Historical Nansen Cast Temperature Profiles and the Ocean Heat Content Changes since 19th Century

Viktor Gouretski, Lijing Cheng, Juan Du

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Correspondence to: Viktor Gouretski (viktor.gouretski@posteo.de); Lijing Cheng (chenglij@mail.iap.ac.cn); Juan Du (dujuan10@mail.iap.ac.cn)

Abstract. Ocean heat content (OHC) estimates rely on subsurface temperature observations whose abundance, quality, and accuracy have changed over time as oceanographic techniques and methods have developed. The majority of the available global OHC time series begins in the 1950s when the International Geophysical Year 1957-1958 (IGY) observational campaign resulted in a rapid improvement of the observational basis. However, a significant amount of subsurface temperature observations had already been accumulated before the IGY. Extending knowledge back in time beyond current estimates provides further insight into global changes in the ocean-atmosphere system. We perform a three-step quality control (QC) of the Nansen cast temperature profiles with a focus on the time period before 1970. The QC includes automated and manual control of temperature observations, along with the quality control of temperature anomalies. We use the World Ocean Circulation Experiment/Argo Global Hydrographic climatology, based on high-quality temperature observations between 1990 and 2025, as the reference. To interpolate temperatures vertically, we use the temperature anomaly method by Frankignoul (1981). This method yields smaller interpolation errors than linear interpolation and the Reiniger and Ross (1968) parabolic method. Corrections for depth bias in Nansen cast profiles are also suggested. The oldest temperature profiles from the 1870s are treated separately, with the data from the "Challenger" expedition used as reference. Finally, we estimate the OHC changes between the 1870s and the 2020s for six selected multi-year time periods. We find that for the layer 0-2000 m the OHC increase between 1920-1944 and 1945-1969 amounts to 32.7 ± 6.8/37.5 ± 3.8 % of the total OHC increase between 1920-1944 and 2010-2025 for original and depth corrected profiles respectively.

1 Introduction

Recognition of the crucial role of the global ocean for the Earth climate system and experimental evidences of the ocean warming highlighted the importance of the global archive of temperature profiles as it became clear that the OHC evolution can be estimated from the historical data (Levitus et al., 2005; Gouretski and Koltermann, 2007; Cheng et al., 2025; Pan et al., 2026). The majority of OHC time series based on subsurface hydrographic observations begins in the late 1950s, as the observational campaign during the IGY 1957-1958 for the first time permitted a quasi-global coverage of the global ocean with hydrographic observations.

It should be noted that a significant number of temperature and salinity profiles has been accumulated over the time period preceding the IGY. The "Challenger" expedition 1873-1876 marks the beginning of the observational physical oceanography (Thomson and Murray, 1885; Roemmich et al., 2012). Guided by a group of scientists the expedition resulted in a large set of deep temperature profiles obtained with the accuracy and vertical resolution sufficient for the quantitative analysis of secular changes of the ocean thermal state. After the "Challenger" expedition

several important improvements in oceanographic practice have been introduced. One is related to the further development of reversing mercury thermometers (Negretti and Zambra, 1874) which soon replaced the minimum/maximum Six's thermometers used on the "Challenger" cruise (Austin and McConnel, 1980).

The other improvement is related to the thermometric method of thermometer depth estimation (Brennecke, 1921). The method found the first extensive application during the RV "Meteor" German Atlantic Expedition 1925-1927 (Wüst et al., 1932). The technique of taking water samples and measuring temperature at a number of depth levels is known under the name "Nansen cast" in commemoration of the Norwegian oceanographer Fridtjof Nansen who improved the method of serial hydrographic casts and developed the brass bottles (Nansen bottles) for taking water samples at depth levels (Helland-Hansen and Nansen, 1909; Nansen, 1906). Being gradually replaced by electronic profilers since late 1960s the Nansen cast method remained in use until the beginning of the 1990s as the main observational method on the ships of the former USSR - the country which contributed a significant part of all hydrographic profiles between the 1950s and the 1990s. Since 1940 mechanical bathythermographs (MBT) (Spilhaus, 1938) began to contribute to the archive of temperature profiles. However, the accuracy of MBT temperature measurements is much lower compared to that of Nansen casts and the maximum instrument depth is limited by 100-200 meters. The MBT data were found to be biased (Gouretski and Koltermann, 2007; Gouretski and Cheng, 2020).

Only a small number of the global-scale OHC estimates exists for the time period before the 1950s. Zanna et al. (2018) produced OHC time series since 1871 based on global reconstructions, which link the OHC with the observed sea surface temperatures (Rayner et al., 2003) using Green's functions derived from the data assimilation experiments (Wunsch and Heimbach, 2007). Fu et al. (2025) provided the update of the above reconstructions. Roemmich et al. (2012) and Gebbie and Huybers (2019) used data from the "Challenger" expedition to estimate OHC changes after the 1870s. Gouretski et al. (2012) provided estimates of the global temperature change between the "Challenger" expedition and the 1950s for the upper 400 m layer using in-situ hydrographic profiles.

In this study we describe the quality controlled and adjusted global archive of Nansen cast hydrographic profiles. Based on this dataset we assess the change of the ocean heat content of the upper 2000 m layer over a 150-year time period between 1870s and 2020s.

2 Data

For the current analysis we use only Nansen cast temperature profiles obtained between 1873 and 2025, with the focus on the time period before 1970. The World Ocean Database (WOD) (Locarnini et al., 2023) (update as of August 2025) served as the main data source providing 2,8 million profiles. The profiles from the WOD include *observed level data* from the Ocean Station Data (OSD) instrument type. A closer inspection of profiles suggests that some of them might represent vertically interpolated data, whereas another group of profiles may be a blend of data from both observed and interpolated levels. Unfortunately, in most cases no respective metadata exist regarding the possible application of the vertical interpolation. Apart from the WOD archive we used four temperature profile datasets which have been digitized using the published station logs. These include:

1) Temperature profiles of the British "HMS Challenger" 1873-1876 expedition (Thomson and Murray, 1889);

2) Temperature profiles from the cruises of the United States steamer "Tuscarora" in the North Pacific 1871-1874 (U. S. Hydrographic Office, 1874);

3) Temperature profiles from the German "Gazelle" 1874-1876 expedition (Hydrographisches Amt der Admiralität, 1888);

4) Temperature, salinity, and oxygen profiles from the German Atlantic Expedition obtained by the ship "Meteor" in 1925-1927 (Wüst et al, 1932). This dataset was used earlier for the estimation of secular changes in temperature and salinity in the South Atlantic Ocean (Gouretski et al., 2012).

