Dear reviewer,

we are very grateful for your highly useful comments and suggestions, which address the most important aspects of our research. Please find below our step by step response (in blue) to your comments (in red):

The authors do not include the mapping error in the error estimate (lines 439-442), which is, on average, the largest uncertainty in global OHC reconstruction (10.1175/JCLI-D-15-0801.1). This is an important caveat that should be mentioned in the abstract and conclusion (lines 534-536) at least.

We are grateful for your criticism of this caveat in our research! In the revised version we estimate the OHC uncertainty due to insufficient temporal and spatial sampling by implementing the frozen grid method to the GECCO3 reanalysis. Respectively the sampling error contributes now to the total uncertainty. The details of changes in the revised version are described below following your detailed comments to the error section.

Fig. 6: The caption states that the blue circles are observed levels that are close to a neighbour level. However, there is no neighbour level close to the blue circles below 1000 m in the top-left and top-right panels.

We checked the plotting routine and updated Figure 6. The blue circles show the pairs of levels with a small vertical gap between them. The small gaps may occur throughout the entire vertical profile or may occur only for a limited number of level pairs.

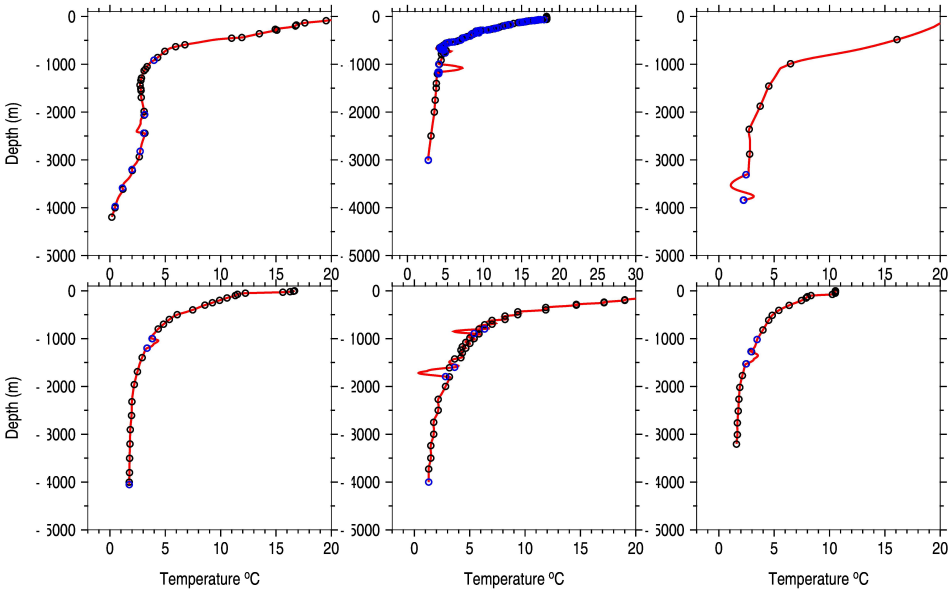


Figure 6. Examples of temperature profiles with uneven distribution of observed levels resulting in unrealistic overshooting of interpolation polynomials in RR method: observed levels (black circles), observed levels with a small distance to a neighbor level (blue circles), interpolated profile (red).

Fig. 7: I wonder if the label of the x-axis should be something like the upper bound of the re-sampling depth based on the text in section 4.2? If the x-axis is the depth of the gap center, the gap itself (y-axis) should be close to zero when the depth of the gap center is close to zero.

Thank you for noticing this inconsistency. We replotted Figure 7. The label of the x-axis indicates now “Gap upper bound (m)” in accordance with the description in the text.

