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 #1:

This paper aims to estimate the meridional heat transport (MHT) at transatlantic sections throughout the Atlantic Ocean. Mainly, it uses hydrographic and satellite data via a Bayesian hierarchical model (BHM) to calculate the ocean heat content (OHC) tendency. The latter is then combined with air-sea heat flux data product to derive the ocean heat divergence and the MHT (as a residual from heat budgets). Accurate MHT estimates are critical for understanding the ocean's role in our climate system. Overall, the paper reads well, and the results are presented. However, there is potential confusion about the goals and motivations of this study, which would make it hard to follow what is presented and what one can learn from it. I recommend it for publication after the following minor comments are addressed.

We thank the reviewer for their positive assessment of our manuscript and their comments and suggestions.

Main comments:

The main goal of this paper is to provide MHT estimates that maximize the use of hydrographic and satellite data, via a new framework (BHM). However, I have difficulties in understanding the argument of not using the RAPID data to derive the MHT at other latitudes. Instead, the authors make assumptions about the MHT at the northern boundaries, which introduce uncertainties in the MHT estimates across all latitudes. In my opinion, it undermines the deliverables from this study.

We agree with the reviewer that the argument for not using the MHT from RAPID to derive the MHT at other latitudes could have been made more convincingly in the original manuscript. We would like to assure the reviewer that we considered this issue carefully during the development and testing of the Bayesian model. Based on numerous tests, we concluded that assuming no variability in MHT at 65º N leads to more reliable estimates at high latitudes.

It is important to note that setting the MHT at 26º N equal to the RAPID-derived MHT can also introduce errors at all latitudes, as we discussed in the manuscript. These errors arise for two main reasons. First, any biases in the RAPID-derived MHT, such as the spurious drift reported by Volkov et al. (2024), will propagate throughout the Bayesian estimates at all latitudes. Second, inconsistencies between the RAPID MHT and the Bayesian solutions can also introduce significant errors, especially at higher latitudes where the MHT variability is smaller. Even small mismatches at 26°N can appear as large errors at higher latitudes. This is precisely what our tests demonstrated.

In one such test, we compared the two Bayesian estimates of MHT at several latitudes in the North

Atlantic, and also compared both estimates with the OSNAP-derived MHT at 60° N. Following the reviewer's comment, we have added the results of this comparison to the revised manuscript, including a new Figure (Fig. 4) and the following paragraph:

> *"At 26°N, the two Bayesian MHT estimates closely match each other in both magnitude and phase. However, the differences between them become more pronounced at higher latitudes. Specifically, the magnitude of the MHT variability in the Bayesian solution constrained by the RAPID-derived MHT remains nearly constant across latitudes. This contradicts direct observational evidence from the RAPID and OSNAP arrays, which shows that MHT variability at 26°N is more than four times greater than at 60°N. In contrast, the variability in the MHT derived from Eq. (20) decreases progressively towards higher latitudes and closely matches the magnitude of the MHT variability from OSNAP at 60°N, despite assuming zero variability at 65°N. These findings highlight two key points: (1) the approximation inherent in Eq. (20) holds to a good approximation; and (2) the solution obtained using Eq. (20) is more accurate than the one based on the RAPID-based constraint."*

These results together with the good agreement of our estimates with the RAPID-derived MHT at 26°N give us confidence in the robustness of our estimates.

Another source of uncertainty in the MHT estimates is from surface heat flux. However, surface heat flux itself likely contains larger uncertainty than the heat divergence derived from this study – as that is indicated by the discrepancies between BHM solutions. I would suggest that the authors provide a thorough uncertainty estimate that takes into consideration errors in surface heat flux.

We would like to clarify that observation errors are already accounted for in all data sources, including the surface heat flux (HF) data. It is not entirely clear to us what additional steps the reviewer is suggesting beyond what is already implemented. As noted in the manuscript, none of the HF products provides uncertainty estimates, thus we derive a measure of uncertainty from the spread across three HF products. See Section 2.4 for more details. These uncertainties, along with those from other data sources, are incorporated into the BHM, which enables a comprehensive treatment of uncertainty with rigorous error propagation.

