Review on : "On the ocean's response to enhanced Greenland runoff in model experiments: relevance of mesoscale dynamics and atmospheric coupling"

Torge Martin and Arne Biastoch
Correspondence: Torge Martin (tomartin@geomar.de)

General comment:

I thank the authors for the submitting this manuscript, they present here a new ensemble of freshwater forcing (FWF) experiments around Greenland with four simulations each in a different models configuration. The paper reads well and show a great knowledge of the processes governing North Atlantic ocean's dynamics. Figures are clear and response of the increased freshwater is nicely described as well the role of atmospheric feedbacks and mesoscale dynamics. Summary and conclusions were particularly well-written. That being said, the paper is quite long and it would benefit from being synthesized. I also propose below some specific and technical comments in order to improve the paper.

We thank the reviewer for their overall positive feedback on our manuscript and the detailed comments, which certainly help to improve the paper. The specific comments on forcing, AMOC, internal variability and overall length of the paper are well taken and we certainly will consider all of them for the revised version.

Specific comments:

1 The forcing:

It unclear to me why the forced simulation were prescribed an atmosphere with a transient forcing while the coupled one have a preindustrial one. Comparing an historical forced simulation to a preindustial coupled simulation mixes the role of the atmospheric feedback and anthropogenic warming, why constant forcing was not used for the forced configuration?

The climate forcing indeed differs and we were weighing our options to match various project goals. The coupled, non-nested model simulations were first conducted for the multi-model comparison presented in Martin et al. (2022) and the coupled-nested simulations were executed to match these, i.e. running pre-industrial control runs to (a) clearly identify model-specific mean state and internal variability and (b) isolate the effect of Greenland meltwater input. This is standard procedure for coupled sensitivity experiments.

Another goal was to compare such experiments with typical ocean-only hindcast simulations, such as in Böning et al. (2016). While repeat-year forcing would have appeared to enable straight comparison with the pre-industrial control runs of the coupled model, we think that the historical forced simulations are advantageous for two reasons: (1) they enable a more direct comparison to the study of Böning et al. (2016) and (2) include realistic interannual to decadal atmospheric variability which would be excluded by repeat-year forcing and internal climate variability would appear largely damped. Further, we note that even by repeating the coupled experiments under historical greenhouse gas forcing would not necessarily yield a better comparability with the forced ocean-only runs because timing of global warming effects may not be exactly the same in coupled experiments.

Since there seems not to be an ideal solution, we considered internal variability the more important factor and thus ran the ocean-only experiments with variable forcing but wanted the coupled experiments comparable with the multi-model comparison presented in Martin et al. (2022, GRL).

2 The AMOC:

The AMOC has a large decadal to multidecadal to multicentennial variability (Ortega, 2015) depending on the model. The long term period chosen (100 years) is thus a rather short, AMOC could be experiencing a trend (see DOI: 10.1175/ JCLI-D-13-00651.1, their figure 4). By comparing the period 51-100 years to the mean state 1-100 in the coupled simulations, you are mixing the response of the FWF and internal variability. Same period would appear to be a more clear comparison rather than artificially increasing the signal by changing the period. Additionally, AMOC could not respond to FWF the same way if it is on its stronger or weaker phase, so it would be useful to have a figure showing the time series of the annual AMOC, and what was its state when the perturbation was added. Last, the mean response stays within the limits of the internal variability, so it should be mentioned that it is not significant.

We agree that AMOC internal variability exists on multi-centennial time scales, of course. Computing trends for various subsections of the reference runs, we found weak AMOC trends for the two coupled reference runs for periods of 100 to 200 years (coupled: 0.025-0.07 Sv/decade, coupled-nested: -0.05 to +0.03 Sv/dec) but a considerably wider distribution when using periods shorten than 80 years. We thus consider a 100-year long reference period as sufficient to limit the imprint of internal variability. Regarding the AMOC strength at the onset of the freshwater experiment, the coupled-nested run starts close to the long-term mean AMOC strength (+0.5 Sv) whereas the coupled experiment presented here starts in a low phase (-2.2 Sv). However, the coupled run is part of a small ensemble discussed in Martin et al. (2022), who show a robust AMOC weakening over the last 50 years independent of the starting condition. See also response and figure further below.