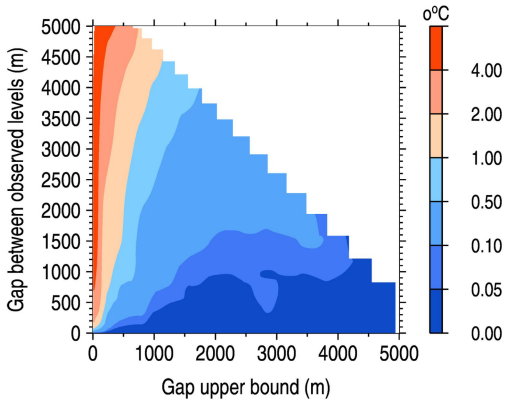


Figure 7. Global root mean squared interpolation error versus depth and the vertical gap size. Error refers to the center of the gap.

Lines 230-231: What is the time dimension of the validated dataset? Is it time mean or a snapshot of a specific time?

We have changed the caption to Figure 5: it is indicated that the outlier percentage is for the time period 1870 to 2025.

Figure 5. Outlier percentage for OSD temperature profiles between 1870 and 2025: a-c) automated statistical quality control; d-f) manual quality control. Outlier percentage is shown for depth/year bins (a, d), depth/temperature bins (b, e), 1° × 1° latitude/longitude boxes (c, f).

Lines 240-241: "the FR method results in a smaller interpolation error compared to the RR method" Fig. 9a shows that the FR method has a larger error than the RR method for the 'big gap' case, which contradicts the statement at L240-241.

We checked Figures 9a-9b. In both cases the interpolation errors for RR method (shown in orange and red) are larger than interpolation errors for FR method (shown in blue and cyan).

Line 251: How is the interpolated data derived? Is it based on the FR or RR method?

It is indicated in the beginning of the section 4.2, that interpolation was done using RR method.
“Non-linearity of temperature profiles interpolation over too big gaps might lead to significant errors in temperature. In order to get a measure of the interpolation error due to vertical gaps we used the climatology temperature profiles from the updated WAGHC climatology (Gouretski, 2018). The WAGHC gridded profiles give temperature values at 65 unevenly spaced levels between the surface and 6500 m. Each WAGHC profile was interpolated on 10-meter levels using the Reiniger-Ross method.”

Fig. 9 caption: The panel number (c) is missing in "Mean gap size versus ...".

Thank you: we added “c)” in Fig. 9.

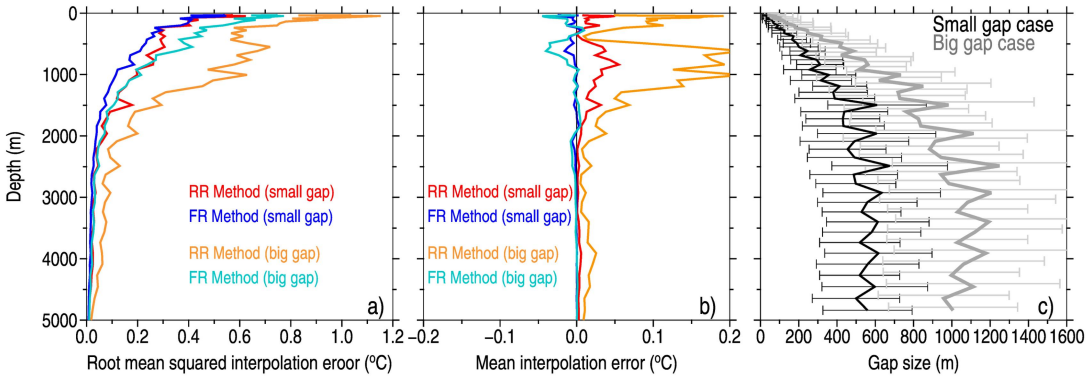
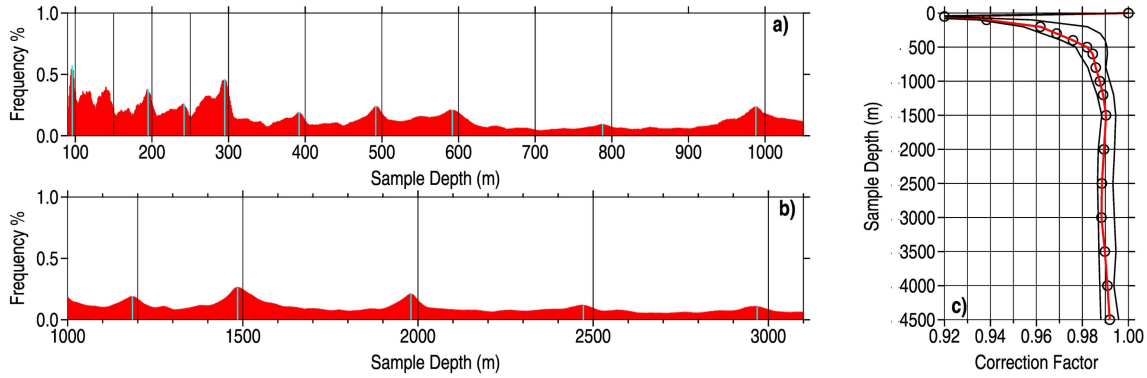


