The authors thank reviewer#1 and #2 for their constructive comments and evaluation on our study, here the point-to-point replies are provided in blue, the comments are in black, the modified texts for the manuscript are shown in orange.

Referee #1's comments and replies:

This manuscript presents the latest edition of the IAP OHC data set, including a comprehensive description of methodological advancements and evaluation. The paper is informative for developers of similar data sets as well as users and will be a useful reference for the community. It is generally well written, but I have a number of comments and suggestions for clarification and improvement. In addition, the section on the sea level budget appears half-baked to me. That section requires more substantial revision and elaboration, or it could possibly be removed given the paper is already quite long.

Re: Thanks for the evaluation. We appreciate your helpful comments and suggestions. We have tried to address all of them. Please find the replies and the revisions introduced below. We are grateful that these comments helped improve our paper's quality.

Regarding your specific concern about the sea level budget section, we have decided to maintain it as it provides useful metrics for evaluating the dataset. However, as we agree that the material needs major work, we have rewritten it to illustrate better the impact of IAPv4 on sea level budget closure.

Specific comments:

L44: The authors postulate consistency of IAP OHC data with EEI. Given the only moderate correlation and the only modest visual agreement in Fig. 17 the authors should specify based on which metric they conclude "consistency".

Re: This sentence has been revised to be more specific for the consistency: "the trend of ocean warming rate (i.e., warming acceleration) is more consistent with the net energy imbalance at the top of the atmosphere than IAPv3", so it mainly refers to the trend of EEI (trend of ocean warming rate)

L328-331: I understand that even post 2005 monthly data are actually based on 3-month windows. This is important to note more explicitly. It also has implications for the variance of the time series as presented in table 2 – in fact this method will reduce the monthly variance compared to other data sets which might represent truly monthly data. Regarding time- and depth varying windows: Could this be illustrated with a time-depth Hovmoeller diagram displaying the employed window length? Also, was the impact of time-varying window length on signals assessed with synthetic data?

Re: This is a good point, instead of the Hovmoeller diagram, we included this information as Supplementary Table 1 to list the time- and depth-varying windows. A table is used instead of a figure to increase the transparency.

This method will reduce the monthly variance in temperature and OHC time series, but it is an improvement compared with other datasets (we respectfully disagree with the statement that other data sets might represent truly monthly data). As noted in Trenberth et al. (2016) and other studies, the month-to-month in all OHC datasets explored in that study is likely spurious compared with CERES data. Boyer et al. (2016) and others (e.g. Meyssignac et al. 2019) also noted that the current observation system is still too sparse to monitor the physical month-to-month variability: too high noise level. Thus, it is desired to reduce the month-to-month variability. Besides, the increase of windows with depth is also physically meaningful because of the reduced temporal variability with depth. Furthermore, a careful investigation of inter-annual variability (ENSO) of OHC by Cheng et al. (2019) also indicates that the choices give reliable estimates of inter-annual variability compared with other data (e.g., Roemmich and Gilson 2011). A sentence added in section 2.6 "The use of a time window will reduce the monthly variance compared to other datasets, which is likely too high compared with independent Earth's Energy Imbalance data at the top of the atmosphere (Trenberth et al. 2016)."

References:

Trenberth, K. E., Fasullo, J. T., Von Schuckmann, K., and Cheng, L.: Insights into Earth's Energy Imbalance from Multiple Sources. J. Climate, 29, 7495-7505, https://doi.org/10.1175/jcli-d-16-0339.1, 2016.

Roemmich, D., and Gilson, J.: The global ocean imprint of ENSO. Geophys. Res. Lett., 38, https://doi.org/10.1029/2011GL047992, 2011.

Cheng, L., Trenberth, K. E., Fasullo, J. T., Mayer, M., Balmaseda, M., and Zhu, J.: Evolution of Ocean Heat Content Related to ENSO. J. Climate, 32, 3529-3556, https://doi.org/10.1175/jcli-d-18-0607.1, 2019.

Boyer, T., Domingues, C. M., Good, S. A., Johnson, G. C., Lyman, J. M., Ishii, M., Gouretski, V., Willis, J. K., Antonov, J., Wijffels, S., Church, J. A., Cowley, R., and Bindoff, N. L.: Sensitivity of Global Upper Ocean Heat Content Estimates to Mapping Methods, XBT Bias Corrections, and Baseline Climatologies, J. Climate, 29, 4817–4842, https://doi.org/10.1175/JCLI-D-15- 0801.1, 2016.

Meyssignac, B., Boyer, T., Zhao, Z., Hakuba, M. Z., Landerer, F. W., Stammer, D., Köhl, A., Kato, S., L'Ecuyer, T., Ablain, M., Abraham, J. P., Blazquez, A., Cazenave, A., Church, J. A., Cowley, R., Cheng, L., Domingues, C. M., Giglio, D., Gouretski, V., Ishii, M., Johnson, G. C., Killick, R. E., Legler, D., Llovel, W., Lyman, J., Palmer, M. D., Piotrowicz, S., Purkey, S. G., Roemmich, D., Roca, R., Savita, A., Schuckmann, K. von, Speich, S., Stephens, G., Wang, G., Wijffels, S. E., and Zilberman, N.: Measuring Global Ocean Heat Content to Es- timate the Earth Energy Imbalance, Front. Mar. Sci., 6, 432, https://doi.org/10.3389/fmars.2019.00432, 2019.

L343: Does that mean the influence radius changes with depth? Is this a physically based choice or is this pragmatic owing to data availability?

Re: The influencing radius changes with depth for the first iteration (2,000 km for the upper 700 m and 25,000 km at 700–6000 m), and no change for the second and third iteration (800 km and 300 km). This is based on a test provided by Cheng&Zhu (2016) paper (their Fig.3), which subsamples the recent decadal data based on past data locations to test the different choices of the influencing radius. The results show that a ~20-degree influencing radius can minimize the

reconstruction error for the upper 700m. The 700-2000m radius is further determined by Cheng et al. (2017) with the same approach.
Although it is more of a statistical result, it is physically meaningful because the spatial decorrelation distance is longer in the deeper ocean than in the upper ocean, and the decadal to multi-decadal variability of ocean temperature is generally associated with large spatial patterns.

L345-346: "real forcings" is perhaps a bit overconfident. Better say something like "reconstructed".
Re: You are right. We removed the "real forcings", so the sentence reads as follows: "For each month, IAPv3 used 40 model simulations (historical runs) from the Coupled Model Intercomparison Project phase 5 (CMIP5) to provide a flow-dependent ensemble……"

L360: Please explain E_i and M. If "E" is instrumental error, is it meant to represent a bias (which could simply be subtracted) or random error?
Re: Good point. We have included an explanation and basic assumptions for this error: "***Ei*** is the instrument's precision for each individual observation, assuming random error (the basic assumption is that after bias correction, the systematic errors can be eliminated).".

Table 1: Instead of saying doi: "YES" I suggest to simply state the doi.
Re: Done, doi provided here.