3 The periods of comparison and internal variability:

The choice of the periods of comparison is key to this study because it impacts all the results. This question is discussed only in the appendix while it seemed rather central to me, and it could benefit for being a bit more structured and clarify. Figure A2 shows averages over periods, but we are lacking some times series to give us an idea of the decadal variability in the Labrador sea for instance. The response in the coupled simulations could be a result of the internal variability: as the system is chaotic, changes in initial conditions could lead to another state. Seems to me that taking a long-term time mean is not enough to take out the several feedback effects from the system (Swingedouw, 2007b). I understand that this paper have chosen to do one member (run) per configuration but please maybe add a paragraph discussing and clarifying this issue (forced signal from FW versus forced signal from GW etc...).

The Appendix is intended to collectively address the issue of distinguishing the response to the freshwater from internal variability and we appreciate the additional suggestions to discuss this in a more comprehensive and convincing way in the revised version. For the coupled experiments we chose 50-year means to exclude influence by internal variability as best as possible considering the overall experiment length of 100 years and a multi-decadal adjustment phase. Martin et al. (2022) show that for most of the subpolar North Atlantic a quasi-equilibrium of the response can be assumed after a few decades, such that an average over years 50-100 seems like a reasonable approach. Further, three ensemble members of the coupled configuration are discussed in that paper showing that despite internal variability interfering with the response signal the changes discussed in the present manuscript and also their magnitudes are robust (for the same configuration). We assume the same holds for the additional configurations discussed here. Because of the already extensive main part of the paper, we remain with our decision to focus this discussion in an appendix. However, we added a dedicated hint towards the appendix at the end of the introduction section.

Technical corrections:

1-Introduction

l.20: add a citation after “decay”- “as well as” → “as well as is”
l. 58: “to shedding” → “to shed”
done

l.60: description of (a), (b) and (c) experiments is not clear, please describe the whole experiment in one time, for example: you will compare 4 simulations with freshwater forcing (FWF) to the same simulations without FWF and those 4 simulations are : one coupled, one forced, both with and without nest.
good point; done

l.67: introduction the question of the mean state question is a bit abrupt, maybe explain a bit before line 66 why it is has to be addressed with one citation
Done. We address the issue of the model mean state citing Stouffer et al (2006) and Swingedouw et al. (2013) as examples in the previous paragraph now: "While earlier studies suggest no systematic dependency of the AMOC response on its reference mean strength \citep{Stoffer2006, Swingedouw2013}, we do consider a potential sensitivity to the general ocean mean state."

l. 71: "by" → "from"
done

2-Model configurations and experiment

Table 1: the term "spatially varying" is bit misleading: the spatial distribution is kept constant in time right (*cf "The perturbation is constructed from the monthly-mean runoff plus discharge fluxes of Bamber et al. (2018) by averaging the period 1992–2016", line 129*)? Please clarify that either in the legend of the Table 1 or in the text
The table caption now reads: "A freshwater flux (FWF) of 0.05 Sv is added as seasonally varying runoff using a spatially heterogeneous but time-invariant pattern along Greenland's coasts in the perturbation experiments."

l.107: specify or give a bit more information about what "model parameter" you are referring to
Our statement refers to diffusion and viscosity parameters set in the namelist, which typically scale with both grid resolution and time step, such as rn_ahtbbl: base/RHO^2, rn_aeiv_0 set to 0 in nest, rn_aht_0: base/RHO, rn_aht_m: base/RHO, rn_ahm_0_blp: base/RHO^2, rn_ahm_m_blp: base*RHOT/RHO^4, and rn_ahm_m_lap: base*RHOT/RHO^2 where RHO and RHOT are the spatial and temporal refinement factors.
We add "…, which mostly affects viscosity and diffusion settings." to the respective sentence.

