Thank you very much for your positive comments and constructive feedback, you addressed some important points. Your clarifications helped to make the manuscript clearer for the reader. Our responses are provided in green (changes made in the manuscript are written in ***bold***) together with your original comments in black.

We really appreciate your time and insight in reviewing our manuscript!

Kind regards,
Susanna (on behalf of all co-authors)

Reviewer #2

Major Comments

1. As I understand it, the authors use the OSNAP data as gold standard for the evaluation of OHT in different reanalyses. However, they repeatedly conclude that most of the differences could be attributed to a lack of observations at specific depths and areas along the section, as OSNAP uses constant fields in the interior and relies on end-point dynamic height moorings to capture the total integrated transport and its variability (e.g. of conclusion at l.213, l.229, l.233, l.289, l.295, l.316, l.394, l.411). In this context, can the authors comment on the choice of OSNAP data as a gold standard to evaluate these reanalyses while there are not enough OSNAP observations in the interior to conclude anything on the possible causes for the total OHT differences? In a way, the reanalyses might assimilate more observations in the interior than the OSNAP product, meaning that their OHT distribution could be considered as closer to the truth than the one from the OSNAP product. Hence, how can we reliably assess inconsistencies between reanalyses and OSNAP?

   We appreciate this thoughtful comment. We would like to clarify that we do not intend to use OSNAP as a gold standard in the sense of providing a definitive "truth". Rather, it represents the most comprehensive direct observational estimate of heat transport across the subpolar North Atlantic currently available. Our analysis is therefore framed as a **comparison** (not a validation) between two fundamentally different approaches: direct observations with incomplete spatial coverage, and reanalyses that assimilate a range of datasets but rely on imperfect model dynamics and parameterizations. The purpose is not primarily to determine which product is closer to the truth, but to identify where and why these approaches diverge, thereby highlighting regions and processes that remain uncertain.

2. A specific example of my main comment #1 is the discussion of an anticyclonic eddy in the NAC region at lines 309–317. Can the authors clarify their conclusion: is the inconsistency coming from the resolution of the reanalyses at the boundaries or from a lack of observations in the available OSNAP dataset? If

these two points are valid, how a validation of one dataset as compared to the other can be convincing? Related to this specific point, I am not sure to understand why the anticyclonic eddy can be observed in the glider data (as discussed in Lozier et al., 2017) but not in the final OSNAP product?

As clarified in our response to Comment #1, we do not use one product to validate the other. Rather, our analysis is framed as a comparison between two complementary approaches in order to identify where and why they diverge.

We emphasize that the paragraph at lines 309–317 refers to two distinct circulation features that serve different purposes in the discussion:

Lines 301–312 refer to a narrow, intensified northward current anomaly near Hatton Bank in 2015, observed in the **eastern** OSNAP glider region. This feature is directly linked to the 2015 heat transport peak and reflects a localized strengthening of the NAC that is more pronounced in the OSNAP observations than in any of the reanalyses.
The source of the inconsistency is difficult to attribute uniquely to either reanalysis resolution or OSNAP sampling with the information available.

Lines 313–317, by contrast, describe a separate anticyclonic eddy in the **western** glider region, which is explicitly stated to be unrelated to the 2015 heat transport peak. This second example is included only to illustrate the strong mesoscale variability in the glider regions and their sensitivity to both (i) model resolution/representation and (ii) observational sampling and product construction. **In this case, the inconsistency arises primarily from the construction of the gridded OSNAP velocity product rather than from limitations in reanalysis resolution:** This anticyclonic eddy discussed by Lozier et al. (2017) was observed by a glider transect between June and November 2015 with distinct temperature and salinity characteristics. However, the OSNAP velocity field (as used here) is primarily determined by moorings spaced hundreds of kilometers apart and does not directly incorporate the glider-derived velocities. As a result, the OSNAP velocity field is not designed to resolve fine horizontal structures such as eddies located **between** moorings. E.g., in the region where the eddy was identified, the velocity reconstruction relies on the M3 and M4 dynamic height moorings to estimate geostrophic shear, which cannot represent the horizontal structure between them (see Fig. 9 in Lozier et al., 2017). Consequently, while the eddy is evident in the glider observations it is not retained in the monthly gridded OSNAP velocity product used here. The hydrographic property fields, by contrast, show broadly consistent behavior between the OSNAP product used here and the one from Lozier et al. (2017).

