Response to Reviewer #1

We thank the reviewer for their very constructive comments and suggestions as well as for the time spent reviewing the paper. In the following, the reviewer's comments are in black, and our responses are in red.

Thank you for your document "Simulated and Observed Transport Estimates across the Overturning in the Subpolar North Atlantic (OSNAP) Section". It is a very useful contribution to our understanding and obviously contains a lot of analysis. My recommendations are mostly about exploring some relationships a bit more, putting results into context with some other studies and improving the clarity of the results. My suggestion is for major revisions – I do not think the revisions themselves are particularly serious, but am aware that they might take some time.

Thank you for your kind words, appreciating the usefulness and value of the manuscript for the community. We have revised the manuscript, following most of your suggestions. Please see below for details of how we addressed each of your comments.

Major

1. There are a couple of places where the authors make the point that we don't know that OW dominates OE for decadal and longer variability (L243, final sentence). This is true, but an obvious question is what do the models show? Do any/many of the models suggest that stronger variability can be seen at OW? Is there a relationship between decadal variability at OW and shorter timescale variability/mean state which might let us infer what the observations might show?

   First, to clarify, we make statements such as "we cannot yet address the relative roles of the LS vs. the eastern subpolar gyre in AMOC variability on decadal and longer time scales" within the context of available observational datasets. Since the observational record is short, i.e., only 8 years for OSNAP, we cannot assess whether OW (or the Labrador Sea side) dominates low-frequency variability. A discussion and summary of what the models show had been already provided in paragraphs 3, 5, and 6 of the Introduction. Specifically, "in many simulations, the associated low-frequency variability mechanism involves surface buoyancy flux anomalies primarily over the LS region, but also with contributions over the Irminger Sea, that arise from long episodes of the positive phase of the North Atlantic Oscillation (NAO)." As stated in the 8th paragraph (second to the last) of the Introduction, our study focuses on providing a basic, yet much-needed evaluation of simulated transports against OSNAP observations for the 2014-2022

period using a large set of forced ocean – sea-ice simulations. Our effort thus aims to produce an assessment of model fidelity as well as a benchmark for simulated transports and related properties at the OSNAP sections in comparison to these observations for only this specific period when OSNAP is available. As such, a detailed assessment of low-frequency variability in these simulations and its relationship with shorter timescale variability is beyond the scope of the present study. However, we are pursuing exactly these topics as part of a separate ongoing study.

2. You split your models into high and low resolution, but it seems to me that there are 3 groups of about 6 models each – low resolution (>0.25 degrees), medium resolution (~0.25 degrees), and high resolution (<0.25 degrees). Does this split change anything about your conclusions of resolution?

   Thank you. Following your suggestion, we have now split the simulations into 3 groups: Low Resolution (LR; near 1°), Medium Resolution (MR; ~0.25°), and High Resolution (HR; 0.1° or finer). We have redone all the relevant figures accordingly. This new split does not change any of our conclusions.

3. Your timeseries plots are difficult to read because it's difficult to tell some of the colors apart. Some possible options: use different line styles as well as colors; split the models into low, medium and high resolution, with legends for each separately.

   In response to item 2, we have followed your suggestion of splitting the simulations into LR, MR, and HR groups. Accordingly, all the relevant timeseries plots are redone with fewer lines in each panel, making them easier to see. We now also use a single color for the simulations from the same group, making it easier to compare them across resolutions. Similarly, the scatter plots use different symbols for LR, MR, and HR, but with the same color for the same group. We appreciate these suggestions; they certainly improve clarity and readability of the figures.

4. The paper is quite long, however there is quite a lot of space taken up by description of plots, such as values for particular models etc. This might be important to highlight in some cases, but some parts of the text feel like a long description of the figure. Ideally this information could be read off the plot for someone interested, and more space could be used to discuss the more interesting findings such as: common model biases; differences with resolution; relationships between variables; possible causes and implications of specific differences (eg the authors do discuss the implications of having a better overflow

representation in some models); how results fit with other findings. This is done sometimes, but not always. I've made some suggestions in the points below.

Based on our experience with writing similar multi-model intercomparison manuscripts, we find that it is important to provide values of fields for some simulations wherever appropriate, helping the reader. We also think that it is good practice to introduce a figure before its discussion to set the stage, except for some details which are mentioned in the figure captions. To the extend feasible and practical, as the reviewer also indicates, we include discussions on the listed topics. We have now included additional discussions following your suggestions as detailed below. In particular, we now split the models into three groups: low-, medium-, and high-resolution following your suggestion. Accordingly, all relevant figures have been replotted to distinguish these groups. We have also added a new figure (new Fig. 7) to show west-to-east cumulative transports across the OSNAP section in response to another reviewer's suggestion. We have removed Appendix B.

Minor

L52 Remove 'general' – I'm not sure what it means here.

Removed.

L75 These reviews are not recent anymore!

Deleted "recent". Also, added a reference to Jackson et al. (2022) review article.

L88 comma after weakening

We think that this sentence should not use a comma (or commas), because its intent is to emphasize all mentioned aspects (weakening, stability, and collapse) equally, noting also that "weakening" refers to "AMOC" which is at the end of this sentence.