L416-423: As a reader I would like to see a number of how strong the effect of the VC on OHC trends is (especially as a number from earlier works is provided).
Re: Good point. The impact of VC on global and basin OHC is tested here. Some texts have been added in the revised manuscript: "Since the open ocean accounts for the vast majority of the global ocean volume, the influence of the VC method on the global OHC trend is small. For example, the upper 2000 m OHC trend with VC is ~0.15% (~0.45%) smaller than without VC from 1958-2023 (2005-2023) for IAPv4. However, it can significantly affect regional OHC estimates, especially in regions with complex topography. For example, the Maritime Continent region's 0-2000 m OHC trend is reduced by 6.9% (4.2%) after applying VC from 1958-2023 (2005-2023) (Jin et al. 2024)."
Moreover, a previous study (Jin et al. 2024) indicated that the VC could be very important for some data products, for instance, EN4, thus this correction is recommended, I quote from Jin et al. 2014: "*The VC fixes the overestimated volume of EN4 and thereby reduces its OHC variability (Figure 1b; standard deviation adjusted from 2.50 to 1.73 ZJ), achieving a better agreement with those of IAP and Ishii (1.50 and 1.47 ZJ; Table S1 in Supporting Information S1). In this sense, the VC also reduces the uncertainty in OHC variability.*"

Reference:
Jin, Y., Li, Y., Cheng, L., Duan, J., Li, R., & Wang, F. (2024). Ocean heat content increase of the Maritime Continent since the 1990s. Geophysical Research Letters, 51, e2023GL107526. https://doi.org/10. 1029/2023GL107526

Fig. 6: Here and in other instances I suggest to bring in the Lyman and Johnson (2023; LJ2023) data (https://doi.org/10.1175/JTECH-D-22-0058.1) as they state generally improved quality over IAPv3 data. Good agreement with the LJ2023 data (which are derived in a different fashion than the IAP data) would strengthen the confidence in state of the art OHC data sets.

Re: Thanks for the suggestion; we have added Lyman and Johnson 2023 data into Fig.6 and also included some brief discussions about LJ2023 data. Lines 518-527 are rewritten

"The annual cycle of the OHC above 2000 m of IAPv4 is compared with IAPv3, ISH, EN4, RG and RFROM (Fig. 6 and Fig. 7) for 2006–2020. There is a consistent annual cycle among different datasets for the global and hemispheric oceans. Globally, the ocean releases heat from boreal spring to autumn and accumulates heat from boreal autumn to spring, which is dominated by the southern hemisphere due to its larger ocean surface area (Fig. 6). The two hemispheres show opposite annual variations in OHC, associated with the annual change of solar radiation and different distribution of land and sea. For the global OHC above 2000 m, IAPv4 shows a positive peak in April and a dip in August, with the magnitude of OHC variation of 60.4 ZJ for IAPv4 (66.9 ZJ for IAPv3), consistent with other datasets: 53.2 ZJ for ISH, 58.1 ZJ for EN4, 69.2 ZJ for RG and 56.6 ZJ for RFROM (where 1 ZJ = 10^{21} J)."

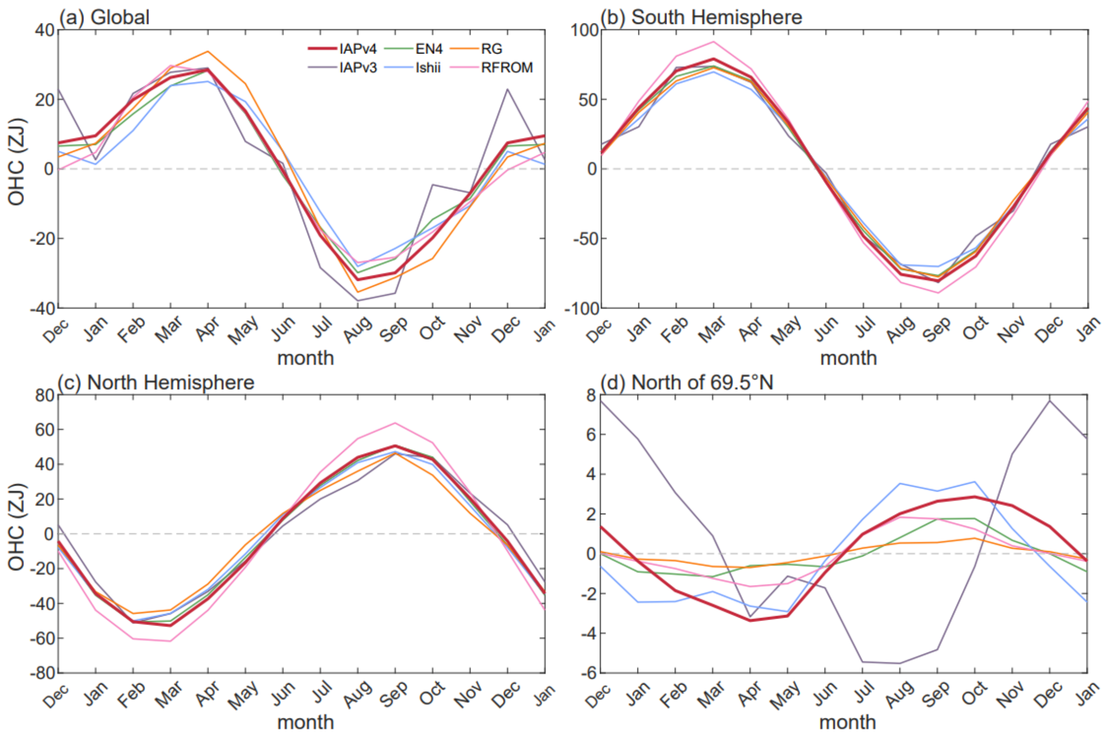

Figure 6: Annual cycle of OHC of upper 2000 m for (a) the global oceans, (b) the Southern Hemisphere, (c) the Northern Hemisphere and (d) the oceans north of 69.5°N. Six different data products are presented, including IAPv4 (red), IAPv3 (blue), ISH (purple), EN4 (green), RG (orange), and RFROM (pink).

Fig. 7: IAPv4 clearly looks more plausible than v3, but it is still only qualitative. Is there a way to really validate the data, e.g., with data from ice-tethered profilers?

Re: This is difficult because we merged all available data into IAPv4, so the ice-tethered profilers are not independent. Nevertheless, we have done two additional checks:

1) compare the IAPv4 data with some independent data found in the central Labrador Sea. A paragraph is added: "Furthermore, the reconstruction of IAPv4 is compared with completely independent observations in the central Labrador Sea (see Data and Methods section for details; Yashayaev, 2007; Yashayaev and Loder, 2017) for the 200-2000 m mean temperature time series (Fig. 15). The direct observations show a substantial decadal variation in the central Labrador Sea, with negative anomalies 1970-2003 and 2015-2020, and positive anomalies 1963-1972 and 2004-2014. Reconstructed based on data from WOD, IAPv4 can well represent this decadal variability. The largest difference occurs in 1989, where direct observations show nearly zero anomaly while IAPv4 shows a big negative anomaly; this difference is likely caused by using a time window in IAPv4, which has a smoothing effect on the time series."