If the goal of this paper is to prove the efficacy of the new BHM framework that combines hydrographic and satellite observations, should it be compared with one that just uses hydrographic data? That would highlight the advantages of the BHM.

We agree with the reviewer's suggestion and have responded by including a comparison between the RAPID-derived MHT at 26° N and the MHT derived using the budget approach based only on the hydrographic data. The results are discussed in Section 5.2 and include a new figure (Fig. A1).

This is related to comment#3. Much of this paper is centered on the discrepancies between BHM solutions (see Figures 4, 5, 6, 7 and the related text). Those comparisons are useful as an evaluation of how different surface heat flux data impact the MHT estimates. But such an evaluation itself is not well motivated. In addition, the MHT estimates are validated against Trenberth et al. (2019). But it is not clear we gain from this analysis that is distinct from Trenberth et al. which uses atmospheric reanalyses (surface heat flux) and hydrographic data (the OHC tendency).

As mentioned earlier in this response letter, the revised manuscript includes only estimates based on a new surface HF product. Following a comment by Reviewer #2, the comparison with Trenberth et al. (2019) is now limited to the time-mean MHT.

To answer the reviewer's question, our estimates differ from those of Trenberth et al. (2019) in that they are based on hydrographic observations constrained by satellite data in a statistically rigorous manner, rather than on ocean temperature fields from reanalyses. Additionally, we use a more recent surface heat flux product and provide quarterly estimates, whereas Trenberth et al. (2019) report 12-month filtered values.

Other comments:

Line 90: TS and HS are anomalies relative to the climatology density. Other terms should also be anomalies? Please be specific about each term.

Done.

Line 133: 'interesting oceanographically' reads odd.

This sentence has been reworded.

Line 148: How large is the volume transport? If it is large, it affects the mass conservation and thus the MHT estimate. Such effects on the related sections need to be discussed.

Originally, we did not exclude the Gulf of Mexico (GoM) and the Caribbean Sea from our analysis. However, during testing of the BHM, we identified issues with the ISAS20 data in these regions. For example, as shown in Fig. R1, the thermosteric sea level from a grid cell in the Caribbean Sea displays anomalous variability prior to 2011, suggesting potential problems in the data. While the BHM may partially correct for such issues through constraints from satellite observations, we ultimately chose to exclude these regions to avoid introducing spurious signals.

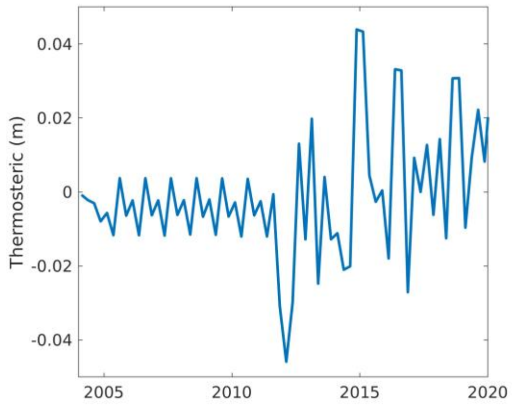

Figure R1. Thermosteric sea level from one of the grid cells in the Caribbean Sea.

In response to the reviewer's comment, we have assessed the impact of this exclusion by comparing MHT estimates from two BHM solutions: one that includes the GoM and Caribbean Sea, and one that excludes them as in the paper (see Figs. R2 and R3). The differences between the two are minimal, both in terms of MHT variability (Fig. R2) and time-mean MHT (Fig. R3), indicating that excluding these regions has a negligible effect on our results.

