

Reply to Anonymous Reviewer #1

Dear Reviewer,

Thank you very much for your detailed and thorough review of the submitted manuscript. Please, find below our response to the issues raised in your review. All pieces of the text changed according to your comments are written in blue.

1. Reconcile internal numerical inconsistencies in the OHC results. The OHC increase between the 1920–1944 and 1945–1969 periods is given as 149 ZJ in Section 8.5 (line 424) but as 139 ZJ in the Conclusions (line 484). Similarly, the total warming since 1932 is 456 ZJ at line 424 but 457 ZJ at line 444. These are presumably the corrected vs. uncorrected values or simple typographical slips, but as written the paper contradicts itself on its headline numbers. Please reconcile these figures throughout the Abstract, Section 8.5, and Section 9, and state explicitly whether quoted values are for original or depth-corrected profiles.

We have reconciled the internal consistency of the OHC results. It is now stated where quoted values are for original or for the depth-corrected profiles. Please note that we recalculated all OHC estimates and error bars using a specific heat capacity of 3991.87 J/kg °C for seawater, as the original estimates used the value for freshwater.

2. Correct the standard-depth list in Section 8.5 (lines 436–438). The list of 67 standard levels contains apparent duplications and omissions: “30” appears twice (“15, 30, 25, 30, 35...”), “2000” appears twice (“2100, 2000, 2300”-presumably 2200 was intended), and “3200” appears twice (presumably 3100). As listed, the levels do not sum to 67 unique depths, and the claim that they represent 1.2% of all 1-m levels cannot be verified. Please correct the list and re-verify the 49.5% statistic if affected.

The standard depth list was corrected (duplicate levels deleted) and moved into the supporting information section.

3. Resolve citation-year and author-name mismatches between text and reference list. (a) Reiniger and Ross is cited as 1966 throughout the text (lines 20, 163, 471) but dated 1968 in the reference list; (b) Chapman is cited as 1955 (line 100) but listed as 1959; (c) Zanna et al. is cited as 2018 throughout but listed as 2019; (d) “Spillhouse, 1938” (line 50) should be Spilhaus, consistent

with the reference list; (e) “Gourteski” at lines 30 and 85 should read Gouretski. These are easily fixed but essential for a data paper that readers will mine for sources.

Thanks for the comments. We have inserted respective corrections in the reference list and in the text.

4. Make the Challenger expedition dates consistent. The expedition is variously dated 1872–1876 (lines 35, 74), 1873–1875 (line 320), and 1873–1876 (lines 413, 578). The voyage departed in December 1872 and returned in May 1876; please adopt one convention (or clarify that the dates at line 413 refer to the analysis period rather than the voyage) and apply it throughout.

The Challenger expedition dates were made consistent using the period 1873–1876. This agrees with the period given by Thomson et al. (1885).

5. Check Figure 15 legend against the stated analysis periods. The legends in Fig. 15 appear to read “2010–2017” for the most recent period, whereas the text and Section 8.1 define the period as 2010–2025. If the figure was carried over from an earlier analysis, please regenerate the legend; if the plotted period genuinely differs, the discrepancy must be explained. Relatedly, in Section 8.4 the error-estimation procedure is clearly described, but the resulting uncertainty magnitudes are never stated numerically in the text—adding one or two sentences quantifying the basin- and global-scale standard errors (or pointing to the error bars in Figs. 15–16 with representative values) would materially strengthen the section at negligible cost.

Fig. 15 was corrected. We rewrote Section 8.4, also taking into account the second reviewer's comments. We added characteristic temperature uncertainties in Section 8.4, as you suggested.