All digitized profiles represent data at observed levels. The total number of temperature profiles between 1873 and 2025 retained for the analysis is 2,838,855, with 1,224,404 profiles for the time period before 1970. Quality controlled salinity and dissolved oxygen data are not used in the current analysis but accompany the temperature data when available. Automated quality control procedures for salinity and dissolved oxygen data are described by Gouretski (2018) and Gouretski et al. (2024).

2.1 Time and spatial distribution of profiles

The yearly number of profiles mirrors the development of observational physical oceanography resulting in a rapid growth of the data volume (**Fig. 1a**). Two periods (1914-1918 and 1940-1947) which include the two World Wars are characterized by the drop of the number of observations at all levels. For all time periods the number of data rapidly decreases with depth. Before the 1920s shallow profiles prevail, leaving the deep ocean below 1-2 km essentially unsampled. Before the 1920s a significant part of deep temperature measurements was obtained by the three expeditions in the mid-1870s (**Fig. 1b**) mentioned in the previous section.

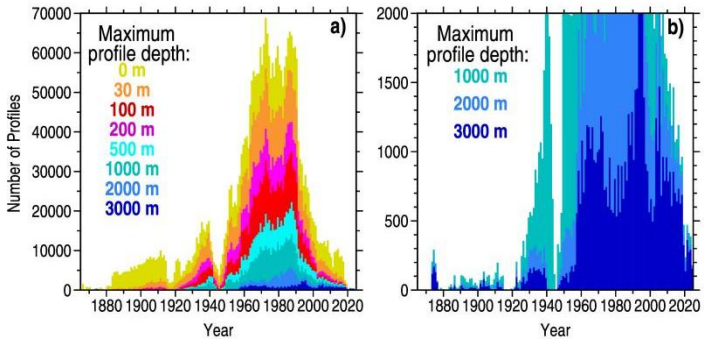


Figure 1. Yearly number of hydrographic profiles used in the study: a) all profiles; b) profiles deeper than 1000m.

The spatial distribution of profiles for selected time periods and for several maximum profile depths is given in **Fig. 2**. The two decade-long time period between 1950 and 1970, which includes the campaign of the International Geophysical Year (Chapman, 1959), is characterized by a quasi-global coverage even for the deep profiles. Before 1950 a relatively good spatial coverage (also with deep profiles) was achieved between 1920 and 1949, with the Atlantic Ocean exhibiting the densest sampling. The sampling between 1900 and 1920 was almost completely confined to the Atlantic Ocean. The largest data gaps are found in the remote central parts of the Pacific and Indian oceans. A coarse quasi-global array of ship tracks with subsurface temperature observations is available for the time period 1870-1900 with the most important contributions from the cruises by "HMS Challenger", "Gazelle", and "Tuscarora" in the mid-1870s.

2.2 Characteristics of the vertical sampling

The accomplishing of Nansen casts becomes more difficult with increasing cast depth. Deep casts require more wire, a larger number of bottles and thermometers, and a much longer cast time. For these reasons shallow casts are prevailing (**Fig. 3a**) with only 10 %, 5 % and 2.5 % of all casts being deeper than 1100 m, 1600 m, and 2500 m

respectively. The dynamical load on the sounding line (a hemp rope on the earliest cruises and a steel wire since the late 19th century) puts limitations on the maximum number of bottles which can be attached to the wire. Typically, not more than 10-12 bottles were used (Gould et al., 2013). Moroshkin (1951) reported about using up to 17 bottles during a cast. Therefore, for deep hydrographic stations several casts had to be accomplished to provide a sufficient vertical resolution throughout the water column. **Fig. 3b** shows the histograms of the number of observed levels. The average number of observed levels versus cast depth **(Fig.3c)** indicates that on average about 23 observed levels are available for the profiles with the deepest level around 5000 m. The vertical gap between a pair of neighbour bottles typically increases with depth from one or several tenth of meter near the ocean surface to 500 m or more in the deep part of the profile (**Fig. 3d**).

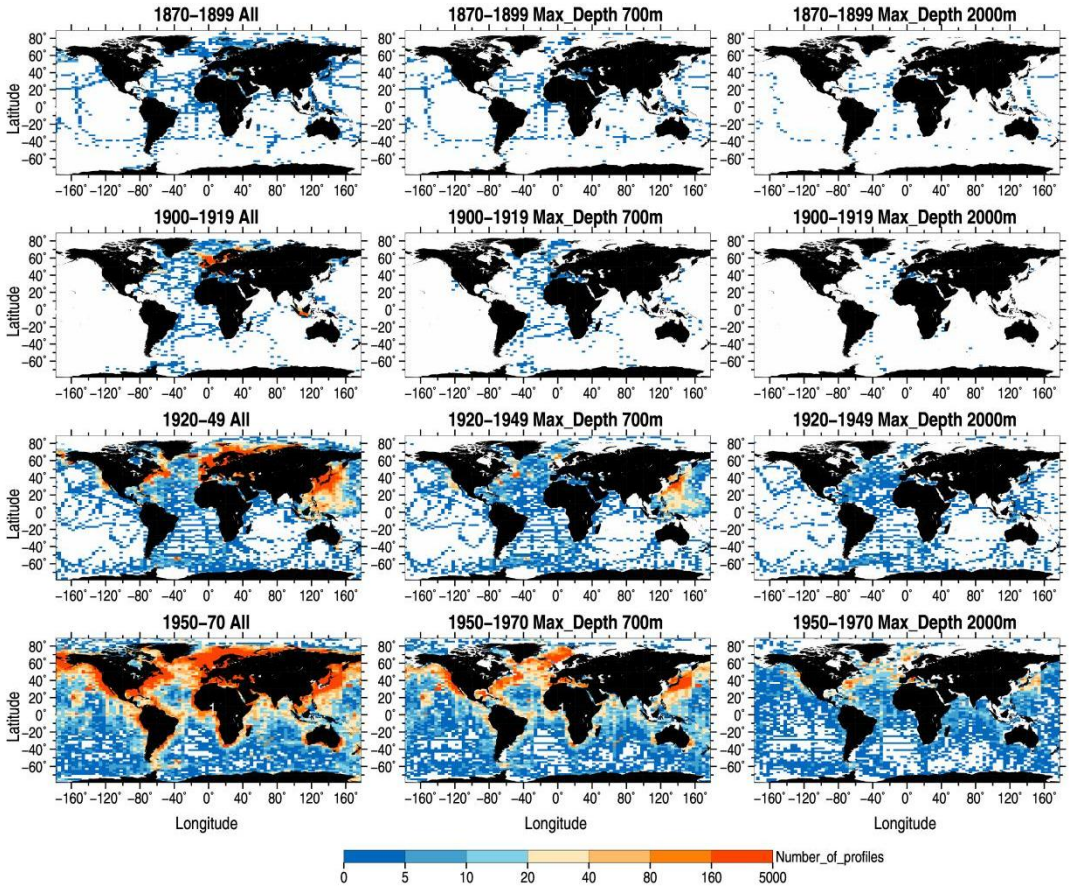


Figure 2. Number of profiles within 2° × 4° Latitude/Longitude boxes for several time-periods and profile maximum depths.