We have added the following sentence to the revised manuscript to clarify this point:

> *"We have tested the impact of excluding data in these regions from on our estimates of MHT and found it to be minimal."*

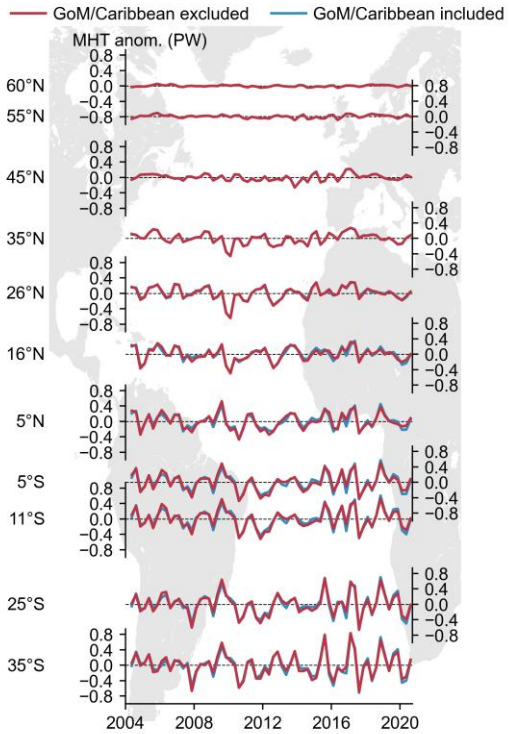

Figure R2. Quarterly (3-month-averaged) time series of MHT anomalies across the latitudinal sections denoted on the vertical axis as estimated using the BHM for two cases: 1) excluding data from the Gulf of Mexico and the Caribbean Sea (red line); and 2) including data in these regions (blue line).

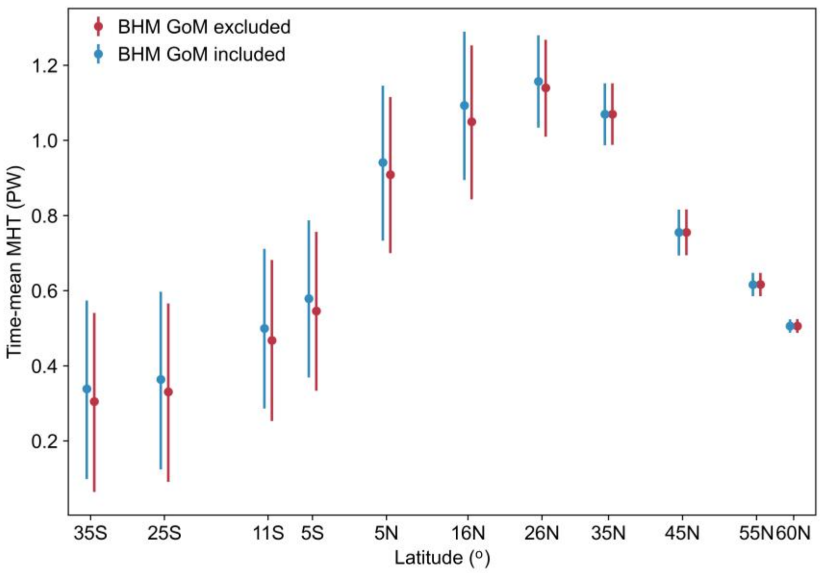

Figure R3. Time-mean MHT along with the associated 5-95% CIs (whiskers) at multiple latitudes of the Atlantic Ocean derived from the BHM for two cases: 1) excluding data from the Gulf of Mexico and the Caribbean Sea (red line); and 2) including data in these regions (blue line).

Line 172: Uniform l= 100m spatially and vertically? How valid are such assumptions?

The decorrelation length scale for errors in the gridded temperature and salinity (T and S) fields is not provided in the ISAS20 product, which only includes error variances but not covariances. While the vertical error correlation structure is unknown, some degree of vertical dependence is expected, as the T and S data are derived from vertical profiles. We therefore consider it reasonable to set the vertical decorrelation length scale equal to the length scales used in the objective analysis of the profiles themselves. We believe that this choice provides a sensible estimate in the absence of explicit error correlation information.

Line 215: Are the two reanalysis products only used onward 12/2017? If yes, how?

