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

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)."

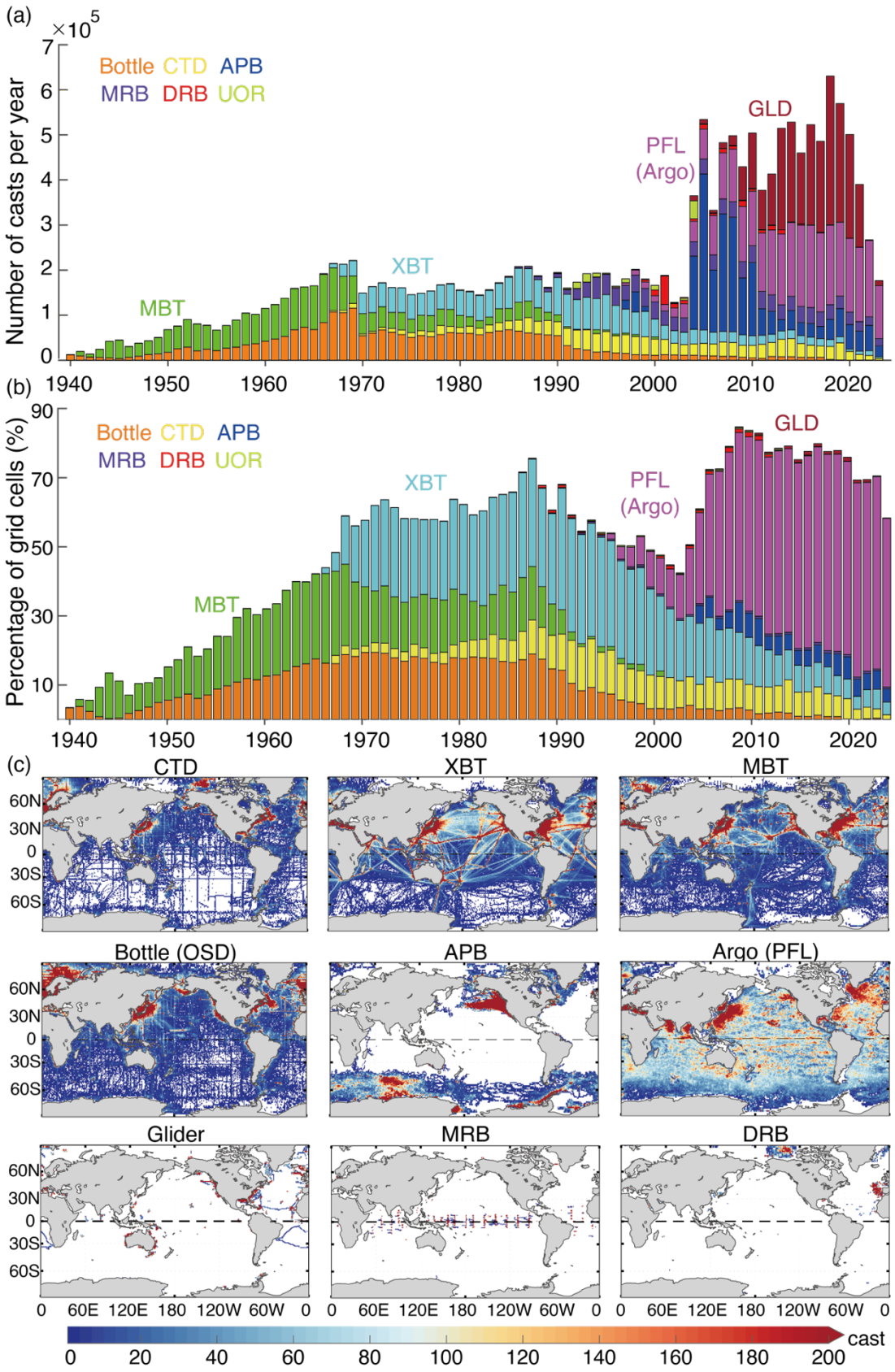

Figure 1: (a) Yearly number of temperature casts for different instruments; (b) percentage coverage (%) of ocean data for each instrument, which is calculated by the ratio between the number of 1° × 1° × 1 year grid cells observed by each instrument and the total number of

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.

