Answers to RC1

General Comment

The paper calculates ocean heat content and its changes over time starting from a principal components analysis. The approach adopted is convincing and supported by the results. The authors' work makes an important contribution to one of the most debated issues in the scientific and general context. The article therefore deserves publication after some minor corrections.

We thank Giuseppe M.R. Manzella for his encouraging comments.

Specific Comment

The section requiring further clarification concerns the principal components used to calculate the fingerprint. The authors state that they use 75 components (out of 236, I assume—but the text should explain this better) accounting 80% of the total variance. A figure (better) or a comment on how the variance is distributed in the various modes would be helpful in understanding whether the remaining 20% can be considered just noise.

Thank you for the comment. Apparently our description was not clear. We introduce 236 mass and 200 steric fingerprints in our inversion setup. The former are subdivided into 68 glacier fingerprints, one for each major glacier region, 16 fingerprints for Greenland, 27 for Antarctica, 75 for terrestrial hydrology and 50 for internal ocean mass variations. Fingerprints for Greenland, Antarctica and the major glacier regions are not derived using a model and applying PCA decomposition, but instead by directly introducing prior information, e.g. from ice-altimetry derived melting rates in cases of Greenland and Antarctica into the Sea Level Equation (SLEQ). For terrestrial hydrology and internal ocean mass variations a hydrology and an ocean model, respectively, have been decomposed using PCA and retaining the modes corresponding to at least 80% of the explained variance, each. Since the hydrological model EOFs are only defined over land, we then also transform these to sea level fingerprints by applying the SLEQ. The individual steric components thermosteric_upper700m, thermosteric_deep, halosteric_upper700m and halosteric_deep have been processed similarly with retaining at least 80% variance each leading to 75, 75, 50 and 50 fingerprints, respectively.

We have updated the overall description by explicitly stating in the beginning “In total we fit 486 fingerprints, subdivided into 236 mass and 250 steric fingerprints, to the input data.” We further expanded description for the individual subcomponents throughout the paragraphs to better describe the overall fingerprint setup, as described here above.

Although of little practical importance, there is a formal difference between Singular Value Decomposition and Empirical Orthogonal Function. In 2.2, the authors define PCA and write EOF on 124 line. Again, this has no practical consequences, but formally, it is inaccurate.

We agree, but the difference is more a matter of naming and conventions. PCA is the method that is used to decompose a dataset into principal components and EOFs by means of singular value decomposition as described in Section 2.1. Nonetheless, we will rephrased the mentioned sentence by no longer mentioning “PCA” to avoid confusion.

Specific Comments

In Formula 1, it would be better to include all the coordinates so as to emphasize the fact that the integration is performed only on the vertical coordinate.

Done

In the same formula, it is necessary to indicate the value of -H in the integral.

The ORAS5 model covers a maximum depth of about 6000m. In order to derive fingerprints for two different depth regimes (0-700m) and (below 700m) this means the integral boundaries change accordingly; similarly when considering processing the Argo profile data, which we limit to 700m to reliably cover the first depth level, as the “deep” ocean down to the sea floor (or to be more precise the maximum ORAS5 model depth in context of fingerprint representation) is not fully covered by Argo floats, limiting the possibilities for utilizing these for our “deep” ocean layer.

We have expanded the description of Eq. 1 to be more clear as “In case of the ORAS5 reanalysis, H represents the maximum model depth of ~6000m, while when we analyze Argo data, a depth of 700m is used. In other words, the integral limits are adapted depending on the utilized data, e.g. model or Argo, throughout the manuscript.”

The use of CERES (which are not completely independent data) can only be used for comparison of results. This needs to be made clearer in the discussion.

We agree with the reviewer. So far our discussion only briefly mentioned that the CERES dataset is not entirely independent. We will add to this making it more clear that CERES data should not be seen as a completely independent reference, but more like one way to derive OHC among the other approaches.

We have added the following sentence to the corresponding paragraph in the discussion section “As such it can only be used for comparison of results, but does not serve as a completely independent reference.”