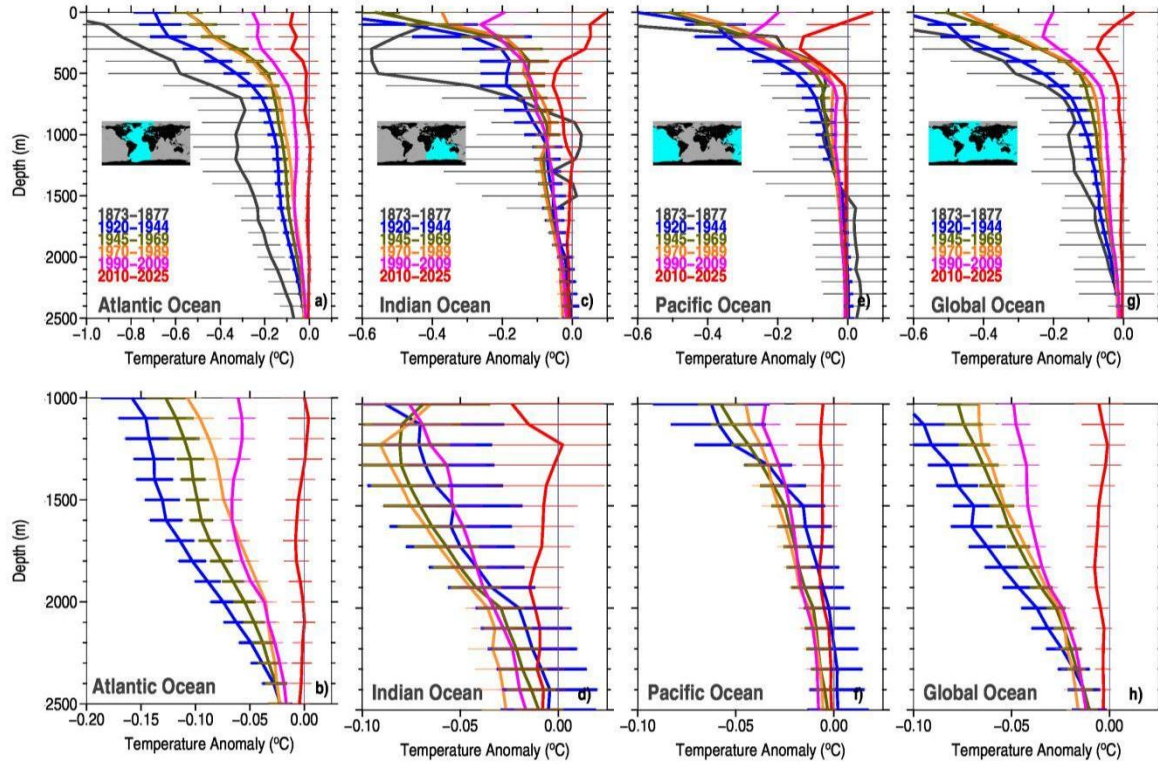


Figure 15. a, c, e, g) Temperature anomaly vertical profiles for the three main oceans and for the entire global ocean and for selected six time periods with standard error bars; b, d, f, h) same but for the depth range below 1000 m. Note the different temperature scale for the Atlantic Ocean.(revised)

“8.4 Error estimation

The box-averaged temperature anomalies in each 100 m layer are calculated using all observed anomaly profiles within the respective $5^\circ \times 5^\circ$ latitude/longitude box. Instrumental errors and mesoscale variability introduce noise, which biases the estimated box-averaged anomalies. Boyce (1966) estimates the instrumental error for reversing thermometers as $\pm 0.04^\circ\text{C}$. Errors associated with the mesoscale variability are much larger. According to **Fig. S7**, the standard deviation of temperature anomalies within a $2^\circ \times 4^\circ$ latitude/longitude box exceeds 3°C in the western boundary currents and in the equatorial belt in the upper 200-meter layer. We assume temperature anomalies within a $5^\circ \times 5^\circ$ latitude/longitude box are independent because the observations span 20 years or longer, and each mean box anomaly is associated with its respective standard error. For the basin- and global mean anomalies, the errors are obtained by the summation in quadrature of within-box errors. In order to estimate the number of degrees of freedom, we calculated spatial autocorrelations in east-west and north-south directions using box-averaged anomalies for the best-sampled period 1990-2009 (**Fig. S10**). The meridional correlation radius is determined by the zero-crossing and is taken to be 1 meridional distance (555 km) between the boxes. The correlation along longitude has a long tail, so we define the correlation radius as the distance at which the correlation function falls below 0.1, which averages two box separations in the zonal direction. We note that for the earliest

periods (1873-1877 and 1920-1944) global and basin means are estimated from observations (box-averaged values) that cover only a small fraction of the respective area (see Fig. 14), leading to a mapping error. This kind of error was not estimated in the current study and is not considered in Figs. 15-16.”