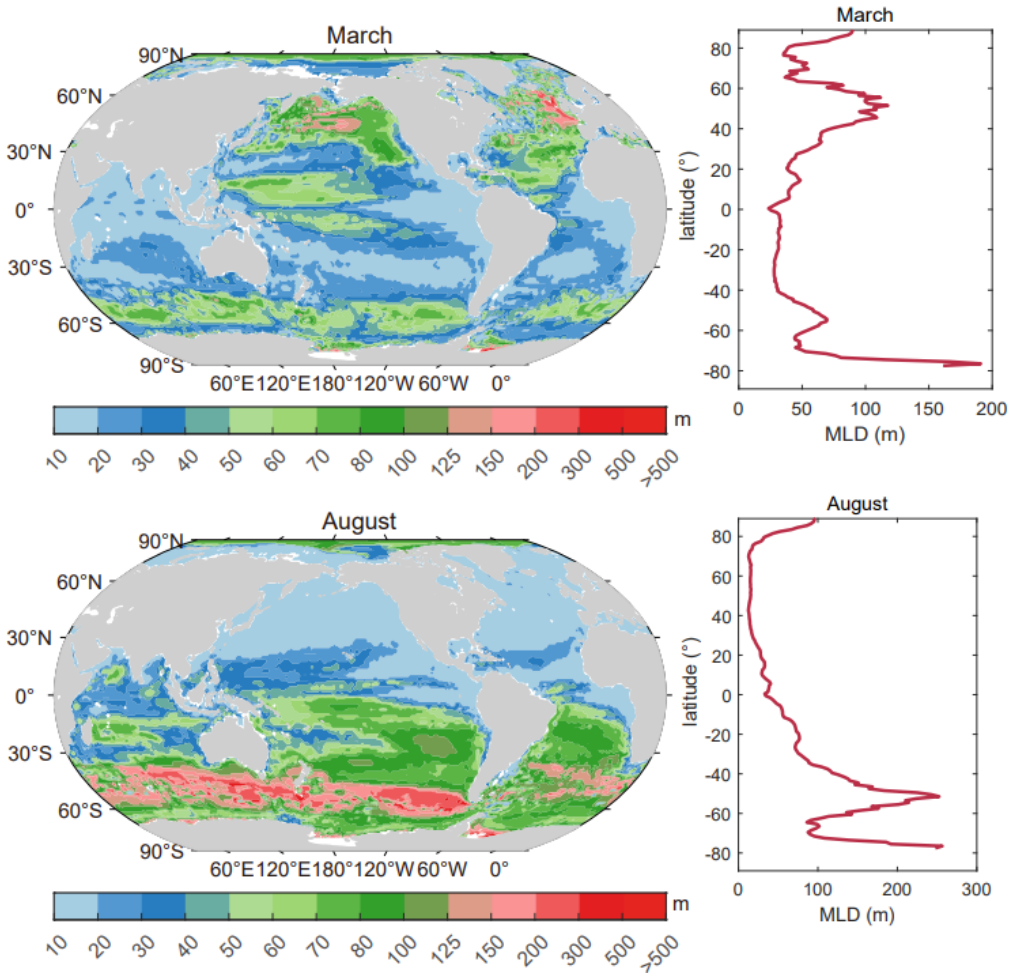

Figure 8: Spatial pattern of the climatological mean MLD (left panels) and zonal mean MLD (right panels) in March (top) and August (bottom) estimated from the IAPv4. Here, the MLD is calculated using the temperature difference criterion of ΔT = 0.02°C between the surface and 10-meter depth.

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.