Figure 9. Global root mean squared (a) and global mean (b) error for interpolated temperature values obtained by means of RR method (red and orange) and FR method (blue and cyan). Red and blue profiles correspond to the small gap case, orange and cyan correspond to the big gap case. Explanation Is given in the text. (c) Mean gap size versus depth for the small (black) and big (grey) gap case respectively. Horizontal bars correspond to one standard deviation.

Lines 294-295: "After some experimenting, we applied a Gaussian kernel width of 20 m to prevent over-smoothing of the histograms" How sensitive is the correction factor to changing the width from 20 m to other values, e.g. 10 m and 50 m? It would be useful to show the histogram without any smoothing in the supplementary.

Thank you for these valuable comments! We followed your suggestion and produced smoothed histograms using different smoothing kernel window width between 1 and 50 meters. The final choice of the depth correction factor is now represented by the median over these 50 estimates. Figure 11 and the table of depth correction factor values in the Appendix were changed respectively. Now the impact of smoothing kernel width on the depth correction factor is seen in the Figure 11, where minimum and maximum values of the depth correction factors are shown along with the median value. The figure caption was changed respectively. The new piece of the text is:



"Fig. 11a, b show frequency distributions of sampled depths around standard levels. These histograms were obtained after removing the standard levels and the subsequent smoothing of the histograms shown in Fig. 4 for the years 1920-1970 using Gaussian kernel with the width of 10 meters (the frequencies for standard depths are not zero due to the Gaussian smoothing). For all standard levels a shift of the maximum frequency depth (Zmax) relative to the respective standard level (Zstand) can be seen (Fig.11a, b). The smoothed histograms were constructed for 50 values of the kernel width varying between 1 and 50 meters and the depth correction ratio R = (Zmax)/(Zstand) was calculated for standard levels at or below 50 m (Fig.11c). Levels above 50 m were not used as they mostly fall within the upper mixed level or a seasonal thermocline. The final depth correction ratio corresponds to the median of 50 estimates. Depth corrections are also given in Table A1. As indicated by Fig.11c original sample depths slightly overestimate the actually reached

depths. We note that Nansen cast bottle depths are reported in round meters imposing a limitation on the relative accuracy of the histogram method in the upper part of the water column."

"Table A1. Depth bias orrections for historical Nansen casts

Original	Depth Correction factor
0050	0.9200
0100	0.9383
0200	0.9617
0300	0.9688
0400	0.9759
0500	0.9820
0600	0.9845
0800	0.9858
1000	0.9876
1200	0.9890
1500	0.9903
2000	0.9894
2500	0.9885
3000	0.9883
3500	0.9898
4000	0.9909
4500	0.9919
5000	0.9919

Corrections are applied to standard level depths listed in the Table S1."