6. Clarify data access in Section 10. The dataset DOI requires email registration and login. ESSD policy expects data to be publicly and freely accessible without barriers wherever possible. Please either remove the registration requirement, or briefly justify it and confirm that access is granted automatically and without restriction, so the editor can verify compliance with the journal’s data policy.

We have also made the data available on the Zenodo platform (<https://zenodo.org/records/22764102>), in addition to <http://doi.org/10.12157/IOCAS.20260520.001>, which requires an extra login to access the data. The Data Availability section in the manuscript has been modified accordingly.

7. The manuscript requires a careful language edit. The science is clear, but grammatical errors and typographical slips occur throughout and occasionally rise to the level of technical error. The most important instance: “Jung’s modulus” at line 459 should be Young’s modulus (and note the same sentence refers to “hemp wire” where “hemp rope” is meant, and “four time higher” should read “four times higher”). A non-exhaustive list of additional corrections follows; I would encourage the authors to have the full manuscript professionally proofread rather than relying solely on this list.

- Line 11–12: “observations which abundance, quality and accuracy changed” → “observations whose abundance, quality, and accuracy have changed.”
Done.
- Line 35 and line 483: “marks the begin of” / “the begin of the ocean warming” → “beginning.”
Done.
- Line 45: “Fritjoff Nansen” → “Fridtjof Nansen.”
Done.
- Line 49: “hydrographic hydrographic profiles” — duplicated word.
Done.

- Line 62: “we access for the first time the change” → “we assess.”

Done.

- Line 126: “An disproportionately large number” → “A disproportionately large number.”

Lines 137, 421, 472: “undergone” used as a simple past (“the data ... undergone three steps”; “the global ocean undergone”; “profiles undergone”) → “underwent” (or “have undergone”).

Done.

- Line 197: “performed for a the resampled profile” — delete “a.”

Done.

- Line 260: “Here we imply this method” → “apply.”

Done.

- Line 292: “In a deviation to the widely used World Ocean Atlas climatologies” → “In contrast to.”

Done.

- Line 309: “mimimum/maximum” → “minimum/maximum.”

Done.

- Line 415: “with wights equal to” → “weights.”

Done.

- Line 417: “There several significant differences” → “There are several...”

Done.

- Reference list: “littele ice age” (Gebbie and Huybers), “Warming of heworld ocean” (Levitus et al.), “Pysics and Chemistry” (Thomson and Murray 1889), “Washigton D.C.” (U.S. Naval Oceanographic Office), “Manual fort hydrological works” (State Oceanographic Institute), and “Her Majesty Stationary Office” → “Stationery Office.” The Frankignoul (1981) entry should also be reformatted to match the journal’s reference style (year at the end).

Done.

We conducted a careful language edit and made corrections according to your detailed list above – thank you so much for your careful reading!

Please note that we substantially reworked Section 4.3, because the old comparison of RR and FR interpolation methods was not adequate. In the old comparison method, the interpolation by the two methods was applied to the same climatological (WAGHC) temperature profiles, so that the RR

method already “knows” the shape of the climatological profile. For this reason, we used a different approach to compare the RR and FR methods. Now we use the original observed level quality-controlled profiles, which are interpolated vertically using both methods. Before interpolation, artificial data gaps are created by removing data levels. The error is defined as the difference between the original (omitted) temperature and the interpolated value. The subsequent comparison of the interpolated values with the observed (artificially removed) values shows that 1) the FR method has a smaller global root-mean-square interpolation error and 2) the RR interpolation introduces a positive temperature bias (global mean temperature error). For this reason, only the FR method is used. We updated Fig. 9 and rewrote Section 4.3.

The new text of Section 4.3 is (Lines 225-259 in the revised manuscript):

“As shown in **Fig. 7**, the global root-mean-square interpolation error in the RR method increases with gap size. To reduce the error for large vertical gaps, the climate anomaly method suggested by Frankignoul (1981) is used (FR method). In this method, deviations of individual temperature profiles from the mean climatology are linearly interpolated over depth, and the interpolated value at a level is the sum of the mean climatological temperature and the interpolated anomaly. **Fig. 8** shows several selected observed profiles with big vertical gaps along with the respective interpolated profiles. For large gaps, temperature differences between the two methods can exceed several degrees. For large gaps, the RR method tends to follow simple linear interpolation across the gap, resulting in a significant deviation from the climatological profile.