Since Nansen casts take samples and measure temperature at a number of discrete levels, the choice of the levels is made to permit a finer vertical resolution in the more variable upper part of the water column. Shown in **Fig.4** is the sample depth histogram for the period 1920-1970. A disproportionately large number of observations comes from a limited number of levels corresponding to the so called 'standard depths' which were usually chosen as target depth levels for sampling. The set of standard depths was recommended in several hydrographic manuals (Borishanski, 1954; State Oceanographic Institute, 1977; U.S. Naval Oceanographic Office, 1968).

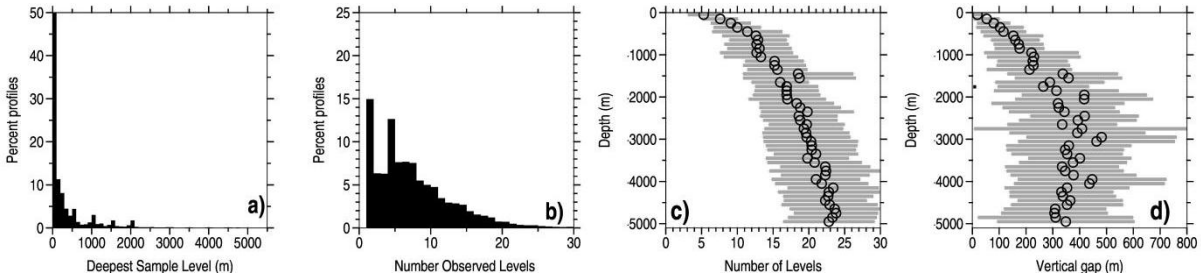


Figure 3. a) Deepest sample level histogram; b) number of observed levels histogram; c) mean number of observed levels versus cast depth; d) mean vertical gap size versus depth.

3 Data quality control

Imperfection and malfunctioning of instrumentation, especially in the early period, errors in reading temperature values, lack of the proper thermometer calibration, errors in sample depth lead to both random and systematic errors in temperature. The data used in the study have undergone three steps of quality control: statistical automated quality control, manual quality control, and statistical quality control of temperature anomalies. The steps are described in sections 3.1, 3.2, and 6.2 respectively.

3.1 Statistical quality control

First, the original temperature profiles were put through the automated quality control procedure described by Gouretski (2018). The procedure consists of a suite of independent quality checks with parameter thresholds for each check based on statistical properties of the entire data population. Outlier diagnostics for the automated quality control are shown in **Fig. 5a-c**. The overall percentage of the observed level temperature values identified as outliers is 4.2 %, with outliers found by 13 % of profiles. The outlier statistics suggest a gradual improvement of the data quality over time. The data before about 1920 are characterized by the highest percentages of outliers, which we attribute to the imperfectness of the early observational technique.

3.2 Manual quality control

Secondly, temperature profiles were subjected to manual quality control. The data have been screened for remaining outliers in 507 5° × 20° Latitude/Longitude boxes (**Fig. S1**). The manual control was based on the visual inspection of temperature profiles and temperature anomaly profiles within each box. The visual outlier check is focused mostly on the deep parts of profiles (typically below 500-1000 m) where seasonal changes attenuate and the temperature/depth relation is tight. Outlier diagnostics for the manual quality control are shown in **Fig. 5d-f.** Outliers are identified visually using observations which passed the automated quality control as a background reference. An increased percentage of manually identified outliers is clearly seen below 1000-1500m, that is in the depth range below the main thermocline. Examples of manually rejected observations are shown in **Fig. S2** for six randomly selected boxes. Temperature profiles for observed level data which passed both the automated and manual quality control are shown in **Fig. S3** for selected boxes. Overall, the manual quality control identified 0.17 % of all temperature observations as outliers with 0.85 % of the profiles affected by the manual check. We note, that in spite of improved accuracy and reliability of the modern oceanographic instrumentation the manual (expert) control and data evaluation remain a necessary step in the preliminary analysis of the hydrographic data (Swift, 2010).

4 Vertical interpolation of the observed level data

The choice of the interpolation method is important for historical profiles due to a sparse and uneven resolution over the water column. The originally crude vertical resolution further deteriorates due to the exclusion of observed levels during the quality control. As indicated by the **Fig. 3c** even the deepest profiles have on average about 23 observed levels. The distances between the observed levels (vertical gaps) increase with depth from typically 5-25 m near the surface to more than 300 m below about 1500 m (**Fig. 3d**).

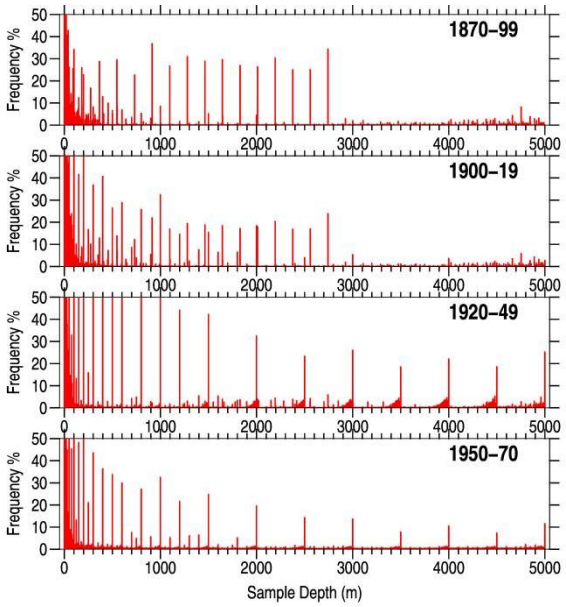
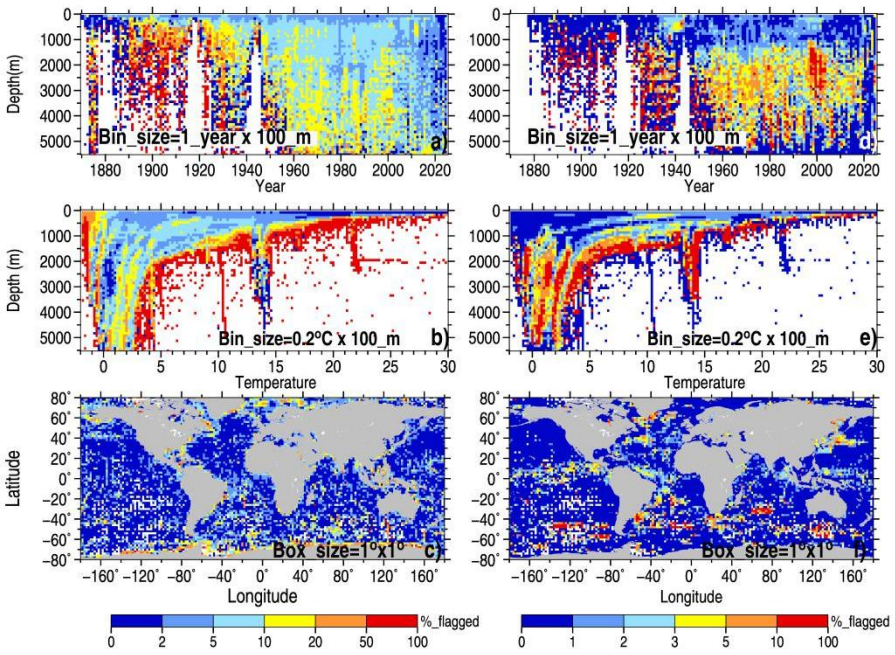