Answers to RC2

The manuscript introduces a budget method to reconstruct global and regional ocean heat content (OHC) and uptake (OHU) rates over 2005–2024. Although budget methods have been used to estimate global OHC time series widely, it is less commonly used to derive regional OHC changes. The manuscript provides a new dataset for understanding OHC, which is valuable for the ocean community and fits the scope of ESSD perfectly. However, I think the authors should provide the uncertainty of their indirect OHC datasets. The uncertainty of this method is closely associated with the challenges of regional sea level budget. The uncertainty of regional sea level and ocean mass changes is much larger than that of global mean time series. Thus, some studies have pointed out that sea level budgets are not balanced in some basins, such as North Atlantic. The authors should add some discussions. This could influence the confidence of the work and should be addressed before publication. The authors should be cautious when they compare their results with other products, such as observations and ocean reanalyses. The calculation of the ocean reanalyses ensemble mean in this study is not acceptable. The ocean reanalyses products do not have the same periods, resulting in inconsistent errors. I believe that this manuscript must address these two major issues before publication.

We are grateful for the valuable feedback, which helped us to improve our manuscript.

We agree that the uncertainty information has been missing from our dataset. As we recently submitted our gridded dataset to the GEWEX EEI intercomparison in terms of OHC [J] with corresponding errors (also in J) we replaced our previous version given in J/m^2 with the one used in the GEWEX intercomparison. As a consequence the unit has changed by multiplying each grid cell with its respective area and we added the uncertainty. The dataset submitted to PANGAEA and the cloud repository used for the review process has been updated, too.

We generally agree with reviewer that the uncertainty of the regional OHC estimates, i.e. each grid point or subregion, is directly connected to the uncertainty of the corresponding sea level budget, i.e. individual mass and steric components, at these locations. This is certainly true for every product regardless if one utilizes Argo only, an ocean reanalysis, thermo-steric sea level derived from combining GRACE(-FO) and altimetry (and in this case also Argo) observations. However, since this manuscript focuses on OHC, it would be out of scope to do an extensive regional sea level budget discussion.

Consequently, we have added a (appendix) figure (Fig. B2 in the updated manuscript) and corresponding discussion regarding the major regional sea level budget components (total, mass, steric and residual). In addition, we added a figure (Fig. 7 in the updated manuscript) showing the regional GAAMR OHU and OHU trend errors to the main part of the manuscript.

Please see also the detailed comments below.

Major comments:

- Previous studies have highlighted the challenges of closing regional budgets, which means there is still great uncertainty in sea level, ocean mass, and steric sea level. The budget gaps can be induced by satellite altimetry, GRACE/GRACE-FO, or Argo observations. But it seems that this method mainly adds the budget gaps into steric sea level. This is not reasonable. I know it is hard to separate budget gaps into three parts, so I hope the authors can discuss the uncertainty of this method.

We are not entirely sure what the reviewer is referring to here. Our inversion method partitions the altimetry observed total sea level change into individual mass and steric contributions by fitting pre-defined fingerprints (spatial patterns) to satellite altimetry, GRACE(-FO) gravity and Argo in-situ temperature and salinity profile data; the latter is first converted to steric sea level change at the Argo positions. We provide budget closure estimates of the summed individual contributors (total) versus satellite altimetry, as it is customary done for sea level budget closure. We did not mean to give the impression that all of the (small) budget residual signal is attributed to errors in the thermo- and halosteric sea level estimates, as this is not the case. On the contrary, by definition in the least-squares inversion, the budget “gap” is distributed according to the suspected error levels of mass and steric contributions. If the reviewer is referring to the errors for the global mean sea level budget in Table A1, the formal error level of the “mass sum” and “Steric sum” components, i.e. the sum of all mass and steric sub-components, respectively, are very similar in magnitude. Consequently, the overall error is composed by roughly 50% from steric and mass contributions, depending on the analysis period. Of course, regionally this error distribution might vary significantly considering sea level change in some regions is mainly driven by ocean mass change (e.g. Baltic Sea) while other regions are dominated by steric variations (e.g. Bay of Bengal).

