

As submitted for discussion, the paper introduction does follow a conventional structure (context, motivation, previous work and its shortcomings, and goals of the present manuscript). We also believe the abstract succinctly summarizes the main story. However, it seems this structure isn't coming out clearly for the reviewer. For this reason, we added some additional guiding sentences and words to make the logical argument clearer to all readers. Examples of the changes we made include:

'The goal of this study is to directly address all of these limitations...' 'We start our discussion of the experimental design with a brief overview of the model configuration...' 'Here we discuss the heredity of our simulations'

-Figure 1, is a very nice figure but it is not really discussed and not clear why you used it. I think it deserves more explanation.

We have added additional contextual information to better make use of Fig. 1 as well as expand the introduction to address the first issue the reviewer raises. As well, to increase the utility of Fig. 1 we have merged in the runoff flux figure from the supplement and placed it on the same time scale so as to better discuss our results in the context of the last deglaciation.

-The model description should be improved (e.g., you need to clearly specify that you use a global model and how many vertical levels your model has). Next, discuss the forcing. Then, explain the initialization of experiments, the control runs, number of spin up years (exact numbers), and total simulation years. Last, explain the experiments with all the needed details. In the current version, you might have given most of these information but it is done in a messy way.

We note the global nature of the model grid as well as the vertical level count in our revisions, as well we have made the heredity and durations of each of the experiments clearer via a table in the supplemental material.

We agree with the reviewer that the Experimental Design section could be structured better. However, we don't find the structure the reviewer suggests to be very helpful. Instead, we add additional information to the model description as requested, and then discuss the control simulations (initialization and forcing), followed by the freshwater injection runs (initialization and forcing). Since not all of the simulations are of the same duration, these types of details bog the text down. Instead, we've added a table to the supplement that specifies such details for each run. Furthermore, as with the introduction we have added additional guiding sentences.

-I think your Figure S2 should be discussed in this section and be used as a main figure and not a supplementary.

We have merged the useful information of Fig. S2 with Fig. 2.

- I wonder why there is no summer sea ice in the Arctic in the region above (north of) Greenland? As far as I know that is a region that is covered by sea ice in summer (for present day condition).

We have noted this in our discussions as well. To the best of our knowledge, it is not an artefact of the surface forcing, which was the most likely candidate, but rather an artefact of the time-domain averaging chosen. We previously used time max/min, and so values of zero sea ice at any given point in the last 5 years would result in what appears to be a cell with zero sea ice. We have addressed this via using monthly means averaged over the last 5 years with February corresponding to the maximal extent and August corresponding to the minimal extent. A slightly lessened sea ice concentration is also coincident with increased vertical mixing in the area (this can be seen in the mixed layer depth contour of Figure 2) but it is unclear if this is a result or the cause of this feature. However, digging further into this question lies outside the scope of this paper. Finally, we note that our sea ice extents are largely consistent with reconstructions (e.g. de Vernal, et. al., 2005)

A. de Vernal (2005) Reconstruction of sea-surface conditions at middle to high latitudes of the Northern Hemisphere during the Last Glacial Maximum (LGM) based on dinoflagellate cyst assemblages, Quaternary Science Reviews, https://doi.org/10.1016/j.quascirev.2004.06.014.

- Figure S5: you show only 1 year (the last year of simulation). Please choose similar intervals and same number of simulation years for all the figures (e.g., 5-y mean of year x to y). Easily done. All figures are shifted to reflect the last 5 simulation years.

Same for Figure 2, what is meant by single day? For your study, yearly-mean values should be fine but take the average over several simulation years.

We used quite literally a single day (daily mean to be more precise, we have added this language to enhance clarity in the manuscript) to more readily demonstrate the turbulent nature of the model. We had multiple versions of that figure using various other temporal averaging schema but none conveyed the point as well as a single day. Eddies and variations in the coastal boundary currents are readily 'averaged away' when considering any sort of time-averaging.

- Would be interesting to see the timeseries plot for the MLD in Labrador and Nordic seas.

We had generated similar plots but they did not convey useful information at the time (anomaly does not show a trend), upon revisiting this idea with a smaller domain as per the reviewer comment this conclusion is unchanged.

- Figure S6-AMOC: You initialize the model from the experiment by Hill and Condron (2014), right? But why your experiment's AMOC is about 6 Sv at year -10 while it should be larger given the AMOC in Hill and Condron (2014)? If I am mistaken, please explain this part better.

Not quite, our run was initialized from a re-run of the Hill and Condron control simulation (very early versions of this work required additional data not available from the Hill and Condron simulations so a re-run was

required). After this rerun, we used the temperature and salinity fields from the 20th year of the LGM simulation to initialize the Younger-Dryas-like configuration (a direct restart was not possible due the significant bathymetry changes). The Younger Dryas configuration was run forward for 10 additional years, after which we then branched the two control runs. We make this information clearer in the revised version to address this comment as well the previous comment regarding experimental design.

Overall, the AMOC in your study seems to be smaller than some similar studies (https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015GL064583), right?

Between what other comparable studies we have found (e.g. Hirschi et al. 2020) and what the reviewer has referenced we do find our AMOC is among the weaker values we have found in the literature. We do note that our use of 26N to correspond with RAPID and Condron and Windsor (2012) makes our model seem weaker relative to some other studies as that is not the peak of the stream function in the Atlantic as is commonly used for studies where AMOC is a useful climate metric. When we use the peak of the stream function in the Atlantic we find our values to be ~9.5-10Sv for the last 5 years of the CBS control vs. ~3.5Sv at 26N.