Figure 4. Sample depth frequency histograms for OSD temperature profiles for selected time periods.

Two methods were compared to vertically interpolate temperature profiles: the Reiniger and Ross (1968) interpolation method (RR method) and the climate anomaly method by Frankignoul (1981) (FR method). The RR method constructs two three-point parabolas giving two interpolated values. The final interpolated value is the distance-based weighted average of the two estimates, aiming to prevent artificial extrema. The FR method uses the local climatological temperature profile in addition to the observed temperatures and interpolates temperature anomalies (deviations of individual temperature values from the respective climatological values). We do not use a simple linear interpolation here and opt for the RR and FR methods instead. As shown by Li et al. (2022) using linear interpolation leads to a positive temperature bias for sparsely vertically sampled historical temperature profiles.



183 **Figure 5. Outlier percentage for OSD temperature profiles between 1870 and 2025: a-c) automated statistical quality**
184 **control; d-f) manual quality control. Outlier percentage is shown for depth/year bins (a, d), depth/temperature bins (b, e), 1°**
185 **× 1° latitude/longitude boxes (c, f).**
186
187 ### 4.1 Case of uneven distribution of observed levels over depth
188 The straightforward application of the RR method to observed level profiles was found to result in unrealistically
189 large excursions on the interpolated temperature profile for a not-negligible number of profiles. Such cases typically
190 occur when the observed levels are distributed extremely unevenly so that neighbour vertical gaps differ by more than
191 an order of magnitude. Such profiles occur for instance when the entire temperature profile at a hydrographic station is
192 composed of two or more hydrographic casts, each covering a part of the entire water column. When such profiles are
193 merged together in the final station table the deepest observed depth of the shallower cast might be very close to the
194 shallowest observed level of the deeper cast. In other cases, small gaps may be due to the inclusion of the interpolated
195 levels along with the observed levels in the resulting station table, or when big gaps on profiles appear due to the
196 rejection of outliers. Since the interpolating polynomial must pass through each point of the original temperature profile
197 unrealistic over-shootings appear on the interpolated profile. **Fig. 6** shows several examples of such profiles. In some
198 cases, the overshooting amounts to several degrees. To avoid the overshooting the original profiles are checked for pairs
199 of levels which are very close to each other before interpolation. Such pairs are then replaced by one level referred to
200 the middle point and temperature is taken as the average of temperature at the two levels. Another option is to use a
201 simple linear interpolation in such cases.
202

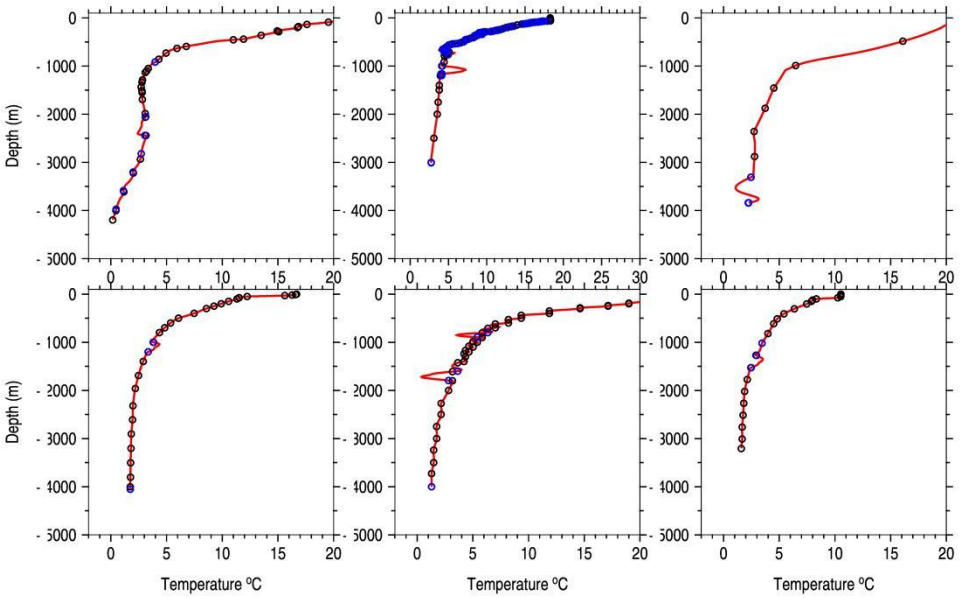


Figure 6. Examples of temperature profiles with uneven distribution of observed levels resulting in unrealistic overshooting of interpolation polynomials in RR method: observed levels (black circles), observed levels with a small distance to a neighbor level (blue circles), interpolated profile (red).

4.2 Dependence of interpolation error on distance between observed levels.

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. The vertical interpolation was then performed for the resampled profile. Resampling was done by deleting a number of levels below the selected level so that the gap between the level under consideration and the next level below was increased iteratively. For each climatological temperature profile (i) the interpolation error $\varepsilon(z,g)_i$ for a given WAGHC depth level z and the gap size g is defined as the difference between the temperature value from the full profile $T(z)$ and the temperature value from the interpolated resampled profile $\check{T}(z,g)$: $\varepsilon(z,g)_i = T(z) - \check{T}(z,g)$. The root mean squared value of $\varepsilon(z,g)_i$ over all WAGHC grid nodes between 70°S and 70°N gives a measure of the global root mean squared interpolation error $\varepsilon(z,g)$ (**Fig. 7**). The error refers to the center of the gap and increases with increasing gap size and decreases with increasing depth. We note that **Fig. 7** serves as a guideline and characterizes the global mean interpolation error. Regional interpolation errors may be different due to variations in the vertical temperature stratification over the global ocean.