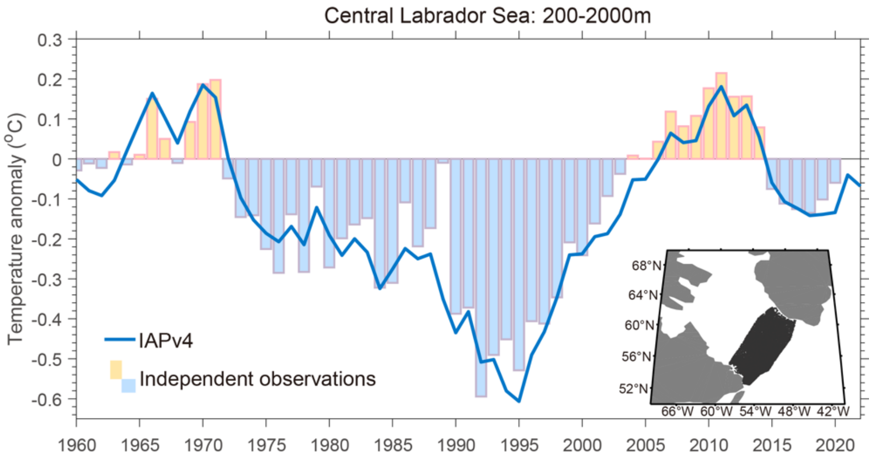

Figure 15: Comparison of IAPv4 data with independent observations in the central Labrador Sea (304-310 °E, 55-61 °N) from 1960 to 2020. The 200-2000 m averaged temperature anomaly time series is shown, and the baseline is 1960-2020. The inner box shows the locations of the independent observations in black dots (showing a total of 49849 profiles).

2) Provide the gridded averaged observations without any interpolation and calculate the RMSE between IAPv4 and gridded averages, IAPv3 and gridded averages. This comparison is more quantitative. See the updated Fig.7. It is evident that IAPv4 shows smaller RMSE than IAPv3 in the polar regions.

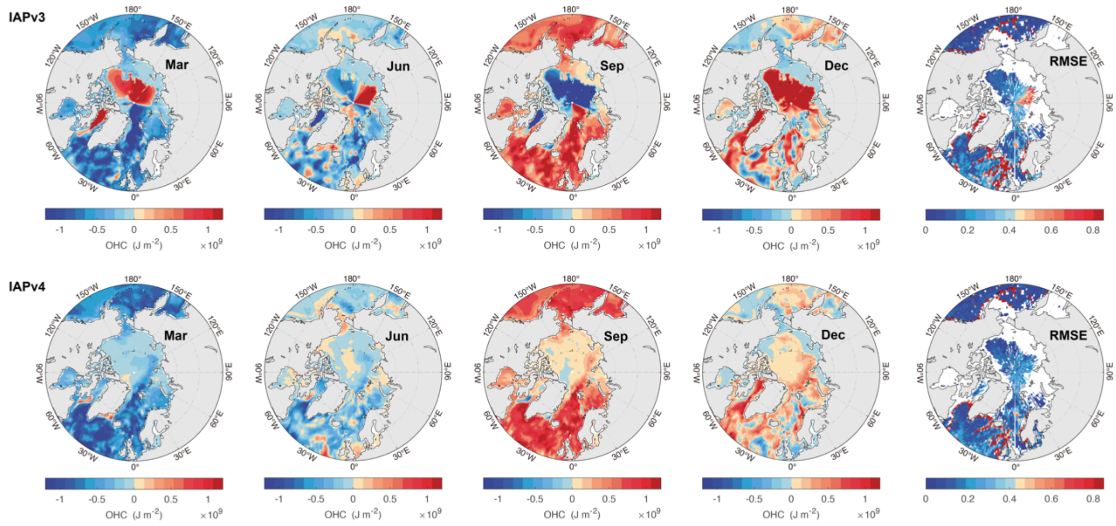

Figure 7: Seasonal distribution of monthly mean upper 2000 m OHC anomalies and root mean square error (RMSE) of OHC 0-2000 m between gridded data and in situ observations. For OHC anomalies, four months are shown: March, June, September, and December. The OHC anomalies are relative to the 2006 – 2020 annual mean. The upper and lower panels are for IAPv3 and IAPv4 products, respectively. The panels in the last column are for annual RMSE for IAPv3 (upper) and IAPv4 (lower), respectively.

L575-584: Given that cooling SST trends in the eastern equatorial and south-eastern Pacific have received quite some attention recently, I suggest to discuss more potential causes of the non-existing cooling trend in that region in the IAP data.

Re: Thanks for the suggestion. We have explored a bit more the SST trends in the two regions you mentioned (Fig. X). From the time series, it appears that the tropical SST changes show substantial variability from inter-annual (mainly ENSO) to decadal (mainly PDV) scales. The cooling trends from the 1990s to the present are mostly associated with the PDV phase change. Thus, the long-term trends (e.g. from the 1950s to the present) are subtle: indeed, according to our calculation, the tropical SST trends within 1955-2020 are mostly not significant (based on our LOWESS-based approach in Cheng et al. 2023). Based on these results, we tried not to delve into the explanation of the trends in Fig.10, instead, we added a discussion suggesting that the trends are mostly not significant because of the strong inter-annual to decadal variability.

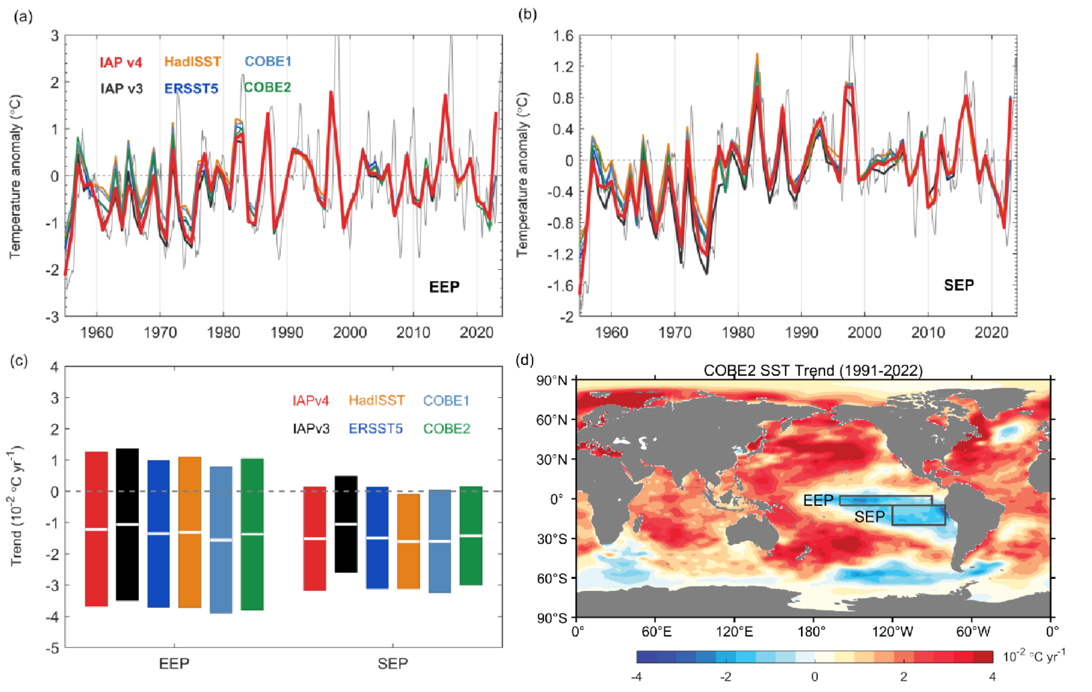

Fig.X2. SST time series (a, b) and trends (c) in two regions denoted in the boxes in (d). Two regions: Equatorial Eastern Pacific (EEP) and Southeast Pacific (SEP) are shown.

Fig. 11: This is a good figure, but in addition I would like to see a figure including other OHC data sets (similar to Fig. 9 for SST).