l.117: "hight" → "height"
corrected

l.127: "most" → "mostly"
This change would give the sentence a different meaning. We decided to rather remove "most".

l. 134 "62 and 100" → missing the word "years"
added; thanks for noting these small glitches

l. 138: not clear why "maximum runoff in June to August" is simplification, I guess this relates to the line l.144: "shifting the seasonal peak" maybe reformulate to make it easy to follow and specify what should be the real maximum month for the runoff to get out of the fjord into the open sea
The respective paragraph begins with listing three simplifications of which the seasonal timing is one. In the reminder of the paragraph we provide further detail ("meltwater runoff is first entrained into the fjord circulation, which causes both a vertical redistribution and a temporal delay of several weeks before entering the open ocean") and arguments for the feasibility of the simplification ("prescribed freshwater … also shifting the seasonal peak by a month."). We consider this sufficient detail but note here, that the Bamber et al. data set has peak runoff in June to August, which is why it is a simplification to simply prescribe this runoff field without further treatment of the seasonal timing.

The delay between runoff (and calving) into the fjord and meltwater being exported into the open ocean varies depending on fjord circulation and topography. It is thus not possible to provide a "real maximum month".

l. 148: is the error calculated here computed from he loss of tracer concentration along the experiment? Please specify
The sentence now begins with "Using the passive tracer concentrations, we compute an error …"

3-Results

l.154 and 155: "variations in internal variability" → "internal variability"
"variations in" removed

l.155: "which evolve freely within the preindustrial boundary conditions provided" → "which atmosphere evolves freely under preindustrial forcing" the term " boundary conditions" is used for regional modelling, when we prescribe values at the spatial boundaries of the model domain. For climate simulation, better to use the term "forcing".
agreed and corrected

l. 157: suppress: "which are the same for each simulation", already said line 98 and in Table 1
done

l.159: "can only be expected to exist after several decades" justify this choice of time frame, maybe by adding a citation
done, citing Swingedouw et al. (2013) and Jackson and Wood (2018) for example

3.1 Ocean mean states and responses

AMOC

Table 2: "Denmark Strait (DS) overflow potential density" → is it the annual mean?
Yes, in fact also decadal running mean. We rephrased the sentence: "Correlation coefficients (Pearson's r) between annual-mean AMOC strength and Denmark Strait (DS) overflow potential density ($p_{ref}=0$) as well as March-mean mixed-layer depths (MLD) in the Labrador Sea after applying a decadal boxcar averaging filter to all time series"

Figure 4: The caption is unclear, please explain what are the dark blue histogram and maybe add "(light blue)" after "perturbed states".
The blue perturbed state histograms are transparent, i.e. "dark" blue shading indicates underlying black histogram. We added "(blue, transparent)" after "perturbed states".

l.180: you are not coupling to the same atmosphere, the slower AMOC in the coupled simulations could be the results of the transient forcing
true, could be. Sentence rephrased: "Furthermore, the coupled configurations simulate a stronger AMOC than their forced counterparts, which could either be related to coupling with an interactive atmosphere as in \citet{Hirschi2020} or to comparing the historical (forced) with the pre-industrial climate state (coupled)."

Large-scale upper ocean salinity and freshening

l.199: add reference to Figure 1 to show transportation of FW
done

l. 201: Salinity is decreased a lot along the western coast of Europe in the coupled non-nested simulation. Are the FW leaking towards the subtropical gyre as seen in other hosing experiment (Swingedouw, et al. 2013; Devilliers et al , 2021), maybe showing a larger map could answer that?