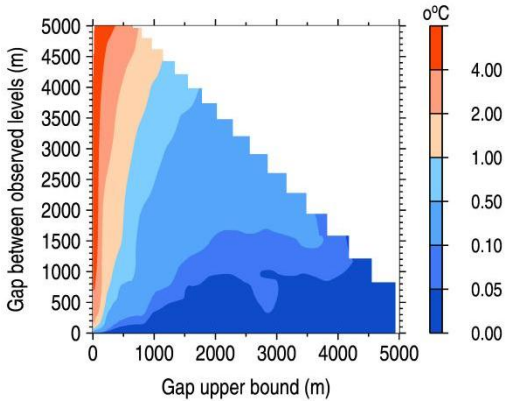


Figure 7. Global root mean squared interpolation error versus depth and the vertical gap size. Error refers to the center of the gap.

4.3 Comparison of the Reiniger-Ross interpolation method and the climate anomaly method

As suggested by the **Fig. 7** the global root mean squared interpolation error in RR method increases with the gap size. To reduce the error for big 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 represented by the sum of the mean climatological temperature value 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. The RR method in case of large gaps tends to follow closer to a simple linear interpolation over a 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. In order 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 the level k levels *k-1* and *k+1* are omitted as well. Vertical interpolation of reduced profiles is performed on the depth of the removed level *k*, and the difference between the observed and interpolated temperature gives the estimate of 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. Interpolation is performed twice, using both RR and FR methods respectively. Finally we calculated the overall rms (**Fig. 9a**) and mean (**Fig. 9b**) interpolation errors within the WAGHC standard depth layers. **Fig. 9c** shows the mean vertical gap size for the two cases. Our calculations demonstrate that FR method results in a smaller interpolation error compared to the RR method. The layer-mean errors (**Fig. 9b**) show that the RR method results in a systematic positive bias which is especially strong in the layer 500-1000m with maximum values up to 0.05/0.2 °C for the small and big gap case respectively. In contrary, for the same layer the FR method indicates a maximum negative bias of only -0.01/-0.04°C for the small and big gap case 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**) both for the FR and RR methods. Within the main thermocline (798-878 m) the RR method exhibits the largest positive errors up to 0.4 ºC for 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.

5. Estimation of the Nansen cast depth bias

Only under ideal conditions the wire with attached Nansen bottles takes a vertical position over the entire wire length. Because currents and wind cause ship drift during the occupation of the hydrographic station the wire deviates from the vertical so that the length of the wire paid out is greater than the actual depth at which the temperature was measured. Brennecke (1921) suggested to use paired pressure protected and pressure unprotected mercury in glass thermometers for the calculation of the actual bottle depth. The unprotected thermometer shows a higher temperature compared to the protected one with the difference being dependent on hydrostatic pressure which can be re-calculated into depth. The method was first implemented during the German Antarctic Expedition 1911-1912, but the first full-scale implementation took place during the German Atlantic Expedition 1925-1927 (Wüst, 1932).

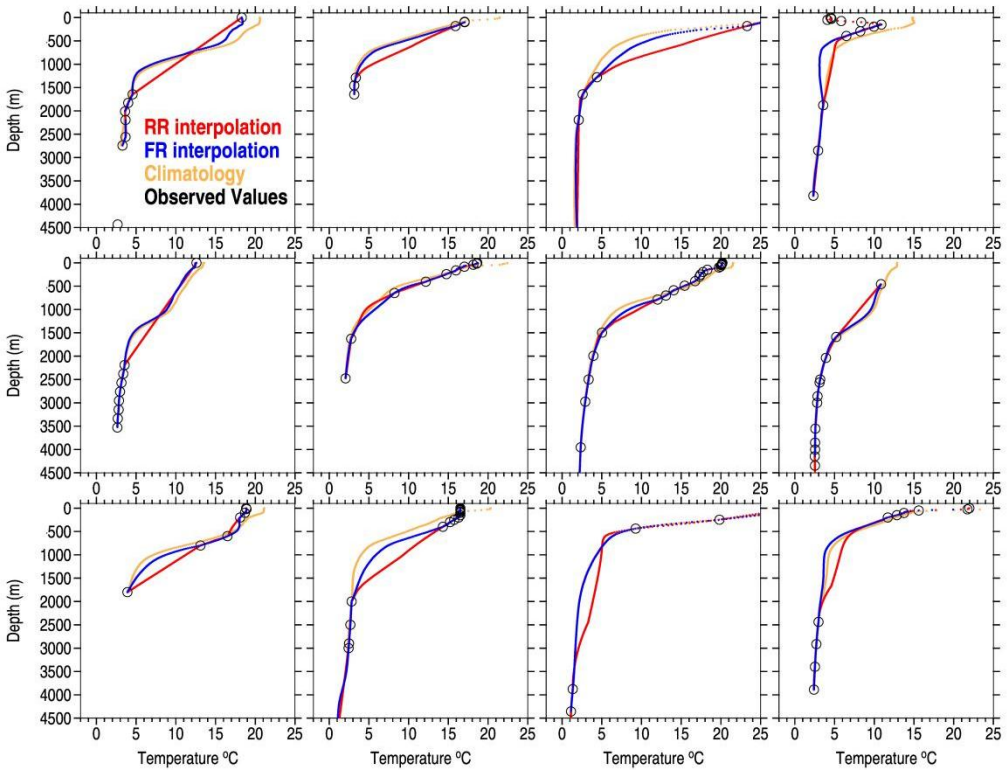


Figure 8. Comparison of RR (red) and FR (blue) interpolation methods for 16 arbitrarily selected temperature profiles with large vertical gaps. Observed level values are marked with black circles, WAGHC climatology temperature profiles are shown in green.

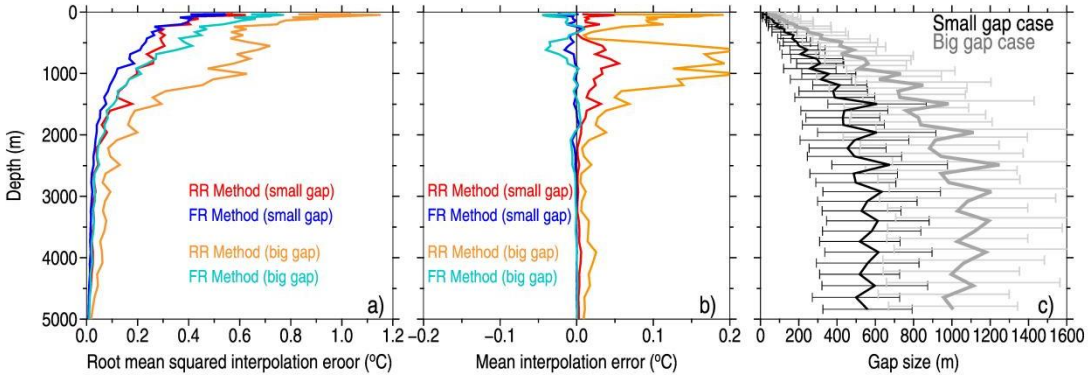


Figure 9. Global root mean squared (a) and global mean (b) error for interpolated temperature values obtained by means of RR method (red and orange) and FR method (blue and cyan). Red and blue profiles correspond to the small gap case, orange and cyan correspond to the big gap case. Explanation Is given 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.