Re: A new figure has been created (Fig.12) in the revised manuscript to compare the IAPv4 time series with other data based on the plot shown in the Cheng et al. 2022 review paper. Some data have been updated to 2023. It is evident that despite the differences in datasets, IAPv4 shows a stronger long-term trend since the 1960s than almost all datasets, showing the impacts of XBT/MBT/Bottle bias corrections. Also, stronger warming occurs for IAPv4 than IAPv3 and many other data, mainly because of the QC.

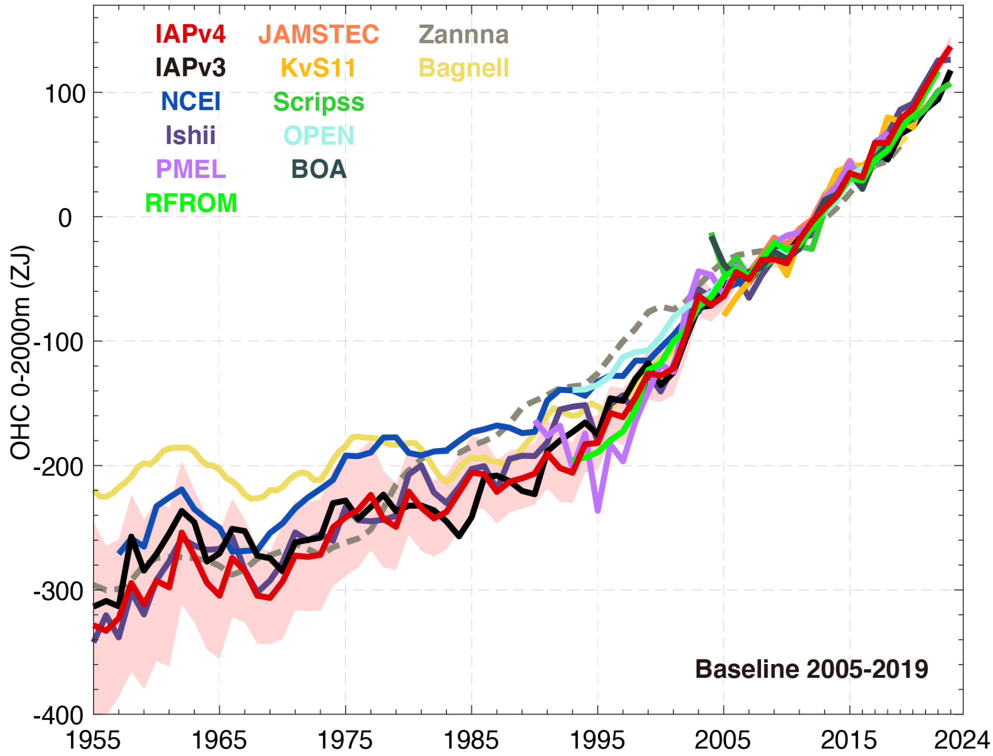

Figure 12: A comparison of annual mean OHC 0-2000 m time series from different data products. Solid and dashed lines represent direct and indirect estimates, respectively, and shading indicates the IAPv4 90% confidence interval (pink shading). OHC anomalies are relative to a 2005–2019 baseline. The plot is updated from Cheng et al. (2022a).

Table 2: The annual results suggest a variance ratio between IAPv4 and CERES of >2, while Lyman and Johnson (2023) get a ratio of 1.3 for their data. This should be mentioned. A reason for this might be the fact that LJ (2023) seem to actually apply a stronger than annual smoothing (their annual OHC variation is obtained by differencing subsequent annual means) to their data, but reading your lines 792-793 it appears you are doing the same for this comparison? Needs to be clarified.

Re: We have mentioned this issue in the revised manuscript "In addition, Lyman and Johnson's (2013) data suggest a yearly variance ratio of 1.3 between annual RFROM and CERES data from 2008 to 2021. Using the yearly mean OHCT indicates a ratio of 1.4 at the same period between IAPv4 and CERES, which is similar to that of RFROM.".

We mark here that in Table 1, we have applied a 12-month running smoother to all-time series, while Lyman and Johnson (2023) used an annual time series (for their 1.3 ratio result). That's why we have made additional calculations using annual time series of IAPv4 and CERES data to derive the ratio, which is 1.4, similar to Lyman and Johnson 2023 data.

Table 2: Why is no CERES trend provided?

Re: OHC trend in CRESE record is adjusted by Johnson and Lyman dataset (0.71 Wm^{-2} within 2005-2015) as described in Loeb et al., (2018). Here, we calculate the mean CERES net flux within

2005-2022, consistent with the period of other data, and assuming 90% of heat stored in the ocean. This number is added to the table as 0.77 Wm^{-2}. A sentence has been added in figure caption of the revised manuscript "The OHC trend for CERES is calculated as the mean of net TOA radiation flux within 2005–2022 multiplied by 0.9, assuming 90% of the EEI stored in the ocean."

Fig. 13: Please comment why the eastern Pacific cooling signal in upper 300m OHC is so much more prominent than in SST?

Re: Thanks. That is because Figure 13 is for the 1991-2023 period, but the SST figure in Fig.10 is for a longer period of 1955-2022. The previous plot, Fig. X2, shows the big inter-annual and decadal-scale variability in SST, so the trends for different time periods are different. For OHC, there is also substantial inter-annual to decadal scale variability (Fig. X3, below), so if one calculates the 1991-2023 trend, the eastern Pacific shows a negative trend for OHC0-300m.

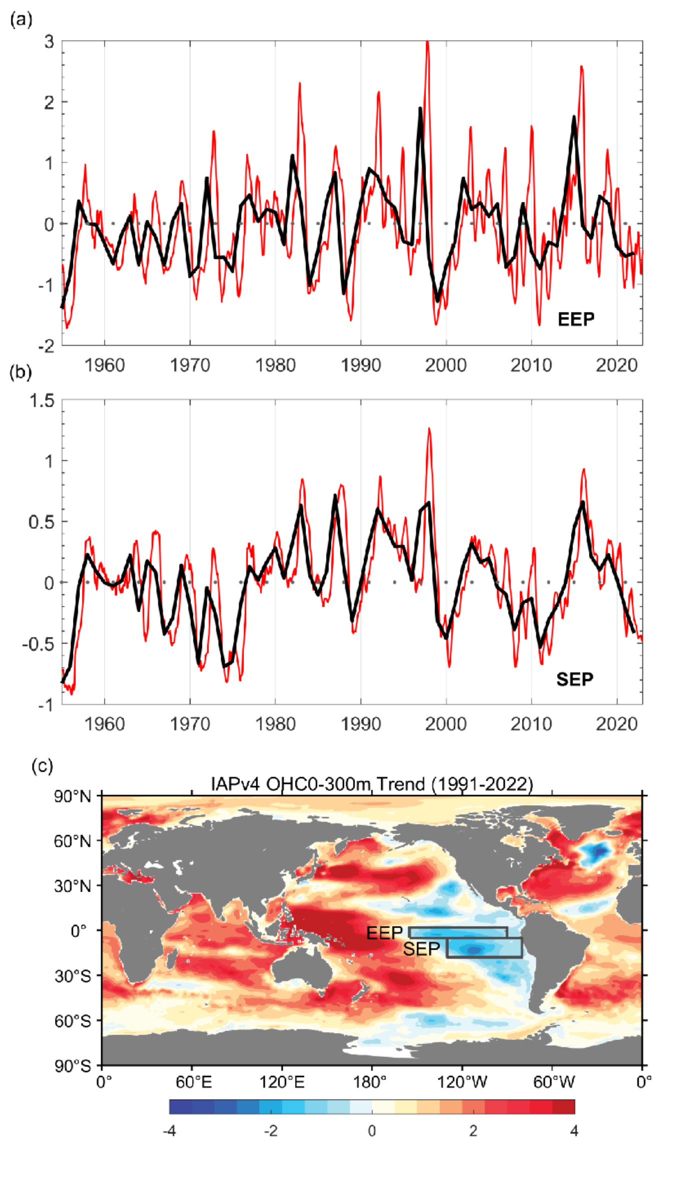

Fig.X3. OHC0-300m time series in the Equatorial Eastern Pacific (EEP) and Southeast Pacific (SEP) regions in (a) and (b) respectively. (c) shows the spatial trend map for OHC0-300m from 1991 to 2022.