In the previous version of the manuscript, we did not apply depth correction to the oldest Challenger, Gazelle, and Tuscarora profiles. We find this approach unsatisfactory as similar to the Nansen casts the hemp rope used on those cruises might have experienced deviations from the vertical and so might lead to depth overestimation. In the revised version we apply depth corrections derived for Nansen casts noting however that the corrections derived for Nansen casts with steel wire might not fully correspond to the behavior of the hemp rope used on the three early expeditions. However, we take into account the elasticity of the hemp rope: Joung's modulus of the hemp rope is two orders of magnitude smaller compared to a steel wire. The increase of the hemp rope length under load is calculated based on the known Joung's modulus for a hemp rope, the known rope diameter and weight of the load attached to the end of the rope. The calculations suggest a one percent increase of the hemp rope length. Respectively, all original sample depths for Challenger, Gazelle, and Tuscarora profiles were increased by the factor 1.01 before the application of the depth corrections derived for Nansen casts. The following text was added in the end of the Section 7:

"No additional contemporary temperature profiles exist which could be used to estimate possible depth bias in "Challenger", "Gazelle" and "Tuscarora" data. For this reason, we apply depth corrections derived for Nansen casts in Section 5 realizing that the use of these corrections entails a certain error because instead of a steel wire a hemp rope was used on these early cruises. However, we take into account the elasticity of the hemp rope, which Joung's modulus is two orders of magnitude smaller compared to the steel wire. Abdulla et al. (2021) report the values of 1.20-1.47 Gpa for hemp ropes. The increase of the hemp rope length under load is calculated using the known rope circumference (25.4 mm) and the weight of the load (63.5 kg) attached to the rope on

"Challenger" cruise. The calculations suggest an approximately one percent increase of the hemp rope length. Respectively, all original sample depths for "Challenger", "Gazelle" and "Tuscarora" profiles were increased by the factor 1.01 before the application of the depth corrections derived for Nansen casts."

Fig 12: Please show depths using positive numbers as in other figures.

We deleted minus signs on the vertical axis.

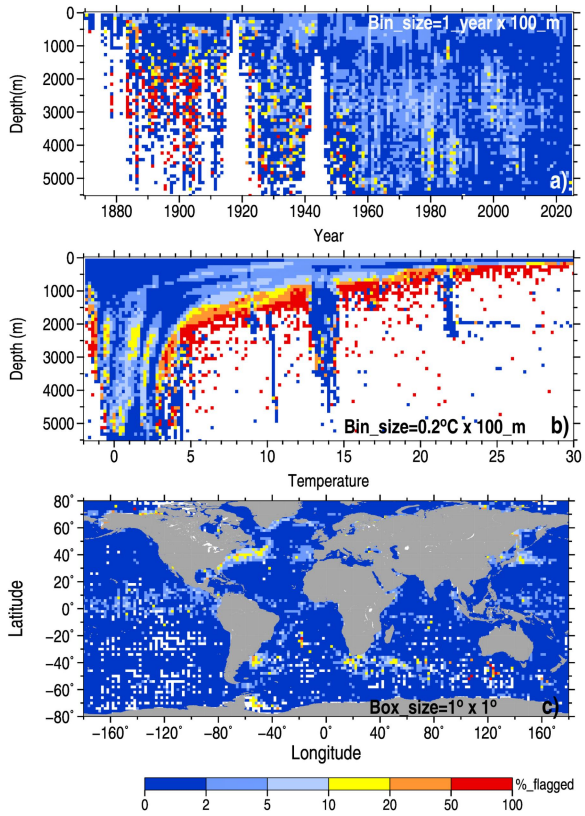


Figure. 12. Outlier percentage for temperature anomalies: a) percentage in depth/year bins, b) depth/temperature bins, c) 1° × 1° Latitude/Longitude boxes.

Section 8.4

I appreciate the added detail about the instrumental error and the mesoscale variability. However, I still find this section difficult to follow. In particular, it remains unclear to me how different types of errors are combined.

