RC1

General assessment and recommendation

This study investigates the response of the Subpolar North Atlantic, the AMOC, and selected global climate indicators to a freshwater perturbation using the eddy-rich global climate model EC-Earth3P-VHR. By combining high spatial resolution with a spatially and temporally distributed freshwater forcing around Greenland, the authors provide a detailed analysis of boundary-current pathways, boundary–interior exchanges, and the time evolution of the AMOC response.

The manuscript is rich, technically sophisticated, well-written and novel in several respects. In particular, the use of an eddy-rich global model allows the authors to resolve narrow boundary currents and mesoscale processes that are poorly represented in most previous hosing experiments. The analysis provides valuable insight into how freshwater anomalies are advected along boundary currents, gradually penetrate the Labrador Sea interior, and ultimately lead to a weakening of the AMOC. The exploration of associated atmospheric responses further broadens the scope of the study.

Overall, the paper addresses an important and timely question concerning the sensitivity of the AMOC to Greenland freshwater input, and the results have the potential to make a meaningful contribution to the literature. However, several key methodological and interpretational aspects require clarification before the conclusions can be considered fully robust. In particular, the effective magnitude and realism of the freshwater forcing, the distinction between imposed and background freshwater fluxes, and the physical mechanisms underlying the relatively rapid AMOC response need to be more clearly documented and discussed. Addressing these issues is essential to ensure that the simulated AMOC weakening is interpreted correctly and placed in an appropriate physical and observational context. For these reasons, I recommend major revision.

> We thank the Referee for the time dedicated to review our paper and for the constructive feedback. The suggested changes have largely improved its scientific clarity and the overall quality of the article. Answers to each comment can be found below in blue.

Major comments

A central issue that needs to be addressed more explicitly concerns how much freshwater is actually added to the ocean in the experiments, and how this compares to observed Greenland freshwater fluxes. While the manuscript states that a freshwater flux of 0.04 Sv is applied based on Greenland meltwater estimates, this flux appears to be added on top of the model's background freshwater inputs, including runoff and iceberg (calving) fluxes already present in the control simulation. However, the magnitude and temporal evolution of these background fluxes are not documented. As a result, it is currently unclear what the total freshwater input experienced by the ocean actually is in the hosing experiments.

This point is critical for two reasons:

- without explicitly documenting the control-run runoff and calving fluxes, it is not possible to determine whether the AMOC response corresponds to an additional 0.04 Sv or to a substantially larger total freshwater perturbation. Given that observed Greenland freshwater fluxes around 1950 are closer to ~0.03 Sv or lower (runoff plus solid ice discharge), the applied forcing may significantly exceed realistic values once background fluxes are included.
- all simulations are conducted under constant 1950 forcing, rather than transient historical.

Combined with the relatively strong freshwater input, this implies that the experiments should be interpreted as an idealized sensitivity study, rather than as a fully realistic representation of recent or near-future Greenland melt conditions. This distinction should be made explicit throughout the manuscript, including in the abstract and conclusions.

In addition, given the slow advective timescales of deep water masses, it would be helpful to clarify which mechanisms allow the AMOC and associated northward transports to respond within approximately 10 years. In particular, the authors should distinguish more clearly between:

- dynamical circulation adjustments (e.g. thermal wind balance, pressure-field and boundary-current adjustments), and
- the physical propagation of deep water mass anomalies, which occurs on much longer timescales.

Clarifying this distinction would significantly strengthen the physical interpretation of the results and avoid potential confusion between transport changes and tracer propagation.

To address these issues, I strongly recommend that the authors provide time series of Greenland runoff and iceberg (calving) freshwater fluxes in the control ensemble, clearly state the total freshwater flux applied in the hosing experiments (background + additional forcing), explicitly compare this total flux to observed estimates, and clearly frame the experiments as idealized sensitivity experiments under constant 1950 forcing, rather than fully realistic.

Addressing these points would substantially improve the transparency, physical consistency, and interpretability of the manuscript.

We thank the referee for the suggestions to improve the clarity of our protocol. We agree that documenting the background freshwater fluxes from Greenland in the control simulation is important for transparency, and have added this information to the manuscript. For that, we have computed the corresponding monthly climatologies and standard deviations of the river runoff and iceberg calving fluxes from Greenland, which are the only freshwater fluxes coming from Greenland in the model, whose time series are shown below in Fig. AC1.1 and summary metrics have been added to the methodology section. Their climatological annual means ± standard deviation in the last 53 years of the control experiment (data from previous years was not saved) are 0.0047 ± 0.0016 Sv (148 ± 50 $km^3\ yr^{-1}$) for the river runoff and 0.0205 ± 0.0060 Sv (650 ± 190 $km^3\ yr^{-1}$) for the iceberg calving fluxes, therefore giving a combined background flux of 0.0252 ± 0.0062 Sv (800 ± 200 $km^3\ yr^{-1}$), against the imposed hosing of 0.0419 Sv (1322 $km^3\ yr^{-1}$). The total average flux in the freshwater hosing is therefore 0.0671 ± 0.0062 Sv (i.e., 2120 ± 200 $km^3\ yr^{-1}$). For context, this is 2-3 times higher than the average of observed estimates product from Bamber et al. (2018) for the period 1958–2000, and 1.5–2 times higher than the average for the period 2000–2016. We have added this comparison to the manuscript: "Taking into account the climatological river runoff and iceberg calving fluxes in the control simulation (0.0047 ± 0.0016 Sv and 0.0205 ± 0.0060 Sv respectively, computed over 53 years), the total annual average freshwater injection around Greenland in the hosing experiments increases up to 0.0671 ± 0.0062 Sv (i.e., 2120 ± 200 $km^3\ yr^{-1}$; see Fig. B1a for associated monthly climatologies). This amount is 1.5-3 times higher than the historical estimates of Bamber et al. (2018) for 1958-2016." We note that, since the control experiment is already in equilibrium with the background fluxes, the responses identified in the water hosing experiment only respond to the extra forcing of 0.0419 Sv included. That is why that value is more prominently mentioned in the abstract and the conclusions. Nevertheless, we agree that the total flux is the relevant quantity for contextualizing the experiment against observations, and for that reason we also report it in the methodology section.

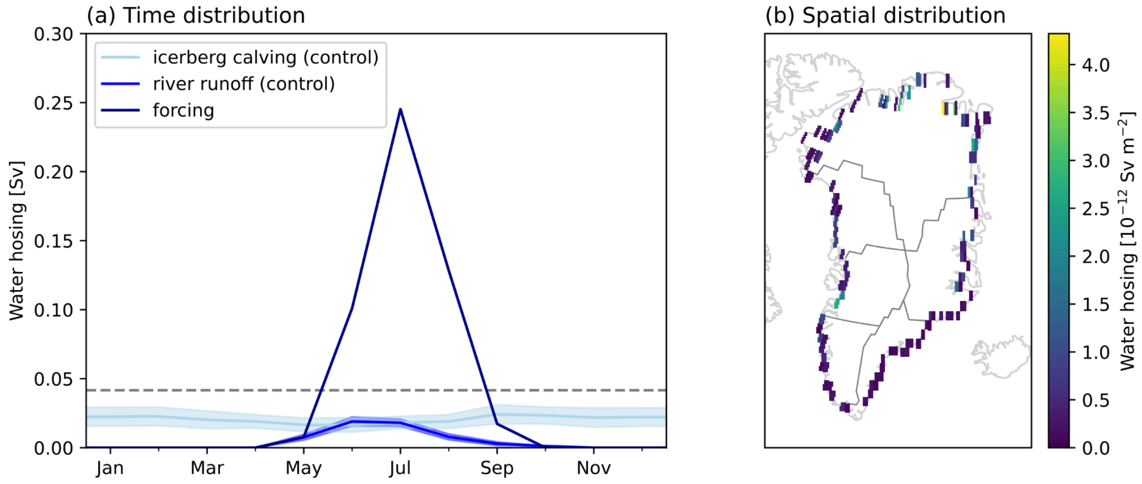

***Figure AC1.1**. (a) Total monthly freshwater forcing time distribution, monthly climatologies of river runoff and monthly climatologies of iceberg calving fluxes and (b) annual averaged freshwater forcing runoff spatial distribution. Forcing values are those given to the model as an input. The dashed line in (a) indicates the annual average of the forcing. The shadow areas in (a) represent the standard deviation of monthly river runoff and iceberg calving fluxes. The grey lines in (b) indicate the separation between the six sub-drainage basins.*

Regarding the idealised nature of the experiment, we agree with the Reviewer that our experimental setup (i.e. perpetual 1950 radiative forcing conditions with a freshwater flux that exceeds the historical Greenland melt estimates) should be framed explicitly as an idealised sensitivity study rather than a historically realistic simulation. We also recognise that our use of the word "realistic" in the original manuscript might have been misinterpreted in this regard. We have therefore revised the Abstract, Conclusions and relevant methodological paragraphs accordingly. In particular, we now clearly state in the Abstract and Conclusions that the experiments use a perpetual 1950 radiative forcing (see answers to the related minor comments). We have also rephrased the mentions to the forcing "time/temporal distribution" to be "monthly distribution" throughout, and have rephrased the expression "derived from observations" to "derived from a product based on observations and model simulations".

Finally, we thank the Reviewer for raising the point on the mechanisms behind the fast AMOC response and the distinction between dynamical and advective processes. We have added some text in the discussion section clarifying the two potential mechanisms at play.

Minor Comments

Abstract

The abstract states that the model is forced for 21 years, add if this forcing is applied only at the surface or also distributed at depth.

We have included the following sentence at the end of the first paragraph: "An annual average close to 0.04 Sv is released together with the simulated model river runoff, <u>which is vertically distributed in the coastal points connected to each hydrological basin</u>".

The abstract mentions "an acceleration and a cooling". Please clarify what is accelerating (e.g. boundary currents, gyre circulation) and specify the spatial scale. Similarly, please describe the origin of the cooling (e.g. reduced northward heat transport, enhanced stratification), as the freshwater itself is not necessarily colder than the ambient ocean.

> We have split the sentence into two and written the following in the first part: "Since the beginning of the experiments, these boundary currents also experience an acceleration and cooling. The cooling arises because freshwater-induced stratification suppresses vertical mixing, reducing the entrainment of warmer subsurface waters into the surface layer of the boundary currents". The specific currents refer to the Greenland and Labrador currents that are mentioned in the previous sentence.

The phrase "along the currents" is vague. Please specify which currents are being referred to (e.g. Greenland boundary currents, Labrador Current, North Atlantic Current).

> The new sentence mentioned in the previous comment solves that.

The statement "lead to a rapid weakening of the AMOC" would benefit from clarification. Surface cooling alone would tend to increase density and potentially strengthen the AMOC; please clarify how the combined effects of freshening, stratification, and cooling result in a net AMOC weakening.

> The second part of the split sentence now says: "The meltwater fluxes also lead to a rapid weakening of the AMOC at subpolar latitudes via thermal wind balance, with the salinity-driven density reduction outweighing the temperature-driven density increase".

The sentence "By the end of the simulation, the AMOC has weakened by almost 3 Sv at subpolar latitudes (i.e. a decrease of around 20 %), with an average relative decrease of 10 % for the whole Northern Hemisphere" is unclear. Please specify the latitudes at which these values are computed and provide corresponding absolute AMOC values to avoid ambiguity.

> We have replaced the previous sentence by this new one: "By the end of the simulation, on decadal average, the AMOC has weakened by almost 3 Sv at 60.2° N in density space (equivalent to almost a 20 % of the climatological value in the control, 14.9 Sv), with an average decrease of 1.3 Sv (equivalent to a 10 % of the control reference, 13 Sv) for water masses between 1027.4-1027.6 kg m^{-3} and 10-65° N".

For clarity, the abstract should explicitly state the total freshwater flux applied (including background runoff and calving) and how this amount compares to observed Greenland freshwater fluxes, as this directly conditions the magnitude of the AMOC response.

> As indicated in the response to the first comment, we believe that the relevant value for the Abstract is the water hosing imposed, that it's the only one that determines the model response. We have, however, revised the sentence at the end of the first paragraph, which now reads as: "An annual average close to 0.04 Sv is released together with the simulated model river runoff, which is vertically distributed in the coastal points connected to each hydrological basin."

In addition, the abstract should explicitly state that the simulations are conducted under constant 1950 forcing, rather than transient historical, as this strongly conditions the interpretation of the results and the realism of the freshwater forcing.

> We have expanded the sentence in the abstract to be: "Our study fills that gap with a set of three hosing experiments using perpetual 1950 radiative forcing with the global configuration of the eddy-rich EC-Earth3P-VHR climate model".

Introduction

General remark: the Introduction tends to make strong causal assumptions, sometimes blurring the distinction between what is directly observed, what is inferred from models, and what remains speculative. A more cautious framing of these mechanisms would improve the scientific clarity and consistency of the manuscript.

> We have carefully revised the Introduction to specify which results come from model experiments and which ones from observations and to avoid speculative phrases.

l.18: The statement "The associated changes in salinity modify the surface density" would benefit from a citation. It is not clear that this link has been directly observed in the context discussed here, so please support this statement with an appropriate reference.

> We have rephrased to change this and the next sentence for:
> "The increased freshwater is expected to reduce surface density, enhancing vertical stratification in the Subpolar North Atlantic (SPNA), thereby inhibiting deep mixing (Böning et al, 2016). This reduces the deep water formation, which is linked to the weakening in the Atlantic Meridional Overturning Circulation (AMOC) (cites)".

l.19: The link between freshwater forcing and deep water formation is presented in a very direct way. Consider explicitly mentioning the role of upper-ocean stratification and deep convection processes, so that the physical mechanisms and underlying assumptions are clearly stated.