L695: This is only true for the tropics, not for higher latitudes.

Re: Thanks, to avoid potential issue of the quantification, this sentence is modified to "The ocean meridional heat transport (MHT) is fundamental to maintaining the earth's energy balance."

L730ff: Please add "Pacific" everywhere (also when stating correlations) to be clear you are not discussing full zonal averages

Re: Done.

L759: add "based on our data"

Re: Done.

L768: to me "EEI" is a rate of change, not an accumulated value. So maybe better to add "accumulated" before "EEI"

Re: Yes, modified.

L772-773: validation method (1) does not appear very meaningful to me, as the integrated CERES value depends on the one-time global adjustment for the EBAF product. Changing this adjustment to match the IAPv4 average OHC increase (as apparently done by the authors) does enforce the agreement seen in Fig. 16. I am unconvinced this is a meaningful approach and recommend to only keep method (2).

Re: Yes, the trend of the accumulated EEI depends on the adjustment, but the variation of the accumulated EEI does not, that's why the RMSE between accumulated EEI and IAPv4 time series is meaningful, which could be a useful metric for the EEI/OHC consistency in unit of ZJ (heat, not rate of change). This is the reason why we prefer to maintain both approaches. We believe this explanation makes things clearer.

L795: "consistency" in which sense? E.g., is the correlation significant?

Re: The sentence has been modified to "The IAPv4 and CERES estimates show inter–annual variability with a correlation of 0.44. The higher correlation of IAPv4 versus CERES than IAPv3 increases confidence for the new data (correlation of only ~0.15 for IAPv3)."

Fig. 17: The CERES series does not look de-trended.

Re: It is not de-trended. We stated in the figure caption that "The long–term mean is removed for all-time series.", so the baseline is removed, but the trend remains.

L815ff a: The methods are not clear to me. In my understanding, only steric sea level can be derived from IAPv4 data. It would be useful to note how the conversion from T/S profiles to sea level is performed? A reference would be useful.

Re: This has been made clearer in the revised manuscript: "The updated IAPv4 data is used to assess the sea level budget for 1960-2023 in combination with other data, including IAP salinity data, glaciers, Greenland, Antarctic ice sheets mass loss from Frederikse et al. (2020) and altimetry sea level record (see Methods section for details).".

And the description of the conversion from T/S to steric sea level is introduced in the Methods section: "To derive steric sea level, IAP salinity data is used (Cheng et al. 2020). The temperature and salinity data are converted to steric sea level based on the Thermodynamic Equation Of Seawater – 2010 (TEOS-10) standard (McDougall and Barkerv, 2011). "

L815ff b: It is confusing in table 3 that for all terms there is an IAPv4 entry, although everything but steric stems from other sources. Specifically, I understand that GMSL as well as "sum of contribution" (which itself should be explained better) are taken from Frederikse et al., but the values still differ. What is the reason for this? Is it only because of the different approach for trend computation? Can the authors provide the sensitivity to the trend estimation method based on IAP data?

Re: We agree that Table 3 is confusing. The sea level section has been rewritten and the Table 3 has been removed. In the revised manuscript, we have added two clean tests on the impact of IAPv4 on sea level budget closure: 1) we replace the steric sea level component in Frederikse et al. (2020) by IAPv4 and then test the residual error and RMSD; 2) we replace the thermosteric sea level component in IPCC-AR6 (Gulev et al. 2021, they do not have a halosteric sea level estimate) with IAPv4 and check the residual error and RMSD.

Because the other components are the same and the only difference is the steric sea level data, these two tests can isolate the impact of the new T/OHC data on sea level budget closure.

L815: 1991 or 1993?

Re: It is 1993, corrected.

L852-854: This is an important result, and it seems to suggest that the stronger warming in recent years as indicated by IAPv4 is more realistic. This should be stated more clearly. Also, here it would be useful to make a link to Fig. 11 or a potential new OHC figure including other OHC products.

Re: Great point, we added some discussion at the end of this paragraph "This suggests that the stronger warming in recent years, as indicated by IAPv4, is more realistic. As discussed in Section 3.4, the QC is mainly responsible for the increased warming of IAPv4 compared with IAPv3 since ~2015 (Fig. 11).

Many traditional QC procedures use a static climatological range check to filter out outliers, which does not account for the increase of extreme events with climate change and removes too many extreme (positive) values during the recent period. Thus, we strongly recommend that data product generation groups revisit the QC procedure. Furthermore, as the stronger long-term OHC trends since ~1960 in IAPv4 than in IAPv3 are mainly attributed to the bias corrections for Nansen Bottle, MBT, and XBT data, it is also recommended that international groups to revisit the biases in ocean observations."

L916: Is there a reference for the sampling bias of CERES on monthly time scales?
Re: A reference "Loeb et al., 2009" is added. The issue is the number of samples per day to get the diurnal cycle and hourly variations in clouds. It was clearly inadequate in the early days, with 2 samples per day. After 2002 or so, there were 4 per day, but that was also inadequate for clouds. They have a major imbalance that is not physical. Part of that may be a diurnal cycle, but part of it is likely errors in angular corrections, etc. Loeb et al. (2009) stated the issue, and later, the CERES group just used OHC for adjustment, but the bias is real.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget. J. Climate, 22, 748–766, https://doi.org/10.1175/2008JCLI2637.1.

Typos/edits:

L41: suggest replacing "first" with "uppermost"
Re: Modified
L107: "support the follow-on studies on climate assessments" – This does not read very smooth. Please rewrite.
Re: Modified to "Upgrading the product with new developments is important to better support the ocean/climate research and climate assessments.".
L137: "this paper" à perhaps better say "the presented product"
Re: Modified
L222: is it a warm or cold bias?
Re: The XBT bias is generally positive but with time variation. The statement is modified to: "The XBT bias was found to be generally positive, as large as ~0.1 °C before 1980 on the global 0–700 m average, diminishing to less than 0.05 °C after 1990".
L232: "systematic biases" is redundant: either say "biases" or "systematic errors"
Re: Revised to "biases".
L243: here and in several other instances the references have the parentheses wrongly placed. Please revise.
Re: Corrected.
L270: I am not sure that "adjustive" exists
Re: Modified to "adjusted".
L282: It is unclear what you mean by "such a choice"
Re: Sentence modified to "Recent developments from other groups, such as Li et al., (2022), include the choice of a short-period climatology.".
L303-304: This alone is not necessarily a problem. But I assume salinity data are not always available?
Re: This sentence has been removed in the revised manuscript.
L435: I assume "monthly" climatology?