Unfortunately, the thermometric method was not always implemented in oceanographic practice. Respectively, for a substantial number of Nansen casts the reported sample depths in reality represent target (standard) depth levels (see **Fig. 4)**. Typically, only several bottles of a hydrographic cast were equipped with paired thermometers. Since the pressure effect becomes noticeable only below a certain depth the paired thermometers were used at depths greater than several hundred meters. The potential depth bias for Nansen casts was investigated by Gouretski et al. (2022), who used collocated pairs of Nansen cast and temperature profiles from Conductivity/Temperature/Depth (CTD) profilers. Such collocated data are available only for the time period after 1970 since the introduction of CTDs. The Nansen casts were found to be biased in depth with the sample depth reported by station logs overestimating the actually reached depth. Gouretski et al. (2022) suggested corrections for depth bias and also pointed to another method of depth bias estimation, based on the analysis of the sample depth histogram. Here we apply this method to derive depth corrections for old Nansen casts obtained before 1970. As illustrated by the **Fig. 4** the disproportionately large number of observations comes from a limited number of the so-called standard levels which were chosen as target depth levels for sampling. The non-standard depth levels can be linked to the profiles where the thermometric method was used or where corrections were applied to account for the deviation of the wire from the vertical.

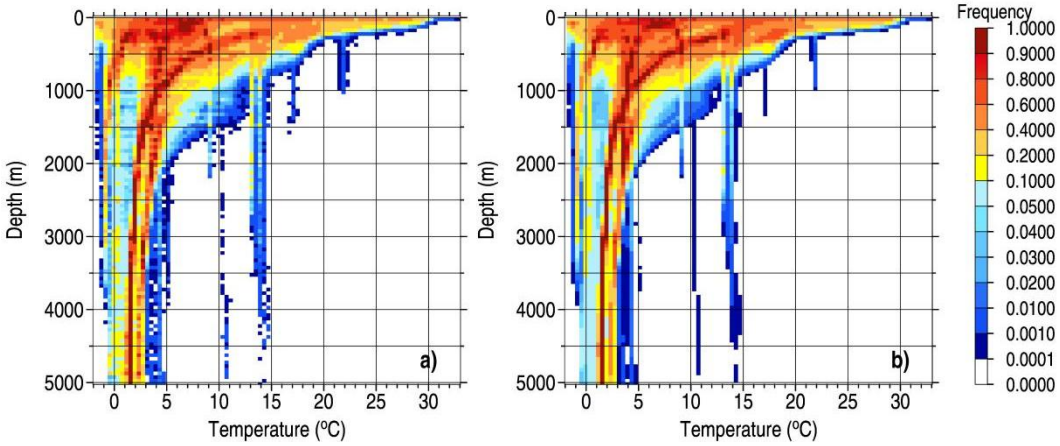


Figure 10. Normalized temperature-depth histograms for validated observed level data (a) and for interpolated data (b).

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 (Z_{max}) relative to the respective standard level (Z_{stand}) 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 $R = (Z_{max})/(Z_{stand})$ 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.

Depth corrections were determined for the time period 1920-1970. For this time period no collocated CTD profiles are available. We note that depth corrections should not be applied for the profiles for which the usage of the thermometric method is confirmed by the respective metadata (for instance the data of the German Atlantic Expedition 1925-1927 (Wüst et al, 1932). The impact of the depth corrections on OHC estimates is discussed in section 8.5.

6 Calculation of temperature anomalies

One of the goals of the current study is to estimate changes in ocean heat content before the International Geophysical Year 1957-1958. Such estimation is based on temperature anomalies calculated relative to a base-line climatology.

6.1 WOCE-Argo Global Hydrographic Climatology – the reference ocean thermal state

The updated WOCE/Argo Global Hydrographic Climatology (WAGHC) (Gouretski, 2018) was used to calculate temperature anomalies. This climatology represents climate monthly values of temperature on a 0.5° grid and is based on the extended set of Argo, bottle, and CTD profiles (status as of June 2024) (the original version (Gouretski, 2018) included the temperature profiles before 2018). The climatology is based on 201,962 OSD profiles and 714,716 CTD profiles between 1990 and 2024, and 3,035,127 Argo profiles between 2001 and 2024. **Fig. S5** shows the annual climate temperature field at selected levels. In contrast to the widely used World Ocean Atlas climatologies (Garcia et al., 2018; Locarnini et al., 2023), the WAGHC climatology provides information about the median year of climatology temperature field (**Fig. S6),** which varies in the range 2014-2015 and 2003-2006 above and below the Argo maximum depth respectively.

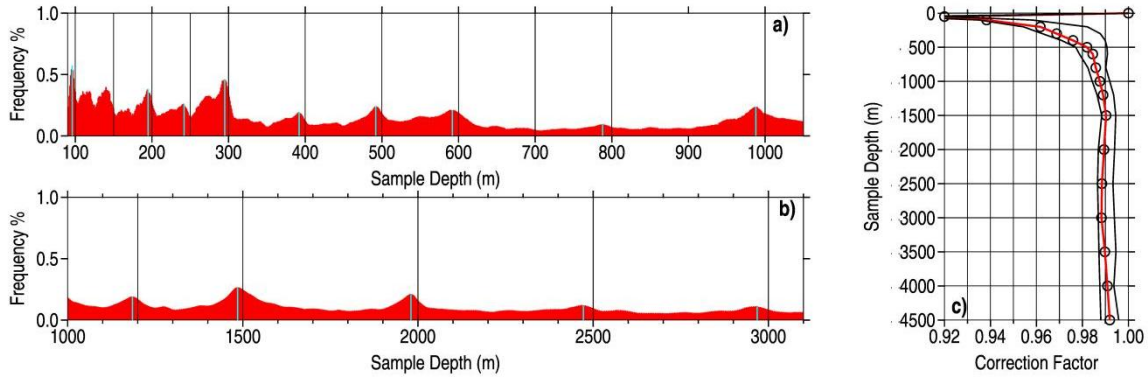


Figure 11. a, b) Smoothed sample depth histograms with deleted standard levels for the smoothing kernel window of 10 meters. Black vertical lines correspond to the standard depths; cyan bars denote maximum frequency. Differences between the standard depth and the depth of the respective local maximum frequency represent the depth bias; c) median depth correction factor versus depth (red); thin black lines correspond to the minimum and maximum value over 50 estimates obtained by varying smoothing kernel width between 1 m and 50 m.