Yes, freshwater leaks into the subtropical gyre as described in the named references. For most of our manuscript we prefer to keep the focus on the subpolar North Atlantic. The exchange with the subtropical gyre in non-eddying and strongly eddying simulations is subject of another forthcoming paper. However, we have expanded Figure 13, the maps of meltwater tracer concentration, southward to include and illustrate both the eastern upper and western deep export routes.

l. 203: I do not see a more realistic Gulf Stream separation in the coupled nested response than in the coupled response (Fig5 b, left) Please correct the statement.
The improved Gulf Stream location in the nested configuration is shown in Figure 3, e.g. by the reduced warm bias at the US coast (35-40˚N). The sentence addressed here describes what happens in the simulation and what is later discussed in conjunction and supported by Figure 13. We add references to Figures 1 and 2.

l.210: 1 → 1 psu
we consider salinity to be unit-less

l.219: Add a figure of the sea-ice, or "(not shown)"
reference to Figure 7 added at end of sentence referring to both sea ice and mixed layer changes

l. 229 ENA is defined later (line 247)
indeed, thanks for catching this, abbreviation now defined here

l. 230: "Nordic Sea" → "Nordic Seas"
corrected

l. 237: "The two nested experiments both feature an overall stronger inflow of Atlantic water into the Nordic Seas" I do not see that in the figure, please explain
True, this is not directly shown here, we thus add "(not shown)" to the sentence. The statement is based on the fact that the subpolar gyre circulation is generally stronger in the nested configurations. While SPG strength is not discussed in detail, we do note in subsection "Boundary currents" a stronger western boundary current for the nested setup.

Water-mass transformation:

This subsection is 6 pages itself, far larger than AMOC, salinity and temperature responses (1 to 2 pages each), please consider to reduce it or making it a 3.2 section to have some equilibrium
We have shortened this section along with other parts of the paper.

l. 243: "sights" → "sites"
done

l. 251: which is due to a weaker AMOC in the non-eddying simulations
No, this is due to the misrepresentation of the NAC, it's too zonal placement (c.f. Fig. 2 and 5). We added this reason for the fresh bias to the sentence.

l. 257: but coupled simulations also present with a stronger AMOC, bringing more warm water into SPNA
Good point. We considered this aspect more carefully, also in reference to Figure 2, and changed the statement according to your comment: "With 10–11˚C the potential temperature is very similar for all model configurations except for the coupled, non-eddying one, in which the ENA region is strongly influenced by the cold bias with respect to late 19th century reanalysis (see \reffig{fig_bias}b). The forced experiments must thus be considered relatively cool running with historical atmospheric forcing but having a weaker AMOC and hence less northward heat transport."

l. 264: question ?
comment unclear

l. 269: “source waters” → “water sources”
We consider “source waters” correct oceanographic terminology here, since we address the source water masses of the T,S properties found in the upper Labrador and Irminger Sea.

Figure 8: add the Labrador sea shelf region to be coherent with Figure 6
Comment not quite clear: do you mean Figures 8 and 9? we tried adding a frame for the Labrador Sea shelf to Figure 8 but the plot became too crowded. Instead we add a more precise definition to the caption of Figure 9: “The shelf is defined as areas shallower than 500~m in the region 62-46˚W and 56-65˚N, i.e. within the same geographical box as the deep, interior Labrador Sea”

l. 270: Figure 9 shows that density seems more different between Labrador and Irminger sea in forced non-nested than in coupled nested
l. 273: “Moreover, the coupled runs exhibit a stronger salinity”: I see that only for the coupled nested simulations
l. 273: “thus density gradient” → “thus stronger density gradient”
l. 275: “more detail” → “more details”
this paragraph has been rewritten: “In both regions the coupled reference simulations are a little saltier than their forced counterparts and the same holds for the comparison between eddying and respective non-eddying simulations. We suggest a lack of offshore Ekman transport in the coupled configurations and generally stronger deep convection in the nested ones as causes and discuss these further below. Another more obvious salinity and thus density difference is found between the boundary current (small purple circles) and the interior Labrador Sea (purple). This difference is significantly smaller in the non-eddying than in the respective eddying simulations and can be related to a insufficient exchange across the shelf break in the latter (more details in \refsec{sec_meso_dyn}).”

