Response to Reviewer #2

We thank the reviewer for his 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.

This work provides a valuable benchmark for the modelling community for evaluating simulations at the OSNAP section. It draws on six different ocean GCMs and a wide range of model simulations. All models are consistent with the earlier findings of Lozier et al. (2019), showing that OSNAP East largely determines the strength and variability of the AMOC. This further supports the conclusion that convection in the Labrador Sea may not, as previously thought, be the primary driver of AMOC variability.

Although all models reinforce this central message, the model spread at OSNAP is, unsurprisingly, substantial. The analysis is comprehensive and highly valuable, and the paper is clearly structured and professionally written. The manuscript would be suitable for publication in GMD following the authors' responses to the minor comments listed below.

Thank you for your kind words, appreciating the relevance and value of the manuscript for the community.

line 55. worth mentioning here that they cancel each other at constant depth and density.

Done.

line 116: Do you have any citations for this? I am unsure whether this statement is correct. One can certainly say many things about coupled versus forced AMOC simulations. Xu et al. (2019), if I remember correctly, reported different observations in their paper 'On the variability of the Atlantic meridional overturning circulation transports in coupled CMIP5 simulations' (https://doi.org/10.1007/s00382-018-4529-0). Gent (2018) also discussed this issue in 'A commentary on the Atlantic meridional overturning circulation stability in climate models' (https://doi.org/10.1016/j.ocemod.2017.12.006).

The statement is indeed correct. It comes from comparisons of AMOC transport timeseries from forced simulations (e.g., Danabasoglu et al. 2014; 2016) to those participating in CMIP5 (e.g., Cheng et al. 2013) and CMIP6 (e.g., Weijer et. 2020 – supplemental Information). Figure 3 of Xu et al. (2019) also supports this statement for simulations participating in CMIP5. Gent (2018) is not directly relevant here. As requested, we have added several citations to this sentence, including Xu et al. (2019).

line 126: the findings by Ortega et al. 2012, 2021 were further confirmed using AWI-CM by Sidorenko et al. in AMOC variability and watermass transformations in the AWI climate model. Journal of Advances in Modeling Earth Systems, 13, e2021MS002582. https://doi.org/10.1029/2021MS002582

Added a citation to Sidorenko et al. (2021).

line 186: also in FESOM forced ocean by Sidorenko et al. (2020). AMOC, water mass transformations, and their responses to changing resolution in the Finite-volumE Sea ice-Ocean model. Journal of Advances in Modeling Earth Systems, 12, e2020MS002317. https://doi.org/10.1029/2020MS002317

Added a citation to Sidorenko et al. (2020).

line 188: I am a strong supporter of the work by Megann et al. (2021), which you cite above. Their interpretation is that Subpolar Mode Water formed in the northeastern Atlantic, which initially retains relatively high buoyancy, is advected into the Irminger Sea, where it loses buoyancy, and is subsequently transformed in the Labrador Sea—following further buoyancy loss—into Upper North Atlantic Deep Water before being exported southward. This view is consistent with McCartney and Talley (1982). Megann et al. (2021) also introduced a buoyancy-loss–ramped accumulation index that successfully reproduces the decadal variability of the AMOC, which could be a valuable approach to consider in the proposed follow-up study.

We concur with you that it would be valuable to consider the approach used in Megann et al. (2021) in a follow-up study.

line 353: How would you expect the results to change if instantaneous sampling were used to compute the AMOC in density space?

We have added the following text in the manuscript:

"Finally, we employ monthly mean fields for both OSNAP and model simulations as this temporal frequency is what is available from all model simulations. Compared to use of higher frequency sampling, e.g., instantaneous, in transport calculations if they were to be available, we do not expect our results to change in any significant way for the spatial scales considered here, supported by the findings of Ballarotta et al. (2013) for transport calculations in an eddy-permitting regime in the Southern Ocean."

line 590: If the transports within each bin are provided in Sv, I would suggest plotting the thin blue, red, and black lines in Figures 7–10 in a stepwise manner, since the transports are already aggregated within the respective bins (depth ranges).

We prefer to keep these figures the way they are already plotted. This is because any discretized variable, e.g., potential temperature, salinity, velocity components, can be interpreted as representing the spatial mean for a given grid cell in a model. Nevertheless, we conventionally plot all model fields not in a stepwise or block manner.

line 708: Just a question: to what extent would a more realistic simulation of the Arctic Ocean improve the representation of the OSNAP transports?

This is a difficult question to answer. While a more realistic simulation of the Arctic Ocean would produce transports and water masses that agree better with limited observations, properties of these waters will likely be impacted by surface fluxes and lateral and vertical mixing processes by the time they reach the OSNAP latitudes. A forced ocean experiment in which a model's potential temperature and salinity fields are restored to their (limited) observed counterparts may be helpful to shed light on this question.

line 785: Here you use the depth range above 700 m, whereas in the section above you considered depths deeper than 500 m for the T–S diagrams, in order to exclude the larger surface biases. How would Figures 17 and 18 change if deeper biases were considered instead?

Section 7 (original Figs. 13-16) and Section 8 (original Figs. 17 and 18) serve different, but complimentary purposes. Specifically, Section 7 is about deep and abyssal ocean water mass properties. In contrast, Section 8 concerns the upper-ocean biases as they directly impact deep convection in model simulations. So, deeper biases should not be included in Figs. 17 and 18. In response to both this comment and another one by Reviewer #1, we have modified a couple of sentences in this section to clarify its purpose, including the very first sentence.

line 834: More generally, in the context of model bias analysis, this suggests that the AMOC may transport water masses with similar (and less biased) densities that nonetheless occupy different regions of T–S space. This behaviour appears to be common across many regions of the global conveyor belt.

Original Fig. 18 supports this summary for our analysis region.

lines 874–876: Here, $\sigma 0$ should also be explicitly mentioned and a reference to Fig. 6 added; otherwise, the reader may become confused.

Done.

line 886: The same correction as in the abstract applies here: both flows cancel each other at constant depth and density.

Done.

I wonder whether a figure similar to Fig. 2 in Lozier et al. (2019) would show comparable behavior across the models. Such a comparison could potentially serve as a useful benchmark. The same applies to their Fig. 3. Some brief discussion of this point would be nice to have.

Thank you for this suggestion. We have added a new figure (Fig. 7) along with its discussion in Section 4. The figure reproduces the top panel of Fig. 2 from Lozier et al. (2019) for all simulations. Indeed, it provides a useful benchmark, showing that the HR simulations show better agreement with the OSNAP observations than those of MR and LR. We decided not to include an additional figure akin to that of Fig. 3 of Lozier et al. (2019), because similar information is already provided in Fig. 4. Furthermore, since these simulations use the same wind product, the Ekman components would be rather similar among the simulations.