> We have rephrased it as requested, as indicated in the previous comment.

l.22: The phrase "whether the projected AMOC reduction will result in its complete shutdown" is ambiguous. Please clarify whether this refers to continuous anthropogenic forcing in general or specifically to freshwater forcing.

> After a general review of the Introduction, we have decided to remove the sentence, as it does not provide any useful information for the carried out analysis or discussion.

l.39: The expression "using idealized model experiments" to describe Devilliers et al. (2021, 2024) is misleading. These studies use realistic, spatially and temporally distributed freshwater fluxes derived from observations rather than idealized hosing. Please revise this wording accordingly.

> We have removed the word "idealized" from the sentence, to refer in general to model experiments.

l.60: The sentence "In order to make the most of the finer resolution, a protocol with a greater focus on the imminent transient response than on long-term equilibrium is needed" does not clearly follow from the previous sentence and introduces a new topic. Consider separating this into a new paragraph and reformulating to clarify that the long-term response is difficult to explore at such high resolution due to computational constraints. It would also be useful to briefly discuss the role of internal variability and the need for large ensembles to separate forced signals from internal variability on long timescales, compared to the shorter timescales targeted here.

> We have split the paragraph into two and added a new sentence; now it says:
> "Exploring long-term responses is challenging due to the significant computational needs to run climate-scale simulations, and the need to consider ensembles to disentangle the forced signals from those of internal origin. The vast data volumes generated by eddy-rich models is another

important constraint. Therefore, to make the most of the finer resolution, a protocol with a greater focus on the imminent transient response than on long-term equilibrium is needed".

l.61: The sentence "An overly idealized Greenland hosing configuration can result in an unrealistic distribution of injected freshwater" would benefit from clarification. Please briefly explain what aspects of the hosing configuration are emphasized in Goldsworth (2026) to obtain a more realistic response, and consider also referencing and describing the recent protocol proposed by Schmidt et al. (2025, https://doi.org/10.5194/gmd-18-8333-2025).

We have extended the sentence for clarity and added some additional text afterwards referring to the protocol from Schmidt et al (2025):
"An overly idealised Greenland hosing configuration, e.g. applying a uniform freshwater flux across the North Atlantic rather than distributing it according to the observed drainage basins and seasonal cycle, can result in an unrealistic spatial distribution of the injected freshwater and potentially bias the modelled ocean response (Goldsworth, 2026). A potential way forward is the new community protocol presented in Schmidt et al. (2025), which provides observation-based datasets of absolute and anomalous freshwater fluxes from the Greenland and Antarctic ice sheets, broken down by drainage basin and freshwater pathway (runoff, subglacial discharge, and solid ice discharge), intended for implementation in models without interactive ice sheets.

l.63: Replace "This" with "This protocol" for clarity.

Changed

2 Methodology

2.1 Experimental set-up

l.83: The statement "based on observations from Bamber et al. (2018)" is not fully accurate. The dataset in Bamber et al. (2018) is derived from a combination of observations and high-resolution regional models. Please correct this wording here and consistently throughout the manuscript.

We have replaced the sentence by "We apply the freshwater hosing according to a spatial and monthly distribution derived from Bamber et al. (2018), which is based on observations and high-resolution regional model runs, see Fig. 1".

l.85: The conversion from 0.04 Sv corresponds to approximately 1261 $km^3\ yr^{-1}$, not 1322 $km^3\ yr^{-1}$. Please correct this value.

The Reviewer's value of 1261 $km^3\ yr^{-1}$ corresponds to a continuous 0.04 Sv forcing applied year-round. However, the actual injection is 0.1 Sv ($1\cdot10^{-4}\ km^3\ s^{-1}$) applied only in May-September (153 days for each year), which is mentioned before in the paper and gives the reported values of:

$$1\cdot10^{-4}\ km^3\ s^{-1} \cdot 86400\ s\ day^{-1}\ 153\ day\ yr^{-1} = 1321.92\ km^3\ yr^{-1}$$

Note that this corresponds to an annual average flux of 0.0419 Sv. To properly address this discrepancy, we have replaced the reference of 0.04 Sv to 0.0419 Sv in this sentence and the conclusions, while in the Abstract we now say that the forcing is "close to 0.04" Sv for readability.

l.86: The sentence “close to estimations for the values from 2013–2016 (Bamber et al., 2018)” is ambiguous. Please clarify whether this refers to the total Greenland freshwater flux or only to specific components (e.g. solid ice discharge plus runoff, including ice and tundra contributions).

We have removed the reference for the estimated values as the quantity in the estimation refers to the total quantity of freshwater flux which is greater in our experiments counting the river runoff and iceberg calving fluxes climatologies. We have added that the quantity is “higher than historical records estimate” at the end of the sentence.

l.88: The statement “We include the freshwater as an additional term to model river runoff” raises an important point that needs clarification. The manuscript does not specify the magnitude of runoff and iceberg (calving) freshwater fluxes in the control run. If the hosing flux is added on top of these existing fluxes, the experiment evaluates the impact of 0.04 Sv plus the model’s background freshwater input, rather than 0.04 Sv alone. Please specify the control-run freshwater fluxes and consider providing a time series of runoff and iceberg discharge for the control experiment.

We have included the values for the annual climatological river runoff and calving fluxes from Greenland in the text. Even though the total freshwater is bigger than the forced value of 0.0419 Sv, the control experiment dynamics are already in equilibrium with the calving and river runoff, and thus the response in the hosing experiments is just caused by the introduced anomaly of 0.0419 Sv. We have included the following sentences:
“Taking into account the climatological river runoff and iceberg calving fluxes in the control simulation (0.0047 ± 0.0016 Sv and 0.0205 ± 0.0060 Sv respectively), the total annual average freshwater injection around Greenland in the hosing experiments increases up to 0.0671 ± 0.0062 Sv (i.e., 2120 ± 200 km^3 yr^{-1}; see Fig. B1a for associated monthly climatologies). This amount is 1.5-3 times higher than the historical estimates of Bamber et al. (2018) for 1958-2016”.

l.90: The phrase “as well as vertically” is vague. Please describe the vertical distribution of the freshwater input (e.g. depth range, uniform or non-uniform distribution).

We have replaced the old text “We include the freshwater as an additional term to model river runoff (Fig. B1 shows the extra runoff input to the model) that is evenly and instantaneously distributed along several ocean coastal points connected to each hydrological basin in the vicinity of the major river outlets, as well as vertically to avoid numerical and physical problems (Gurvan et al., 2017), at 0 psu and local seawater temperature.” by this new text: “We include the freshwater as an additional term to model river runoff (Fig. B1 shows the extra runoff input to the model) that is evenly and instantaneously distributed along several ocean coastal points connected to each hydrological basin in the vicinity of the major river outlets, at 0 psu and local seawater temperature. The runoff is uniformly distributed throughout the upper ocean until a given depth in order to avoid numerical and physical problems (Gurvan et al., 2017). More specifically, most of the water is added to the surface within the upper 0.94 m, with an average depth at which runoff is injected of 1.78 m, reaching 243.76 m at some coastal points.”

l.97: The terms “weak, moderate and strong AMOC states” would benefit from quantitative definitions. Please provide representative values (e.g. in Sv) at standard latitudes such as 26° N and/or 45° N to clearly distinguish these regimes.

We used the time series of the AMOC at 45° N to determine the initial conditions to be used for the three hosing experiments. The three annual average values of the control for the starting year are 11.4 Sv, 12.5 Sv, and 14.5 Sv for weak, moderate, and strong AMOC states, respectively. The average AMOC at that latitude is 12.8 Sv. We note that this classification into weak, moderate and strong AMOC states does not necessarily hold at lower latitudes, as Martín-Martínez et al. (2025) already showed a weak cross-latitudinal coherence for the AMOC in EC-Earth3P-VHR, the model configuration used in our water hosing experiments. For example, at 26° N, the initial-year averages are 14.5 Sv, 13.8 Sv, and 15.6 Sv for the “weak”, “moderate”, and “strong” AMOC states, respectively. Note that the purpose of selecting these three states is to ensure diversity in the initial phase of internal variability. Due to these differences, we avoid discussing changes based on the initial state and discuss the aggregate of the three members instead. We have included the values at 45° N in the paper.

l.99: The sentence “To avoid the effects of a strong initial model drift” is unclear. Since the control and hosing simulations use identical forcings, it is not obvious why the control run would drift relative to the spin-up. Please clarify the origin of this drift.

We have included the underlined sentence for clarification: “To avoid the effects of a strong initial model drift, derived from the HighResMIP spin-up of 50 years, …”

l.101: The statement that “The three hosing runs use the same […] constant 1950-forcing” is key information and should be clearly stated in the abstract. The experiments are conducted under fixed control forcing rather than transient forcing, which is not fully consistent with the term “realistic conditions” currently used. This distinction should be made explicit in the abstract.

We have expanded the sentence in the Abstract to be: “Our study fills that gap with a set of three hosing experiments using perpetual 1950 radiative forcing with the global configurations of the eddy-rich EC-Earth3P-VHR climate model.”

In addition, freshwater fluxes around 1950 were substantially lower than 0.04 Sv (approximately 900 km^3 yr^{-1}, i.e. < 0.03 Sv, for runoff plus solid ice discharge in Bamber et al., 2018). In the present experiments, 0.04 Sv is applied in addition to the model’s background runoff and calving, resulting in a total freshwater input that is significantly larger than realistic 1950 values. This discrepancy should be clearly acknowledged and discussed, as it conditions the interpretation of the results.

As we have acknowledged in previous comments, our experiments are idealised and do not pretend to reproduce a fully realistic AMOC response to the observed Greenland meltwater fluxes. The text has been revised accordingly throughout to make this more clear.

2.2 Results evaluation

This subsection would benefit from being rewritten to clarify the evaluation methodology and improve overall readability.

We have improved the description of the evaluation methodology following the suggestions below. We have also rewritten the paragraph describing the bootstrap method. We moved the first paragraph to the end to highlight the actual methods rather than the software used.

l.109: The sentence “The anomalies between the hosing and control experiments are computed by matching the years between both experiments after the initialization” is unclear. Please specify what is meant by “matching the years” (e.g. identical simulation years after initialization, fixed lag, or alignment relative to the hosing onset), and explicitly state how many years are matched.

We have added the following clarification after the highlighted sentence “, i.e., the hosing member initialised in 1987 is matched with its respective 21-year period in the control (1987–2007) year-by-year, and the same is done for the other members with their respective control years”.

Similar methods are used in other freshwater hosing papers which also branch from control simulations, e.g. Diamond et al. (2025):

For all maps, unless otherwise indicated, we present the time-mean of the u03hos simulation over years 50–100, with the time-mean of the piControl simulation over the same period subtracted. [...] For timeseries, unless otherwise indicated, we show the u03hos simulation result with the piControl simulation result subtracted, year by year, as in Andrews et al. (2019), to account for unforced model drift in the simulations without assuming the drift is linear.

l.110: The phrase "to remove any remaining model drift" would benefit from clarification. Please specify the origin of the drift being removed (e.g. background model drift, adjustment to freshwater forcing, or both).

We have rephrased it "to remove any <u>residual</u> model drift".

l.110: The expression "align the initial internal climate variability" is unclear. Please clarify what is being aligned (e.g. phase of internal variability, ensemble mean state) and how this alignment is performed.

The purpose is to align the phase of internal variability, we have replaced the word "initial" by "phase of". The details on how the alignment is achieved are provided in comment of l.109.

l.115: The reference to "the previous sample of anomalies" is ambiguous. Please clarify which sample is being referred to and how it is defined.

We have fully rewritten the paragraph to make it more easy to understand. The new full paragraph is:
"Statistical significance of time-mean anomalies is assessed using a bootstrap method. For each grid point, the sample consists of annual (or seasonal/monthly for variables computed in specific seasons/months such as sea ice or mixed layer depth) anomalies pooled across all three ensemble members, which provides three times the number of averaged years. For example, the average of surface salinity anomalies in Fig. 4a is based on a sample of 30 elements (10 annual anomalies x 3 members). We then generate 1000 bootstrap replicates by resampling this pooled sample with replacement and computing the mean of each replicate. Finally, we compute the 95th percentile of the 1000 bootstrap means to determine whether the difference is significant at the 95 % confidence level."

l.120: The choice of a three-year threshold ("for at least three years") appears arbitrary unless motivated by previous studies. If this choice is methodological, please state it explicitly, for example by rephrasing to "we assume here that three years is sufficient to…" or "we propose to use a three-year threshold to…".

It is a methodological choice. We have rephrased the last sentence to "However, as alignment between the three members may frequently occur by chance, <u>we assume here that three consecutive years of coherent, physically explainable signal is sufficient to distinguish a forcing-induced response from a spurious alignment due to internal variability</u>".

3. Results

General remark: the term 'volume overturning streamfunction' is technically correct but somewhat redundant, as overturning streamfunctions are by definition volumetric. Consider simplifying to 'overturning streamfunction' for clarity and consistency with standard usage.

While we agree that overturning streamfunction commonly refers to the volume streamfunction, as streamfunctions are defined as integrals of velocities, many models actually provide the mass overturning (CMOR variables, msftyz or msftmz) which is not strictly identical and come with different units (kg s^{-1} instead of Sv). Therefore, we prefer to retain the term volume to avoid ambiguity. .

3.1 Meltwater impact on the AMOC

l. 124: “the time evolution of the maximum overturning” on annual means? The caption says “Monthly maximum” which is unclear.

We compute the maximum overturning in the z-axis in monthly means and perform a 36-month moving average to plot the data. Annuals means are not used in this figure, we have rephrased the caption:

- “Monthly maximum volume overturning streamfunction” -> “Maximum volume overturning streamfunction”
- “The data is filtered with a 36-month (3-year) moving average.” -> “The values have been computed with monthly averages and filtered with a 36-month (3-year) moving average.”

l. 122–124 + Fig. 2 caption: The latitudes 33.8° N (depth space) and 60.2° N (density space) are selected based on where the maximum change occurs during the last 10 years. Since these latitudes are chosen a posteriori, it would be helpful to clarify that they are intended as diagnostic locations rather than representative of the basin-wide AMOC response, or to briefly discuss the sensitivity of the results to this choice.

