We thank Referee #2 for their encouraging statement and the constructive comments that improved our manuscript. Please find our full response to the comments, and the changes we have made to address these comments, in the attached pdf file. Referee #1 and #2 similarly asked for the clarification of the terms diapycnal mass flux as well as a more in detail explanation for the differences in water mass definitions in observations and an ocean model. We thank you for these remarks, as the explanations were enhanced through this revision.

In the following the review is copied in black with our comments and answers in blue just below each remark.

Review on “Major sources of North Atlantic Deep Water in the subpolar North Atlantic from Lagrangian analyses in a high-resolution ocean model” by Fröhle et al.

Summary

Fröhle et al. investigated the major sources of the North Atlantic Deep Water transports at the southern exit of the Labrador Sea using Lagrangian particle experiments in a high-resolution ocean model. They quantified contributions from different processes, including diapycnal fluxes (in and out of the mixed layer) and overflow from the Greenland-Iceland-Scotland ridge, to the total deep water transports. For each source, the associated pathways and transit times were discussed.

Overall, I find results from the study quite interesting as they show, in a model, what the subpolar deep western boundary current is composed of from a Lagrangian perspective. The manuscript was overall clearly written and the particle experiments were reasonably designed, supporting the major conclusions. Therefore, I recommend this manuscript for publication after addressing the following comments.

Major comments

[1]. The method sections, especially section 2.3, are very dense and hard to understand without looking at the results. Could they be merged into the Results section, along with where the results/figures are presented?

Thank you for this valuable comment. We adapted your proposition in the following way: The subparagraphs are now highlighted with fat letters for the key analysis techniques to enhance the readability and guide and make it easier to find certain techniques in the manuscript. We will keep the separation between methods and results in order to stick to a clear structure of the document.

[2]. Differences between current study and previous literature need better explanations. In the Discussion section, it was mentioned that the LSW/INADW transport ratio differed from 53N array observations because of a different water mass definition in this study. But I am having a hard time understanding the explanations. I thought the LSW and INADW layers were defined using a fixed isopycnal 27.86 kg/m^3. In this sense, shouldn't the LSW (26.7 Sv) and INADW (3.4 Sv) transports derived from the particles equal the Eulerian transports in the corresponding density layer? Does this large transport ratio suggest a model bias in simulating the overflow waters? A relevant comment is that the authors claimed the 5.7 Sv NADW from the Greenland-Iceland- Scotland ridge was consistent with overflow observations of 6 Sv. However, the NADW transport in the study is mostly contained in LSW layer, whereas the INADW transport, which should be used to compare with the overflow observations, is as small as 0.6 Sv.

The LSW / INADW boundary is defined based on the hydrography in the Labrador Sea, assuming that LSW is majorly formed within the Labrador Sea. However, in the presented experiment it is revealed that (at least in the model) LSW is not solely constituted of water masses formed within the Labrador Sea, but there are contributions from different sources and processes. Thus, the used definition is in this context probably insufficient. This has been discussed in the Discussion section.

To better avoid confusion, a paragraph about this problem has been added in the methods section, where the boundary isopycnals are introduced. “ The water mass boundaries are defined as the mean density value over the complete model output, covering 1958 through 2019. Contrary to the dynamically defined upper bound of NADW, the definition of the boundary between LSW and INADW is based on the hydrography in the central Labrador Sea [Handmann et al. 2018]. Even though this method works fine with observations and yields the distinguished densities of the three NADW water masses, we show in this study, this does not necessarily hold for a water mass distinction in the classical sense in an ocean model. This is partly related to the unrealistically large diapycnal mixing in regions where dense waters descend topographic slopes, producing lighter water Willebrand et al. (2001). This spurious mixing is dependent on the vertical and horizontal resolution of the ocean model and is a typical model artifact.”

Furthermore, within the results section, the terms LSW and INADW are replaced by the corresponding density ranges. The terms themselves are discussed in the Discussion section, where the corresponding section has been moved to the beginning of the Discussion to emphasize the importance of this finding.

As for the overflow transports, the Eulerian estimates reported in this study are given at the sills themselves, not at 53°N. The estimate for the Lagrangian experiment presented is based on the transport assigned to each particle. As explained in the manuscript, the transport is conserved along the particle trajectories. Thus, particles amounting to 5.7 Sv cross the GSR and reach 53°N. The volume is not changed, but due to mixing the density decreases and the particles mostly end up in the LSW layer at 53°N. However, at the GSR itself, the estimate based on the Lagrangian analysis presented and the Eulerian estimates reported from other studies [Biastoch et al. 2021] are comparable.

[3]. Finally, I am trying to understand the consistency/difference between diapycnal mass flux inferred from the Lagrangian particles and the classical diapycnal water mass transformation from a Eulerian perspective. I guess the two cannot be compared directly but they should be ultimately linked. For example, observations show that 7 Sv of lighter waters are transformed into denser layers by surface buoyancy loss in the Irminger and Iceland basins, as reported by Petit et al. (2020). Some of these waters might travel across the gyre and reach the boundary current at 53N, which will be counted as part of the diapycnal mass flux discussed in this study. This was only briefly mentioned in the Discussion. I suggest the authors to elaborate a little bit more.