l. 277: “The freshwater perturbation leads to a freshening and cooling in the ENA and on the ENA shelf in all configurations” → I disagree: Fig 9 shows a warming in the ENA shelf for the non nested simulations and in the ENA for the forced-nested (comparing circle and cross)
We do not agree with this statement and think there may have been a mix-up of the color coding of ENA shelf and Nordic Seas. It is true though, that there is a slight but non-significant warming on in the ENA shelf in forced_nested configuration.

l. 291: “similar pattern” →“similar pattern to ENA”
added

l. 295: it is consistent with the reduction of the convection activity in the Lab. Sea (Fig 7, b)
right, noted in text

l.305: “a consequence of the shallower deep convection in the forced configuration” → “a consequence of the shallower mixed layer in the forced configuration (see Fig. 7)”
changed accordingly

l. 318: “but the 1/10◦ ones without though” → reformulate
sentence reformulated: “As will be further discussed below this is related to running the ocean model at 1/2\textdegree grid resolution with an eddy parameterization but at 1/10\textdegree without. However, the higher resolution …”

l. 326: “not shown”, isn’t it shown in Figure 5 a)?
true indeed; reference to Fig. 5a added, thank you

l. 330: “(not shwon)” → “(not shown)”
done

l. 331: “the least” → “less”. Figure 13 is cited before Figure 12, please exchange figure numbers.
wording changed; Figure 12 is referenced first just 3 lines above (was line 328)

l. 341: Figure 7 a) does not show that "mixing across the SPNA [...] is enhanced compared to the non-eddying configurations", not for the coupled one at least
statement removed as part of shortening the entire "Water mass transformation" section

l. 342: "in both experiments" : which ones? It decreases more in the forced than in the forced-nested
see above

3.2 Mesoscale dynamics

l.348: "This is" → "These are"
corrected

l. 352: you would need the same figure at 1/2 degree to compare to use the word "improves", please change to "display a realistic..."
done, see next comment

l. 354: "For example": This is not an example of why "the finer resolution [...] is inadequate to simulate the full dynamical mesoscale spectrum." please re-organize
first part of paragraph restructured: "The ocean-grid refinement yields realistic dynamics in the nest region (Fig. 1a). We find a strongly eddying ocean where the 1/10\textdegree grid sufficiently resolves the Rossby radius, which is the case south of approximately 50°N. In higher latitudes the finer resolution yields stronger and more focused boundary currents, such as in the Nordic Seas and the Labrador as well as Irminger Sea. For example, the western boundary current transport in the Labrador Sea at 53\textdegN of the coupled model amounts to 33~Sv and that of coupled-nested to 53~Sv, which is much closer to observations. The grid refinement significantly improves mesoscale variability over large parts of the SPNA (Fig. 2a) but is inadequate to simulate the full dynamical mesoscale spectrum north of 50°N. Nevertheless, we find individual WGC eddies-"

l .356: "much closer to observations", a citation is need here to justify the numbers
sentence rephrased and citation added: "For example, the western boundary current transport in the Labrador Sea at 53˚N (below 400~m) amounts to 19.4~Sv and 39.3~Sv in the coupled and coupled-nested configurations, respectively, where observations yield an estimate of 30.2~Sv \citep{Zantopp2017}"

l. 360-361: "over/underestimation" is not the best term since there is no comparison to observation here so we do not know if the deep mixing is over/underestimated maybe use "stronger/weaker" instead?
text rephrased accordingly

l.362: "in the nested perturbation experiments", please add the depth you are referring to (50 meters I guess)
added: "… over the entire water column but most pronounced at 50~m depth in Figure 13."

l. 363: "highlight the necessity of using at least 1/20 ◦ grid resolution" → "suggest that the resolution may not be high enough with this model"
done

l. 370-373: add references to figures.
added

l. 390: "ocean below 1000 m." → "ocean below 1000 m for the configurations with eddy parameterization."
added but this is "for the nested configurations."

l. 395: "stronger meridional density gradient in the NAC region", add a reference to Figure
This is somewhat visible on the salinity and temperature fields presented in Figures 5 and 6. However, the statement is rather speculative and we decided to remove the sentence.

