Response letter

We thank the reviewers for their valuable and constructive comments. The reviewers have raised important points that provide us with an opportunity to clarify a number of aspects relating to our results and the Bayesian model. In the following, we explain how the manuscript has been changed and provide a point-by-point response to each of the reviewers' comments. For clarity, our responses are in standard font whereas any text from the reviews is denoted by blue font. Text from the manuscript is italicized.

The main changes made to the manuscript are outlined below:

1. We use a new observation-based surface heat flux data set and focus our analysis on the Bayesian estimates derived from this new data set, in response to a comment by Reviewer #2. Accordingly, we no longer include Bayesian estimates based on ERA5 and NCEP data.
2. We now account for errors in the ocean mass from GRACE due to the GIA correction, geocentre motion and Earth oblateness.
3. We provide a clearer justification for our assumption of zero MHT at 65ºN, in response to a comment by Reviewer #1.

Referee #2:

This study tackles the challenge of estimating variations in Atlantic meridional heat transport (MHT) using satellite data and in situ temperature and salinity profiles across 12 latitudinal cross-sections. The authors propose a method to estimate MHT by combining changes in ocean heat content (OHC) with surface heat flux (HF) data. Recognizing the limitations imposed by sparse hydrographic observations, the study advances traditional methods by integrating hydrographic data with satellite altimetry and gravimetry within a joint spatiotemporal Bayesian framework. This fusion enables the generation of probabilistic MHT estimates from 2004 to 2020 across 12 Atlantic latitudinal sections, from 65°N to 35°S. The methodology effectively leverages the comprehensive spatial coverage of satellite data to compensate for the uneven distribution of in situ observations, thereby improving the quality of MHT estimates. Validation against independent measurements from the RAPID array at 26°N (which were not used in the derivation of the MHT estimates) shows good agreement in both the magnitude and timing of variability (with correlations of 0.77 for the raw series and 0.93 for the smoothed series), as well as in the mean transport value (1.17 PW). Results at other latitudes are consistent with prior estimates.

This work addresses the critical issue of variability in Atlantic meridional heat transport—a key component of global and regional ocean heat transport influencing climate. Continuous and precise measurements of Atlantic MHT are essential but limited by the high costs and logistical challenges associated with direct ocean observation systems, which currently provide data at only a few latitudes (e.g., through the RAPID and OSNAP arrays). The authors manage to overcome this limitation here by infering ocean heat transport convergence (HTC) as a residual from the imbalance between OHC changes and surface heat fluxes, using all available data to estimate OHC (i.e., satellite altimetry, gravimetry, and hydrographic observations) within a Bayesian framework. As such, this study is original and highly relevant to the climate science community. The ocean energy budget approach used to derive HTC is not new, nor is the combination of satellite altimetry, gravimetry, and hydrography to estimate OHC. The novelty of this study lies in its application of a Bayesian statistical framework that explicitly accounts for uncertainties in each dataset.

While the overall approach is sound and the results are promising, there are several significant limitations in the current version of the manuscript that must be addressed for the study to be fully convincing.

We thank the reviewer for their valuable and constructive feedback, which has helped us to improve the manuscript.

The reviewer writes that “the novelty of this study lies in its application of a Bayesian statistical framework that explicitly accounts for uncertainties in each dataset.”. We appreciate the reviewer’s recognition of the Bayesian framework used in our study, but we respectfully note that this comment does not fully capture the novelty of our approach. The core innovation lies not only in accounting for uncertainty in each dataset but in how we integrate multiple data sources (hydrographic, altimetric, and gravimetric) within a joint spatiotemporal Bayesian hierarchical model. This framework allows us to account not just for individual uncertainties but also for spatial and temporal dependencies within and across variables, something that previous studies focused on ocean heat content have not done.

Earlier approaches have typically combined hydrographic and satellite data in a pointwise manner, treating each dataset independently and often ignoring both error structures and spatial correlations. By contrast, our approach enables sharing of information across space, time, and data types in a statistically rigorous way. This leads to more robust and spatially coherent estimates of ocean heat content, heat transport convergence, and ultimately meridional heat transport, with improved quantification of uncertainty.