My specific questions are:

*) How is the "standard error" at line 433 related to the instrumental error and the standard deviation mentioned earlier?

*) Is the "standard error" at line 433 the same as the within-box error at line 434?

*) How is the correlation radius used when computing the global- and basin-mean errors?

I suggest including equations here to make sure that the above details are clearly documented.

It seems that the error estimate explained in 8.4 is relevant to the results shown in 8.2 (the within-box error) and 8.3 (the global- and basin-mean errors). If that is correct, I suggest moving 8.4 to before 8.2 and 8.3.

Lines 486-487: For a given time period (e.g. 1874-1932), did you work out the annual heating rate first and then compute the time mean over that period?

We greatly appreciate your criticism of this most important aspect of our research and tried to follow your suggestions in the revised version of the manuscript. First, we reordered the subsections of Section 8: in the revised version, and the error-estimation section is now Section 8.2. We have changed the text of the section 8.1, where the methodology of the calculations is described:

"**8.1 Calculation of temperature anomalies**

Because of a sparse sampling we estimate the mean thermal state of the entire global ocean and of the main three oceans for 5-25-year long time periods, rather than for distinct calendar years. The selected time periods are: 1873-1876, 1920-1944, 1945-1969, 1970-1989, 1990-2009, 2010-2025. The earliest period spans the quasi-global oceanographic survey conducted by "Challenger", "Gazelle", and "Tuscarora". For the calculation of basin-scale and global temperature anomalies, we use 10-meter level profiles described in Section 4. The 10-meter level anomaly profiles are then averaged within the 100-meter layers and the layer mean anomalies are calculated for 5° x 5° Latitude/Longitude boxes and for each selected time period. The box-mean anomalies provide the basis for the calculation of the heat content anomalies for distinct ocean basins and for the global ocean."

Our total uncertainty is represented by the sum of two errors: 1) error due to insufficient within-box sampling and 2) error due to incomplete geographical sampling. The within-box uncertainty is calculated by subsampling the data within each 5x5-degree box and comparing the estimated OHC values. The geographical sampling uncertainty arising due to unsampled areas is estimated by resampling the GECCO3 reanalysis using the "frozen-grid method".

The new text of the Section 8.2 is":

"8.2 Error estimation

We consider two sources of errors in global and basin-wide OHC estimates.

The first source of error is due to within-box sampling uncertainty (Eb). Both instrumental uncertainty and meso-scale variability introduce noise characterized by the within-box standard deviation. Boyce (1966) estimates the instrumental error for reversing thermometers as ± 0.04 ^{o}C. Errors associated with the meso-scale variability are much larger. According to Fig. S7, the standard deviation of temperature anomalies within a 2° × 4° latitude/longitude box exceeds 3 ^{o}C in the western boundary currents and in the equatorial belt in the upper 200-meter layer.

The insufficient sampling may bias the estimated box-averaged anomalies. The normally used error propagation (summation of individual box anomaly variances) leads to unrealistically high Eb, because the within-box anomaly variances are dominated by the real oceanic variability within relatively large boxes and long time periods. In order to estimate Eb we calculated global OHC using boxes with at least 10 profiles. Within each box individual anomaly profiles were subdivided in ten groups each including 1/10 of the data. Respectively ten global OHC estimates were obtained for each time period using box-mean values based on sub-sampled profiles. The standard deviation from the mean over 10 estimates is taken as the estimate of Eb for each time. As the number of profiles for the time period 1873-1877 is very limited, we used all boxes with at least two profiles with the subdivision in two groups only.