We have modified the sentence as follows: “
which is the latitude where the AMOC exhibits the strongest response to the forcing over the last 10 years (Fig. 3a), <u>that is, when the magnitude of the decadal anomaly is maximum. We note that while this choice helps for illustrative purposes, the selected latitude may not necessarily represent a basin-wide response</u>”.

l.123–127: the AMOC is described as “relatively stable” during the first 7-10 years, while differences between ensembles are said to be consistent with internal variability. The term “stable” may be ambiguous here. A formulation explicitly referring to the absence of a statistically significant signal relative to internal variability might improve clarity.

We have modified the beginning of the sentence to say: “This AMOC index <u>shows no statistically significant slowdown</u> during the first 7–10 years”.

l.128: it is stated that the AMOC response in density space is “clearer and much earlier” than in depth space. While this is convincing qualitatively, the argument would be strengthened by a more quantitative criterion (e.g., emergence time or a significance threshold).

We have rephrased the first sentences of that paragraph to make the statement more clear and fair including the values of when the signal starts being coherent “The maximum change in the AMOC in density space happens at much higher latitudes, 60.2° N. The AMOC anomaly at that point starts being sign coherent in year 5, much earlier than the signal at 33.8° N in depth latitudes for which it happens in year 12. To contextualize...”. See Fig. AC1.2a below (included as a supplementary now).

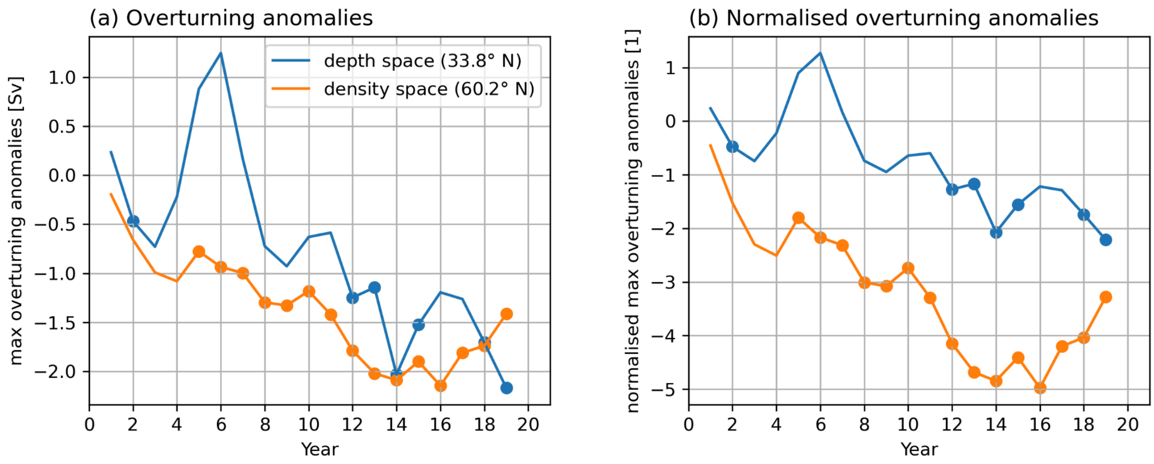

Figure AC1.2*. Maximum volume overturning streamfunction and in the Atlantic basin for (a) annual data with 3-year moving average and (b) annual data with 3-year moving average divided by the control 3-year moving average's standard deviation. The values have been computed with annual averages and filtered with a 3-year moving average. The latitudes were selected where the corresponding maximum change in the last 10 years happens, see Fig. 3. The gray dots show the annual average maximum volume of the overturning stream function in the control during the first year.*

l.133: the method used to quantify this ratio of the signal to internal variability is not specified. Please clarify how internal variability is estimated (e.g., ensemble spread or temporal variance) and how this ratio is computed.

> We agree that the method for quantifying the signal-to-internal-variability ratio was not sufficiently described. While both the ensemble spread and the temporal variance are reflections of internal variability, we prefer to use the second due to the very small ensemble size. In particular, internal variability is estimated as the temporal standard deviation of 3-year running mean in the control simulation. and the signal-to-noise ratio as the hosing anomaly divided by this standard deviation . We have replaced the original sentence with "As a result, the hosing signal in density space deviates from the control from the beginning of the simulation ensemble, and when normalised by the temporal standard deviation of the control, the signal-to-noise ratio is 1 to 3 times higher than at 33.8° N in depth space (Fig. B2), indicating that density coordinates provide a clearer separation between the forced response and internal variability".

l.128–138 and Fig. 2: comparison between depth space and density space responses involves different latitudes (33.8° N versus 60.2° N). As a result, the contrast reflects both differences in vertical coordinate and latitude. A brief clarification disentangling these two effects, or explicitly stating this limitation, would strengthen the interpretation.

> We agree that the comparison is incomplete For this reason we have rewritten the text comparing both time series to refer also to the latitude, as follows:
> "The delayed response in depth space at 33.8° N indicates that the freshwater injection needs some years to significantly impact the lower latitudes. Thus far, comparisons have focused on contrasting responses in two different coordinates and latitudes, which provides an incomplete picture of the large-scale response. From now on, we will focus most of our analysis on the last 10 years of the hosing simulation, to remove the initial years where the signal of the forcing in the AMOC is too small compared to the noise from internal variability. This approach will also allow us to compare depth and density spaces consistently across latitudes".

l.150: the statement that the stronger coherence in density space is expected because "water flows along isopycnals and not isobaths" is broadly correct, but somewhat simplified since diapycnal transformations are also central to the AMOC, a more nuanced phrasing is needed.

> We have rewritten the sentence to say: "This is to be expected, given that, once the dense waters are formed and the contributions of water mass transformation are minimal, the newly created water masses flow along isopycnals and not isobaths (Foukal and Chafik, 2024)."

l. 146: when stating that 21 years of hosing produces a detectable AMOC weakening, it would be useful to briefly contextualize the magnitude of this response relative to the imposed freshwater flux (add the amount of the total freshwater fluxes imposed to the model) and to previous freshwater hosing studies (Swingedouw et al 2022, Jackson et al 2023, Van Westen 2024).

> We have expanded the text, which now reads: "This shows that 21 years of hosing with an annual average freshwater anomaly of ~0.04 Sv is sufficient to produce a detectable weakening in the AMOC, although this remains far from representing a full AMOC collapse. This response is weaker than in the study of Swingedouw et al. (2022), who report a weakening of about 2 Sv after 13 years of a comparable freshwater forcing (~ 0.035 Sv anomaly). Studies applying substantially larger forcings, such as Jackson et al. (2023) with 0.1 Sv, or operating on centennial timescales with time-varying freshwater inputs (Van Westen et al., 2024), are less directly comparable given the differences in experimental design, but collectively support the notion that the magnitude of the AMOC scales with the intensity and duration of the freshwater perturbation".

l. 154: regarding the link between LSW, deep convection, and AMOC weakening, it is plausible but remains qualitative at this stage. This interpretation needs to be supported by additional diagnostics of deep convection (e.g., mixed-layer depth or buoyancy fluxes), which are addressed in the following section, so it may be a bit too soon to describe it here.

> We have rewritten the paragraph, removing the first sentence and rephrasing the rest, which now reads: "In the next Section, we investigate buoyancy changes in the main deep convection areas of the North Atlantic to understand their potential impact on the LSW and, subsequently, on the AMOC. For that, we examine how different thermodynamic variables have changed in the Subpolar North Atlantic (SPNA) in the last 10 years of the simulations."

Figure 3a: adjust the scale to -2 → 2 so the differences appear clearer. Usually significant values are marked with dots, not the other way around. caption: since the hosing → after the hosing.

> We understand that plotting the dots in the significant areas is the most common practice and what many readers will expect, which can lead to confusion. However, we believe that masking the non-significant areas with the dots strongly improves the visibility of the significant values, which, in the end, is what we will analyse. In particular, adding dots over significant areas make them appear darker than they really are, see differences between Fig. AC1.3 and Fig. AC1.4. For that reason, we prefer to keep the representation of significant areas as it was. We have modified the color scale in Fig. 3 and changed the caption as requested.

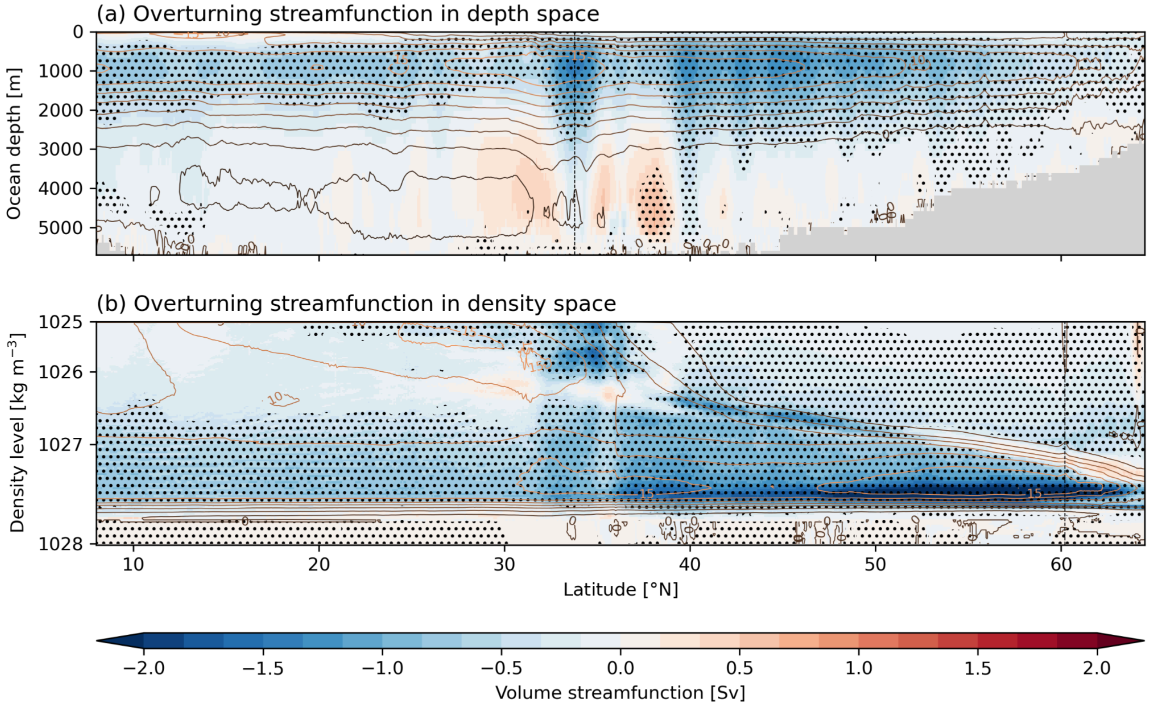

Figure AC1.3. Response to the hosing in the last 10 simulated years (i.e. years 12 to 21 after the hosing is started) for the overturning streamfunction in (a) depth space and (b) density space; filled coloured contours represent the ensemble-mean anomalies of the hosing with respect to the reference control and the contour lines the climatology of the control. Significant values as identified by the bootstrap methodology are covered with dots. Vertical dashed lines mark the latitude where the maximum difference happens; 33.78° N in depth space and 60.21° N in density space.

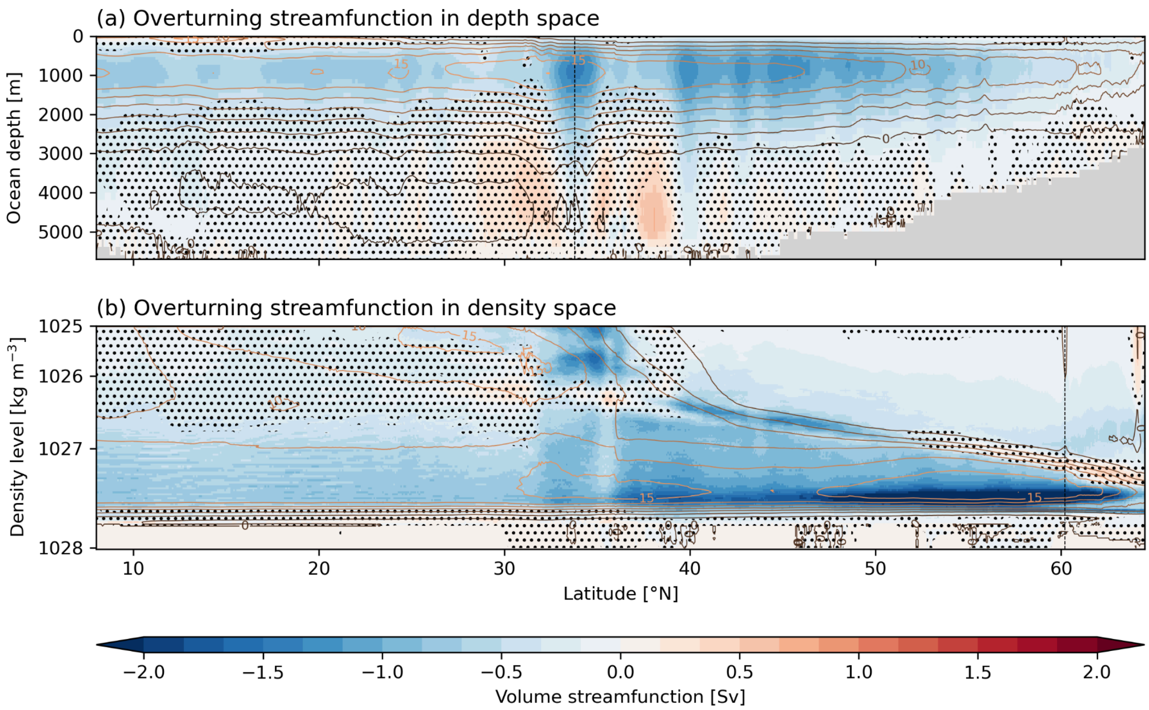

Figure AC1.4. Same as Figure AC1.3 but masking the non-significant anomalies.