6.2 Temperature anomalies relative to the WAGHC climatology

For each observed temperature profile, climatology values are obtained by means of spatial interpolation of 0.5° resolution climatology temperature fields. Above/below 200 m monthly/annual climatology fields were used respectively. The standard deviation of temperature anomalies within 2° × 4° latitude/longitude boxes (**Fig. S7**) reveals a well-known pattern with high variability areas confined to the western boundary current systems (Gulf Stream, Kuroshio, Agulhas, Malvinas), to the Antarctic Circumpolar current, and to the equatorial and tropical bands in the Pacific, Atlantic, and Indian Oceans.

An additional quality control check is applied to individual temperature anomalies. First, the overall anomaly limits versus depth are calculated based on all temperature profiles. The minimum/maximum anomaly limits are defined as the 1st and 99th quantiles respectively (**Fig. S8**). To arrive on a local threshold at the geographical location (Lat, Lon) and at the depth level Z the overall thresholds are modified by the multiplicative factor $\sigma(x,y,z)/\sigma_{MEAN}(z)$, where $\sigma(x,y,z)$ is the anomaly standard deviation within the respective 2° × 4° Latitude/Longitude box, and σ_{MEAN} is the global mean of $\sigma(x,y,z)$. **Fig. 12** shows outlier diagnostics for the anomaly check. This check rejects 0.55 % of all observations as outliers, with about 2.10 % of profiles being affected.

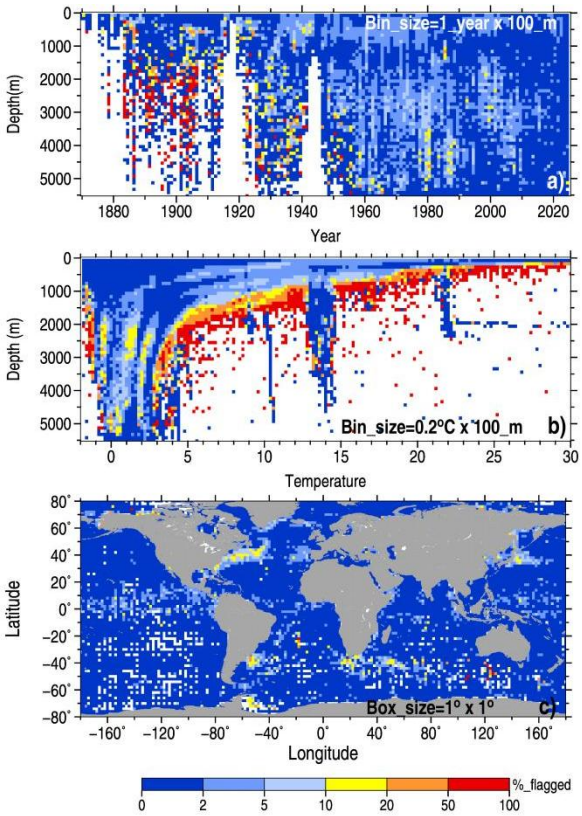


Figure. 12. Outlier percentage for temperature anomalies: a) percentage in depth/year bins, b) depth/temperature bins, c) 1° × 1° Latitude/Longitude boxes.

7 Adjustment of the oldest temperature profiles

Whereas the first subsurface temperature measurements available in the WOD database date back to the cruises from the end of 18th century, the voyage of "Challenger" (1873-1876) represents the first large-scale oceanographic expedition with temperature measurements conducted by scientists. The comparison of the "Challenger" temperature data with Argo-era temperatures revealed a profound warming of the global ocean along the entire track of the expedition (Roemmich et al., 2012). Almost simultaneously with "Challenger" voyage two other major expeditions were conducted by the German ship "Gazelle" (Hydrographisches Amt der Admiralität, 1888) and the US ship "Tuscarora" (Belknap, 1874). Both cruises did not have scientists on board. The location of stations for all three expeditions is shown in **Fig. 13a.** Analysis of temperature anomalies revealed significant systematic differences between the expeditions, with the most pronounced disagreement found between "Challenger" and "Tuscarora" temperature measurements. Because of a larger size, better vertical resolution, and a wider geographical scope of the "Challenger" data, along with the fact that the measurements and their analysis being conducted under scientific supervision, we use the "Challenger" data as the reference to which temperatures from "Gazelle" and "Tuscarora" cruises are adjusted.

The choice of stations for comparison is defined by the maximum distance between profiles of 2000 km. For "Challenger" and "Tuscarora" this area includes the majority of "Challenger" and "Tuscarora" stations in the North Pacific (**Fig. 13c**), whereas for "Challenger"-"Gazelle" comparison profiles in all three oceans are used (**Fig. 13b**). Distribution of point temperature anomalies over depth (**Fig. 13 d, e)** and binned anomalies **(Fig. 13f, g)** suggests a linear correlation of anomaly differences between the expeditions. Linear regressions for binned anomalies below 500 m are: $\Delta T_{Gazelle} = 0.223 - 9.195*10^{-5} Z$; $\Delta T_{Tuscarora} = -0.130 - 1.358*10^{-4} Z$, where Z is the "Challenger" sample depth in

meters, and ΔT is the respective difference relative to "Challenger" data in degrees Celsius. These corrections were applied to all temperature profiles from "Gazelle" and "Tuscarora". **Fig. S9** shows the distribution of temperature anomalies versus depth for quality-controlled and adjusted "Challenger", "Gazelle", and "Tuscarora" temperature profiles.

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 Young'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.