We have added a figure (Fig. B2 in the updated manuscript) showing the regional sea level trends for the major components (total, mass, steric and residual) together with their regional errors. We discuss these in context of other published regional estimates, including those mentioned in this review as part of our sea level results. Please see also our answers to the specific comments below.

- The authors give the comparison among different products to validate their results. This comparison is important to increase the reader’s confidence. But it is not wise to use ‘overestimate’ or ‘underestimate’ in the observations when they are compared with the indirect method. I think the uncertainty of observations is smaller than the method. If the authors can point out the limitations of observations, such as instrument bias and other problems, I think the description is acceptable.

We agree with the reviewer and have rephrased the sentences in our discussion. We never meant for this statement to be interpreted in a way that our GAAMR results are the only truth. We merely wanted to point out that some of the investigated products, i.e. those mentioned, exhibit larger global OHU compared to the other products, which seems to mainly result from differences in the Indian ocean. We meant “overestimate” in the sense of “larger compared to other products” and not “relative to a reference or truth”.

Consequently, we have rephrased this in order to avoid misinterpretation. 1) We rephrased “We also found a clear overestimation ...” to “We also found larger estimates …”. 2) We also removed this phrasing from the discussion of the Hovmöller diagram.

Please see also our answers to the specific comments below.

- I am confused about the resolution. The output of this manuscript is 0.25 degree. I know the satellite altimetry and GRACE have the resolution of 0.25 degree, but the salinity observation only have a resolution of 1 degree. Can the authors provide more details?

Thank you for this comment, apparently we have not been clear on that. All of our fingerprints used for generating the results are evaluated on a 0.25 degree grid, which represents the highest resolution from the available input auxiliary data. For the output of the OHC data we then utilize only the thermosteric part, i.e. monthly grids of thermo-steric sea level change. The resolution of our fingerprints defines the output resolution but has to be viewed separately from the observation resolution: 1) along-track altimetry, 1Hz data ~5km along the track, GRACE(-FO) spherical harmonics up to d/o 120, ~300km and the Argo data, which is globally not homogeneous and observation density is strongly regionally dependent.

We have updated our inversion description to make this more clear by adding the following to section 2.3.3: “We select 0.25 degree resolution to evaluate all our fingerprints, based on the highest available spatial resolution of the model data used from constructing those fingerprints.”

Minor comments:

L94: Another challenge is that the uncertainty of regional sea level observations are much larger than that of global mean time series.

We agree with reviewer that the larger regional error levels will also significantly affect the regional OHC uncertainty.

We have added this to the manuscript as: “In addition, uncertainties of the involved sea level contributions are significantly larger at regional scale compared to their respective global averages, which might add to the uncertainty of OHC at regional scale..”

L123: It is more proper to use EOF. PCA is used in time series decomposition generally, and EOF involves spatial and temporal decomposition.

We politely disagree here as the difference is more a matter of naming and conventions. PCA is the overall technique that decomposes a dataset into EOFs and principal components (PC)s. Nonetheless, we have rephrased that line by removing mentioning “PCA” in order to avoid confusion.

L159: Why choose 700m as a divided range? Some studies use 700m as a specific depth because it is the maximum observation range of XBT (the major source of ocean temperature before Argo). The Argo observations can cover upper 2000m ocean. Maybe 2000m is a better choice.

We also chose 700m based on previous studies and potential to add XBT measurements as additional observations to the inversion in the future without the need to redefine the steric fingerprint database. So far we limit the profile data to Argo profile floats only. While 2000m would provide a theoretically better depth coverage for the Argo observations, we would have to filter out many more data points as not all floats always reach that depth despite the bathymetry indicating that they should. However, we plan to expand on this future inversion versions, especially considering the increasing number of available deep Argo floats in recent years, which will allow to better separate individual depth levels. However, this requires extensive rebuilding of the corresponding steric fingerprints, which is out of scope of this study.

L160: Please clarify how to remove artificial trend and salinity drifts

We simply enforce the global halosteric trend to be zero over the GRACE(-FO) period when computing our (halo)-steric fingerprints.

We have added this information to the text as “This is done by enforcing the global halosteric trend to be zero during the GRACE(-FO) era before EOF decomposition.”