3.2 Large-scale changes in the subpolar region

The title may be misleading, as the AMOC is a basin-scale circulation, while this section focuses on regional changes in the subpolar North Atlantic. Consider changing it to "3.2 Changes in the subpolar region".

> We have modified it to "Changes in the subpolar region".

l.164–169: The statement that changes in surface buoyancy responsible for the AMOC slowdown are "probably happening in the Labrador Sea" is plausible but not yet directly demonstrated at this stage. Consider softening the wording, e.g. "suggesting that the Labrador Sea may play an important role in the AMOC response", or explicitly framing this as a hypothesis supported later by mixed-layer depth diagnostics.

> We have rewriten the sentence to say: "This also suggests <u>that the Labrador Sea may play an important role in the AMOC response</u>".

l.168–180: The attribution of freshening and cooling in the North Atlantic Current to the already developed AMOC slowdown remains qualitative. It would be helpful to briefly clarify whether these anomalies are interpreted as direct advection of freshwater from the Labrador Current or as an indirect dynamical response to reduced overturning and northward heat transport.

> The sequence of events supports an interpretation in which the NAC anomalies develop primarily as an indirect dynamical response rather than through direct freshwater advection from the Labrador Current. The AMOC weakening and Labrador Sea mixed layer shoaling emerge consistently around year 7 at subpolar latitudes (60° N, Fig. 7d,e), prior to the NAC freshening and cooling, which only appear around year 9-10. This temporal ordering is consistent with a dynamical signal originating in the subpolar gyre, where deep convection is suppressed, and propagating southward along the western boundary, reducing northward heat and salt transport into the NAC region. Furthermore, the SST anomaly in the Labrador Current is considerably weaker than in the NAC (Fig. 4b), arguing against the Labrador Current being the dominant freshwater source for the NAC cooling. We have added a sentence to the text to make this interpretation explicit.

l.180: The acceleration of boundary currents is clearly shown but not physically interpreted. A short explanation linking this response to enhanced lateral density gradients and geostrophic adjustment associated with surface freshening would strengthen the discussion.

> This acceleration of boundary currents is physically interpreted later in l.224, based on density gradient. As we don't show density changes here but later in Fig. 5 we preferred to keep the interpretation to that point.

l.185: The lack of a significant mixed layer depth response in the Irminger and Nordic seas is an interesting result. A brief discussion of possible reasons for this regional contrast would improve the physical interpretation.

> We agree with the Reviewer that the lack of significant impacts in Irminger and Nordic seas is itself an interesting result. However, we think that focusing it only in the mixed layer depth may not be appropriate as the climatological mixing in the Nordic Seas for the control experiments is small compared to the Labrador Sea. The average mixing in the control for the interior of the seas defined in Fig. 6 is 478 m for the Labrador, 312 m for the Irminger, and 123 m for the Nordic. We have linked it to the response to comments made in l.164, where we discuss it as smaller

penetration of freshwater in Irminger and Nordic seas where the mixing is also shallower. We now say: “Interestingly, we note a lack of response in the Irminger and Nordic seas, the two other deep convection areas in the NH. This may be due to the lack of significant freshwater penetration, as previously mentioned. In addition, the Nordic Seas has very shallow climatological mixing due to a positive bias in sea ice (Martin-Martinez et al. 2025, Moreno-Chamarro et al. 2025), which limits the potential impact of the freshwater anomalies”.

l.190–193: The absence of a clear Subpolar Gyre slowdown despite mixed layer shoaling contrasts with previous studies. A short comment on why this relationship may not hold in the present experiments (e.g. forcing magnitude or timescale) would be valuable.

The mentioned studies do not include freshwater hosing, just radiative forcing, which we have indicated, and also added a second sentence mentioning the potential counteract of the boundary currents acceleration: “a connection that has been identified in previous studies (Moreno-Chamarro et al., 2017; Ghosh et al., 2023) that investigate the response to radiative-driven global warming. In our study, the expected deceleration of the SPG in response to the reduction in deep mixing may be counteracted by the acceleration of boundary currents caused by the strong freshwater-driven density gradient”.

l.215: The interpretation of isopycnal-aligned freshening as evidence of weak diapycnal mixing could be slightly softened. Consider rephrasing to emphasize the dominance of along-isopycnal transport rather than excluding diapycnal processes.

We have rephrased the second part of the sentence to: “suggesting a prominent role of the along-isopycnal transport by the SPG circulation”.

Figure 4: The use of March mixed layer depth is reasonable but the authors may consider also examining a winter-mean mixed layer depth (e.g. JFM) to assess the robustness of the results to the choice of month.

Figure AC1.5 shows the comparison between the March and JFM mixed layer depths for the maps in Fig. 4c. Both figures lead to similar qualitative conclusions regarding the reduction of mixing in the Labrador Sea and no significant changes in the eastern part of the SPNA. However, since March experiences larger changes, it has been kept in the manuscript.

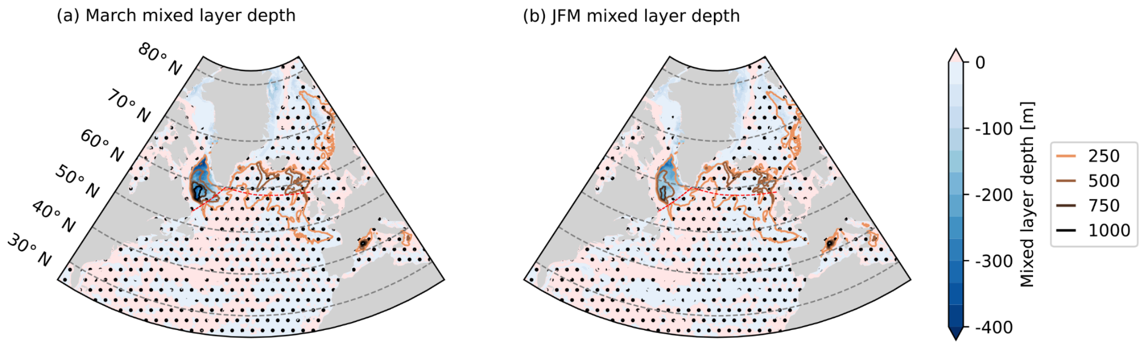

***Figure AC1.5**. Response to the hosing in the last 10 simulated years for (a) March mixed layer depth and (b) JFM mixed layer depth; filled coloured contours represent the ensemble-mean anomalies of the hosing with respect to the reference control and the contour lines the climatology of the control. Non-significant values as identified by the bootstrap methodology are masked with dots to improve the visibility over the significant areas. Red dashed lines mark the OSNAP section.*

3.3 Time evolution

l.235: This robustness criterion relies again on the three-year threshold discussed above (see comment on l.120). Please ensure that the rationale for this choice is clearly stated once and applied consistently throughout the manuscript.

Done.

l.240: The weaker salinity signal in the Labrador Current is attributed to mixing with neighboring water masses. This is plausible, but remains qualitative; the wording could be softened or briefly supported by additional diagnostics.

We have replaced it with "One possible cause of this could be the horizontal mixing of negative salinity anomalies from the Greenland Current as they circulate around the Labrador Sea".

l.245: Brine rejection is mentioned as a possible contributor to salinity changes along the Labrador Current. Please clarify whether this process is explicitly diagnosed in the model or inferred qualitatively.

This is inferred qualitatively, we have softened the sentence, now it says "At the same time, the stronger winter sea-ice growth in the hosing experiments could also contribute to the salinity anomalies along the current due to brine rejection".

l.252: Two mechanisms are proposed for the delayed freshening of the Norwegian Current. Since the delayed SPG weakening is dismissed based on the barotropic streamfunction, it would help to clarify whether this diagnostic alone is sufficient to rule out changes in salt transport.

The first mechanism is still plausible, that is, the transport of the freshwater anomalies by the mean SPNA circulation, is still plausible based on our results. We now indicate this at the end of the paragraph. "This second hypothesis, however, is not supported by the barotropic streamfunction, which does not show a clear large-scale reduction in the SPG (Fig. 4f)<u>, suggesting that the transport of freshwater anomalies by the mean SPNA circulation is the most likely cause of the Norwegian Current freshening</u>".

l.259: Freshening and cooling in the NAC region are linked to the AMOC slowdown. This interpretation is reasonable, but the causal link could be phrased more cautiously, as other contributions (e.g. direct advection or air–sea fluxes) cannot be excluded.

Note that the current sentence already says "in the NAC region <u>may</u> just emerge". We have also added the following sentence afterwards to mention other potential effects "Other processes could also contribute, such as the direct advection of fresh and cold waters by the SPG or changes in air-sea fluxes."

l.270: A thermal wind response is invoked to explain the behavior at 60° N. A short clarification of why this mechanism would dominate there compared to 40–50° N would improve readability.

We thank the Reviewer for this suggestion. We have added a physical clarification to the manuscript. At 60° N, the freshwater anomaly in the early years of the hosing is largely confined to the boundary currents, preserving a strong cross-current density gradient with the denser interior. This supports a rapid thermal wind adjustment of the boundary current transport that does not depend on deep convection or interior spreading. The freshwater has spread more broadly into the subpolar gyre interior by the time it reaches the 40-50° N latitudes, weakening the cross-current density contrast, and the overturning response is instead governed by the slower reduction in Labrador Sea deep convection. This contrast can explain why the mixed layer depth changes lead the overturning signal at 40-50° N but lag it at 60° N. The revised text now reads: "The changes in the mixed layer depth seem to lead by 2 to 4 years the changes in the maximum overturning streamfunction at 40° and 50° N, but lag by 2 years the AMOC changes at 60° N. At 60° N, the faster response is consistent with a thermal wind adjustment because the freshwater anomalies remain largely confined within the boundary currents, maintaining a strong cross-current density gradient with the denser interior. By the time they reach 40-50° N, the freshwater anomalies have spread more broadly into the subpolar gyre interior, reducing the cross-current gradient, and the overturning response is more dominated by the reduction in deep convection in the Labrador Sea, which acts on a longer timescale."

l.279: The temperature–salinity compensation during the first seven years is well described and likely delays the emergence of a clear density signal; stating this explicitly would help link this discussion to the delayed mixed layer depth response.

We have added "This competition between opposing salinity and temperature contributions likely explains the delayed emergence of a consistent negative density signal in the upper ocean".

l.283: The explanation that heat anomalies are advected more efficiently than freshwater anomalies is plausible but speculative. Consider softening the wording or clearly framing this as an interpretation.

We have replaced the "is" by "may be".

l.286: The stratification-driven shoaling of the mixed layer depth is convincing. For clarity, it may help to briefly restate that salinity dominates the density signal in the later years.

We have rephrased the end of the sentence to be "which leads to a more stratified density distribution in the later years when salinity dominates the density signal (Fig. 8d)".

l.292–295: The comparison with other freshwater injection studies is useful and could more explicitly emphasize the role of the spatial distribution of freshwater forcing, rather than its magnitude alone.

We agree that the mechanistic link between forcing distribution and SPNA response could be made more explicit. We have revised the paragraph to clarify that the key factor is the proximity of the injection location to the boundary current system that connects Greenland's margins to the Labrador Sea convection region. Freshwater entering this pathway is efficiently transported to the deep convection sites, whereas forcing applied outside it as in Wei and Zhang (2024), or uniform forcing as in Jackson et al. (2023), follows different pathways and can produce qualitatively different responses. This mechanistic framing now precedes the citations rather than being left implicit.
The revised paragraph reads as follows: "The spatial distribution of the forcing appears to play an essential role in this pathway: freshwater injected close to the boundary current system is efficiently transported along the SPG boundary and can reach the Labrador Sea convection region, whereas forcing applied far from these pathways may have a very different or even opposite effect. This is illustrated by Wei and Zhang (2024), who produce a contrary response when injecting freshwater only in the south of the Nordic Seas, and by Ma et al. (2024), who show that among four injection regions in the North Atlantic, freshwater applied in the Irminger Basin, which is directly upstream of the Labrador Sea along the boundary current, is the most effective at suppressing deep convection there. Idealised experiments applying uniform forcing across the North Atlantic, such as Jackson et al. (2023), may therefore not capture the spatially

selective nature of this mechanism and could produce a less realistic representation of SPNA dynamics.”

3.4 Global Impacts

l.297-302: Since this paragraph summarizes results from previous studies rather than findings from the present simulations, it may be better suited for the Introduction.

We have removed that paragraph as the intro already contains that information.

l.309: Wintertime cooling over the Labrador and Nordic Seas is strong (>5 °C locally). A brief clarification on whether this reflects mean seasonal anomalies or episodic extremes amplified by air–sea coupling would improve interpretation.

The >5° C cooling over the Labrador and of about 2.5° C over the Nordic Seas reported in the text reflects local maxima in the ensemble-mean DJF anomaly, not episodic extremes. We acknowledge that Fig. 9a may have contributed to the confusion, as the colour scale saturates at ±2° C, obscuring the full magnitude and extension of the signal in these regions. See below in Fig. AC1.6 a zoom over the SPNA with the boxes taken for Fig. AC1.7. As shown in the additional diagnostics in Fig. AC1.7, the annual-mean (monthly data with 12-month moving average) cooling is more modest, while the DJF ensemble mean is amplified, locally exceeding 5° C for the northeast of the Baffin Bay box. We have added a clarifying sentence to the text “The anomalies to the north-east of Baffin Bay are significant throughout the whole year (not shown), reaching maximum strength during winter”, modified the name of the regions by “north-east of Baffin Bay” and “Greenland Sea next to Svalbard” and included Fig AC1.6 with the supplementary figures (without the red boxes) to better convey the full amplitude of the DJF signal.