3.3 Atmospheric coupling

Figure 14: there is one extra parenthesis in the caption. Seems like the coupled-nested displays values on a coarser grid than the forced-nested?

extra parenthesis removed. Yes, surface fluxes in the coupled configurations is computed on the coarser atmospheric grid. We mention this in what was originally line 414f: "The block-like structure in the SHF output of the coupled configurations is due to the surface fluxes being computed on the coarser grid of the atmospheric model at a horizontal resolution of about 1.9°"

l. 417: "In the non-eddying configurations," → "In the coupled non-eddying configuration,"

No, this holds for both non-eddying configuraitons (see Fig. 14a). In fact, remnants of this feature are visible in the long-term mean of the forced-nested configuration as well.

l. 421: "can adjust to changing" → "can adjust"

done

l. 426: "the upper ocean cooling [...] reduces the temperature difference between ocean and [...] atmosphere" → you mean that in the forced model, it is the upper ocean who adjusts to the atmosphere to reach equilibrium? Maybe add a little more details about the surface heat flux estimation in a forced model, or a citation where this is explained

No, the upper ocean cools in the perturbation experiments as a consequence of AMOC weakening. Since (1) surface heat fluxes are driven by the temperature difference between ocean and atmosphere, (2) in winter the atmosphere is colder than the ocean and (3) the SST decrease but atmospheric temperature is unchanged in the forced experiments, the decrease in SST drives a reduction in surface heat fluxes.

l. 455: "this results": you should mention you are referring to the response to FWF

done (actually by adjusting the previous sentence)

l. 457: "southward expansion of the sea-ice edge" : the extension is not very clear and wind response has a lot of noise, maybe worth to be mentioned

Since we look at 50-year averages this would need to be wind variations on almost centennial time scale. The sea-ice edge position and associated surface heat flux changes are the only systematic changes we found and they coincide with wind stress decrease over Davis Strait. We agree that the sea-ice change is rather small but so is the wind-stress change.

l. 458: "The particular reinforcement of the onshore Ekman transport" is that a stable feature in the coupled-nested configuration or period dependent? Is it more or less constant along the simulation, have you tried different time-slices?

Yes, this is a stable feature.

4 Discussion

l.475: "decade" → "decades"

changed to "last decade"

l. 487: "to hindering" → "to hinder"

corrected

l. 488: "Potentially in consequence thereof, enhanced deep convection in the Irminger Sea has certainly offset any impact of recently enhanced runoff from Greenland on deep water formation." I do not understand that statement as enhanced deep convection means impact on deep water formation

Enhanced runoff from Greenland is expected to reduce deep convection, first and foremost in the Labrador Sea. A coincidental increase in deep convection in the Irminger Sea could compensate for the lack of deep water formation in the Labrador Sea and hence offset the impact by enhanced Greenland runoff. The sentence was rephrased: "Recently enhanced deep convection in the Irminger Sea \citep{Ruehs2021} may have compensated a lack of deep water formation in the Labrador Sea and hence offset an impact by recently increased runoff from Greenland."

l. 493: why the plan of the results is not kept here? As: first mesoscale eddies and second atmospheric coupling
In the present order the last presented results are revisited and discussed first.

l. 503: "support stronger deep convection" you mean "support stronger reduction of deep convection"
No, we actually mean that the deep ocean heat and salinity bias help to maintain the ongoing deep convection in the coupled model. Sentence is rephrased: "This may help to maintain the mode of recurring deep convection making the coupled ocean less susceptible to the prescribed moderate freshwater perturbation."

l. 504: "surface heat loss is less than 10%" this is because atmosphere is adjusting along the simulation, not so sure this questions the importance of a positive feedback
good point. Statement changed to "However, we cannot exclude a significant influence by the ocean and climate mean state, which differs between coupled and forced experiments."