L160: The authors should point out the depth range. The ORAS5 covers the full-depth ocean, but some of the Argo-based observations (such as SCRIPPS) only provide the upper 2000m.

The fingerprints cover two depth levels upper 700m and everything below (== deep). We limit the fit of Argo observations to the fingerprints of the upper 700m. The deep ocean steric part is computed from the overall combination of GRACE, altimetry and upper 700m steric observations.

We have added this to the description as follows: “Despite the deep ocean steric contribution not being directly constrained by observations, it can still be derived from the combination of altimetry, GRACE(-FO) and upper 700\,m steric information.”

L169: ORAS5 has a resolution of 0.25 degree, but the Argo-based observations generally only have 1 degree.

Please see our discussion regarding major comment 3.

In addition, we have added the following to the Argo data description in Sect. 2.3.2: “These have a resolution of about 1 degree globally, but their density varies regionally (Fig. B3).”

L178: It is unusual to use along-track satellite altimetry. I do not think it is damage, but I think the authors should point out why they do not use the gridded products and the corrections used in altimetry (such as GIA, wet troposphere correction).

1) We use along-track altimetry data because these are the actual observed sea level estimates. Gridded products have to fill the gap between altimtery tracks by interpolating or model projection/scaling. As such, grid points not supported by actual along-track data basically do not provide any real observation. From our perspective, utilizing gridded altimetry data introduces another unnecessary potential source of error into our inversion results.

We now mention this in section 2.3.2 of the updated manuscript as “We prefer using along-track altimetric data over gridded products, since not all grid points are observed regularly and gaps in these products need to be filled, which would introduce another unnecessary potential source of error into our inversion.”

2) Our along-track input altimetry represents sea level anomalies (SLA). As such these are corrected for all instrumental, atmospheric and geophysical corrections. For the GIA correction we made sure to utilize the same model (ICE6G_D) for GRACE(-FO) and altimetry data.

We have updated the description to clarify any misunderstandings by “All corrections applied to the observations, such as GIA corrections etc., are consistently selected across all datasets.”

L351: The number is odd. Maybe the authors want to express “$(1.34 \pm 0.12)*10^{22}$”. I recommend using “ZJ”.

We agree, thanks for spotting this. We are did not type the 10 before the exponent of 22. This happened with a careless replacement of a latex “unit”-command, not supported by the official template, throughout the manuscript before submission.

This has been corrected.

L354: Point out the computed period.

The period is 2005-01 till 2015-12, this has been added to the text.

L355: The number is odd.

See answer on L351 above.

Fig. 3: I do not trust the ocean reanalyses time series in this figure. For example, it seems the ocean reanalysis time series has a fake increase in 2024 in the Atlantic Ocean. The increase is induced by the number of ocean reanalysis ensembles. I recommend using the ocean reanalyses, which cover the full analysis period, such as CIAGR, SODA4 and ORAS5.

We agree this needs to be better explained what is pictured in Fig. 3. These residual numbers are heavily influenced by 1) the analysis period and 2) the fewer available products and, thus, ensemble members outside of the analysis period. As stated in the text we investigate two time period (2005-01 till 2015-12 and 2005-01 till 2024-12). For Fig. 3 trends accelerations and seasonal signals are derived over the first period for all estimates and then removed from the entire dataset. As discussed in the text, the majority of the data products used for comparison is not available over the longer 20y time period, which is the main motivation for investigating the shorter 11y period, which is covered by all data products. As a results of removing the trend/acc/seasonal derived from the 11y period from the entire dataset and in addition having fewer ensemble members will significantly impact the results, especially after 2016 leading to the residuals becoming artificially large. When analyzing the longer 20yr period (2005-01 till 2024-12) we consequently limit the number of data products involved to those covering the full time period.

We have briefly touched this in line 427 in the manuscript before, but we now also added this to the descriptions of figures 1 and 3 in the updated manuscript as “Note: The residuals are computed over the aforementioned period and removed from the entire time series, leading to artifacts outside of the 11\,yr period that is covered by all datasets. Outside of this analysis period and with fewer ensemble members residuals will become artificially large.”