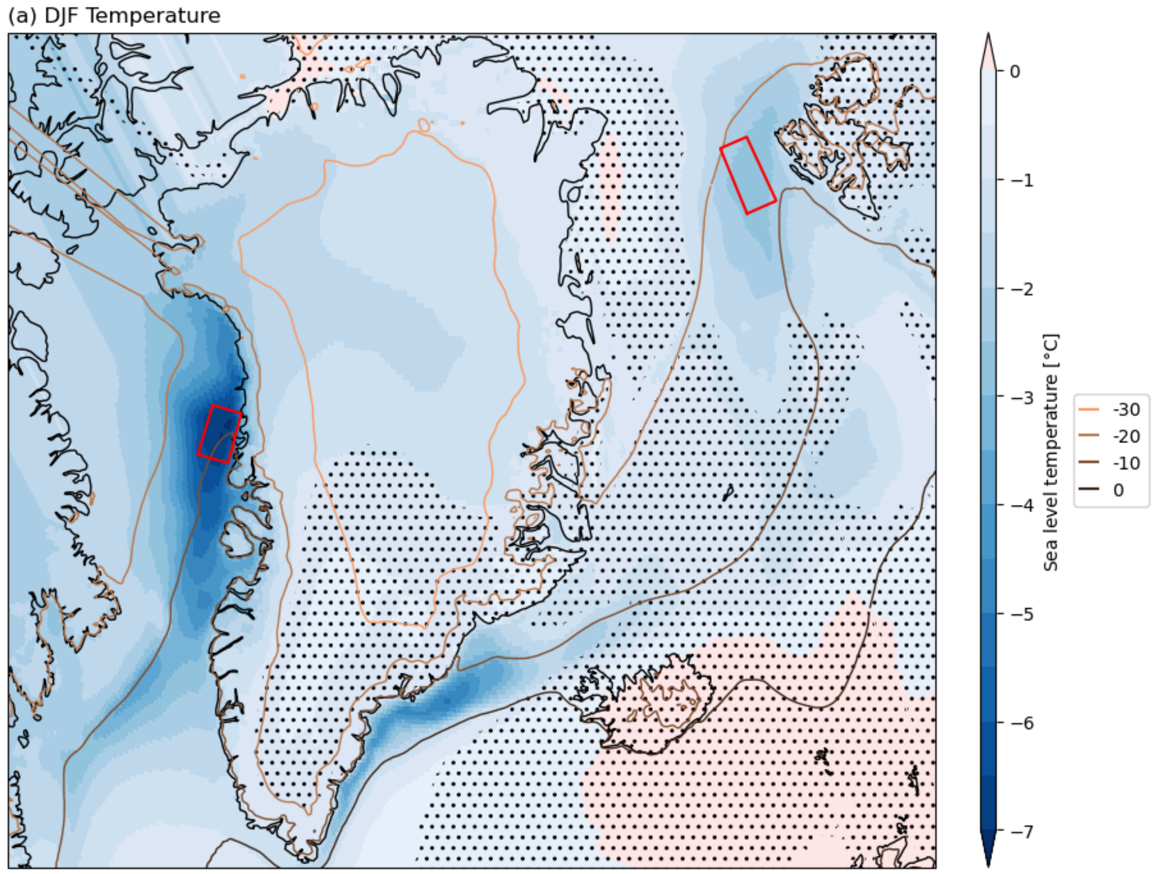

Figure AC1.6. *Response to the hosing in the last 10 simulated years for (a) atmospheric surface temperature in DJF; filled coloured contours represent the ensemble-mean anomalies of the hosing ensemble with respect to the reference control and the contour lines the climatology of the control. Non-significant values as identified by the bootstrap methodology are masked with dots to improve the visibility over the significant areas. Red boxes are those taken for values in Fig, AC1.7.*

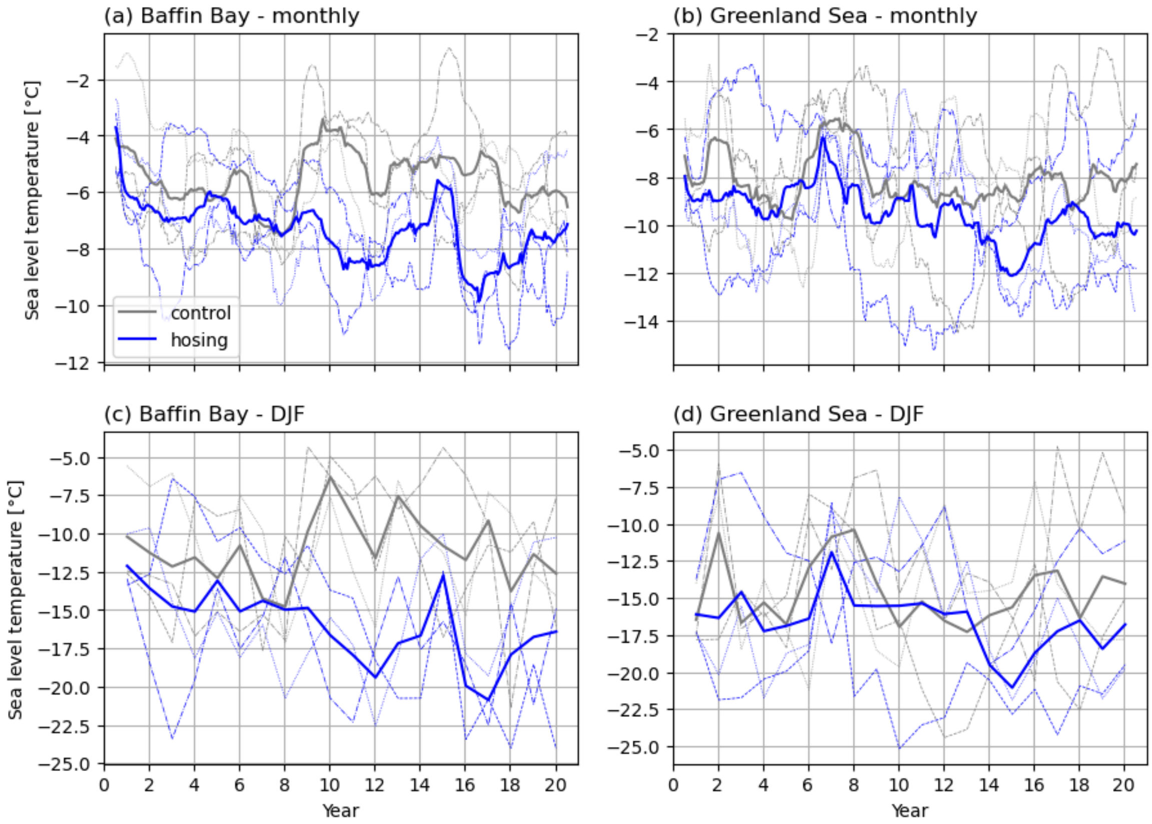

***Figure AC1.7**. Time series for (a-b) monthly values with 12-month moving average and (c-d) DJF values of atmospheric surface temperature for the (a, c) Baffin Bay and (b, d) Greenland Sea boxes in FIg. AC1.6.*

l.313: Warming over the Amundsen and Ross Seas is consistent with previous studies. Given the short integration time, it would be helpful to stress that this response likely reflects atmospheric teleconnections rather than oceanic heat transport.

> We have added a new sentence after: "Due to the short time of integration, these anomalies are most probably caused by an atmospheric teleconnection rather than to adjustments in ocean heat transport".

l.320: Interpreting the sea level pressure anomalies in terms of NAM and SAM is plausible, but remains qualitative. Please clarify whether these modes are diagnosed explicitly or inferred from spatial pattern resemblance.

> It is inferred from spatial pattern resemblance. We have added the following to the sentence "These changes are consistent with a positive Northern Annular Mode (NAM)-like response in the NH and a negative Southern Annular Mode (SAM)-like response in the SH".

l.330: Summer sea level pressure anomalies over the North Atlantic are described as a positive NAO phase: as for the NAM/SAM, it would be useful to indicate whether a formal NAO index is computed or whether this interpretation is based on pattern similarity.

> The NAO is inferred from the pattern, we have rewritten the sentence for clarity: “There is a consistent increase in sea level pressure over the central North Atlantic region and a decrease over Greenland. This spatial pattern is roughly consistent with a positive summer North Atlantic Oscillation (NAO)-like response, a pressure pattern that has been shown to be more frequent under larger hosing experiments (Bellomo et al., 2023)”.

l.339: Briefly contrasting this with stronger hosing experiments could further emphasize the role of AMOC amplitude.

> We have added the following sentence: “This suggests that the amplitude of the ITCZ shift scales with the magnitude of the AMOC weakening”.

4. Conclusions

l.345: Given the importance of the experimental design please briefly restate in one sentence that the simulations are performed under constant 1950 forcing and that the freshwater flux is applied in addition to the model background runoff.

> We have added “The additional freshwater is added on top of the coupled model river runoff and iceberg calving fluxes. The hosing experiments share the same perpetual 1950 radiative forcing conditions as the control simulation, ensuring that the simulated responses can be unambiguously attributed to the freshwater forcing” at the end of the paragraph.

l.352–355: Quantifying the AMOC reduction in density space is useful but please add values in depth space, and ensure consistency with earlier sections when citing percentage versus absolute changes.

> We have added all the values in the conclusions

l.364: May be useful to remind the reader that these atmospheric signals are less robust across ensemble members than the oceanic response.

> We have added the following sentence at the end of the paragraph: “In general, these atmospheric signals are less robust than the identified ocean signals described earlier”.

l.378: List key differences in forcing magnitude or timescale when comparing with Schiller-Weiss et al. (2024)

> We have added this sentence to explain the differences: While the simulation length is broadly comparable to ours (25 vs 21 years), their freshwater anomaly (derived from the observed post-2000 acceleration of Greenland melt relative to the pre-2000 climatology) is substantially weaker than the constant 0.0419 Sv perturbation applied here.

l.383: The statement that the initial AMOC state does not critically affect the ocean response is not fully demonstrated as there are visible inter-member differences, plus some signals become consistent only after several years plus atmospheric responses show a larger spread. Maybe “…the qualitative ocean response appears robust across different initial AMOC states, although quantitative differences remain.” would be more suited.

We have modified the sentence to say: “We further highlight that the qualitative ocean response appears robust across different initial AMOC states, although quantitative differences remain”.

l. 388: TipMIP → TIPMIP

Changed

RC2

In their manuscript “North Atlantic response to a quasi-realistic Greenland meltwater forcing in eddy-rich EC-Earth3P-VHR hosing simulations” Eneko Martin-Martinez and co-authors present results from a common Greenland hosing experiment though with a more realistic, moderate magnitude of 0.04 Sv and spatial distribution of the additional freshwater flux along the coast of Greenland. Most importantly, they use a climate model of very high resolution in ocean and atmosphere unprecedented for such experiment to my knowledge. Unfortunately, this limits the experiment length to only 21 year. However, they ran three ensemble members for estimating the imprint of internal variability on the response. The presentation of results is focused on the response by the AMOC and potential drivers of its decline.

The experiment is certainly timely as the discussion of Greenland ice sheet melting potentially impacting AMOC strength over the next decades is still ongoing. The experimental design is reasonable though slightly different to previous studies (of which numerous are referenced), which limits comparability and differences to previous studies cannot be attributed to model uncertainty unambiguously. On the one hand, the results align with previous findings, thus are supportive of the state of knowledge. Which is good. On the other hand, this means there is not much groundbreaking novelty inherent to the outcome. Maybe because of this, I have the impression that the authors do not really take full advantage of the unique part of their experimental setup, the model resolution. Overall the results section is largely descriptive, few connections are drawn, less are proven, and the analysis is mostly limited to surface fields. It would be helpful if the authors would connect the responses by indicating responsible processes, even if this is speculative (which should be noted then). And by putting more emphasis on the role of mesoscale features simulated by the high resolution of the model. Arguments related to explicitly resolved eddies, eddy mixing and sharper gradients at fronts are rather qualitative but could in principle be quantified. I wish there would be more specific analysis regarding, for instance, freshwater spreading and associated timescales, changes in surface fluxes (e.g. heat loss), and water mass transformation. How do these shape the initial decline of the AMOC?

> We thank the reviewer for the detailed and constructive comments. We have thoroughly addressed each point in the point-by-point response below (highlighted in blue), and have made a special effort to include new analyses that strengthen the physical interpretation of our results.
>
> Regarding the suggestions related to mesoscale features specifically, several of these were highly relevant and appealing to us. Unfortunately, many of the three-dimensional and high-frequency model outputs required for those analyses (including daily sea surface height) are currently inaccessible to us due to an ongoing issue with our tape storage system, and could not be performed. For instance, daily sea surface height fields would have allowed us to compute eddy kinetic energy (EKE) and track how mesoscale eddy activity along the boundary currents responded to the freshwater forcing and contributed to interior freshwater spreading. Ocean velocity fields at depth would have enabled a more quantitative assessment of eddy-driven lateral exchanges between the boundary currents and the Labrador Sea interior, including estimates of eddy diffusivity and mixing timescales.
>
> With the data currently available to us we have added a new surface-forced water mass transformation (SFWMT) analysis that directly addresses the reviewer’s request for a more process-oriented interpretation of the AMOC decline. This analysis quantifies the respective haline and thermal contributions to buoyancy forcing in the western and eastern subpolar regions, connects them to the changes in overturning transport across the OSNAP-West and OSNAP-East sections, and explains how the lightening of surface waters through freshwater injection reduces the availability of dense water for transformation. This ultimately provides a mechanistic link between the freshwater forcing, the suppression of deep convection in the Labrador Sea, and the warming of the DWBC discussed in the main text. The analysis has been added as a new subsection after the OSNAP cross-section analysis including the new Figures AC2.1 & AC2.2 and Table AC2.1, below.