Geographical sampling uncertainty (Eg) represents the second source of uncertainty and characterizes how accurately the observed boxes represent the entire global ocean. According to Domingues et al (2018) Eg dominates among other sources of uncertainty. We used GECCO3 ocean reanalysis data (Köhl, 2021) with one degree spatial and monthly temporal resolution for the

time period 1957-2018 (62 years in total) and implemented the so-called frozen-grid method to estimate this type of uncertainty. The method was used in meteorological (Brohan et al. 2006) and oceanographic (Gouretski et al., 2012) studies. For each of six selected time periods the global OHC anomaly was calculated using all grid boxes within the time span of the respective time period (OHCall) and using only sampled grid boxes (OHCsamp). The beginning of the sampling period was moved with one-year increment starting with the first year 1957. For each time period N OHC estimates based on all and actually sampled GECCO3 boxes were obtained respectively, where N=62-Np (Np is the number of years spanned by the respective time period). The root-mean-squared difference (OHCall) -(OHC)samp is used as the measure of Eg. Both errors are summed in quadrature to get the total uncertainty Etot2 = Eg 2 + Eg2. Other kinds of uncertainty (for instance uncertainty due to errors in the base-line climatology) are not considered in this study."

References
Some journal names are written in abbreviations (e.g. Geoph. Res. Letters), while others are not (e.g. Geophysical Research Letters). Please make them consistent.

Thank you, we made journal name abbreviations consistent. Please note, that two references have been added to the list: Boyer et al., 2018 (on the importance of sampling error) and Koehl 2020 (GECCO3 description)

We also added four new references to the reference list:

Abdulla, K., Cunningham, L, and Gillie, M.: Out-of-plane strengthening of adobe masonry using hemp fibre ropes: An experimental investigation. Engineering Structures, 245,1-14, https://doi.org/10.1016/j.engstruct.2021.112931, 2021.

Boyer, T., Domingues, C. M., Goog, 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, DOI: 10.1175/JCLI-D-125-0801.1

Brohan, P., J. Kennedy, I. Harris, S. Tett, and P. Jones: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850, J. Geophys. Res., 111, D12106, doi:10.1029/2005JD0065, 2006.

Köhl, A.: Evaluating the GECCO3 1948-2018 ocean synthesis – a configuration for intializing the MPI-ESM climate model, Q J R Meteorol Soc., 146, 2250–2273, https://doi.org/10.1002/qj.3790, 2020.

Finally, we would like to note the following:

1. We refrain from using the standard error of the mean temperature anomaly across all boxes as a measure of uncertainty, since this measure incorporates the effect of the natural variability and thus exaggerates the estimate of sampling uncertainty. Respectively, we deleted **Fig.S10** from the supplementary information (spatial correlation between the boxes).

2. All changes introduced during the revision do not affect the results of the study.

3. **Fig. 16** was replotted using new error calculation results.