We edited L316 to make sure that this statement only applies to the velocity field, not the property field:

However, no such signal is evident in the ***velocity field of the publicly available OSNAP dataset used in this study, which is primarily determined using mooring observations that do not resolve the small-scale spatial structures such as eddies between moorings.***

We also clarified the wording in this section to make it clearer that we are talking about 2 separate events:

As a separate example of mesoscale variability (unrelated to the 2015 heat transport peak), a strong anticyclonic eddy is also present in the western glider region*, clearly visible in GLORYS12V1 and, to a lesser extent, in GLORYS2V4.*

3. To add some clarity in the differences between the data used in this study, I suggest changing the structure of section 2 by adding a 'section 2.1 Data' that would introduce the data used in this study: OSNAP (including lines 101–106 currently in the following section), the reanalyses and the altimetry data. Sections 2.1 would become section 2.2 etc..

   We have restructured Section 2 accordingly by introducing a dedicated **Section 2.1 (Data description)**, which now explicitly introduces the observational OSNAP dataset, the ocean reanalyses, and the satellite altimetry data used in this study.

4. In the new section 2.1, I strongly recommend the authors discussing the differences and similarities between the reanalyses in terms of observations assimilated in these reanalyses. For example, can the authors clarify if the reanalyses assimilate OSNAP observations? Are they assimilating the same observations otherwise (e.g., Argo, altimetry, hydrographic sections…) meaning that their differences in OHT (or for example the results discussed at l.239-241) can be interpreted as a result of different horizontal resolutions and dynamical models only?

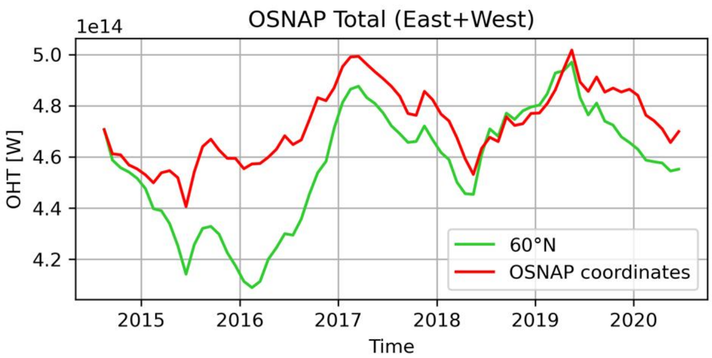
   We have expanded Section 2.1 to clarify both the similarities and differences between the reanalyses in terms of assimilated observations. We now explicitly state that none of the reanalyses assimilate direct velocity observations from the OSNAP array, or from current meters and ADCPs more generally. While temperature and salinity measurements from OSNAP may indirectly contribute via global in situ databases (which is difficult to verify), the OSNAP velocity field itself is not assimilated.

   We clarify that all reanalyses assimilate broadly similar observations, including in situ temperature and salinity profiles and sea level anomalies, which constrain the large-scale geostrophic circulation. However, they differ in horizontal resolution, data assimilation schemes and in the spatial domain over which sea level anomalies are assimilated (e.g., ORAS5 versus the other products). To improve transparency, we have added a new table (Table 1) summarizing the assimilated data and assimilation schemes for each reanalysis. Therefore, differences in ocean heat transport among the reanalyses reflect a combination of model dynamics, resolution, and data assimilation methodology.

5. Finally, I recommend the authors discussing another OHT dataset produced by combining Argo, altimetry and gravimetry data from Calafat et al., 2025. https://doi.org/10.5194/os-21-2743-2025

We thank the reviewer for pointing us to this recently published and valuable dataset. The ocean heat transport estimates of Calafat et al. (2025) provide an important independent perspective on large-scale Atlantic heat transport variability.

However, this product provides heat transport estimates across complete latitude bands, whereas our study focuses on transport across the specific OSNAP section and its eastern and western components separately. Using the Calafat et al. dataset would therefore require approximating OSNAP by a zonal section, which we find introduces non-negligible differences compared to transports calculated along the actual OSNAP geometry (see figure below). In addition, the Calafat et al. product is provided at 3-monthly resolution and does not allow for a separation between OSNAP East and West, whereas the 2015 anomaly discussed here is most pronounced in OSNAP East.