Please note that in the revised version of the manuscript, we compute a single Bayesian solution based on a new observation-based surface heat flux data set. 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. This has been clarified in the revised manuscript.

Line 218: Are the reanalysis products averaged together with DEEP-C? This appears to contradict the previous statement that 'it is preferable to' the reanalysis products.

Please see our response to the previous comment. But to answer the reviewer's question, no, the reanalysis products were not averaged together with DEEP-C in the original submission. Instead, they were used to characterize the uncertainty in the surface heat flux data.

Line 228: 'effective spatial resolution is much lower than what such grids imply' Hard to understand what it means – please reword.

We have reworded the sentence for clarity. What we meant is that although some products are provided on relatively fine spatial grids, this does not mean they can reliably resolve features at those grid scales. The actual effective spatial resolution (i.e., the smallest scale at which meaningful information is retained) is often coarser than the grid spacing suggests. The revised sentence reads:

> *"Both the hydrographic data and the GRACE data are provided on relatively fine grids, but their effective spatial resolution is much lower as these data do not resolve features at the scale of the grid spacing."*

Line 240: If the goal is for an integrated value over the region between two latitudes (11 regions in total, Fig. 1), why does one need spatial grids anyway? Why not consider the enclosed basin as a whole?

We apply the BHM at the spatial grid level because this allows us to leverage cross-variable spatial information (across SL, TS, HS and OM) as well as to account for spatial error structures. This leads to significantly improved estimates of thermosteric sea level at each grid cell. These improved gridded fields can then be used to produce more accurate spatial averages over each budget region.

Our concern with computing regional averages directly from the original, unconstrained thermosteric fields is that such averages are likely to be biased due to the sparseness of the temperature data. Once the data are spatially averaged, it becomes very difficult to correct for these biases as it is no longer possible to exploit spatial information. Working in the gridded domain allows us to apply a coherent spatiotemporal modelling framework that reduces such biases and provides a more principled approach to uncertainty estimation.

We explain this in the manuscript, for example:

> *"By first calculating spatial averages separately for each variable, the procedure ignores any spatial dependencies between the variables and loses the opportunity to leverage cross-variable spatial information, both of which can lead to suboptimal estimates of spatially averaged values. Also, such a modelling choice makes the estimation of uncertainties in the spatially averaged values challenging, often requiring ad-hoc or approximate methods."*

> *"Our approach extends that of Kelly et al. (2016) by accounting for spatiotemporal dependencies between processes (i.e., TS, HS, and OM) and enabling information sharing*

across the various data sets. This is achieved by simultaneous spatiotemporal modelling of the observational fields and their error structures, in contrast to time series modelling of spatially averaged values as done in Kelly et al. (2016)."

Line 281: Setting rhoij= 0 requires justifications. The decorrelation time scale should be evaluated separately for each dataset, which is likely longer than a month.

The decorrelation time scale referred to by the reviewer is only used in our analysis to propagate observational uncertainties from monthly to quarterly values. In response to the reviewer's comment, we have assessed the sensitivity of our Bayesian MHT estimates to different values of ρ_{ij}.

The most conservative assumption, perfect temporal correlation, corresponds to ρ_{ij}=1, in which case the standard errors on the quarterly values are $\sqrt{3}$ times larger than in the uncorrelated case (ρ_{ij}=0). We have compared the resulting MHT estimates for both cases (see Fig. R4). The two estimates are nearly indistinguishable, with the main difference being a modest increase in the uncertainty of the MHT estimates for ρ_{ij}=1, by a factor of 1.07 on average. This limited sensitivity is expected, as changing ρ_{ij} affects only the magnitude of the observation errors, not their spatial structure, which is the dominant sources of constraint in the BHM.

Based on this assessment, we have decided to retain our original choice of ρ_{ij}=0, which simplifies the calculation without significantly impacting the results.