L425: I think the authors should explain the Arctic time series with caution. The sea level observations in the Arctic is limited, which greatly reduces the reader confidence. The authors should point out limited observations here.

We fully agree with this and kindly refer to lines 425-435 in the original manuscript. However, not all data products suffer from a lack of data in these high latitude regions in the same way, as model data generally provide information here. This actually points out one of the potential benefits of our fingerprint inversion. As the fingerprint spatial patterns are assumed to be static over time, the global rescaling from the combined GRACE, altimetry and Argo observations nevertheless does provide "predicted" sea level change, budget and derived OHC in these high latitude regions; even if it is clear that these must be considered with caution.

We have extended our description of this paragraph to make this more evident by adding: "Space geodetic and model estimates are less affected due to high latitude observation coverage and the model domains including these ocean regions. For the inversion, the fingerprint spatial patterns are assumed to be static over time, the global rescaling from the combined GRACE, altimetry and Argo observations nevertheless does provide "predicted" sea level change, budget and derived OHC from GAAMR in these high latitude regions. However, these still need to be interpreted with caution. "

L565: I do not think so. The sea level budget is an indirect method to estimate OHU, and it is hard to be a baseline to judge other methods. If the authors can give evidence to show that the uncertainty of their method is smaller than that of observations, the description is acceptable. In Meyssignac et al., (2023) supplement, the Indian Ocean exhibited a great budget residual, and the residual is mainly to be included in the steric sea level with such a method. It is not the fact. The sea level and ocean mass observation also possibly contribute the budget gaps. That's why the authors get a higher trend in Indian Ocean.

Meyssignac, B., Ablain, M., Guérou, A., Prandi, P., Barnoud, A., Blazquez, A., et al. (2023). How accurate is accurate enough for measuring sea-level rise and variability. Nature Climate Change, 13(8), 796–803. https://doi.org/10.1038/s41558-023-01735-z

The statement in line 565 has been made in context of all available estimates not in terms of taking our GAAMR approach as the truth. We wanted to point out that some estimates stray quite significantly from the mean cluster (the mentioned ORAS5 and Glorsys models), while other estimates, including our GAAMR approach, seem to agree better to each other and to the CERES-based estimate. In this context, we used the word "overestimate" in the sense of "larger compared to other products" and not "relative to a reference or truth".

In the updated manuscript we have rephrased it to make it more clear that we talk about the magnitudes of the mentioned products in comparison to all other products. Please see also our answer to major comment 2.

Regarding the sea level budget residuals, there seems to be a misunderstanding here. We never discussed our SLB residuals in this context and thus never claimed all of it being related to steric error. In fact, we find very low residual trends all over the ocean from our inversion method, see Fig. B2 in the updated manuscript. In the updated manuscript we will include regional sea level budget results for 2005-01 till 2015-12 covering the major components (total, mass, steric and residuals) together with

their respective errors and compare these to other published estimates including those mentioned here.
1) We added a paragraph to the sea level budget discussion:

"Regionally, we find good agreement between our trend maps for 2005-01 till 2015-12 (Fig. B2) and recent publications (Royston et al., 2020; Jeon et al., 2021). When comparing to results by Bouih et al. (2025) we find similarities between the patterns of total sea level change and Argo-based steric sea level and our results (Fig. B2), despite utilizing a slightly different time period for the derived trends. However, for the ocean mass change component Bouih et al. (2025) report significant trend anomalies in the North Atlantic ocean, also been mentioned by Mu et al. (2024), where the same region is also found to exhibit large residual trends, when subtracting the sum of all components from the total sea level observed by altimetry. From our fingerprint approach (Fig. B2, C), we can not confirm these residual trend signals, similar to Jeon et al. (2021), who do not see these North Atlantic trend signals in the majority of their investigated products, including fingerprints, mascons and spherical harmonic solutions, except for the CSR spherical harmonic solution which shows slightly negative trends. EOF decomposition of our residual sea level (not shown here) reveals that there are no major trend signals left within our residual trends (Fig. B2, G), as these are dominated by the major current regions (Gulf stream, Aghulas current, etc.), which are not modeled as part of our fingerprint approach. Similarly, we do not find significant positive residual trends in the Indian ocean as have been observed over the 2005-2015 period by others (Meyssignac et al., 2023, suppl. figure 5)."