Re: Yes, “monthly” added here.

L462: please spell out “CERES-EBAF” once.

Re: Yes, “CERES-EBAF” spelled out here “Clouds and Earth's Radiant Energy Systems (CERES) Energy Balanced and Filled (EBAF)”.

L663: Please add a reference to the subsection where this is explained

Re: Added.

L937: What is meant by “T/OC”? Do you mean T/OHC?

Re: Yes, corrected to “T/OHC”

Referee #2's comments and replies:

The manuscript presents the description of the technical methods employed for the creation of the temperature and ocean heat content estimate IAPv4 and a basic assessment of the product in comparison to some other products. Additionally independent data such as sea level change or meridional ocean heat transport are employed to verify the product.

IAPv4 is an update with respect to its predecessor IAPv3 and a great deal of the manuscript is dedicated to the changes and impact between these different products as the reader would expect to see.

The manuscript is well written and lacks only few information. Detailed comments and suggestions are as follows:

Re: Thank you for the evaluation. We have addressed all your comments, which greatly improves the quality of this study.

L 55 I assume "based on gridded products" is more appropriate

Re: Revised.

L 78 Maybe a newer citation to point at the current product.

Re: For their gridded time series dataset, the NCEI/NOAA group did not have an updated reference, unfortunately.

L 124-125 Would it be possible to have for Fig.1a something like observed number of grid cells/months in addition to the casts? Fig.1a suggest the dominating importance of GLD while they provide very high resolution (in time and space) data which your product is not really be able to benefit from so much.

Re: Great point. We have added a new panel, Fig. 1b, to show the statistics of the observed number of grid cells by each instrument (see the figure below: Fig. R1). The new panel complements the current ones. Some texts are added in the revised manuscript "MBT, XBT, Nansen Bottle and CTD data are the major instruments before 2000 (Fig. 1a, b). The spatial coverage of these data increased to >30% in 1960 and >70% in the late 1960s for 1° × 1° × 1-year resolution. After 2005, there is a huge number of GLD and APB data, and they are mainly distributed in the polar regions (APB) and coastal regions (GLD) (Fig. 1a), their spatial coverage is usually less than 5% for 1° × 1° × 1 year resolution. By contrast, the Argo data cover most of the global open ocean since ~2005 (Fig. 1b)."

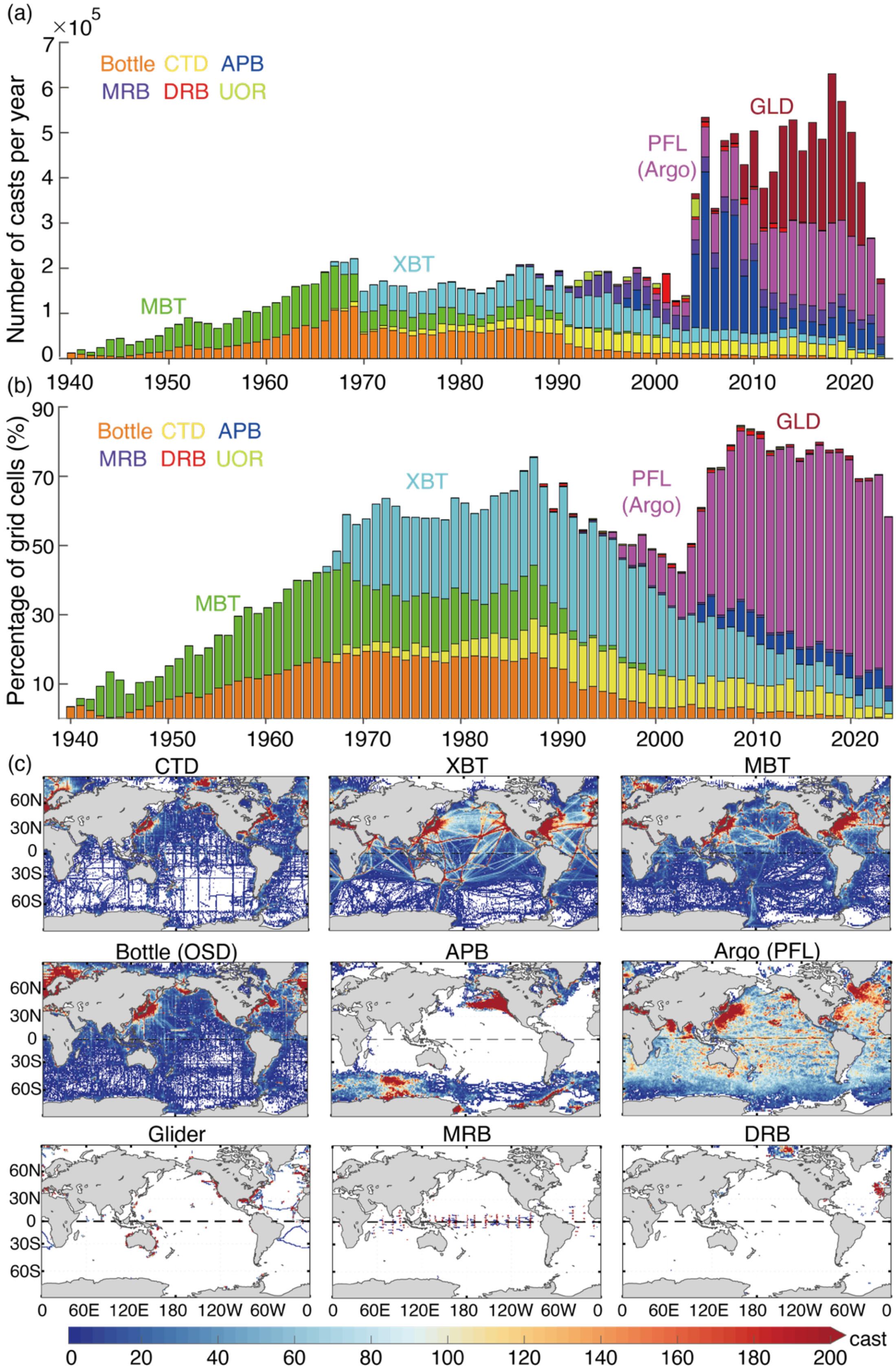

Figure 1: (a) Yearly number of temperature casts for different instruments; (b) percentage coverage (%) of ocean data for each instrument, which is calculated by the ratio between the number of 1° × 1° × 1 year grid cells observed by each instrument and the total number of

ocean grids; (c) number of subsurface temperature casts in 1-degree grid box from 1940 to 2023 collected by different instruments: CTD (Conductivity/Temperature/Depth), XBT (eXpendable BathyThermographs), MBT (Mechanical BathyThermograph), Bottle, APB (Animal mounted Pinniped Borne), PFL (Profiling Floats, i.e. Argo), GLD (Glider), MRB (Moored Buoy), and DRB (Drifting Buoy).

L 137 Which are the sources. That may be interesting to know for users that are looking for data.
Re: Thanks. We have to refer to our recent publication of the CODC-v1 dataset, where the sources of data are referred to and discussed. A dedicated paper will be published with respect to this dataset. The following texts are added in the revised manuscript "To complement the WOD with relatively less data in the Arctic and coastal regions of the Northwest Pacific, this presented product also uses data from other sources. The majority of these data are from the Chinese Academy of Sciences Ocean Science Data Center (Zhang et al., 2024), and some data are rescued from the old documents of marine survey. All these data will be publicly available. There are a total of 85,990 additional temperature profiles, about 0.50% of the data, which is expected to improve the reconstruction in these data-sparse regions (compared with IAPv3 and other products)."