Major Concerns

Surface Heat Flux (HF) Datasets:

The datasets used to estimate HF are not state-of-the-art. The authors rely on outputs from atmospheric reanalyses, whose surface flux estimates are known to suffer from inconsistencies and large biases due to the weak observational constraints on the short-term forecasts used to generate them. A more robust alternative involves estimating surface fluxes from the atmospheric energy budget using CERES observations at the top of the atmosphere (TOA) and computing the divergence of atmospheric energy transport from reanalysis fields (e.g., winds and temperature), which are more strongly constrained through data assimilation than the short-term forecasts. This approach has been adopted by Mayer et al. (2017, 2021, 2022, 2024) and Meyssignac et al. (2024), and is now widely accepted as yielding net surface fluxes with smaller large-scale biases than reanalysis output-based or satellite-derived model outputs. The authors are strongly encouraged to apply this method, which would substantially increase the reliability of their MHT estimates.

We thank the reviewer for this comment. We would like to note first that, to the best of our knowledge, the DEEP-C product that we used in the original submission was derived using the approach that the reviewer recommends. However, DEEP-C stops in 2017, which is not ideal. For this reason, and following the reviewer’s comment, we are now using a new surface heat flux product derived using the approach recommended by the reviewer. Specifically, the net surface HF is calculated by combining top-of-the-atmosphere (TOA) radiative flux with the divergence of vertical integral of total energy flux and the tendency of vertical integral of total energy as described in Mayer et al. (2022). The TOA flux has been obtained from the Clouds and the Earth’s Radiant Energy System–Energy Balanced and Filled (CERES-EBAF) Edition-4.2.1 monthly data product (Loeb et al., 2018), whereas the vertically integrated atmospheric energy quantities are those of Mayer et al. (2022).

Note also, as mentioned earlier in this response letter, that the revised manuscript focuses on estimates of MHT derived using this new surface heat flux product. That is, we no longer include Bayesian estimates based on ERA5 and NCEP data. The ERA5 and NCEP heat flux products are used only to obtain a measure of uncertainty in the surface heat flux data; they are not used directly in the BHM.

Uncertainty Estimation in GRACE Data:

Given the central role of uncertainty quantification in this study, it is concerning that the uncertainty associated with space gravimetry data is only partially addressed. The authors rely on uncertainties from the mascon product, which do not account for critical error sources in GRACE data, such as the glacial isostatic adjustment (GIA), geocenter motion, and C20 corrections. These components are known to dominate the error budget in ocean mass estimates from GRACE (see Quinn & Ponte 2010; Blazquez et al. 2018; Uebbing et al. 2019) and significantly impact thermal expansion estimates. The authors should incorporate these additional uncertainty sources into their analysis.

We thank the reviewer for this comment. As mentioned at the beginning of this letter, we now account for errors in the ocean mass from GRACE due to the GIA correction, geocentre motion and Earth oblateness.

Thermosteric Sea Level (TS) and OHC Relationship:

The assumed relationship between thermosteric sea level and vertically integrated ocean heat content in the process layer involves important simplifications. Specifically, the neglect of the deep ocean (below 1500 m) and the linearity assumption between TS and vertically integrated OHC (see Eq. 17) could introduce significant inconsistencies. These approximations should be explicitly discussed and, if possible, their impact quantified.

Thank you for this comment. First, we would like to clarify that the contribution from the deep ocean (below 1500 m) to thermosteric sea level (TS) is not neglected in our analysis. As described in the manuscript, we account for the full-depth contribution through the relationship SL = TS + HS + OM, where the sea level (SL) from satellite altimetry reflects the total steric signal, including from the deep ocean. We accommodate the potential deep-ocean contribution by inflating the uncertainties in the TS and HS data by 20%. This allows the Bayesian hierarchical model to reconcile the observed altimetric sea level with full-depth TS and HS contributions, including those below 1500 m.

Regarding the second point raised by the reviewer, we agree that, because the thermal expansion coefficient (α) varies with depth, assuming a constant α to relate vertically integrated heat transport convergence to changes in TS is an approximation. However, this simplification is necessary because modelling the relationship exactly would require knowledge of the vertical structure of the heat transport convergence (or the velocity and temperature fields), which is not available. Moreover, our Bayesian framework operates on two-dimensional, vertically integrated fields; incorporating vertical variations into the framework would significantly increase the model's complexity and is beyond the scope of this study.

We tested the sensitivity of our estimates to the depth over which α is averaged and found the results to be relatively insensitive. While this test does not fully rule out the possibility of small biases, it gives us confidence that the approximation is unlikely to introduce substantial errors.