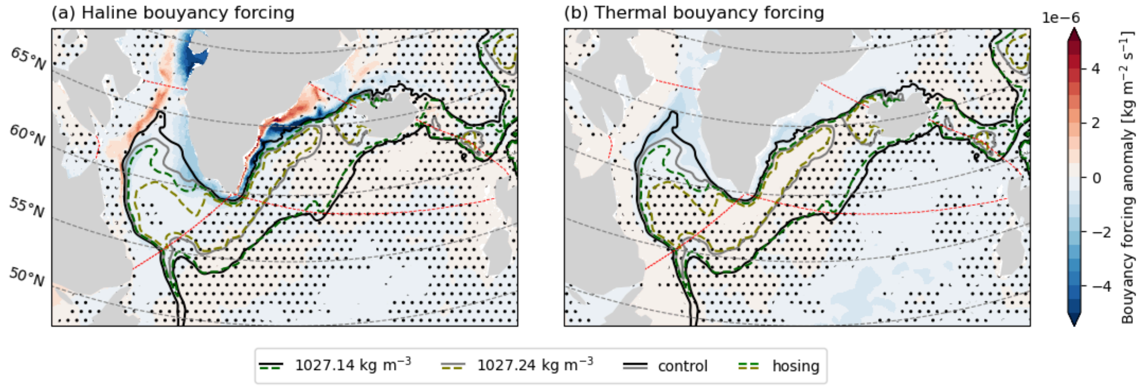

Figure AC2.1. *Response to the hosing in the last 10 simulated years for the (a) freshwater flux and (b) heat flux contributions to surface forced water mass transformation; filled coloured contours represent the ensemble-mean anomalies of the hosing with respect to the reference control. Contour lines show the surface 1027.14 and 1027.24 kg m^{-3} isoline in control (black, grey) and hosing (green, olive). Non-significant values as identified by the bootstrap methodology are masked with dots to improve the visibility over the significant areas. Red dashed lines mark the limits for the analysis in Fig. AC2.2.*

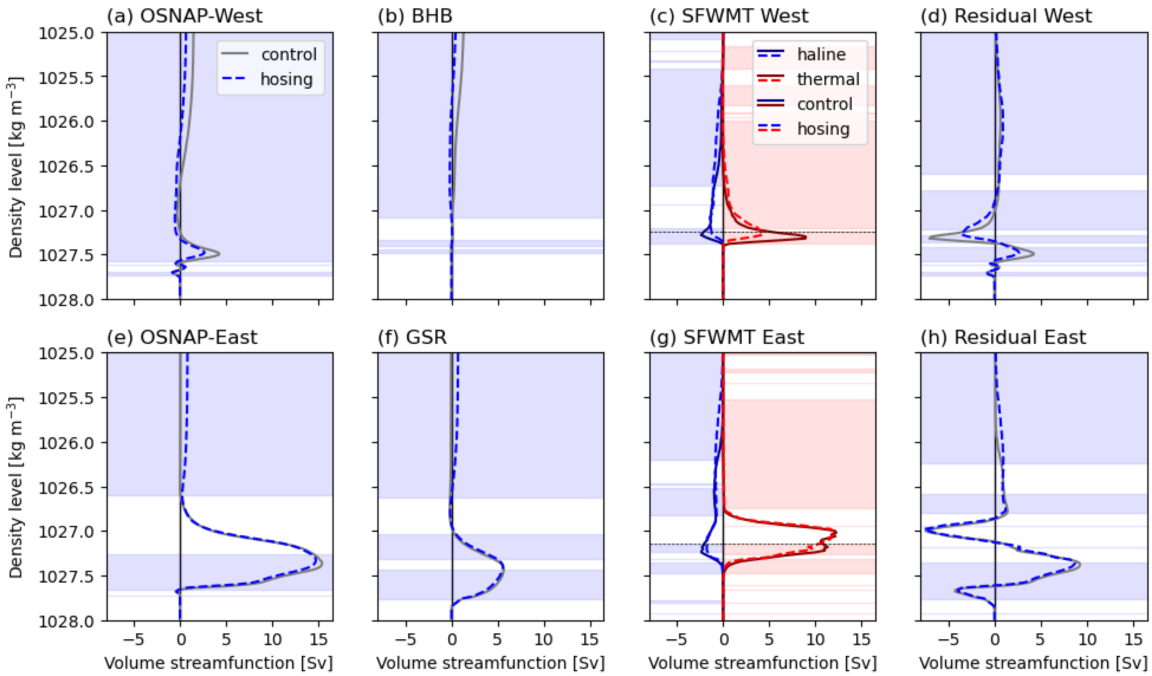

Figure AC2.2. *Response to the hosing in the last 10 simulated years for the overturning streamfunction in density space in (a) OSNAP-West, (b) BHB, (e) OSNAP East, (f) GSR; values in the hosing (blue) and control (grey) ensembles. Thermal (blues) and haline (reds) contributions to the SFWMT in the (c) West and (g) East regions. Residual WMT computed as the net transport out of the region minus the total SFWMT for (d) West and (h) East). The ensemble means are shown by lines, the background is coloured when the anomalies are deemed statistically significant by the bootstrap methodology. The lines in (c) and (g) represent the 1027.14 and 1027.24 kg m^{-3} density levels, respectively, which isolines are plotted in Fig. AC2.1.*

***Table AC2.1**. Maximum value of the volume overturning streamfunction in density space across both OSNAP transects for the last 10 years of the hosing.*

	OSNAP West	*OSNAP East*	*OSNAP Full*
control	*4.3 Sv*	*15.4 Sv*	*17.2 Sv*
hosing	*2.7 Sv*	*14.7Sv*	*15.4 Sv*
anomaly	*-1.6 Sv*	*-0.7 Sv*	*-1.8 Sv*

The other main criticism I have is that there is little validation of the model provided. The authors use the OSNAP section in Figure 5 to show the depth structure of changes in the subpolar North Atlantic. This would be an opportunity to show the mean state of the reference run in comparison to OSNAP observation, at least in the supplementary material. Some discussion of ocean biases is found in Moreno-Chamarro et al. (2025), which is referenced in line 77. Please emphasize this. Nevertheless, a specific discussion of biases with respect to the subpolar North Atlantic would be a great addition to the paper since the model seems to perform rather well (judging from global plots in above reference).

Apart from Moreno-Chamarro et al. (2025), Frigola et al. (2025) also explores the biases in the historical run with the same model, and few extra discussion for the vertical profiles in the Labrador Sea and AMOC mean states are found in Martin-Martinez et al. (2025). These are now reflected in the revised manuscript which also includes the corresponding references in the discussion.

Regarding further validations, we have computed the OSNAP East and West transports in the control simulation and compared them with the observed ones. We note, however, that both datasets are not fully compatible, as the control run uses fixed 1950 radiative conditions and observations cover the period 2014-2022 with time-varying forcing. For EC-Earth3-VHR we get 16.6 ± 2.7 Sv (mean ± std of monthly data) for OSNAP East and 5.6 ± 2.3 Sv, which lay within the uncertainty range of observations (https://www.o-snap.org/data-access/), corresponding to 16.4 ± 2.9 Sv and 2.7 ± 1.3 Sv, respectively. Interestingly, EC-Earth3-VHR reproduces the stark asymmetry between the observed OSNAP eastern and western components, with much stronger transports associated to the eastern side. Other studies using NEMO-based climate models have shown a positive bias in the transport in the OSNAP-West transport, e.g., Petit et al. (2023). Note that our transport biases are smaller than those presented by Petit et al. (2023). In general, the contributions to the SFWMT are comparable, although the maximum transport and SFWMT do not occur at exactly the same density levels.

Considering a major revision and **sharpening of the arguments**, maybe some **additional analysis emphasizing more the role of the explicitly resolved mesoscale dynamics** (beyond reducing biases), this paper could become a valuable contribution and suitable for publication in OS.

Minor comments by line

93 Separating lines of the six sub-drainage basins could be added to Figure 1b

We have included the sub-drainage basins in Fig. B1b, modified the caption accordingly and added a reference in the text.

109 "…by matching the years between …" I am not sure I understand this sentence. Do you mean that you compute differences by subtracting the experiment from control year-wise, i.e. not comparing the experiment to a long-term mean of control? Please rephrase. The interactive atmosphere would lead to independently evolving internal variability. The shortness of the simulations (21 years) suggests however, that low-frequent variability is likely still preserved. This could be noted here.

Yes, we compared with the control year-wise. We have extended the sentence to now say: "The anomalies between the hosing and control experiments are computed by matching the years between both experiments after the initialisation, , i.e., the hosing member initialised in 1987 is matched with its respective 21-year period in the control (1987–2007) year-by-year, and the same is done for the other members with their respective control years. This is intended to remove any remaining background model drift and align the phase of internal climate variability. Although the atmosphere could evolve into different states of internal variability, the low-frequency variability is likely to be preserved due to the relatively short length of the simulations."

124 and 129 The latitudes given (33.8N and 60.2N), are these locations of max AMOC in control or of the maximum change in experiment minus control? This is somewhat unclear here. Please specify.

We have modified the sentence as follows: "which is the latitude where the AMOC exhibits the strongest response to the forcing over the last 10 years (Fig. 3a), that is, when the magnitude of the decadal anomaly is maximum".

132 "The ensemble-mean difference …" Please provide numbers here. Judging from Figure 2 I'd say there is a significant difference with an AMOC weakening of 1.5 and 2 Sv respectively. Also, simply comparing the last year is likely misleading and an average weakening of at least the last 5 year should be given.

We agree that the sentence would have benefited from some numerical comparison, as we are doing a deeper comparison later with the full streamfunction in Fig. 3. We have decided to remove this sentence. Also even though signals are relatively comparable in magnitude their signal-to-noise ratio is very different, see Fig. AC2.3 below (included as a supplementary figure now).

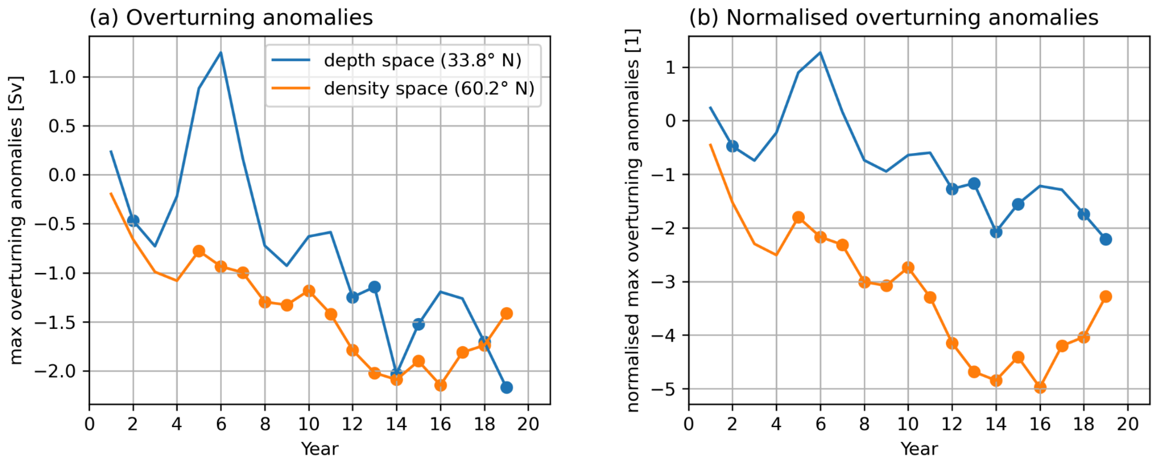

Figure AC2.3. *Maximum volume overturning streamfunction and in the Atlantic basin for (a) annual data with 3-year moving average and (b) annual data with 3-year moving average divided by the control 3-year moving average's standard deviation. The values have been computed with annual averages and filtered with a 3-year moving average. The latitudes were selected where the corresponding maximum change in the last 10* *years happens, see Fig. 3. The gray dots show the annual average maximum volume of the overturning stream function in the control during the first year.*

134 not sure which "ratio" is meant hear. Isn't this a simple difference?

We have rewritten the sentence "As a result, the hosing signal in density space deviates from the control from the beginning of the simulation ensemble, and when normalised by the temporal standard deviation of the control, the signal-to-noise ratio is 1 to 3 times higher than at 33.8° N in depth space (Fig. B2), indicating that density coordinates provide a clearer separation between the forced response and internal variability".

135 replace "happens" with "is located"

Changed.

139 remove last part of sentence: "…, that is when the AMOC…" Due to the shortness of the runs, it remains unclear whether the AMOC response is already fully developed.

We have replaced it by "to remove the initial years where the signal of the forcing in the AMOC is too small compared to the noise from internal variability".

143 add "… by 1.4 Sv on decadal average" at the end of this sentence.

Added.

145 replace "lowest latitudes" with "subtropical region"

Replaced.

146 there appears to be something missing at the end of the sentence: "…, representing a 7%." of what?

We have added "of reduction computed with respect to the average value in the control for the same area" after "7 %".

148 remove "the" from "…AMOC in the density space" and "… than in the depth space".

Removed.