l. 506: "to doubting" → "to doubt"
see last comment above

l. 510: "In" → "in"
but the sentence after the colon is complete by itself and thus starts with a capital letter

l. 536: "The eddies resolved in our model are obviously not sufficient for bringing enough meltwater to the deep convention sites to achieve results comparable to Böning et al. (2016)." this is contradictory to the statement of before l. 534: "larger eddies, [...] which carry relatively fresh water from the boundary current into the interior Labrador Sea", so the resolution is sufficient to carry the freshwater
The resolution is sufficient to carry some(!) freshwater by eddies into the Labrador Sea but not a sufficient amount in total. We adjust the sentence accordingly: "The WGC eddies resolved in our model are not numerous and hence not sufficient for bringing …"

l. 548: "a apply" → "apply"
done

5 Summary and Conclusion

l. 599: "deviates" → "deviate"
corrected

l. 604: "We note, that" → "We note that"
corrected

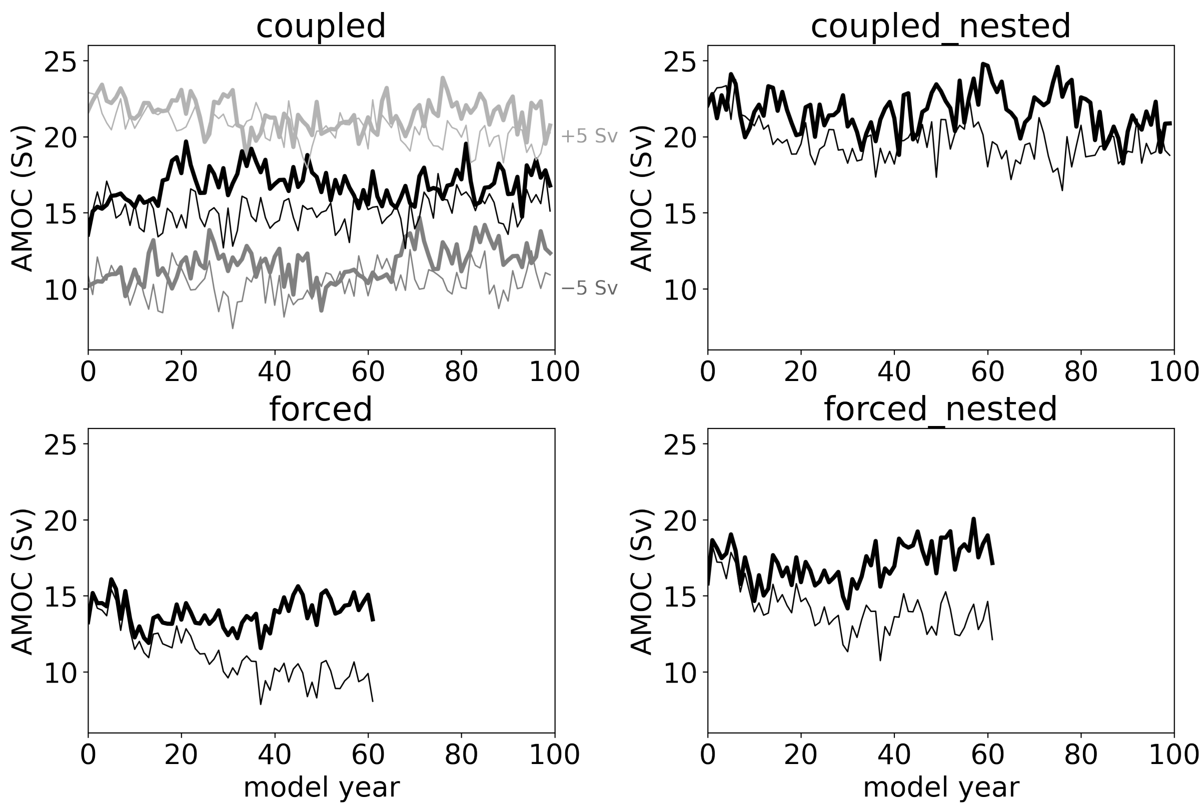
l. 610: "are" → "is"
corrected

Appendix:

Figure A1: These figure are hard to read, please zoom in the forced and forced-nested and add a figure showing the annual five year running mean to display the phase of the AMOC. Explanation of the orange line is unclear, maybe add a formula
The idea of Fig. A1 using the same y-axis scaling for all plots is to emphasize the amount of internal variability adding noise to the AMOC timeseries in all the experiments. Sources of noise are the interactive atmosphere in the coupled runs and mesoscale eddies in the nested ones.