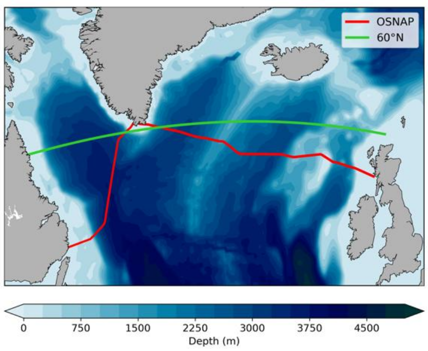


Fig.: Comparison of OHT calculated at the exact OSNAP coordinates (red) and across 60°N (green).

For these reasons, we do not include this dataset in the quantitative analysis, but we now cite and briefly discuss it in the manuscript as an important complementary product for basin-scale heat transport assessments. We added the following in the manuscript:

Calafat et al. (2025) present a novel Atlantic ocean heat transport dataset derived from a combination of Argo temperature profiles, satellite altimetry and gravimetric constraints. This product provides valuable insight into basin-scale heat transport variability across complete latitude bands. However, because it is defined along zonal sections and provided at 3-monthly resolution, it cannot be directly applied to the OSNAP section geometry or used to distinguish between OSNAP East and West, which is central to the present analysis. Nevertheless, such approaches offer an important complementary perspective on large-scale heat transport variability and are highly valuable for basin-scale assessments.

Minor Comments

L. 62: use AMOC instead of MOC, as it was the authors' choice for the rest of the manuscript

Done

L. 108: consider using the term 'derived' instead of 'calculated'

Done

L. 114-115: Consider clarifying the grid that is used for the bilinear interpolation.

We have clarified the description of the interpolation:

To assess cross-sectional biases and RMSE values between OSNAP and reanalyses, all reanalysis sections are interpolated bilinearly ***in the along-section and vertical directions onto the common OSNAP gridded section, defined by the OSNAP "along-section" distance coordinate and depth levels.***

L. 128-130: The sentences don't read properly. Maybe: 'Additionally, the potential temperature and a reference temperature are needed for estimating the heat transport. An unambiguous heat transports require closed volume transports, which is [...].'

To improve readability the sentence was changed to:

Additionally, the potential temperature θ and a reference temperature θ_ref are needed for estimating heat transports.

L. 133: Can the authors describe in few sentences what is the StraitFlux's line integration method?

We have added a brief description of StraitFlux's line integration method:
To avoid interpolation and preserve numerical conservation properties, net integrated transports from the reanalyses are calculated using StraitFlux's line-integration method,

in which transports are integrated directly along the native model grid cell faces that intersect the section, thereby approximating the target section as closely as possible on the native grid.

L. 198: Typo ‘Iceland basin’

Corrected

L. 203-205: EN4 includes a large number of Argo profiles and has probably a better spatial coverage over the subpolar North Atlantic than OSNAP that is missing observations in the basin interiors. However, there are possible issues of data QC in EN4. Consider also discussing in more details what structural and methodological uncertainties in OSNAP can explain these differences. From my understanding, OSNAP uses EN4 in the interior?

OSNAP does not use EN4 data directly in the basin interior. While OSNAP and EN4 may draw from overlapping in situ measurements (e.g. Argo profiles), they differ fundamentally in methodology: EN4 is an objective analysis with spatial smoothing, whereas OSNAP temperatures are derived from its dedicated observing system along the section and its own gridding methodology. Differences between the two likely reflect both EN4 mapping/QC choices and OSNAP uncertainties associated with sparse interior sampling and section construction.

L. 270: Related to my main comment #4, I recommend the authors to clarify in the new section 2.1 if they use independent data sources in the reanalysis.

At line 270, the term “independent data sources” refers to the multiple, distinct datasets used in the indirect (budget-based) heat transport estimates (listed in Table 1), including different atmospheric reanalyses, ocean heat content products, and sea-ice datasets, rather than data within the ocean reanalyses. We have slightly rephrased the sentence:

This holds true across all combinations of datasets and both choke-point approaches, despite the use of ***multiple, independent data sources contributing to the indirect heat transport estimates (see Table 1).***

L. 290-292: I am confused, didn’t the authors say that OSNAP cannot represent the circulation over this portion of the interior array because there aren’t enough observations there?

We have clarified the text to explain that although the detailed circulation structure in the Iceland Basin is not resolved by OSNAP, the opposing transport branches largely compensate, resulting in consistent net accumulated transports between reanalyses and observations.