2) We refer to these results when discussing the larger estimates from some products, mainly found in the Indian ocean in section 4 as "Of course, differences in regional sea level budget modeling may also affect OHU results. As shown above, Meyssignac et al. (2023) found large residual trends in the Indian ocean, while our inversion results (Fig. B2, G) show significantly smaller trend levels and smaller scale features."

L575: The authors can add more discussion on the uncertainty of this method. And in my opinion, the authors should add the discussion about the regional sea level budget (Mu et al., 2024; Bouih et al., 2025). Also, the GIA correction in GRACE/GRACE-FO and limited observation of deep ocean (below 2000m) salinity/temperature are important source of uncertainty.

Mu, D., Church, J. A., King, M., Ludwigsen, C. B., & Xu, T. (2024). Contrasting Discrepancy in the Sea Level Budget Between the North and South Atlantic Ocean Since 2016. Earth and Space Science, 11(8), e2023EA003133. https://doi.org/10.1029/2023EA003133

Bouih, M., Barnoud, A., Yang, C., Storto, A., Blazquez, A., Llovel, W., et al. (2025). Regional sea level trend budget over 2004–2022. Ocean Science, 21(4), 1425–1440. https://doi.org/10.5194/os-21-1425-2025

While we agree that an extensive discussion of our global and regional sea level budgets would be quite interesting, this might be a little bit out of scope of the current manuscript focusing on OHC. Nonetheless, we have added trend maps of the major components (total, mass, steric and residual) to our appendix (Fig. B2) and compare these to published estimates, including those references mentioned here. Please see the answer to the previous comment above.

We agree that we should have discussed the method uncertainty and the propagated error in the previous manuscript version, which we have now done in the updated manuscript. We mention here that GIA correction and limited observability of the deep ocean hinder all approaches, not only ours. We believe it is actually a strength of the least squares inversion to provide error measures; a simple residual approach (deep steric = altimetry – GRACE – shallow steric) does not provide error bars.

1) We have added a figure (Fig. 7 in the updated manuscript) showing the regional OHU and OHU trend errors and a short description to our regional OHC discussions: "By propagating the formal errors of thermosteric sea level from our inversion output throughout our GAAMR method and also assuming a 5% ORAS5 reanalysis error level during conversion to OHC, regional errors of OHU and OHU trends are shown in Fig. 7. These represent the errors corresponding to the GAAMR results in Figs. 5 and 6. Largest errors are detected in major current regions, including the Gulf steam, the Aghulas current and the Kuroshio current. This is expected as thermosteric variability in these regions is large due to eddy and gyre presence."

2) At the end of the discussion section we have also added discussion on the errors and their different sources as

"Major current regions, such as the Gulf stream, Agulhas or Kuroshio current, are found to exhibit the largest uncertainties in OHU. These are expected to a certain extent as these regions also exhibit the largest residual signals (Fig. B2, G), as the highly temporally variable currents and eddy regimes are not well modeled by the static fingerprint approach. Besides propagating the measurement errors of the input datasets itself, combination of inconsistent signals may add to the uncertainties due to differences in corrections applied to individual datasets. However, we preprocessed all input datasets to be as consistent as possible in terms of applied corrections, e.g. utilizing the same GIA model for correcting GRACE(-FO) and satellite altimetry. A third part of the uncertainty budget, which is difficult to quantify, directly relates to the fingerprints employed in our approach, as limited representativeness of the fingerprints directly influences the separation of the total sea level into mass and steric components, which in turn translates to a direct influence on our GAAMR results. Finally, the involved model data, in this case ORAS5, does not provide error information and we simply assume a 5% model error during error propagation within the GAAMR processing. In summary this means that our GAAMR results, while showing good agreement with other datasets, are still affected by different error sources, where some have reduced impact due to meticulous processing."