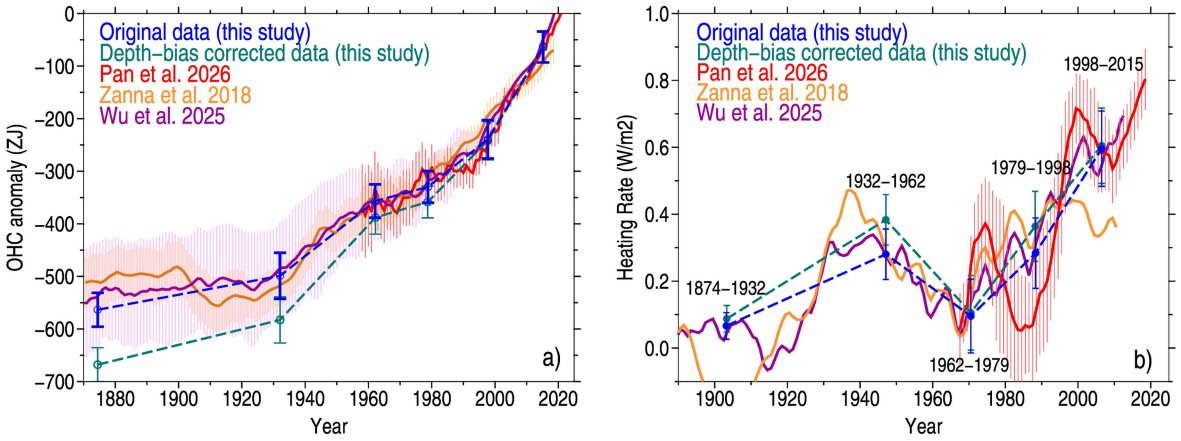


Figure 16. a) Global OHC anomalies for selected time periods for original data (blue) and depth-bias corrected data (cyan), OHC anomaly time series from Pan et al. (2026) (red), Zanna et al. (2018) orange), and Wu et al. (2025) (magenta) are shown for comparison; b) heating rate for five time periods: original profiles (black), depth-bias corrected profiles (blue), Pan et al. (2026) (red). Heating rate time series for Zanna et al. (2018) and Wu et al. (2025) are based on the smoothed OHC time series using a 15-year running window. The following error bars are shown: total error (this study, see description in Section 8.2); 95% confidence interval (Pan et al., 2025); one standard deviation (Zanna et al, 2018); one standard deviation from 73 ensemble members (Wu et al., 2025).

4. All OHC error estimates provided in the manuscript text have been updated respectively. The changes mostly affect the numbers in Section 8.5. The revised version of this section is:

"8.5 Secular changes of the Ocean Heat Content

Finally, mean temperature anomaly profiles for the three main oceans and for the global ocean were used to calculate OHC time series shown in Fig. 16 for the layer 0-2000 m. The following values of parameters were used for the OHC calculations: 3991.87 J/kg°C as the seawater specific heat capacity of the seawater, 1028 kg/m3 as the average seawater density, 6371 km as the mean Earth radius. The volumes of the 100 m layers for which temperature anomalies were obtained are calculated using the GEBCO digital bathymetry (Becker et al., 2009). For the earliest time period 1873-1876 the global OHC estimate is represented by the weighted average of the OHC estimates for Atlantic, Indian, and Pacific oceans, with weights equal to individual ocean areas. Troughout this section, we provide two estimates for the OHC and heating rate: the first for the original data, and the second (in parenthesis) for the depth corrected data.

Firstly, we note a good agreement between our OHC time series calculated for the selected time periods and the yearly time series of the Institute of Atmospheric Physics (IAP) by Pan et al. (2026) (Fig. 16a). There are several significant differences between the two products: the Pan et al. (2026) calculations are based on a much larger dataset, which includes not only Nansen casts but all other instrumentation types as well, with the mapping procedure aiming to fill the data-void areas (Cheng and Zhu, 2017).

Secondly, we find that the global ocean underwent a significant warming before 1960 when the IAP and other observation-based time series begin. The median years for the six time periods indicated in the previous section are: 1874.5, 1932.1, 1962.1, 1978.8, 1997.7, and 2015.2. Between 1932 and 1962 the global ocean heat content increased by 142.2±48.7 (194.9±54.2) ZJ to be compared with 435.4±52.9 (519.3±52.90) ZJ warming between 1932 and 2015, thus amounting to 32.7±6.8% (37.5±3.8%) of the total OHC increase since 1932. Our calculations further suggest that the warming likely took place also between the 1870s and the 1920s amounting to 64.5±54.5 (85.1±54.5) ZJ.

OHC time series for the upper 2000 m from Zanna et al. (2019) and Wu et al. (2025) are also compared with our calculations. These time series are based on the reconstruction of the full-depth temperature changes using temperature anomaly at the sea surface as a passive tracer which is

carried into the ocean interior via ocean transport processes. Our observation-based time series agree well with the indirectly obtained time series by Zanna et al. (2019) and Wu et al. (2025) indicating a significant OHC increase between 1920s and 1960s.