What is the difference in AMOC between CBS and OBS control runs? The AMOC difference between the experiments seems to be small (1 or 2 Sv), and perhaps within the range of model internal variability. I am not sure if the AMOC can give any conclusive view.

As noted, the difference in values is very small relative to the range of model variability. When using the peak of the stream function in the Atlantic averaged over the last 5 simulation years we find that CBS control run has an average of 9.5Sv vs the OBS control run with 8.6Sv. We agree that the AMOC can not give a conclusive view of what is going on in the simulations. Due to this we have presented the AMOC only for context as we anticipated leaving out discussion of AMOC in this manuscript would raise calls for its inclusion. We moved all discussion of AMOC in our simulations to supplementary materials to better emphasise this idea and to address concerns raised by the other reviewer.

-Figure 4: Is it surface salinity? Except in the middle panel, the salinity anomaly shows a downward trend in some of the experiments (for instance the CBS MAK). You need to be careful how you interpret these as the model is not clearly far from equilibrium.

This is indeed sea surface salinity, more specifically the salinity from the top layer of our model (10m thickness). Given the trends, we elected to integrate our simulations further forward within the limitations of our resources (for most runs this amounts to ~5-10 additional simulation years). This additional model integration does not affect most of our conclusions, as the relative ordering of freshening at the different regions investigated remains the same with the exception of a larger freshening effect from the OBS MAK simulation for the NADW region. This additional information is reflected in the updated manuscript.

-One implication of this study is for the Younger Dryas event which is linked to temperature changes, and I was expecting to see a plot for the sea surface temperature (SST). I realized that this an only ocean model study but would still be interesting to see the (indirect) impact of different FW injection on SST.

We include a SST anomaly plot for the 2dSv CBS Mackenzie River simulation using the same time period (last 5 simulation years) below for the reviewer and others interested (this figure is also duplicated in the updated supplemental material). Interestingly the distribution generally follows the salinity anomaly distribution but the extreme values of negative salinity results in warming relative to the control while more saline values results in a cooling, with the threshold between warming and cooling being ~-2PSU. We do not have an immediate use

for this data in this manuscript outside providing some small context and so invite any interested parties to contact the authors at a later date if seeking further discussion.



-I will include a figure similar to Figure 3 but for BS closed in the main paper.

We only have the one closed Bering Strait freshwater forcing simulation and so cannot create something akin to Fig. 3 but with Closed Bering Strait runs. Given the broad scale similarities between the OBS and CBS 2dSv Mackenzie River injection scenarios we will leave the CBS run in the supplement only.

Also Figure S6 is better to be in the main paper.

We reiterate the point that the focus of our work is the transport of freshwater at the surface of the ocean. Including a figure on AMOC (which in our study is not a reliable metric given the short timescales of spinup) would reduce clarity in the manuscript and needlessly focus the reader on one of the more unreliable aspects of the study. As noted by the reviewer and ourselves, the AMOC in our simulations can not give conclusive results. As such we will keep Fig. S6 in the supplementary materials.

Title: needs to be adjusted. Is it really during the last deglacial?

The freshwater forcing fluxes are bounded by those from the Younger Dryas period of the Tarasov GLAC reconstruction, as well the bathymetry is derived from the same. Given the only glacial element of our configuration is the surface forcing we consider deglacial to be the more accurate description. However, the title has been revised given this comment and others made in the discussion phase.

Line 9: You are using paleo forcing and paleo-bathymetry, please correct it.

The line in question is : "We focus particularly on the **prior use of excessive freshwater volumes** (often by a factor of 5) **and present-day** (rather than paleo) ocean gateways..." (emphasis ours) The statement shows that we contrast our work relative to previous investigations who used present-day bathymetry.

Line 35: Sv is the common unit to use, and dSv is not really helping to make things easier.

Indeed Sverdrup is a common unit when discussing the AMOC and hosing. However, 1Sv is an order of magnitude (or more) larger than is reasonable for freshwater outflow from glacial runoff, whereas dSv (1/10 Sv) is the same order of magnitude of fluxes presented in this work as well as upcoming work by the authors. Previous works could reasonably use Sv as many used values of freshwater flow that were O(1Sv) but given reconstructions preclude these values we advocate for using something more easily read (decimal points are quite easy to miss). For clarity and continuity we will continue to use the more appropriately scaled unit, but we have added a line in the text when we define dSv explaining why we prefer this less common formulation.

Line 91-92: Does the model really captures the coastal boundary currents?

Both Yang (2003) (10.1016/S1463-5003(02)00058-6) and Nurser and Bacon (2014) (10.5194/os-10-967-2014) provide ranges of boundary current widths well within the range of our model's horizontal resolution. As noted in the paper this becomes problematic at the highest latitudes due to the significant decrease in the Rossby radius at high latitudes. Despite this we do see the freshwater constrained to very narrow boundary currents even in the Arctic.

Line 104: "The first...": revise

We have clarified this text.

Line 134: "...discussed in Experimental design section": is not discussed

Surface forcing is listed on Line 98 in the Experimental Design section:

"The surface forcing used in that simulation includes winds, precipitation, 2m atmospheric temperatures, short and longwave radiation, surface runoff, and humidity from the CCSM3 working group's contribution to PMIP2 (Braconnot et al., 2007)."

Figure S6: It is strange to use negative time values for the model spin up period.

Given all our runs are relative to each other, the use of negative for a period of time which is only discussed in one figure allows for comparison to Fig. 4 without that figure being required to start at year 10.