This northward–southward circulation feature is not resolved by the OSNAP observations due to OSNAP’s array design that relies on end-point dynamic height *moorings to capture the total integrated transport and its variability in the Iceland Basin.* ***However, because the opposing branches largely compensate each other, the net accumulated heat transport over this segment is consistent between the reanalyses and the OSNAP observations (Fig. 7d).***

L. 295: Consider clarifying if the potential uncertainties in these areas relates to uncertainties in the reanalyses or OSNAP observations?

We have clarified the text:

More broadly, the reanalyses exhibit high variability (Fig. 7c) in regions that lack direct mooring ***observations in OSNAP, underscoring potential uncertainties in the OSNAP transport estimates in these areas.***

L.320: Is the barotropic compensation applied at OSNAP uniform in time or in space (horizontally and vertically) or both? How can it impact the correlation of OSNAP with SLA?

In OSNAP, the barotropic compensation is applied as a spatially uniform (horizontally and vertically throughout the whole section) but time-varying velocity offset to close the net transport budget at each time step. This spatially uniform compensating velocity is not expected to show tight pointwise correlations with SLA within this narrow glider section.

We have clarified this in the revised manuscript:

We note a key methodological difference: OSNAP derives time-varying geostrophic shear from T/S and applies a ***spatially uniform, but time-varying barotropic compensation to obtain absolute velocities****, whereas the reanalyses assimilate SLA and thus include time-varying surface geostrophic flow tied to SLA. Accordingly, the SLA -transport comparison below is intended as a diagnostic of reanalysis consistency and as context for OSNAP.* ***In this framework, tight pointwise correlations between SLA gradients and OSNAP transport are not expected.*** *Nevertheless, because along-section SLA gradients set the surface geostrophic shear, we'd expect a coherent relationship after spatial averaging over broader segments (e.g., the glider regions).*

L. 338-346: Related to my main comment #4, it would be easier to interpret this result with more details on the dataset assimilated by the different reanalyses. Do these two reanalyses (GLORYS2v4 and GLORYS12V1) assimilate the same SLA data? Why not showing the results for the other reanalyses?

GLORYS2V4 and GLORYS12V1 both assimilate multi-mission satellite altimetry derived sea level anomalies from CMEMS, but differ in horizontal resolution and assimilation methodology. We have clarified this in the text. We focus on these two products to highlight the impact of SLA assimilation and resolution on SLA-transport correlations. Other reanalyses are either not directly comparable in this context (e.g., ORAS5 does

not assimilate SLA north of 50°N) or show similar correlations and are omitted for clarity, as including all products would detract from the main focus of the study.
Revised paragraph:

In contrast, GLORYS12V1 and GLORYS2V4 exhibit higher correlations with observed SLA gradients (0.40 and 0.39, respectively) and even larger values when compared to their own SLA fields (0.79 and 0.51). ***Both products assimilate the same multi-mission satellite altimetry-derived SLA observations but differ in horizontal resolution and assimilation methodology (see Table 1).*** *Notably, the higher-resolution GLORYS12V1 shows the strongest correlations overall, consistent with its improved spatial representation of circulation features. Correlations are weaker for the ORAS5 reanalysis (not shown), with a correlation of just 0.25 against observed SLA gradients, likely a result of it not assimilating sea level anomalies north of 50°N.*

L. 347-355: Can the authors clarify why smooth all fields over 1deg resolution while the coarser resolution from reanalysis or altimetry is 1/4deg? In my view, only GLORYS12V1 should be smoothed at 1/4deg.

We thank the reviewer for raising this point. We realize that the original wording was misleading. The 1° smoothing is **not applied to the data prior to the calculation of correlations**. All SLA-transport correlations are computed on the native grids/sections of the respective datasets.

The 1° smoothing is applied **only to the along-section correlation curves shown in Fig. 11**, purely for visualization purposes, in order to reduce small-scale noise and improve readability of the plotted results. The reported correlation values are based on the unsmoothed fields.

We have revised the manuscript to clarify this distinction and avoid confusion regarding the role of smoothing:

To reduce small-scale noise and facilitate visualization, the along-section correlation curves shown in Fig. 11 are smoothed to 1° resolution. This smoothing is applied only for plotting purposes and does not affect the calculation of the correlations themselves, which are performed on the native-resolution fields along the OSNAP section*. This allows us to focus on mesoscale dynamics while suppressing small-scale variability and potential sampling mismatches. Despite this visual smoothing, pointwise correlations across the section remain relatively low for the OSNAP dataset.*