As discussed above historical Nansen casts are subject to a depth bias (Gouretski et al., 2022). Depth corrections suggested in Table 1 were applied to the original sample depths when they belong to the depth levels round to 5, 10, 50 or 100 m depending on depth and/or to the set of "standard" depth levels (see Table S1). These 66 levels represent 1.2% of all 1-meter levels between the surface and 5500 m depth. For the time period before 1970, 49.5 % of all observations are reported at the above standard levels. The impact of the depth corrections on global OHC estimates is shown in Fig. 16a. We found that the difference between temperature anomalies for original and depth-corrected profiles can amount to 5-10 % of the OHC change depending on the time period. Overall, depth corrections lead to the increase of the warming rate because the corrections shift the observed depths upwards leading to a cooling of temperature profiles in case of the prevailing negative temperature gradient. Applying depth corrections increases the total energy gain between 1930 and 2015 by 16.0±7.2%, from 435.4±52.9 ZJ to 519.3±52.9 ZJ. The respective OHC change between 1962 and 2015 increases by 8.5±7.8 % from 293.2±43.3 ZJ to 324.3±43.3 ZJ. The applied depth correction factor (Fig. 11c) varies between 0.92 at 50 m level and 0.99 at the depth of 1500 m and results in a different regional impact as the temperature stratification changes geographically. The impact of corrections is most important in the upper part of the water column with typically stronger vertical temperature gradients. Accounting for similar systematic depth errors in temperature profiles from the expendable and mechanical bathythermographs was found to be important for the OHC calculation (Gouretski and Koltermann, 2007; Gouretski and Reseghetti, 2010).

The average heating rate between the analyzed six time periods is shown in Fig. 16b. The heating rate is estimated as the difference between two adjacent time mean global OHC anomalies divided by the time span between them (the center of each period corresponds to the median year of all profiles). Our calculations indicate a heating rate of about 0.07±0.04 (0.09±0.04) W m-2 before 1930s. Between 1932-1962 the heating rate increases to 0.29±0.08 (0.40±0.08) W m-2 and is close to the value for the period 1979-1998. In agreement with Pan et al. (2026), the fastest OHC growth of 0.62±0.12 (0.63±0.12) W m-2 is found for the latest period 1998-2015. We note that the Pan et al. 2026 heating rate is calculated based on the yearly IAP OHC time series smoothed using the 15-year time window. The averaging period is different from time periods used in the current study. The heating rate for Zanna et al. (2018) and Wu et al. (2025) was calculated using the respective smoothed OHC anomaly time series using a 15-year moving window. Similar to our results the both studies also suggest a local heating rate maximum between 1932-1962."

In addition to the changes introduced into the revised manuscript we now estimate the depth bias for the oldest cruises of Challenger, Gazelle, and Tuscarora during 1873-1877. We apply depth corrections derived for the historical Nansen casts, but with an important deviation. On these cruises a hemp rope was used instead of a of a steel wire typical for Nansen casts. The hemp rope has a two order of magnitude lower Joung's modulus, so that the elasticity of the hemp rope must be taken into account. Before application of the Nansen cast depth correction, we increase the sample depths for the three cruises by the factor 1.01 which is derived for the hemp rope diameter and the attached weight as used during the Challenger expedition. Respectively, a small piece of text was added in the end of the Section 7:

"No additional contemporary temperature profiles exist which could be used to estimate possible depth bias in "Challenger", "Gazelle" and "Tuscarora" data. For this reason, we apply depth corrections derived for Nansen casts in Section 5 realizing that the use of these corrections entails a certain error because instead of a steel wire a hemp rope was used on these early cruises. However, we take into account the elasticity of the hemp rope, which Joung's modulus is two orders of magnitude smaller compared to the steel wire. Abdulla et al. (2021) report the values of 1.20-1.47 Gpa for hemp ropes. The increase of the hemp rope length under load is calculated using

the known rope circumference (25.4 mm) and the weight of the load (63.5 kg) attached to the rope on “Challenger” cruise. The calculations suggest an approximately one percent increase of the hemp rope length. Respectively, all original sample depths for “Challenger”, “Gazelle” and “Tuscarora” profiles were increased by the factor 1.01 before the application of the depth corrections derived for Nansen casts. “