Zhang, B. et al. CAS-Ocean Data Center, Global Ocean Science Database (CODCv1): temperature. Marine Science Data Center of the Chinese Academy of Science, doi:10.12157/IOCAS.20230525.001 (2024).

Fig.1 Define GLD
Re: Done: "GLD (Glider)" in the caption.

L 332-336 Often when too small influence radii are used, the anomalies may become zero and reconstructions fall back to climatology. This can be seen for instance in the earlier years of the EN3 objectively analysed fields. Do you have mechanisms to prevent this from happening, or are zero anomalies being accepted in case of lack of data. How frequent would that happen?
Re: Yes, this is often called "conservative bias" because many analyses (such as EN3, EN4) are not truly global analyses; in large data gaps, climatology (zero anomalies) is infilled, so the long-term warming trends have been under-estimated (Durack et al. 2014; Cheng et al. 2014, 2017, 2019). Durack et al. 2014 estimated that the underestimation could be 24–58% for global OHC, depending on gap-filling approaches. IAP analysis resolves this issue through several strategies (as fully described in Cheng&Zhu 2016; Cheng et al. 2017), some of them are mentioned in the texts, here just a brief introduction:
(1) A localization strategy is applied. The WOA/Ishii/EN4 method uses a radius of less than 900km. Instead, IAP uses 20 degrees for an influencing radius within 0-700m (25 degrees for 700-2000m). The large fractional coverage helps ensure that a near-complete global reconstruction can be reached, so the technique will not bias the reconstructed field toward the first-guess field in data-sparse regions.

(2) Previous products (WOA, EN4 and Ishii) have parameterized the background error correlation between two points as a function which decays in an exponential-like manner with the distance separating the points. This parameterized correlation is always isotropic, however, the covariance should be flow-dependent in the real ocean. The IAP product uses covariance from CMIP5 multi-model simulations. The models have the capability to simulate the general ocean circulation and could provide a better representation of the covariance.

(3) The use of time window to combine several months data together for a monthly estimate. Variable time windows (larger than one month) are used for monthly reconstructions to ensure a truly global analysis (Supplementary Table 1).

IAP data used the above-mentioned strategies to prevent the "conservative biases" and other errors in gap-filling process. Furthermore, a subsample test, in which subsets of data in the data-rich Argo era are co-located with locations of earlier ocean observations, is performed to quantify the sampling error. The subsample test is defined as the difference between the reconstructed and "truth fields". The truth field is taken to be a set of the gridded averaged temperature anomalies during the Argo era. Each truth field is subsampled according to the locations of historical observations and mapped to get the reconstructed fields. The IAP product is evaluated by this subsample test, showing an unbiased mean sampling error and with ocean temperature (or OHC) variability on decadal and multi-decadal timescales that can be reliably distinguished from sampling error.

L 345-350 Unclear what flow-dependent means and how the constraint with observations work, more information is needed here. How do you diagnose which type of flow is present when applying the flow dependent covariances or is this basically just done according to the location?

Re: The flow-dependency is ensured because CMIP5 model simulations are used, which can much better represent the ocean dynamics than traditional statistical Gaussian covariances (used in WOA, Ishii etc.). This is not explicitly parameterised because of the complexity of the covariances.

Optimization is achieved through an Ensemble Optimal Interpolation approach, where observations are combined with the CMIP5 model ensemble to estimate the minimum variance. The detailed formulation can be found in Cheng&Zhu 2016 and the Supplementary material of Cheng et al. 2017.

L360 What is E and i?

Re: It is our oversight. Ei is defined in the revised manuscript "***Ei*** is the instrument's precision for each individual observation, assuming random error (the basic assumption is that after bias correction, the systematic errors can be eliminated)."

Fig 4 "Variance" probably should read standard deviation since the unit is deg C

Re: Yes, it is true; we change "The unit is degree Celsius" to "The unit is °C^2 ".

L 498 What is the relevance of the different land-sea distribution. Maybe you want to point to the amplitude?

Re: This is related to the ocean area/volume: with a similar surface heating rate, the larger the ocean area/volume, the more heat is input into the ocean. That is why the OHC amplitude is larger in the Southern Hemisphere than in the Northern Hemisphere (Fig. 6), and the annual cycle of the Southern Hemisphere dominates the global OHC.

L 518 Check the description: IAPv3 is black in the legend above

Re: Corrected.

L 525-527 Not clear why IAPv4 is considered less physical than IAPv3, there are clearly non-physical features in IAPv3 appearing as rays emerging from the pole

Re: Yes, it is a typo, we rewrite this sentence: “The spatial OHC anomaly distribution in the Arctic region of the IAPv4 is more spatially homogeneous than IAPv3, and IAPv3 appears as rays emerging from the pole which are not physical (Fig. 7).”

L 522 "Anomaly" maximum "change" is from Sep to Dec.

Re: This sentence has been modified to “In IAPv3, the maximum upper 2000 m OHC occurs in December, and the minimum OHC occurs in August. However, for IAPv4, the maximum amounts to 2.9 ZJ in October and decreases to a minimum of −3.4 ZJ in April.”

L 536-537 Why January and July? Maximum MLD is expected to be later in the year: around March and August. Deep MLD in the Labrador Sea is surprising shallow.

Re: Yes, it should be better to present the spatial pattern of MLD in March and August. Figure 8 has been changed as follows, and the sentence in Line 536-537 has been modified to “Spatial distributions of the MLD in March and August are shown in Fig. 8 for IAPv4”. The maximum MLD in the Labrador Sea in March can reach 581 m.