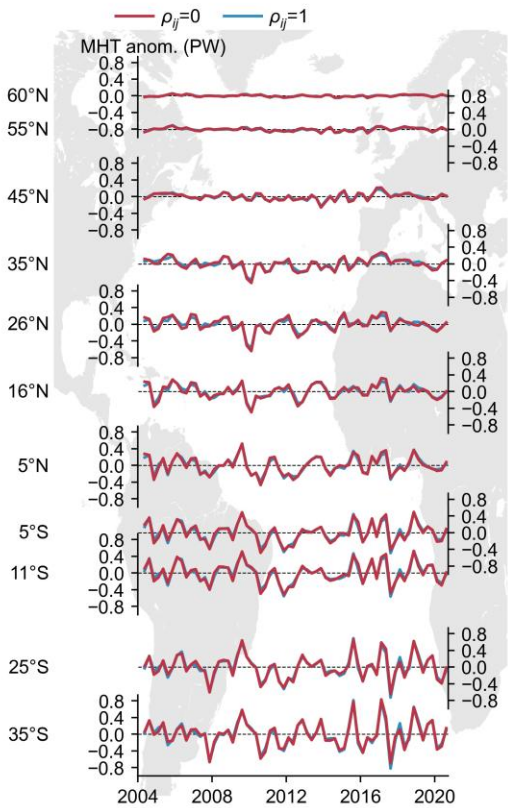

Figure R4. Quarterly (3-month-averaged) time series of MHT anomalies across the latitudinal sections denoted on the vertical axis as estimated using the BHM for two cases: 1) ρ_{ij}=0 (red line); and 2) ρ_{ij}=1 (blue line).

Line 288 and Figure 2: What are the different arrows in Figure 2? For example, those black arrows within the right red box indicate that Q is derived by H minus HTC. But that is opposite to what's described in the text.

We thank the reviewer for this question. In Bayesian diagrams, the arrows represent how we believe the observations were generated. That is, we assume there is an underlying latent process (e.g., the true thermosteric field), and the observations (e.g., instrumental readings) are noisy or indirect measurements of that process. The direction of the arrows reflects the flow of information in the generative process: from parameters/processes to observations. This convention also emphasizes the idea that uncertainty in parameters/processes propagates to uncertainty in the observations. Once we have observed the data, we then use Bayesian inference to reason backward through the arrows to estimate the latent processes.

In response to this comment, we have added the following sentence to the caption of Fig. 2:

> *"The direction of the arrows in the Bayesian diagram reflects the flow of information in the generative process: from parameters/processes to observations."*

Line 298: Why are the reanalysis products used separately? This is related to my comment above.

Please note that this is no longer the case in the revised manuscript. See comment at the beginning of this response letter and also our response to an earlier comment.

Line 315: How exactly are uncertainties determined? It is the key to providing a meaningful estimate.

We thank the reviewer for this important question. In a BHM, uncertainties are explicitly modelled and quantified through the probabilistic structure of the model. Uncertainty in the underlying latent processes is captured by modelling them as stochastic processes. By modelling the latent processes probabilistically as spatiotemporal random fields, we quantify the fact that there are many plausible versions of the underlying fields consistent with the data and prior knowledge. This is what we refer to as uncertainty arising from limited knowledge of the latent process. Uncertainty in model parameters (such as decorrelation length scales) is captured by assigning prior distributions to these parameters. Observation uncertainty is incorporated through the data likelihood, by specifying a distribution for the measurement error. The Bayesian framework allows us to coherently propagate all these sources of uncertainty into the posterior distributions of the quantities of interest.

In response to the reviewer's comment, we have added the following sentence to the revised manuscript:

> *"By modelling the latent processes probabilistically as random fields, we account for the fact that there are many plausible realizations of the underlying processes consistent with the observational data, thereby capturing uncertainty arising from limited knowledge of the latent processes."*

Line 520: What does it mean by 'will be accurate at any latitude'? How to quantify this accuracy? Also, why is the true transport at a given latitude is 'large relative to the true transport at 65N'?

