Reviewer 2

We thank the reviewer for the time he took to make a careful reading of our manuscript and for all the suggestions that are very useful to improve it. In the following the comments of the reviewer are in black, while our responses are in blue.

This paper describes the experimental protocol for a set of coordinated simulations involving oceanic surface freshwater flux perturbations, conducted as part of the international Tipping Points Modelling Intercomparison Project (TIPMIP). As such, it is not a 'normal research paper', but instead provides details so that research group can carry out the needed experiments to contribute to the TIPMIP project. Given the numerical focus of TIPMIP, I do feel it is appropriate for GMD. In fact, I am happy to see such protocols published in the literature rather than just on a webpage or a hard to find report. Making the protocol easily accessible is of value. Additionally, given the amount of work needed to define the protocol, it is appropriate that the authors get the credit for a peer reviewed paper. After reviewing the draft, I would recommend minor revisions. Although generally clear and well presented, there are places where material could be improved.
My specific comments are given below.

We thank the reviewer for his overall encouraging evaluation and for the comments he has provided us with. Below is our point-by-point response.

I prefer the term forced ocean models to ocean standalone models. Especially as it likely that no model is actually ocean only, instead including at least a sea-ice component.
We agree with this comment and have modified the corrected manuscript accordingly.

Line 79 – Likely newer references to the AMOC than just a 2016 paper. Especially given all the work on the topic.
We have added Lozier et al. (2019) and Srokosz et al. (2021) to support this sentence.

Line 80 – Given the recent OSNAP results, this general statement of regions could be modernized and more clearly explained. And in general, more literature on the AMOC and its stability would be good. I'd also like to see more background on freshwater sources, and discussion of Greenland melt, versus sources of freshwater to the Arctic (including rivers) as well as the impact of fluxes through the gateway straits. For a freshwater hosing protocol, the key background concepts need to be presented.
We have extended this paragraph and added a description of freshwater sources following the reviewer's comment (see lines 82-90).

Line 81 – Doesn't a significant amount of the AMOC waters upwell in the Atlantic, before it leaves for the other basins?
Indeed, this has been added to the list of regions.

The discussion of the 8.2 Kyr event focuses on discharge of the glacial lakes through the St. Lawrence into the Atlantic. Yet there is a literature that suggests that there is some debate on where and how the freshwater is released, such as into the McKenzie river into the Arctic Ocean. Even if the authors want to use a Labrador Sea focused protocol, a proper discussion of the freshwater sources to the event should at least be referenced – and whether it leaves by Hudson Strait or the St. Lawrence River likely may mean different freshwater distributions. For example, might the whole sub-polar gyre be better than just the Labrador Sea, as the freshwater likely has to flow around the gyre to reach the West Greenland Current region. Or why not along the Labrador Coast – high resolution studies have suggested that leads to different freshwater fates for this event.

We thank the reviewer for this comment. We agree that the routing of freshwater associated with the 8.2 ka event remains an area of active research. Prior to the final drainage event, overflow from Lake Agassiz-Ojibway is thought to have been directed through different pathways including an eastern outlet towards the St Lawrence system and a Northern outlet towards the Arctic. However isostatic modelling and geomorphological evidence suggest the Northern outlet was closed after 10.6 ka BP (Teller et al. 2004). Continued Laurentide Ice Sheet meltwater was likely routed through several outlets including the Mackenzie drainage basin until 8.5 ka (Süfke et al., 2022), but the vast majority of the melt from the Hudson Bay Ice Saddle would have been distributed between Lake Agassiz-Ojibway (when it was still active), Hudson Strait and the Labrador Sea coast. We have therefore revised the manuscript to acknowledge these uncertainties and provide additional references.

Nevertheless, most reconstructions and modelling studies specifically addressing the triggering freshwater perturbation associated with the 8.2 ka event focus on the drainage through Hudson Bay and Hudson Strait and its impact on the Labrador Sea and adjacent North Atlantic deep-water formation regions. There may have been other pathways, but the Labrador Sea-bound pathway is known with best confidence and likely represents a large proportion of the meltwater delivered to the ocean at this time. For this reason, the purpose of Experiment D is not to reproduce the exact freshwater routing, which remains uncertain, but to provide a standardized perturbation representative of the region most directly linked to the inferred AMOC response during the event and the most likely/majority drainage route for the event.