In response to the reviewer's comment, we have added the following paragraph:

> *"Note that Eq. (16) assumes a constant α to relate vertically integrated HTC to changes in TS. While this neglects vertical variations in α, the approximation is necessary because modelling the relationship exactly would require knowledge of the vertical structure of HTC, which is not available. Furthermore, our Bayesian framework operates on two-dimensional, vertically integrated fields; incorporating vertical variations into the framework would significantly increase the model's complexity. Nevertheless, to assess the impact of this assumption, we tested the sensitivity of our estimates to different choices of α, obtained by averaging over different depth ranges. The results show almost no effect on the phase of the estimated HTC variability and only a small effect on its amplitude. While this test does not fully rule out the possibility of small biases, it gives us confidence that the approximation is unlikely to introduce substantial errors."*

We have also added the following sentence to the Conclusions:

> *"Note also that our Bayesian model operates in two dimensions and therefore uses a vertically integrated thermal expansion coefficient to relate changes in TS to the vertically integrated HTC, which may introduce some approximation error."*

Comparison with Outdated MHT Estimates:

The authors compare their results with an outdated MHT estimate based on ERA-Interim and CERES surface fluxes. ERA-Interim, in particular, is known to suffer from a negative radiation budget at TOA, which inevitably biases surface flux estimates. More recent and accurate estimates using ERA5 are available (e.g., Meyssignac et al. 2024; Mayer et al. 2022; Liu et al. 2020). The authors should compare their results against one of these more recent and reliable datasets.

We thank the reviewer for this helpful suggestion regarding more recent MHT estimates. In response, we have removed the estimate from Ganachaud & Wunsch (2003), as it corresponds to a different period and so is less relevant for comparison. We have also included the time-mean MHT estimates from Liu et al. (2022), which were kindly provided by Chunlei Liu upon request. Unfortunately, the full time series were not available for comparison.

Regarding the other two studies mentioned by the reviewer (Meyssignac et al., 2024; Mayer et al., 2022), we note that they obtained MHT estimates only at 26°N. At this latitude, we already compare our estimates to the RAPID array observations, which offer a more direct and independently measured benchmark. Additionally, to the best of our knowledge, the MHT estimates from these two studies are not publicly available.

Detailed Comments

L204: Is the GIA correction used in GRACE consistent with that used in the altimetry analysis?

We appreciate the reviewer's attention to this detail. In our analysis, the GIA corrections applied to the GRACE and altimetry data do not come from the same model. However, we explicitly account for uncertainty in the GIA correction in both cases. Specifically, uncertainty in the GIA correction is incorporated into the error structure of the GRACE and altimetry observations within the Bayesian framework.

Moreover, the GIA correction applied to the altimetry data is relatively small, with a domain-averaged absolute value of approximately 0.3 mm/yr and a range from -0.57 to 0.37 mm/yr across the domain. Given this, we expect the use of different GIA models to have a minimal effect on our results.

L205: The GRACE error budget is dominated by uncertainties in GIA and geocenter corrections, which are not currently accounted for in the analysis (see major concern #2).

We are now accounting for these error sources. Please see our response to major concern #2.

L217: surface fluxes derived from the output of reanalyses are biased. Use instead a combination of CERES TOA data and vertically integrated atmospheric energy divergence estimated from reanalyses, as in Mayer et al. (2022). (See major concern #1.)

We are now using the surface heat flux recommended by the reviewer. Please see our response to major concern #1.

L240: For the effective resolution of satellite altimetry, refer to the updated analysis in Ballarotta et al. (2019).

Thank you for this reference, which has been added to the revised manuscript.

L271: More satellites are currently operating with ~30-day repeat cycles (e.g., Sentinel-3A/B, AltiKa, CryoSat) than with 10-day cycles. The term "most" should be removed or corrected.

We agree with the reviewer, but the sentence is correct as written. The sentence in question does not state that most satellites have a 10-day repeat cycle, but rather that most operate with a repeat cycle in the range of 10 to 35 days.

L333: Clarify the term "reference datum." If referring to the altimetry reference ellipsoid, note that all altimetry satellites use the same reference ellipsoid. Biases across satellite altimeters' data arise rather because of biases in the radar signal characterization (e.g. biases in the radar signal delay characterisation, biases in the antenna center positioning relative to the center of mass of the satellite, etc…).