Thank you for the opportunity to clarify this point. As shown in Eq. (20), the MHT at any latitude is computed as the sum of the MHT at 65ºN and the cumulative integral of heat transport convergences (HTCs) from 65ºN to the latitude of interest. Therefore, if the second term (i.e., the integrated HTC) is large compared to the (unknown) MHT at 65°N, then setting the MHT at 65°N to zero introduces only a small relative error. This is what we mean by saying the MHT "will be accurate at any latitude where the MHT is large compared to the MHT at 65°N."

This approximation can be quantified at 26ºN by drawing on direct observations from the RAPID and OSNAP arrays. These show that the variability of MHT at 26ºN (from RAPID) is more than four times larger than at 60ºN (from OSNAP), suggesting that at lower latitudes the contribution from the HTCs dominates the total MHT signal. Consequently, any uncertainty introduced by assuming zero transport at 65ºN becomes negligible at latitudes like 26ºN where the integrated HTC is large.

Line 524: ‘four time larger at RAPID than at OSNAP’ – are the comparisons only based on the MHT estimates from this study?

No, the comparison is not based on the MHT estimates from our study. It relies on independent observational estimates derived from the RAPID and OSNAP arrays, as noted in the original submission where we wrote: “estimates of MHT from the RAPID and OSNAP arrays.” These observational datasets provide a direct and independent basis for comparing the variability of MHT at 26ºN and 60ºN.

We have slightly reworded the original text to make it clearer that the comparison is based on independent observations.

Line 529: What kind of error is this referring to? The mean value of MHT1 is not supposed to affect the derived variations.

We appreciate the reviewer’s comment and agree that the time-mean value of MHT_1 does not affect the variability of MHT at other latitudes. However, it does influence the time-mean MHT at those latitudes. As stated in the manuscript, the MHT at any latitude is calculated as the sum of MHT_1 and the integrated HTCs between 65ºN and that latitude. Thus, omitting the time-mean of MHT_1 introduces a bias in the resulting time-mean MHT values elsewhere.

From the RAPID and OSNAP arrays, we know that the time-mean MHT is approximately 1.19 PW at 26ºN (RAPID) and 0.51 PW at 60ºN (OSNAP). Therefore, ignoring the time-mean MHT at 65ºN would result in an underestimate of the mean MHT at 26ºN by over 40%. In contrast, the variability of MHT at 26ºN is more than four times larger than at 60ºN, so the influence of MHT_1 on the variability is minimal.

To account for this, we set the time-mean MHT at 60ºN equal to the mean value from the OSNAP-derived MHT.

Line 532: How representative is the 2014-2018 mean? over a longer period 2004-2020?

The time-mean MHT from the RAPID array is 1.19 PW for the full period 2004–2020, and 1.21 PW for the shorter period 2014–2018. The difference between these two estimates (0.02 PW) is an order of magnitude smaller than the associated uncertainty (approximately 0.2 PW), indicating that the shorter period mean is representative of the longer-term mean.

At 60ºN, the OSNAP-derived time-mean MHT is expected to be even more robust, given that the variability is smaller relative to the mean. Specifically, the standard deviation of the OSNAP time series is only 0.03 PW, while the mean is 0.51 PW, further supporting the stability of the time-mean estimate over time.

Line 536: I am not sure about this assumption that is based on a 4-year time series.

We agree with the reviewer that this assumption might not hold exactly, but there is no definitive way to verify the validity of the assumption. That is why we highlight this as a potential limitation of the estimates in the Discussion. Specifically, we wrote:

> “*it is important to remember that the time-mean MHT at 60º N has been set equal to that derived from the OSNAP array over the period 2014-2018. This assumes a constant time-mean MHT at 60º N over the analysis period (2004-2020), and thus deviations from this premise will introduce an error into the time-mean MHT at all the other latitudes*”

Line 540-546: I found the explanations inadequate – this is related to my main comment above. First, the observed MHT from RAPID is most likely the best estimate one can get, so why does depending on it become an issue? I cannot follow the reasoning behind the second point. Why does the observed MHT from RAPID introduce large errors? It is understandable that the RAPID data may be used to first validate this method. But after that validation, could and should it be used to improve the estimates?