To further illustrate the differences between the two methods, we used the complete validated observed level profile dataset. To estimate interpolation errors, artificial data gaps are created. Error estimation was performed for two choices of the vertical gap size. In a ‘small gap’ case, for each temperature profile with N observed levels, a total of $N-2$ reduced level profiles can be created, with one omitted observed level k between the uppermost ($k=1$) and the lowest level ($k=N$) of the full original profile. In a ‘big gap’ case, in addition to level k , levels $k-1$ and $k+1$ are omitted as well. Vertical interpolation of reduced profiles is performed at the depth of the removed level k , and the difference between the observed and interpolated temperature estimates the interpolation error. The vertical distance between the levels $k+1$ and $k-1$ (‘small gap’ case) and $k+2$ and $k-2$ (‘big gap’ case) gives the size of the artificial vertical gap. We performed interpolation using the RR and FR methods, respectively. Finally, we calculated the overall RMS (**Fig. 9a**) and mean (**Fig. 9b**) interpolation errors for the WAGHC standard depth layers. **Fig. 9c** shows the mean vertical gap size for the two cases. Our calculations demonstrate that the FR method results in a smaller interpolation error compared to the RR method. The layer-mean errors (**Fig. 9b**) show that the RR method produces a systematic positive bias, which is especially strong in the 500-1000 m layer, with

maximum values up to 0.05/0.2 °C for the small- and big-gap cases, respectively. In contrast, for the same layer, the FR method shows a maximum negative bias of only - 0.01/- 0.04 °C for the small- and big-gap cases, respectively. The spatial distribution of the mean interpolation error is shown in **Fig. S4**. For all three selected layers, the FR interpolation method results in smaller interpolation errors compared to the RR method. The difference between the two methods is more pronounced above and within the main thermocline (layers 182-218 m and 798-878 m, respectively). For the layer 182-218 m, the largest errors are found within the strongly stratified equatorial belt (**Fig. S4a, b**) for both the FR and RR methods. Within the main thermocline (798-878 m), the RR method shows the largest positive errors, up to 0.4 °C, across the entire tropical and subtropical belt.

The overall performance of the interpolation method described in the above sections is further illustrated by the temperature/depth histograms based on the observed and interpolated data, respectively (**Fig. 10**). To obtain these histograms, the number for each temperature-depth bin is normalized by the number of observations (observed or interpolated, respectively) in the most data-abundant temperature bin for each depth. A good agreement between the two histograms suggests the absence of spurious features in interpolated profiles.”

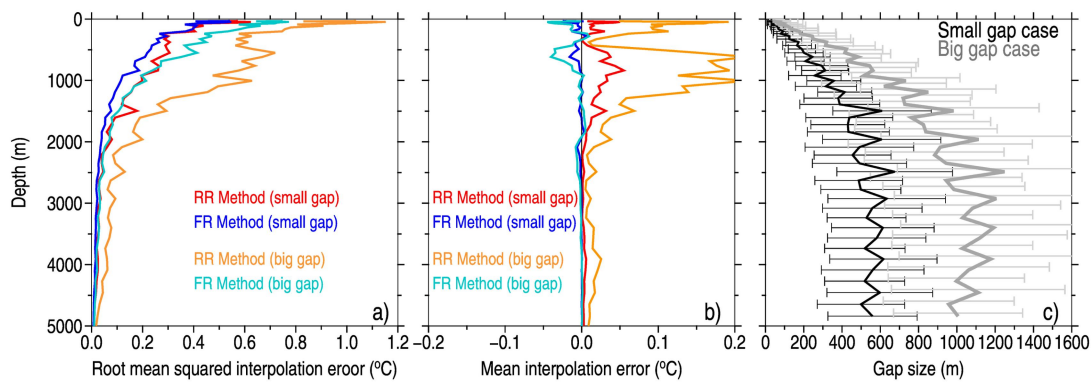


Figure 9. Global root mean squared (a) and global mean (b) error for interpolated temperature values obtained by means of RR method (red and orange) and FR method (blue and cyan). Red and blue profiles correspond to the small gap case, orange and cyan correspond to the big gap case. Explanation is given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