Thank you for this comment. To clarify, when we refer to differences in "vertical datums" we are not referring to inconsistencies between altimetry missions, which indeed use the same reference ellipsoid. Rather, we are referring to potential vertical offsets between the different data sources used in our analysis, specifically thermosteric sea level, halosteric sea level, GRACE-derived ocean mass, and satellite altimetry. Our approach accounts for these potential offsets by including a constant bias term in the Bayesian model.

In response to the reviewer's comment, the sentence in question has been reworded for clarity.

L396: Is c(Rj) computed over the full water column or limited to 0–1500 m? How is the deep ocean contribution addressed? This is important, especially since changes in deep ocean heat content can affect the α–C relationship not only on the time-mean but also over time. (See major concern #3.)

Both α and c are averaged over the top 1500 m of the ocean, although we tested the sensitivity of our estimates to different choices of α and c, obtained by averaging over different depth ranges. Please see our response to major concern #3.

L598: BHM2 is likely biased low due to the negative TOA radiation budget in ERA-Interim. Why compare to such an outdated estimate (Trenberth et al. 2019) instead of more recent ones using ERA5? (See major concern #4.)

As mentioned earlier in this response letter, we no longer include estimates derived from ERA5 or NCEP. We have added a more recent estimate to the comparison, as requested by the reviewer. Please see our response to major concern #4.

References:

Ballarotta, M., Ubelmann, C., Pujol, M.-I., Taburet, G., Fournier, F., Legeais, J.-F., Faugère, Y., Delepoulle, A., Chelton, D., Dibarboure, G., and Picot, N.: On the resolutions of ocean altimetry maps, Ocean Sci., 15, 1091–1109, https://doi.org/10.5194/os-15-1091-2019, 2019.

A Blazquez, B Meyssignac, JM Lemoine, E Berthier, A Ribes, A Cazenave, Exploring the uncertainty in GRACE estimates of the mass redistributions at the Earth surface: implications for the global water and sea level budgets, Geophysical Journal International, Volume 215, Issue 1, October 2018, Pages 415–430, https://doi.org/10.1093/gji/ggy293

Liu C et al (2020) Variability in the global energy budget and transports 1985-2017. Climate Dyn 55:3381–3396. https://doi.org/10.1007/s00382-020-05451-8

Mayer M, Haimberger L, Edwards JM, Hyder P (2017) Toward consistent diagnostics of the coupled atmosphere and ocean energy budgets. J Climate 30:9225–9246. https://doi.org/10.1175/JCLI-D-17-0137.1

Mayer J, Mayer M, Haimberger L (2021) Mass-consistent atmospheric energy and moisture budget monthly data from 1979 to present derived from ERA5 reanalysis. Copernic Clim Change Serv (C3S) Clim Data Store (CDS). https://doi.org/10.24381/cds.c2451f6b

Mayer J, Mayer M, Haimberger L, Liu C (2022) Comparison of surface energy fluxes from global to local scale. J Clim. https://doi.org/10.1175/JCLI-D-21-0598.1

Mayer M, Kato S, Bosilovich M et al (2024) Assessment of atmospheric and surface energy budgets using observation-based data products. Surv Geophys. https://doi-org.insu.bib.cnrs.fr/10.1007/ s10712-024-09827-x

Meyssignac, B., Fourest, S., Mayer, M. et al. North Atlantic Heat Transport Convergence Derived from a Regional Energy Budget Using Different Ocean Heat Content Estimates. Surv Geophys 45, 1855–1874 (2024). doi 10.1007/s10712-024-09865-5

Katherine J. Quinn, Rui M. Ponte, Uncertainty in ocean mass trends from GRACE, Geophysical Journal International, Volume 181, Issue 2, May 2010, Pages 762–768, https://doi.org/10.1111/j.1365-246X.2010.04508.x

Trenberth, K. E., Y. Zhang, J. T. Fasullo, and L. Cheng, 2019: Observation-Based Estimates of Global and Basin Ocean Meridional Heat Transport Time Series. J. Climate, 32, 4567–4583, https://doi.org/10.1175/JCLI-D-18-0872.1.

Uebbing, B., Kusche, J., Rietbroek, R., & Landerer, F. W. (2019). Processing choices affect ocean mass estimates from GRACE. Journal of Geophysical Research: Oceans, 124, 1029–1044. doi 10.1029/2018JC014341