We refer the reviewer to our response to a similar comment earlier in this letter. As explained there, we prefer not to publish the MHT estimates constrained by the RAPID-derived MHT. However, if the reviewer feels strongly that these estimates should be included, we would be happy to make them available on Zenodo alongside the other estimates.

Line 549: Once again, it is unclear why three surface heat flux (Qsfc) datasets are used separately, which are over different time periods.

This is no longer the case, as explained earlier in this response letter.

Line 573: Please justify 'very significant' – what is p-value?

Thank you for the comment. We would like to clarify that the MHT time series in our analysis are derived from a Bayesian posterior distribution. In this context, the concept of a frequentist p-value is not directly applicable, since the Bayesian framework does not involve hypothesis testing in the frequentist sense. However, we agree that "very significant" reads vague and, therefore, we have reworded this to "statistically significant at the 95% confidence level".

Line 575: Why is a discrepancy only occurring in 2020?

We did not mean to suggest that a discrepancy occurred only in 2020. As stated in line 575 of the original manuscript, "*The lower correlation observed during the longer period of 2004–2020 is primarily due to a discrepancy in 2020*." That is, the difference in correlation between the shorter and longer periods is largely driven by the mismatch in 2020, not that discrepancies are entirely absent in other years. In any case, this paragraph has been revised and we hope it is clearer now what we meant.

Line 577: 'This discrepancy is … entire period.' Hard to understand what it means – please rephrase.

This paragraph has been partly rewritten in the revised paper.

Line 588: Figure 4: How are the CIs determined? It is worth a dedicated subsection in Methods on uncertainty in the MHT estimates.

The CIs are computed as the 5th to 95th percentiles of the posterior distribution, based on samples obtained from the BHM. Since the calculation is straightforward once the posterior samples are available, we did not initially include a dedicated explanation. However, in response to the reviewer's comment, we have added the following clarification at the beginning of Section 5.2:

> "*The CIs are computed as the 5th to 95th percentiles of the posterior distribution, based on samples obtained from the BHM.*"

Line 599: As mentioned above, is the difference in the mean MHT between BMH solutions mostly related to Qsfc?

As noted earlier in this letter, we now focus exclusively on the solution derived from the new surface heat flux product, so the comparison between earlier BHM solutions is no longer relevant to the manuscript. However, to answer the reviewer's question: yes, the differences in mean MHT between the previous BHM solutions were primarily due to the use of different surface heat flux products, which was the only distinction between those solutions.

Line 600: 'To complete our comparison' may not be a good motivation. E.g., why apply 5-quarter running averages? How does it help complete the comparison, or how does it help understand the discrepancies?

Thank you for the comment. We have reworded the sentence in the revised manuscript for greater clarity. Our intent in applying a running mean is to extend the comparison across different temporal scales. Agreement at the quarterly scale does not necessarily imply agreement at lower frequencies, and smoothing helps reveal whether discrepancies persist (or are reduced) at longer timescales. This complements the analysis of high-frequency variability and provides a more complete picture of agreement between our estimate and the one from RAPID. In response to an earlier suggestion by the reviewer, we now use a 4-quarter running mean instead of the original 5-quarter window.

Line 609: The data are 5-year averages. What do the differences during 2005-2007 represent?

There seems to be some confusion; our analysis is based on 5-quarter running averages (4-quarter running averages in the revised manuscript), which are effectively 1-year averages. Therefore, the differences observed during 2005–2007 reflect variability at approximately annual timescales, not 5-year averages.

Line 626: For the comparing purposes, why not apply the same 12-month (4-quarter) running averages to the MHT estimates from this study? That would help make meaningful comparisons.

In response to a comment by reviewer #2, we no longer show time series from Trenberth et al. (2019). Note, however, that we now use a 4-quarter running mean instead of the original 5-quarter window.

Line 628: ‘several interesting features’ reads odd.

This sentence has been reworded. Thank you.