L123 I think I'm correct in saying that fluxes may also be important over the Irminger sea as well as the Lab sea?

The sentence has been modified to include the Irminger Sea as well.

L125-131 The text implies that all models show a link between low frequency AMOC variability and the Labrador sea which is not true, certainly in coupled models - eg see https://doi.org/10.1007/s00382-023-07069-y

It certainly was not our intention to imply that all models show a link between the Labrador Sea variability and that of AMOC. We have clarified this point by saying "In many simulations,…." and by mentioning a role for the Irminger Sea as well, following your suggestion above. In addition, we have added a sentence that explicitly mentions connections with the GIN Seas and the Arctic Ocean, citing Zhao et al. 2024), at the end of this paragraph.

L200 I'm surprised that the authors do not also include a discussion of Yeager et al https://doi.org/10.1126/sciadv.abh3592 , which has a different mechanism for the interaction between the regions

Actually, Yeager et al. (2021) was introduced and discussed earlier in the introduction on lines 136-140 of the original manuscript in the paragraph concerning low-frequency AMOC variability. The paragraph on lines 178-207 of the original manuscript covers the OSNAP period. We also note that Yeager et al. (2021) is cited in Section 4 and Section 9.

L221 conclude that coupled models

Modified to read " …. conclude that coupled models analyzed in their study are …. ".

L276 It's not clear to me why MOM5 and 6 are listed as different models, but the NEMO versions are not.

MOM6 is a completely re-written version of the Modular Ocean Model that combines various elements of MOM5 and the Generalized Ocean Layer Dynamics (GOLD) ocean model. As such, there are major architectural and structural differences that include elimination of the B-grid in favor of the C-grid in the horizontal and a generalized Lagrangian-coordinate in the vertical. In contrast, the NEMO versions used here (3.6 and 4.0) differ primarily in their scalability and optimization, physics options, and coupling capabilities among several others. In our view, these differences do not rise to the level of being much different in their architectural and structural aspects.

L286 Define z* or reference something to explain what it is

Added the following sentence: "Both z and z* formulations are utilized in the simulations with depth coordinates where the z* approach enables inclusion of large amplitude variations of the sea surface by distributing such changes over the full water column depth (Adcroft & Campin, 2004)."

L378 I think the dz/dsigma should be inside the dx since the thickness is normally integrated over x as well.

Corrected.

L405 You quote densities as eg 24 kg/m3 rather than 1024 kg/m3 – you need to make this point somewhere.

Since these are sigma densities, by definition they are in values / magnitudes given in the text. Nevertheless, this is now explained in the first paragraph of Section 2.5 where the sigma density is defined.

L465 It looks to me that the black line is plotting the density of the maximum transport rather than a timeseries of the maximum transport itself.

The black lines are indeed the density timeseries of the maximum positive transport. This has been clarified both in the text and the Fig. 3 caption.

L482 Do you mean annual mean of the maximum over depth of the monthly mean transports? It's a bit long, but at the moment it isn't clear what the maximum and average are taken over

These are the annual averages of the maximum transports based on the monthly mean transports in density space. This has been clarified.

L497 Is the weakening trend at OE also seen in most models? How does the overall trend compare to other studies eg https://doi.org/10.1038/s43017-022-00263-2 and references therein

Yes, this weakening trend is consistent with what is seen in other model simulations as well as with those from other estimates. In response, we have added a new paragraph which discusses the low-frequency variations in these transports. We now provide a quantitative trend comparison for the total transports with those trends reported in Jackson et al. (2022).

L558 The reference to 'high densities' seems to suggest relatively high densities for what we are interested in, however it looks like you're referring to densities where the transport is maximum here. Suggest rewording

The sentence is correct as it is written as it is about comparison of profiles for densities greater than 27.7 kg m^{-3} which includes the ranges of densities at which maximum transports occur.

L565 This suggest that in the observations the northwards transports have a narrower range of densities that the models. Is there anything which can be said about why that might be?

We do not think that this discrepancy necessarily suggests a narrower density range in the observations than in the simulations. Indeed, T-S diagrams for the upper 500 m (not shown) reveal largely comparable density ranges among the simulations and with OSNAP. As such, we think that this discrepancy is likely related to differences in current pathways and their T and S properties. We also think that possible under-sampling of the Greenland shelf waters in the observations can contribute. A sentence has been added to list these factors.

L632 What are the deeper/denser levels? Can you say anything about why this might be and the implications?

Quantitative values for "deeper" and "denser" were provided in the previous sentence. Specifically, deeper than 1500 m and denser than 36.75 kg m^{-3}. We do not think that much can be said about sources of these differences and their implications without more extensive analysis of the models which is beyond the scope of this study. We do not think that these differences are simply due to the representation of overflows. Since the main discrepancies arise from northward dense flows, deep gyre circulations need to be also considered.

L650 This isn't obvious to see from the plots – it looks like the scale is saturated for some plots. Could you do a quantitative comparison, eg comparing the mean of the north/south flows?