We rephrased the explanation of the orange line: “This is supported by computing a running mean of the AMOC strength difference between perturbed and reference run using a boxcar window always anchored at the end of the time series and expanding backwards in time.” The expanding running mean informs us about the number of years prior to the end of the experiment that can be included without significantly changing the mean AMOC response occurring towards the end of the run.

The paper is already extensive and includes a number of figures. We thus add here for the purpose of the discussion the requested plot of AMOC strength time series (Note, gray lines in the upper right panel for the coupled experiments depict two more ensemble members shifted by +/- 5Sv for visibility, which are not included in the present manuscript but in Martin et al., 2022). As stated above most perturbation experiments start in a phase of relatively strong AMOC but as demonstrated by the ensemble members for the coupled configuration this has no significant impact on the AMOC response.



l. 653: “varibaility” → “variability”. “we can attribute the larger variability to the explicit simulation of mesoscale eddies” → explain a little bit more maybe how the parametrization of the mesoscale processes leads to such a lower seasonal variability

typo corrected; we add the following condensed sentence to briefly comment on the GM-parameterization: “The eddy paramterization by \citet{Gent1990} adds isopycnal mixing to non-eddying simulations, which otherwise would lack the conversion of potential to kinetic energy from local baroclinic instability, but misses additional sub-grid scale effects and kinetic backscatter, and hence rather acts to smooth variability \citet[e.g.][]{Zanna2017,Hewitt2020}.”

l. 659: “By” → “by”

opposed, complete sentence after colon may start with capital letter

l. 662: “(cf. 2)” → “(cf. Figure A2 or section? 2)”

meant Fig. 2, corrected

l. 663: “stabel” : stable – no I rather see steady decline, since each month AMOC_perturb–AMOC_control<0
typo corrected; stable because difference does not grow further, we rephrase: “stable state of difference from the reference run”

l. 664: again, I do not understand what you are summing here
rephrased, see our reply for Figure A1 above

l. 665: missing parenthesis
added, thanks

l. 666: “30+” → “30”. “we find a relatively stable state for the last 30+ years for the coupled experiments” you mean a stable difference?
yes, see above

l.667: “the adjustment period is likely shorter than in the forced experiments due to the overall weaker response.” I am not so sure about that see general comments
The entire sentence is: “As noted by Martin et al. (2022), the AMOC decline in the coupled experiments is difficult to separate from internal decadal variability but the adjustment period is likely shorter than in the forced experiments due to the overall weaker response.” We clearly acknowledge the difficulty of separating signal from noise in the coupled runs. We thus need to make an assumption for the duration of the adjustment period and consider it as shorter for a weaker AMOC decline. Martin et al. (2022) also present a timeseries of stronger AMOC decline under twice as strong freshwater perturbation, where the adjustment period is clearly longer.

l. 668: “Therefore, we simply use the second half of these experiments to improve statistics.” not comparing with the same period, you are mixing the signals, maybe add a figure comparing the same period to show the difference
There is no “same period” because the experiments are too long and internal variability causes deviations between reference run and perturbation experiment over the course of the perturbation time window. All we can do is to best estimate the noise caused by the internal variability to see whether the deviation due to the perturbation is significant. Therefore, we use as extensive periods as reasonable.

l. 669: “to improve statistics.” : to which statistics are you referring to?
Any mean, difference, standard deviation we use; sentence rephrased: “to reduce noise from internal variability for improved statistics.”

Figure A2: move it before bibliography
caused by the latex template, needs to be done by layout later