Line 628 and the whole paragraph: Those features are related to the similarities and differences between BHM solutions. But it is not clear what we will gain from those comparisons. Please refer to my main comments.

Thank you for the comment. Please note that in the revised manuscript, we no longer include the comparison between Bayesian solutions derived from different surface heat flux products, as our focus is now on the solution based on the new observation-based product.

Line 658: As mentioned earlier, would it be better to use the 12-month smoothed data when comparing with Trenberth et al. (2019)?

As mentioned earlier, we no longer show time series from Trenberth et al. (2019). Note, however, that we now use a 4-quarter running mean instead of the original 5-quarter window.

Line 667: The mean is obtained over different lengths of record and different periods. Given the strong interannual and decadal variations in the OHC and probably in the MHT, the time-mean estimates could be biased and cannot be compared directly to each other. Please justify the choices of those estimates to compare with and discuss the comparisons to avoid misinterpretation.

We should mention first that, in the revised paper, we have removed the estimate from Ganachaud & Wunsch (2003), as this was for a completely different period, and have added an estimate from a more recent paper (Liu et al., 2022).

We agree with the reviewer that differences in the period might explain some of the discrepancies between the BHM estimates and those from other studies. To evaluate this, we have calculated the time-mean MHT from the BHM over the Liu22 period (2006-2013) and compared it to BHM estimates for the period 2004-2020. We have found that the differences are less than 10% of the time-mean MHT at all latitudes, suggesting that the discrepancies with T19 and Liu22 are unlikely to be explained by differences in the time period alone. We have added the results of this analysis to the revised manuscript.

Line 686: Is it because of a similar method (MHT as a residual from heat budgets)?

It is difficult to say with certainty. While both our study and T19 derive MHT as a residual from heat budgets, the approaches differ significantly in how the data sets are combined and how the budgets are evaluated. In addition, the specific data sets used in T19 differ from those used in our study. Therefore, the agreement may be due to a combination of methodological and data-related factors rather than from a shared approach alone. We have added a sentence along these lines to the revised manuscript.

Line 688: Why compare to GW03? What can we learn from this comparison?

As mentioned above, we have removed GW03 from the comparison.

Line 706: What is the main objective of this study? A data set (MHT estimates) or a valid method? Please refer to my main comments.

The main goal of this study is to provide new observation-based estimates of MHT across most of the Atlantic Ocean using a statistically rigorous method that integrates data from multiple sources within a Bayesian hierarchical modelling framework. While the methodology is important, particularly the way it integrates multi-source data and propagates uncertainty through joint spatiotemporal modelling, the primary aim is to generate reliable, probabilistic MHT estimates over a broad range of latitudes. These estimates are intended to complement direct observational records (e.g., from the RAPID array), extending our spatial and temporal coverage. We believe this goal is clearly stated in the Introduction of the manuscript, where we motivate the need for spatially and temporally continuous MHT estimates and describe the advantages of our approach in achieving this goal.

Line 710: It is not clear why three solutions are needed.

Please note that, as mentioned earlier in this letter, we no longer show three Bayesian solutions.

Line 719: This seems to be a hasty conclusion. Those correlations are based on data for different time periods and are based on different assumptions.

We thank the reviewer for this comment. In the revised manuscript, all correlation values are computed over the same time periods as those reported in the previous studies, allowing for a fair comparison. We agree that the estimates are based on different assumptions and methodologies, but that is precisely the point of the comparison. We aim to assess how well our BHM-derived estimates align with RAPID-derived MHT relative to earlier approaches, which necessarily involves comparing methods built on different assumptions and approaches. This comparison is important to demonstrate the added value of accounting for spatiotemporal dependencies and properly propagating uncertainty when combining hydrographic and satellite data. We have revised the paragraph to better emphasize this point.

Line 726: It simply indicates that surface heat flux is a major source of uncertainty in the MHT estimates. Please refer to my main comments.

We agree with the reviewer. Since we are no longer comparing solutions based on different surface heat flux products, this sentence has been removed from the manuscript.