The requested comparisons of the mean (or summed) northward and southward flows / transports had already been provided as the thin red and blue lines, respectively, in the vertical profile panels both in depth (right panel) and density (top panel) space for each simulation and OSNAP.

L654 maximum aggregated transports

Added.

L705 Is there also a relationship of mean strength or monthly variability with annual/decadal variability?

We do not think that such an analysis will produce meaningful relationships on annual variability with only 8 years of data, which certainly exclude decadal variability.

L709 simulated ensemble mean SDs

Added.

L710 and elsewhere. The SD should have units of Sv I believe

Thank you for catching this error. The units have been added.

L732 clarity?

Deleted this part of the sentence.

L738 This sounds like you're saying that the densification is happening on the OSNAP section rather than that the upper/lighter waters are being densified to the north and then returning south. I think this needs a bit of clarification

No. It was not our intention to imply any driving mechanisms associated with these changes as these T-S diagrams cannot really identify mechanisms. We have added a sentence to clarify this. It reads "We note that these T-S diagrams do not necessarily identify mechanisms driving these T and S properties which can result from various processes both locally and remotely, including mixing and advection."

L743 Why might the models be also freshening as well as cooling unlike the obs? Could this be related to too much mixing in overflows? Or something else?

In general, models tend to be *spicy* in their T and S contributions to density. So, it is not surprising to see such behavior, which may arise due to various processes and their combinations. We think that imprints of overflows are more visible in the T-S diagrams at OE. Indeed, the last paragraph of this section discusses impacts of the Denmark Strait and Faroe Bank Channel overflows. We have decided not to include any additional discussions here as our follow-on paper will be looking into some of these in more detail, including a water mass transformation analysis.

L769 Note that UKMO25 captures the density though not T and S.

This information was already included just three lines above on line 766 of the original manuscript.

Fig 16 and others. It is hard to see the deep values – it looks like the deep values are plotted first and then the shallower ones later, and in some cases it looks like they are plotting over the deep values and obscuring them. It might make the figures clearer to plot the shallower data first and deeper data last.

We have checked the plotting package used for creating these figures. Indeed, we have confirmed that the order of plotting is from the shallowest first to the deepest last. We also note that are fewer deeper points due to coarser vertical resolution with depth and presence of bottom topography.

Section 8. There are a lot of points here which are not explored. There are other studies showing a stronger AMOC in more saline models eg

https://doi.org/10.1175/JCLI-D-22-0464.1 https://doi.org/10.5194/os-17-59-2021 https://doi.org/10.1016/j.ocemod.2013.10.005 So discussion about how this seems a robust feature would be welcomed. Also some more discussion about density compensation would be useful. Are there any thoughts about why there is density compensation in some places but not others? What about density compensation in the observations? This paper suggests that there is density compensation for OW https://doi.org/10.1038/s41561-019-0517-1 How does that compare with what you see? Could biases in the models compared with the observations affect the density compensation? Finally you could do with some motivation about why you're looking at Fig 18 as well as 17 and whether anything can be concluded from considering results together.

Following the reviewer's suggestion, we have expanded this discussion by adding two new paragraphs: one at the end of this section and another in Summary and Conclusions right after the related summary paragraph. Robust aspects are now clearly identified. We have also clarified the justifications for this section and the figures.

Fig 18 The caption is a little unclear. Is the y axis the total AMOC across both sections rather than OW and OE separately?

Yes, your understanding is correct. Since the "total" is already defined earlier and used elsewhere, we have not modified the caption.

L886 total velocities -> net transports

Done.

L900 Are there differences in biases with resolution?

No. The answer had been already provided in the last sentences of both paragraphs of Section 8. We have now added a sentence to the end of this paragraph as well.

L929-935 I found this rather confusing – are you always using mean as time-mean and max as max over density? Is the point that you're making here that the annual time mean of the maximum over density is not the same as the maximum over density of the annual time mean?

The answer to both questions is yes. Just to clarify the second question a bit: the annual time-mean of the maximum transports based on monthly mean transport profiles in density space are not the same as the maximum transports based on the annual time-mean of the transport profiles in density space. Our analysis uses both approaches following what is done in the literature. For example, while Fig. 4 uses the

former, Fig. 6 uses the latter approach. What we do is indicated wherever needed in the text. Regarding these particular line numbers referred to by the reviewer, this paragraph (lines 929-942) should be considered in its entirety as further clarifying information is included in the second half of this paragraph. Nevertheless, we made several edits in the first half of this paragraph towards clarification.

L1109 'the same numerical schemes and parameterizations' – but do they not have different ice models? This sentence implies that the overall models have similar set up other than resolution, but this isn't true. Maybe just needs a comment here.

Agreed. The sentence has been deleted.

Appendix B

It really isn't clear why the focus here is on two models only, particularly since the authors not that the differences are likely to be model dependent. What is the aim of this section? If it is to document differences then surely it would make sense to plot them all. Otherwise please explain the motivation/aim of the section.

We agree with the Reviewer's assessment. Accordingly, we have removed Appendix B. This also shortens the manuscript.