151 Which water mass is simulated in this density range?

We computed the average densities shown in Fig. AC2.4. This was done for both the hosing and control experiments, since the injected freshwater could modify the position of the density levels, even though the changes between the experiments are relatively small, and the density range 1027.4-1027.6 kg m^{-3} seems to represent the same water masses in both experiments. For that density range the water mass in the 8-30° N range is located approximately between 1000–1600 m. Between 30° N and 50° N, the range increases to 500-1800 m. The lower limit increases slightly until reaching 1500 m at around 63° N, after which there is a jump to 250-600 m. In high latitudes, the water mass corresponds to the Labrador Sea Water (LSW). We selected this range because it encompasses the maximum change, as the LSW in our model can extend to 1027.2 kg m^{-3}.

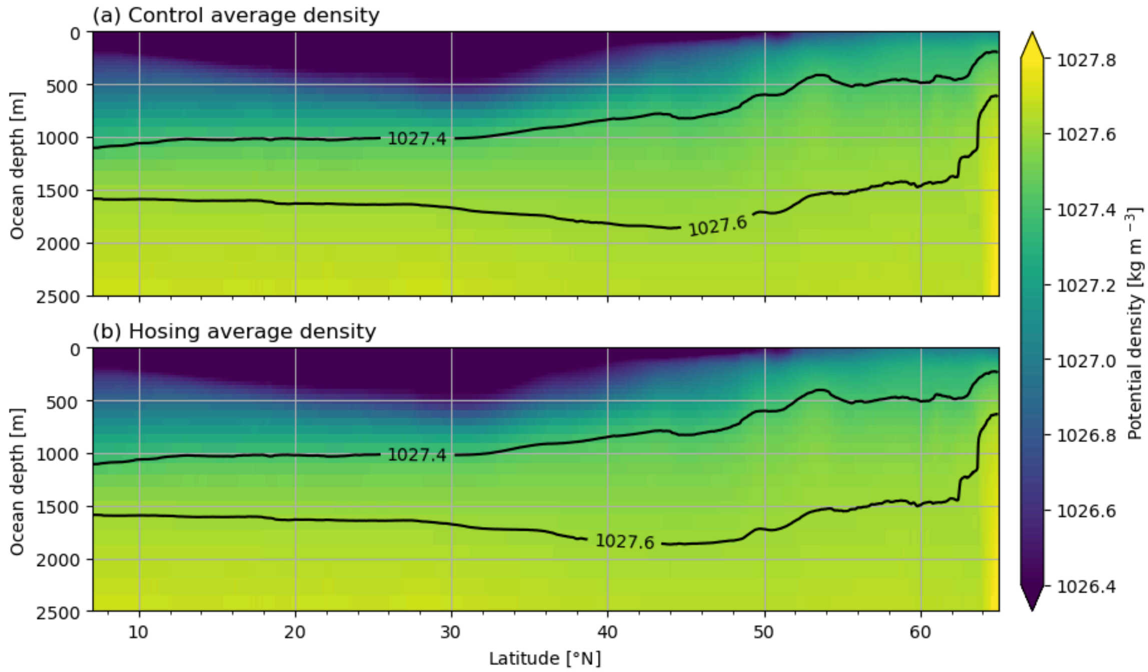

***Figure AC2.4**. Average potential density in the Atlantic basin for (a) control experiment and (b) hosing experiment.*

154 impact “on” not “in”

We have removed that sentence following a comment from RC1.

154 replace first occurrence of “reduction” with “weakening” to avoid repeat wording in same sentence

We have removed that sentence following a comment from RC1.

165 add “-“ to “boundary-interior”

Added.

167 The buoyancy loss could also happen over the path of the NAC in the eastern North Atlantic where the initial densification of the Atlantic water begins.

We agree that buoyancy loss could happen in other regions. As we want to focus on the role of the Labrador Sea in the reduction of the AMOC, we have replaced the sentence by “This also suggests that the Labrador Sea may play an important role in the AMOC response”, following a recommendation of RC1. Later we show the contributions of OSNAP East and West to the overturning changes, which support a role of western part in the AMOC reduction.

175 rephrase last part of sentence: “… Greenland coast, suggesting an absence of any local impact of the meltwater injection.”

Done.

178 suggest to rephrase: “This cooling is unlikely driven by export with the Labrador Current, …” and add “s” to “shows” then.

Changed.

178 SST is not an ideal indicator of the heat content of the Labrador Current. However, I agree that the SST change in the NAC is likely indication of an AMOC weakening.

> We agree that measuring the total heat content of both currents would be more accurate for the discussion. However, the small SST change in the Labrador Current relative to the larger change in the NAC suggests that something more than the Labrador Current is driving the temperature change in the NAC. To be more accurate and acknowledge a potential contribution of the Labrador Current, we have rephrased it adding the word "exclusively", i.e. "This cooling is unlikely to be driven exclusively by freshwater export from the Labrador Current, which shows a weaker SST response than the NAC region".

186 remove "depth" from "mixed layer depth (Fog. 4c)." and remove ",which implies that the impact is relatively high." Leave such judgement to the reader.

> Done.

188 ", where the mixing also weakens" should be dropped for readability of the sentence.

> Dropped.

189 From the thin contour lines in Figure 4c it is very difficult to see whether the controls run has deep convection in the Irminger and Nordic Seas. Please state this more clearly.

> We have made the contours thicker in Fig. 4 and added a legend to more easily identify the strength of the convection in each area. We also added the sentence "In addition, the Nordic Seas has very shallow climatological mixing due to a positive bias in sea ice (Martin-Martinez et al. 2025, Moreno-Chamarro et al. 2025), which limits the potential impact of the freshwater anomalies".

190f The deceleration of the gyre by missing deep convection could be countered by a stronger density gradient to the boundary current, which freshened (but also cooled).

> We have added that to the discussion in the next sentence: "The expected deceleration of the SPG in response to the reduction in deep mixing may be counteracted by the acceleration of boundary currents caused by the strong freshwater-driven density gradient".

193 Please rephrase: "This pattern does not … of the Gulf Stream, as no significant changes in the mean latitudinal position of the current is found when estimating its position from SST and SSH gradients (not shown)."

> Done.

196 "with a slowing of the Gulf Stream" and "temperature and salinity in[!] the NAC"

> Changed.

197 "change of[!] the wind stress" and replace "discards" with "excludes"

> Changed.

198 ", which therefore may only be tied to a reduction…"

> Changed.

230 What is the origin or cause of this warming of the DWBC?

We hypothesise that the warming is linked to a local reduction in the vertical mixing and thus isolation of the DWBC from colder waters. We expose the following in the next sentence: "This may be linked to a local reduction in the vertical mixing (Fig. 5a) which feeds the DWBC". The role of the Labrador Sea may be clear in developing this signal as it does not appear at the Greenland coast. In order to show the reduction of the volume transport in the Labrador Sea, we have computed the OSNAP West and East transport, see Fig. AC2.2. Although the maximum transport is simulated in the OSNAP East transect, as in the observations, the reduction is bigger in OSNAP West (Table AC2.1), which is linked to a reduction of the thermal contribution to the buoyancy forcing (Fig. AC2.2). We have added this table to the main part and the figure to the appendix, and enriched the discussion about Fig. 5):

> "We hypothesize that reduced deep-water mixing in the Labrador Sea could lead to reduced DWBC feeding. This would not only reduce the SPNA AMOC, but also warm the current. This hypothesis is explored further in the water mass transformation analysis in Section 3.2".

32 replace "since" with "from"

Done.

239 How are these numbers (0.4 psu and 1˚C) diagnosed, based on which years?

The quantities represent an estimation of the average difference the whole period. To be more precise we have computed the difference between the mean in both ensembles for the whole period with two decimal, which are now mentioned as follows: "The average salinity reduction for the whole period is of about 0.40 psu, while the temperature reduction is of around 0.94˚ C"

241 replace "since" with "from"

Changed.

246 I agree with the current acceleration but this seems to have a different timescale than the response in temperature and salinity. T and S show a rather rapid adjustment during the first 6 years whereas the Labrador Current exhibits a gradual, cumulative increase in speed. Please explain, eventually speculate, and emphasize this difference.

The two timescales reflect distinct physical mechanisms. The current speed also has a fast initial response, which is consistent with a fast geostrophic adjustment: the freshwater-induced increased cross-current density gradient strengthens the thermal wind balance. The subsequent cumulative increase likely reflects slower ocean-atmosphere feedbacks operating on the subpolar gyre. One plausible pathway follows the adjustment mechanism described by Oltmanns et al. (2020): a progressive freshening of the subpolar region gives rise to a surface cooling, promoting enhanced storminess and wind stress curl, which ultimately reinforces the subpolar gyre circulation in a positive feedback loop. We have added a sentence to the text acknowledging this distinction and its likely cause:

> "It is evident that both currents have also accelerated since the beginning of the freshwater hosing (Fig. 7c). Also, in both currents, while the temperature and salinity response show a rapid adjustment during the first years of hosing, the current acceleration exhibits a more gradual increase. The initial speed-up is consistent with a fast geostrophic response to the enhanced cross-current density gradient, while the longer-term trend likely reflects progressive subpolar gyre adjustments driven by ocean–atmosphere feedbacks (e.g. Oltmanns et al., 2020)"

250 "… 5 years of the simulation (marked by dots)."

Added.

253 and 258 Why is the NAC responding later than the Norwegian Current? Isn't the Norwegian Current a continuation of the NAC and would thus lag the NAC response? Please explain otherwise this appears to be inconsistent.

We thank the reviewer for raising this apparent inconsistency. The earlier response of the Norwegian Current relative to the NAC can be explained by the freshwater forcing applied directly in the Nordic Seas, part of which can escape the Greenland Current and reach the Norwegian Current without requiring long-range advection. The NAC response, by contrast, requires the freshwater anomalies to be integrated into the subpolar gyre circulation and transported along a longer interior pathway to the NAC region, explaining the delayed emergence eleven years.

266 replace "gets consistently reduced over time" with "is permanently reduced after year 6"

Replaced.

267 Shallow mixed layers in the Nordic and Irminger Seas are not worth mentioning since there is no deep convection there in control either – except for there would be an expectation for deep convection to occur under hosing. This would need to be explained though.

The average mixing in the control for the interior of the seas defined in Fig. 6 is 478 m for the Labrador, 312 m for the Irminger, and 123 m for the Nordic Seas. We agree that in particular for the Nordic Seas there is no motivation to discuss a potential shallowing of the mix layer. However, the Irminger Sea may expect some reduction, even if the mixing is not as strong as in the Labrador Sea. We have included the climatological values in the main text.

282 Another reason could be internal variability of surface heat fluxes forced by the coupled atmosphere.

The referee is right. We have produced a time series with the same characteristics (a 12-month moving mean of the spatial average in the Labrador Sea interior) as the downward heat flux at the sea surface (hfds; Fig. AC2.5), overlaying the depth vs. time plots from Fig. 8. We can see that the less cold anomalies around year 6 may be linked to the positive anomalies in hfds. However, we cannot conclude from this that internal variability in local hfds is causing the faster cooling at the beginning. The ensemble-mean signal is negative, but not coherent in sign during the early years, which indicates that at least one member has a positive hfds value that should induce positive temperature anomalies, making temperature signal not coherent across members, which is not what happens. Nevertheless, part of the signal may be explained by this, with other processes counteracting the warming tendency in the member(s) with positive hfds values. We have included a sentence before the mentioned sentence: "Although heat flux loss to the atmosphere could contribute, it does not seem to dominate the faster response of surface temperature in the Labrador Sea interior (not shown). We hypothesize that the fast temperature drop arises because lateral heat advection is more efficient than the advection of salt/freshwater anomalies, which are more susceptible to dilution and mixing along the boundary current, therefore requiring several years of cumulative transport before significantly impacting interior salinity".

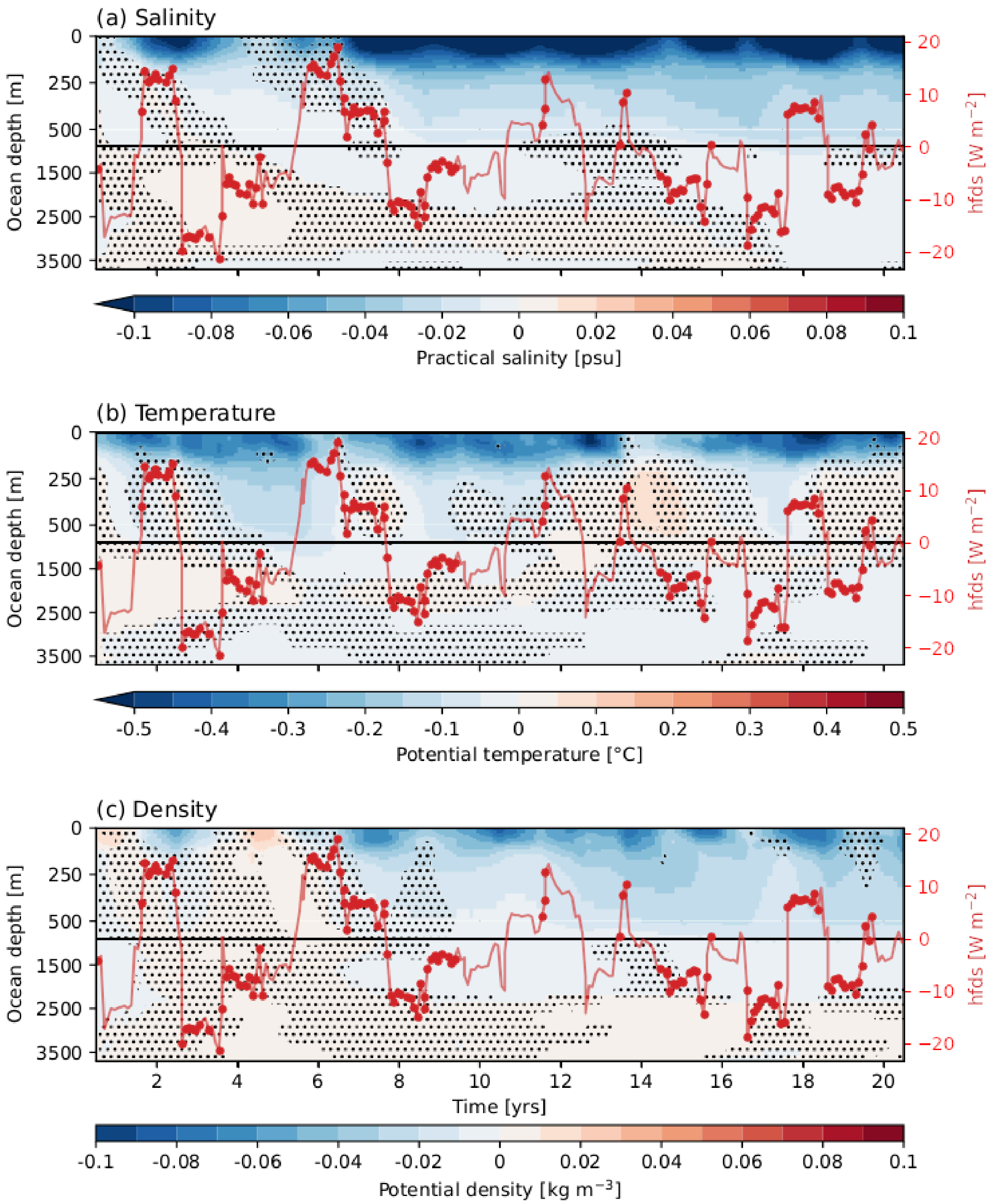

***Figure AC2.5**. Depth vs time Labrador Sea interior ensemble-mean anomalies in (a) salinity, (b) temperature and (c) density as computed between the hosing and the reference control. Anomalies that are not fully consistent in sign between the three ensemble members are masked with dots to highlight the visibility of the fully consistent ones. The Labrador Sea interior ensemble-mean downward heat flux at sea water surface anomalies time series is overlapping in red, positive values indicate warming of the ocean coming from the atmosphere and negative values indicate cooling of the ocean towards the atmosphere. The anomalies are computed as 12-month moving averages.*