The water transformed to NADW density at the surface counts as part of the ML_P class. Only particles that are transferred below the mixed layer from the upper AMOC component to the NADW component count as diapycnal transfer in this study. This was specifically clarified in the section 2.1.3 “Categorisation of particles”: “If particles increase their density during the experiment, from $\sigma_0 < \sigma_{DW}$ to $\sigma_0 \geq \sigma_{DW}$, outside of the mixed layer before reaching 53°N, without contact to the atmosphere, this is referred to as diapycnal mass flux and the particles are classified as DIA_P. Else, if the respective density increase occurs within the mixed layer, with contact to the atmosphere, the particles are classified as ML_P.”. We also state a short version of this in the abstract : “Our experiment shows that, of the 30.1 Sv of NADW passing 53° N on average, the majority is associated with diapycnal mass flux without contact to the atmosphere…. ”.

Additionally, Petit et al. (2020) focus on the recent period, where we average over the period 2010-2019.

Minor comments

[1]. Lines 3-4: Here you mentioned NEADW and DSOW. However, in the manuscript, the sources of NEADW and DSOW transports were not explicitly distinguished and discussed.

Thank you for this comment. As we take all NADW defined as densities larger than the AMOC density at OSNAP at 53°N to investigate the sources. The results show that we cannot define the origins of LSW, NEADW nor DSOW by solely density in the model. We deleted this separation from the abstract since we do not use it later on. We mention this reasoning also in the Discussion (subsection 4.1) : “To conclude, we find that the density interval with the major transports in NADW at 53°N around σ_0 = 27.80 kg m^{-3} is not only associated with one source. Instead multiple sources contribute with different relative importance to similar density regimes, though the DIA_P and ML_P dominate .”

[2]. Line 8: It is better to first report the total transport at 53N in the model, i.e. 30 Sv, before quantifying different sources. Also, please specify in the Abstract that "diapycnal mass flux" refers to the diapycnal flux in the boundary.

The entire sentence in the abstract is changed to : "Our experiments show that, of the 30.1 Sv of NADW passing 53° N, the majority is associated with diapycnal mass flux without contact to the atmosphere, accounting for 14.3 Sv (48%), where 6.2 Sv originate from the Labrador Sea, compared to 4.7 Sv from the Irminger Sea."

[3]. Line 19: "a net downwelling *in density space* of upper AMOC water"

Done

[4]. Lines 66-68: I am not sure if I understand this long sentence. What do you mean by "adding transformed water to a major volume of water…"?

Sentence changed to "Newer research has shown that a major volume of water is transformed along the North Atlantic Current path [Desbruyeres et al. 2019]. This water originates from different transformation processes, which are related to different export time scales [Le Bras et al. 2020]. Hence, the very localized deep convection might only be adding a comparably small amount of transformed water to the NADW overall volume."

[5]. Equation (1): What is "ceil"?

Replaced by mathematic symbols for the ceiling function. And added comment under formula to explain.

[6]. Lines 122-126: So the water mass definitions are based on mean density, but the particle release density varies on daily time scales, correct?

Yes; reformulated the corresponding passage ("The particles are subsampled based on their density at their respective release, i.e. only particles with densities $\sigma_0 \geq \sigma_{DW}$ are considered, resulting in a subset of particles.").

Additionally added the sentence "Contrary to the dynamically defined upper bound of NADW, the definition of the boundary between LSW and lNADW is based on the hydrography and, as shown in this study, does not necessarily hold for a water mass distinction in the classical sense." for clarification of the water mass definitions.

[7]. Line 130: What do you mean by “the same advection time”?

For clarification the sentence was changed to: “For each particle, the trajectory is considered only between the particle's origin, described in detail in the following, and 53°N. Resulting from the definition of the point of origin, the trajectories have varying lengths. In turn these are consequently related to varying transit times.”

[8]. Line 172: Are signs or flow directions considered for the cumulative transport? If particles flow into the bin from different directions, then the cumulative transport should be zero.

Probability density maps are usually computed following the two possibilities described in Van Sebille et al. 2017. Here, however, the probability density map is weighted with the volume associated with each particle. Particles either pass a box or not. The paragraph has been revised.

[9]. Line 179: Is this binned transport (based on point of origin) also converted to the relative transport with respect to the 53N section? Again, are flow directions considered in the binning?

For the point of origin the binned transport is not converted to relative numbers. Thus, integrating over all grid cells yields the total volume transport at 53°N associated with DIA_P and ML_P, respectively. Since only a fixed position is used to create the maps, flow direction is not considered. The paragraph has been revised.

[10]. Lines 191-193: I do not understand how the “volumetric water mass transformations” are calculated here. Please elaborate.

Clarified the explanation paragraph to : “To obtain the ***volumetric water mass transformations*** the particles are grouped by their water mass properties at 53°N and at their point of origin. The considered properties are σ_0, absolute salinity S_A and conservative temperature (Θ), with bin sizes of 0.025 kg m^{-3}, 0.01 g kg^{-1} and 0.2° C, respectively. These properties were computed from the practical salinity, potential temperature and depth tracked along each trajectory using the TEOS-10 toolbox for Python [TEOS10_2015]. The difference between the volume at 53°N and the volume at the point of origin for σ_0, S_A and Θ class then gives the ***volumetric water mass transformation***.”

[11]. Lines 254: The 5.7 Sv of NADW from the ridge is mostly in LSW layer. I am not sure why the authors compare this number with the overflow transport observations. Instead, it is the lNADW transport (as small as 0.6 Sv) that should be compared with overflow observations (6 Sv). Please also see my major comment [2].

The point of this paper is to show that the classical way to define water masses, based on density intervals, does not hold like in observations in this ocean model (and most probably in most other ocean models). Yes, the major part of NADW passing the GSR is in the lighter region though it is still the overflow. Hence we compare it with the overflow from observations - which density vise is defined differently but dynamically comparable to what we have in the model.