We also acknowledge that the location of the meltwater delivery to the ocean can influence the oceanic response , as demonstrated by e.g. Morrill et al. (2014). Applying the meltwater to the St Lawrence outlet, the Labrador Coastline or a box, as we propose here, has implications for the mixing pathway. However, high-resolution ocean studies (Condron and Winsor, 2011) have also demonstrated that the meltwater mixing pathway is highly dependent on the resolution of the models. For this MIP, we chose to align to the hosing region used by most GCM studies (Du et al., 2015; Matero et al., 2017; Morrill et al., 2013; Wagner et al., 2013) to facilitate comparison between models of different resolution and complexity. This aligns with the approach taken by other freshwater sensitivity studies, such as North Atlantic HosMIP (Jackson et al., 2023), which uses a wide idealised forcing region to compare the response across simulations and models, and is highly insightful even without testing the multiple reasonable permutations of the forcing scenario.

We have clarified this rationale in the manuscript (section 6) and added references discussing alternative routing pathways and associated uncertainties.

I didn't notice any discussion about model resolution in the protocol, and wonder whether that is relevant. That may impact the distance criteria to input the freshwater and thus where the freshwater is input in models with such differences in resolution. At the very least, it may impact the locations of discharge around Greenland, especially if any fiords are at least partially resolved. How might that impact the input of freshwater?
For Greenland hosing we use a scheme where the freshwater is distributed with an exponential decay away from the coast. This uses the same length scale for all models so should not be affected by resolution.

For experiment C, there is no thought on considering the rivers that may provide the enhanced precipitation in a future climate?
We have not understood well what the reviewer was suggesting here, specifically for experiment C. We assume that the reviewer is suggesting to add, in the rationale for the large freshwater release experiments, that the rivers in the high latitude might increase their discharge in the future, due to enhanced hydrological cycle. We have now explicitly mentioned the role potentially played by river at line 88 of the introduction

Shouldn't there be an optional for seasonally varying the freshwater input?
This would certainly be an interesting sensitivity experiment, and similar alternatives were discussed extensively among the co-authors as well as during community videoconferences open to a broader audience. However, there was a strong consensus that the number of core experiments should remain limited. Following several rounds of discussion and voting, a number of scientifically valuable experiments were ultimately excluded in order to keep the protocol manageable and maximize community participation. Therefore, although we agree that this would be an interesting addition, we do not believe it is appropriate to introduce further core experiments at this stage, given the consensus already reached among the co-authors. Nevertheless, additional sensitivity experiments performed by individual modelling groups are strongly encouraged, and we have clarified this point in the manuscript (lines 244-246), where we explicitly note that such complementary studies are welcome.

What about for models that include icebergs? How shall the discharge be partitioned between solid and liquid components?
Although in reality freshwater input could be through icebergs as well as liquid freshwater, for simplicity and conformity across models the discharge is considered to be solely liquid. We have added a note making this clear in the text at line 224-225

For experiment B, Tier 1, what initial conditions are recommended for the 1920s?
These are the Restarts one historical simulation at the end of 1919.
We have homogenized the description of the tables accordingly.

Wouldn’t it make sense to be consistent, as the choice of initial conditions may impact the model results.
We are a bit unsure what this comment is referring to. Possibly to the fact that the tables were missing a few pieces of information concerning the initial conditions, which have now been corrected.

Figure 6. Why use the same dark blue for both fluxes in the ramp up and ramp down scenarios? A slightly different shade might help show that there are two scenarios.
We assume this comment rather refers to Fig. 6.
Indeed, in this figure, the dark blue is the same when the experiment reaches 0.3 Sv (tier 1) or 0.6 Sv (tier 2) making the color gradient not fully consistent. Following this comment, we have modified the figure so that the level 0.6 Sv corresponds to a darker blue than the level 0.3 Sv. We thank the reviewer for pointing this out to us.