314 Suggest to summarize the main process of this teleconnection as explained in Diamond et al (2025). The wave train is mentioned in line 336. This is too late.

We thank the reviewer for this suggestion. Rather than restructuring the narrative, which is organised by hemisphere and variable, we have added a brief sentence immediately after the sentence in line 314 to hint at the wave train mechanism and direct the reader to the later discussion where it is described in more detail.

> "Similar positive temperature anomalies in the Amundsen Sea are described by Diamond et al. (2025) in an analysis focused on NAHosMIP runs.These remote Southern Hemisphere anomalies are consistent with a poleward-propagating Rossby wave train forced by anomalous tropical heating associated with the AMOC weakening, a mechanism that is discussed further down in the context of the sea level pressure response."

327ff It is not clear to me how this teleconnection works. As just mentioned, please provide some more insight here likely to be summarized from Diamond et al. (2025).

The teleconnection described by Diamond et al. (2025) may explain the regional response over the Amundsen Sea, which is prominent in JJA but not in DJF. In the austral winter, the SLP response over the Southern Hemisphere is more zonally uniform and is consistent with a negative SAM, which tends to weaken the westerlies and partially counterbalance the broad Southern Ocean cooling. We have extended the discussion of the wave-train mechanism from Diamond et al. (2025), but in the relevant paragraph of the JJA section.
Additionally, the regional anomaly over the Amundsen Sea may reflect an early signature of the tropical-Antarctic Rossby wave train described by Diamond et al. (2025), which originates from Central America and propagates eastward into the Southern Ocean, generating regionally asymmetric pressure and temperature anomalies. We note that our analysis focuses on the second decade of hosing, so this signal is at an early stage of development; confirming the wave train would require upper-tropospheric diagnostics for which we do not have the required 3D outputs.

> "*This last change could explain the positive temperature anomaly previously seen over the same region and season, as the warming is located to the west of the sea level pressure anomaly, where warm southward advection is expected. These changes could be induced by a Rosby-wave train originating in Central America and propagating eastward into the Southern Ocean, as described in Diamond et al. (2025), generating sea level pressure and temperature anomalies with regional asymmetry.*"

334 replace "previously seen over" with "highlighted above for"

Done.

364 "…significant impacts on the short 20-year time scale of the experiments, like …"

Added.

368 I am missing an item summarizing the key findings using a VHR model. What is specifically enabled by having an eddy-rich ocean and correspondingly well resolved atmosphere? This is the unique selling point of the paper!

We thank the reviewer for this important suggestion. We have expanded the first paragraph after the itemized conclusions to better highlight the specific advantages enabled by the eddy-rich model configuration and their connection to the key results of the paper.

"The fidelity of the spatio-temporal distribution of the injected meltwater fluxes and the model capacity to resolve mesoscale ocean processes are two key aspects of our study, both jointly expected to enhance the realism of the simulated responses. Compared to eddy-parameterised models, the higher ocean resolution has enabled a more realistic representation of the narrow boundary currents that quickly carry the freshwaters southward around Greenland, and of the mesoscale eddies that govern the lateral exchanges with the interior of the Labrador Sea (Georgiou et al., 2020; Schiller-Weiss et al., 2024) while keeping the Irminger Sea interior isolated. In the hosing experiments, the water mass transformation analysis has shown that the AMOC weakening is driven by a thermally-mediated reduction in dense water formation confined to the Labrador Sea, with the Irminger and Nordic Seas showing no significant convective response, a regional selectivity that reflects the model's capacity to resolve the eddy-driven boundary-interior exchanges that control how the injected freshwater reaches the deep convection region. Further south, the eddy-rich configuration enables a more accurate Gulf Stream separation, as shown in other studies (Marzocchi et al., 2015; Moreno-Chamarro et al., 2021; Frigola et al., 2025), and sharper frontal gradients, with important implications for the responses in the northward heat and salinity transports and the atmospheric circulation aloft ."

397 (Appendix A) "…, many global 3D temperature (but not SST) data files …" I assume here that SST was available and that this is the reason why your analysis is mostly surface focused, right?

Yes. We have SST data available for the entire globe, as well as 3D temperature data for a small region in the SPNA, as 3D data for this region was preprocessed and saved separately before the loss. Therefore, the OSNAP section in Fig. 5c and the vertical profiles of the Labrador Sea in Fig. 8b are based on the model output temperatures, rather than estimations. However, this limited our potential for analysis in other areas. We do not want to rely heavily on the estimated temperatures, as they may introduce errors that could lead to incorrect conclusions. For example, we would have liked to compute the heat transport by the AMOC, but we cannot proceed due to this important data constraint.

407 "… that the intensity of the anomalies … anomalies are greater." I do not understand this sentence. Please rephrase. Are these two different anomalies? What is smaller, what is greater?

We have rephrased it to:
"However, a test in the OSNAP section (not shown) shows that the estimated temperature anomalies tend to be weaker than the true model outputs. Nevertheless, the sign and significance of the anomalies remain consistent in both cases".

See the values of the comparison of model output (Fig. 4c) and estimated temperature below in Fig. AC2.6.

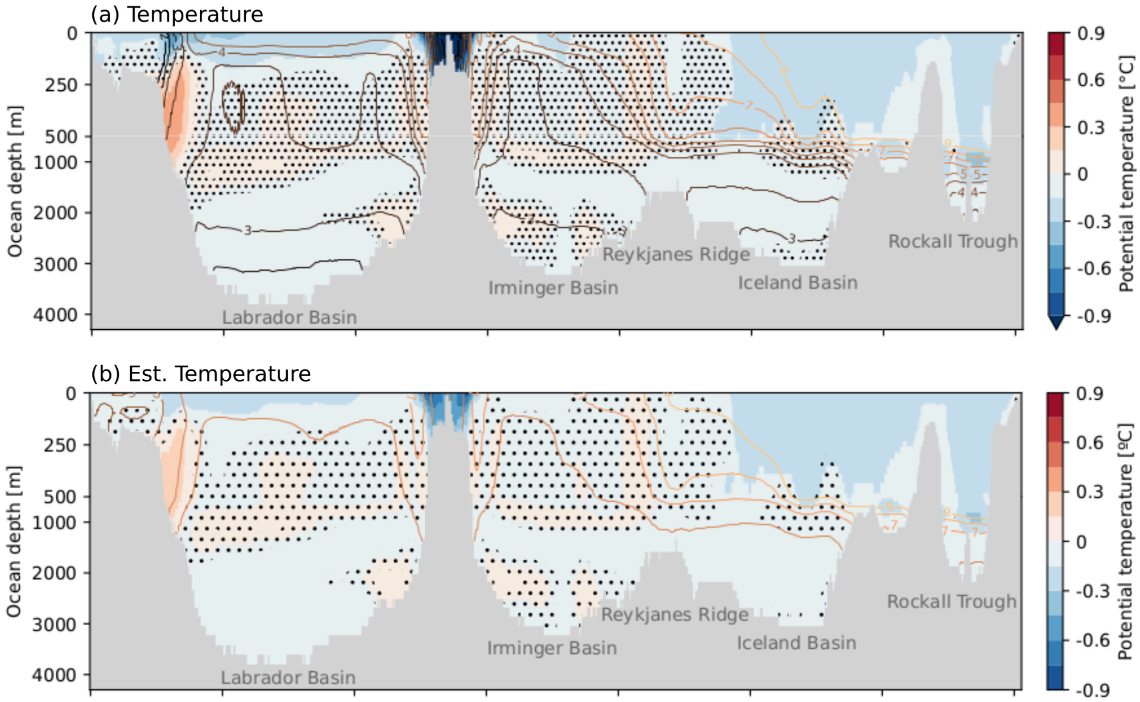

***Figure AC2.6**. Response to the hosing in the last 10 simulated years for the (a) model output potential temperature and (b) estimated potential temperature across the OSNAP section (see red lines in Fig. 4); filled coloured contours represent the ensemble-mean anomalies of the hosing with respect to the reference control and the contour lines the climatology of the control. Non-significant values as identified by the bootstrap methodology are masked with dots to improve the visibility over the significant areas.*

Figures

Figure 2: Please add bold dots for the start values of each of the three experiments in panel (a) to highlight the three different AMOC states chosen.

We have now included the values for AMOC at 45° N in depth space in the description of the methods when mentioning the three states, which are the ones that we took into account for selecting the three members. We think that adding the starting values (first year means) to the figure does not help, as the strong spread at 33.8° N substantially expands the range of the y-axis (See Fig. AC2.7), which hinders the comparability of the timeseries:

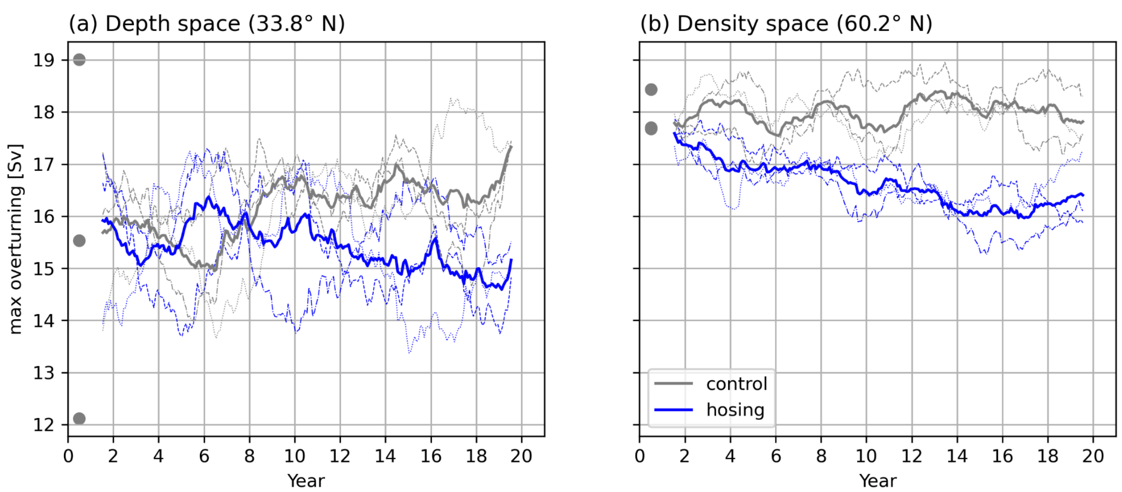

***Figure AC2.7**. Monthly volume overturning streamfunction in (a) depth space at 33.8° N and (b) potential density space at 60.2° N in the Atlantic basin. The solid lines are the ensemble averages, each member and its corresponding control year are plotted with a unique linestyle. The values have been computed with monthly averages and filtered with a 36-month (3-year) moving average. The grey dots denote the starting year average in the control for each of the three members. The latitudes were selected where the corresponding maximum change in the last 10 years happens, see Fig. 3.*

Figure 4 The thin contour lines indicating the control climatology are color coded. Please provide a color legend for this. Alternatively use black lines and provide few labels and an increment (in the caption).

We have updated all the figures with contours to include the legends on one side for a better visualization.

Figure 4d Is the sea ice edge in the hosing experiment really that far south as the stippling indicates? This seems unrealistic.

The sea-ice edge, commonly taken as 15 % of sea ice concentration can be identified by the orange contour line, better seen in Fig. B3 (Fig. B2 in the first version). The far south stippling is produced by small anomalies which produce the difference not being exactly 0.

Figure 7d Showing Irminger and Nordic seas here only makes sense if you also discuss a potential east- and northward migration of the deep convection.

We prefer to keep the values of all the deep convection areas for completeness, irrespective of the signal and significance of the response.

Figure B2 Please increase line width of the contour lines indicating the control state for improved visibility. Similarly for Figure 4 only that in the latter there is less space for this.

We have increased the linewidth of all the contours to be the same in all the figures.