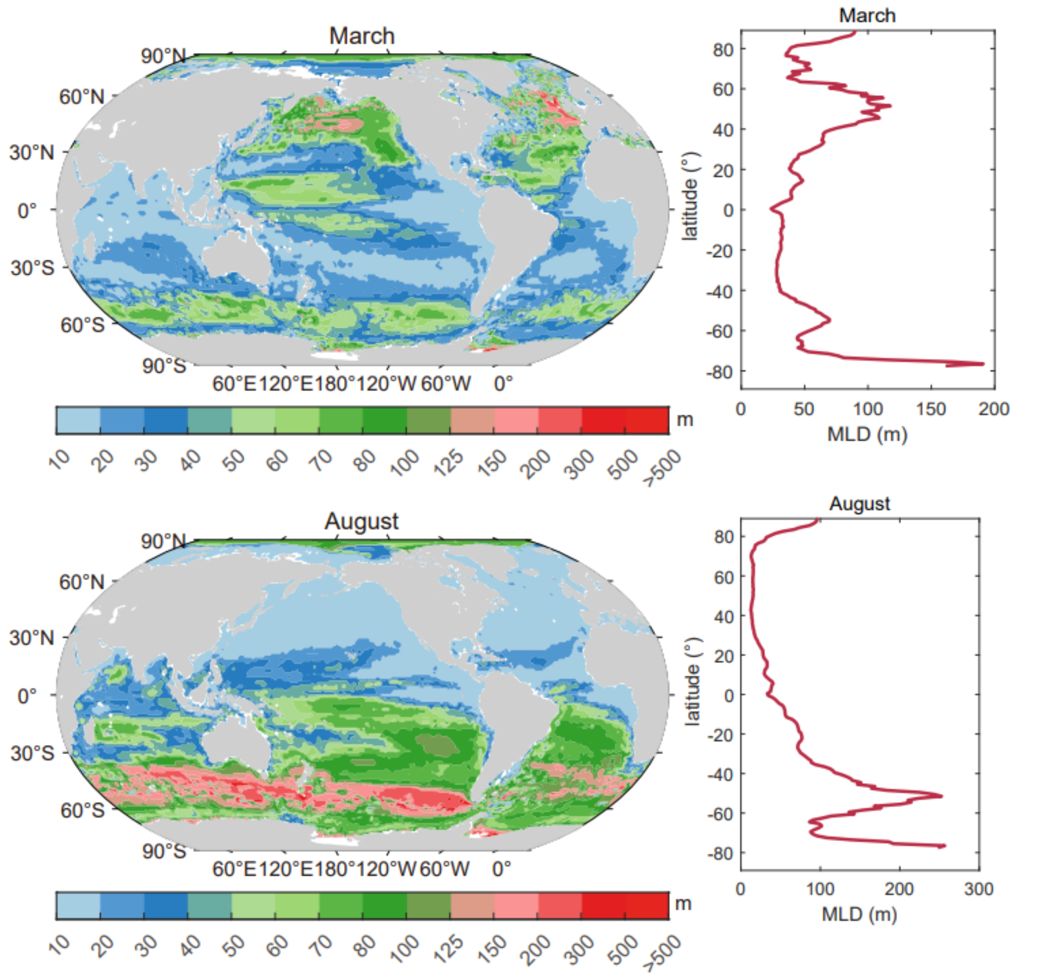

Figure 8: Spatial pattern of the climatological mean MLD (left panels) and zonal mean MLD (right panels) in March (top) and August (bottom) estimated from the IAPv4. Here, the MLD is calculated using the temperature difference criterion of ΔT = 0.02°C between the surface and 10-meter depth.

L 546 "Norwegian Sea", but the maximum appears to be southeast of Iceland which is in the Iceland Basin

Re: This is corrected to "Iceland Basin".

L 552-555 de Boyer Mont.gut et al., pointed out limitations of the delta T criteria. I think it is useful to acknowledged that these limitations also apply for the MLD estimate here.

Re: Yes, thanks. The sentence has been added after Line 555:

"However, as pointed out by de Boyer Montégut (2004), the MLD estimated from the average temperature profiles might lead to an underestimation of MLD by ~25% compared to the MLD computed from individual profiles based on the same 0.2°C criterion method. This potential issue needs further investigation."

L 582 to the south of

Re: Modified.

L 609 Interanual variations are also different

Re: Modified to “Data QC impacts the intra-seasonal and inter-annual variation of the OHC time series”

L 644 Which depth range is used?

Re: The depth range information is added “for the upper 2000 m”

L 679-680 Given the extend of that pattern I would rather call this a negative PDO phase related to the fact that a long warm phase ended in 1999 and since then it is mixed with somewhat more cold phases. Maybe bring this together with your following remarks about PDO

Re: Modified to “A trend pattern mimicing a negative Pacific Decadal Variability (PDV) phase appears in the Pacific for the 0–300 m, 0–700 m, and 0–2000 m OHCs.”

L 687-688 They describe an intensification in the South but a spin-down in the North Pacific

Re: Great, thanks, this sentence is modified to “Broad warming in most regions, but subtropical oceans in the West Pacific and South Indian oceans show a cooling, which is likely related to the subtropical gyre intensification in the North but a spin-down in the North Pacific Ocean (Zhang et al., 2014).”.

L699-701 It would be good to briefly outline how the OHC enters the estimate of the MHT, maybe also give an idea how important OHC is in comparison to Fs

Re: Yes, the information has been added here:

“The ocean MHT can be derived from the OHC and air-sea heat flux data (Trenberth and Fasullo, 2017; Trenberth et al., 2019) as follows: we integrate the OHC and air-sea heat flux from the North Pole southward in the Atlantic Ocean, and solve the energy budget question, the residual at each latitude is the MHT, i.e.,

$$MHT(\varphi) = \int_{\varphi}^{90} \left[Fs + \frac{dOHC}{dt}\right] a\, d\varphi$$

Where a is the Earth’s radius, φ is latitude, Fs is net surface heat flux. Both Fs and OHC are important for the MHT derivation: the integrated air-sea heat flux dominates the magnitude of the MHT while the OHC dominates the variability of the MHT (Liu et al., 2020).

L 734-735 What does it mean released from 20S-5N to 5S-20S? I assume "released" means to the atmosphere, otherwise it would be better to write redistributed., or do you argue is that the redistribution involves release and re-absorption?

Re: Yes it should be better to say “redistributed” because it is mainly the ocean processes.

L 771-774 Why is 90% EEI used in Fig.16? What does this discrepancy mean?

Re: 90% is used because 90% of the EEI is stored in the ocean (increasing OHC); the other 10% of net heat stored in the ocean is used to heat the atmosphere and land and melt the ice. An additional sentence is added here to explicitly state this issue “”

Table 3 What is the difference between GMSL and sum of components, how is the IAPv4 GMSL computed if not as a sum of components?
Re: We have rewritten this section and removed the Table 3. In the revised manuscript, we have added two clean tests on the impact of IAPv4 on sea level budget closure: 1) we replace the steric sea level component in Frederikse et al. (2020) by IAPv4 and then test the residual error and RMSD; 2) we replace the thermosteric sea level component in IPCC-AR6 (Gulev et al. 2021, they do not have a halosteric sea level estimate) with IAPv4 and check the residual error and RMSD.
Because the other components are the same and the only difference is the steric sea level data, these two tests can isolate the impact of the new T/OHC data on sea level budget closure.

GMSL can be observed directly by tide gauge or altimetry. The sum of components is the independent observation of the drivers of sea level rise, including steric sea level, glacier, Greenland ice sheet, Antarctica ice sheet and land water storage.

L 925 Why in particular warm eddies as opposed to cold eddies?
Re: This should include both warm and cold eddies. The sentence is rewritten to "first, are there still real temperature extremes being removed by CODC-QC, such as in small warm/cold eddies?".

Summary: Regarding methods to improve the estimate. Could you comment on the interpolation of the anomalies on isopycnal surfaces rather than depth levels, this could facilitate larger radii and better gap filling without the danger of making the solution overly smooth.
Re: This idea has been explored before by Palmer and Hains (2009). There might be some useful work to do in the future: different mapping approaches should be inter-compared to find whether one is superior. Theoretically, interpolation of anomalies on isopycnal surfaces should have a larger decorrelation length scale, but as introduced in this study (e.g. Fig.2 flow chat), reconstruction involves a lot of techniques and data processing procedures. The difficulty of this strategy would be the identification of isopycnal surfaces, which will add some uncertainty, especially in high latitudes. Nevertheless, a sentence is added in Summary section "Besides, other mapping techniques deserving further investigation include interpolation on isopycnal surfaces (Palmer and Haines, 2009)."
Palmer, M. D., and K. Haines, 2009: Estimating Oceanic Heat Content Change Using Isotherms. J. Climate, 22, 4953–4969, https://doi.org/10.1175/2009JCLI2823.1.