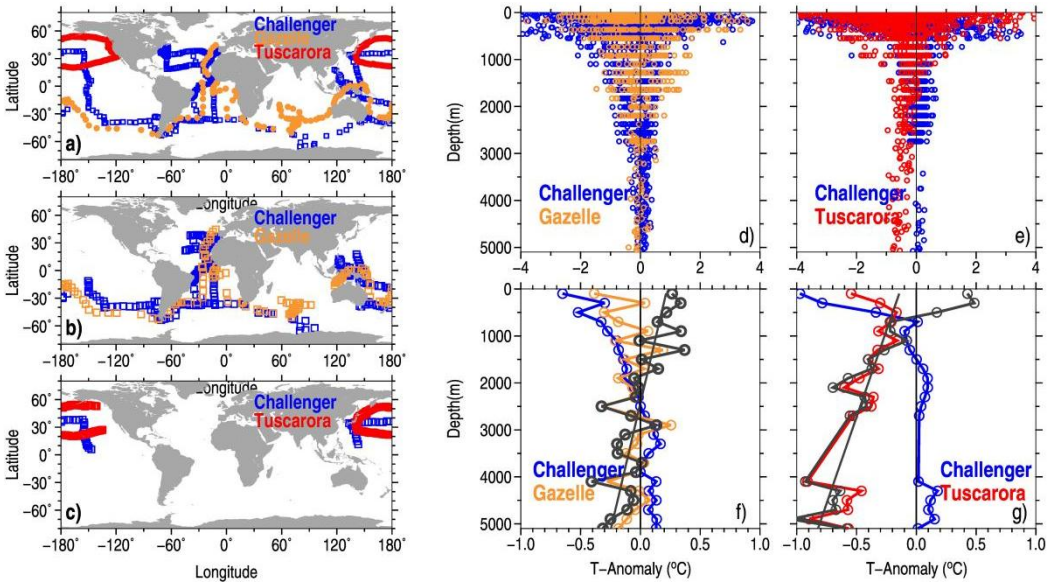


Figure 13. a) Station locations for "Challenger", "Gazelle", and "Tuscarora" expeditions; b) location of "Challenger" and "Gazelle" stations used for data inter-comparisons; c) same as b) but for "Challenger" and "Tuscarora" data; d) comparison of temperature anomalies for "Challenger" and "Gazelle" data; e) same as d) but for "Challenger" and "Tuscarora" data; f) mean temperature anomalies in 200-meter layers for "Challenger" (blue) and "Gazelle" (orange) data and the respective anomaly difference (black); g) same as f) but for "Challenger" (blue) and "Tuscarora" (red) data.

8. Secular changes of the ocean thermal state since 1870s based on in-situ observations

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- to 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 interpolated 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.

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 (E_b). 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 ºC. 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 ºC 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 E_b, 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 E_b, 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 E_b. 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 (E_g) represents the second source of uncertainty and characterizes how accurately the observed boxes represent the entire global ocean. According to Boyer et al. (2016) E_g 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., 2013) 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 (OHC_{all}) and using only sampled grid boxes (OHC_{samp}). 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 $N=62-N_p$ (N_p is the number of years spanned by the respective time period). The root-mean-squared difference (OHC_{all}) -(OHC_{samp}) is used as the measure of E_g. Both errors are summed in quadrature to get the total uncertainty $E_{tot}^2 = E_b^2 + E_g^2$. Other kinds of uncertainty (for instance uncertainty due to errors in the base-line climatology) are not considered in this study.

8.3 Spatial patterns of temperature anomalies

Difference anomaly maps (**Fig. 14)** illustrate temperature changes between the respective time periods in the layers 0-500 m, 500-1000 m, and 1000-2000 m. During the two earliest periods 1873-1876 and 1920-1944, large areas in the central parts of the Pacific and Indian oceans remained essentially unsampled. The general data paucity is also reflected in patchiness of the basin-scale temperature anomaly pattern due to remaining biases and mesoscale noise. However, in spite of a coarse sampling within each layer a progressive warming between 1870s and 1920s can be seen, with the warming most pronounced in the Atlantic and in the North Pacific oceans (**Fig. 14a, f, k)**. The 0-500 m layer features a

cooling in the West Pacific Ocean. For the time period between 1920-44 and 1945-69 (**Fig. 14b, g, l**) the maps indicate a full-depth strong warming of the Atlantic Ocean south of about 30°N accompanied by a cooling in the vast area extending in the south-west/north-east direction between 60-20°W to the north. Large areas of cooling are found in the northern and western parts of the Indian ocean. However, the data basis is rather poor in this region.

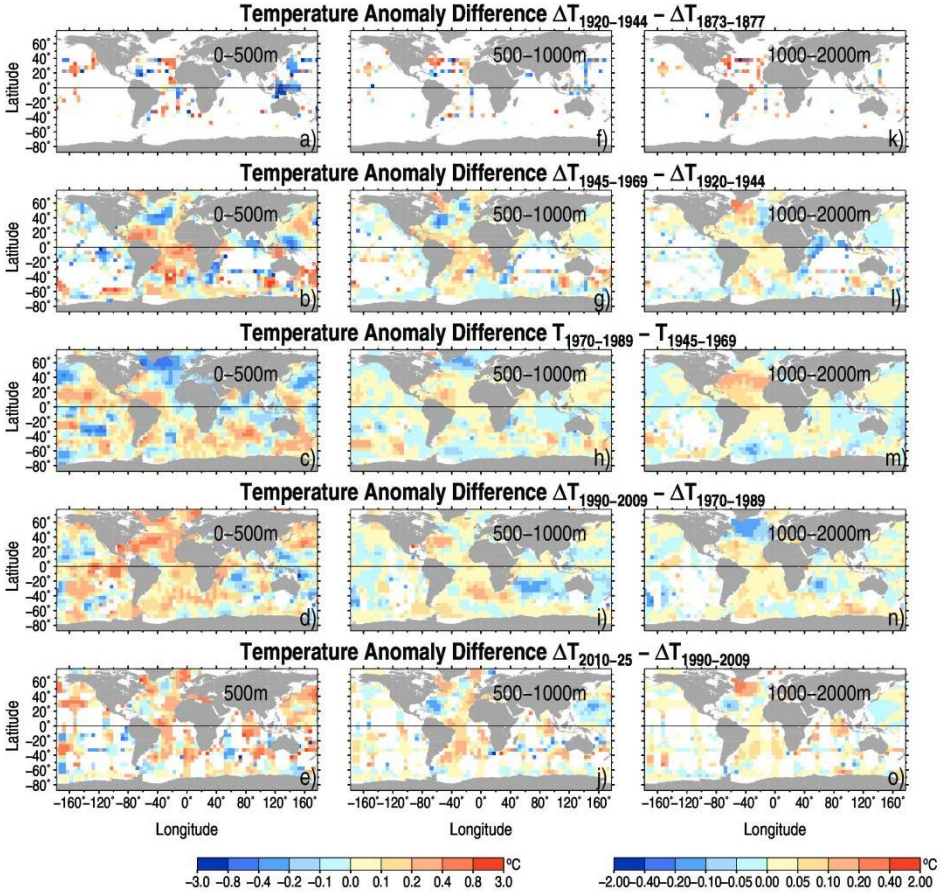


Figure 14. Layer-averaged temperature anomaly differences between selected time periods in 5° × 5° latitude/longitude boxes: 0-500 m (a-e), 500-1000 m (f-j), 1000-2000m (k-o). Note a different color scale for the layer 1000-2000m.

Between 1945-1969 (this period is strongly biased towards the data from the International Geophysical Year 1957-1958) and the most data abundant period 1970-1989, the upper 500 m layer exhibits a prevailing global warming pattern except for the North Atlantic, North Pacific and North Indian Ocean oceans (**Fig. 14c**). In the deeper layers (**Fig. 14h, m**) the strongest warming is confined to the tropical and subtropical regions of the Atlantic Ocean. For the following time period between 1970-1989 and 1990-2009 (the latter is dominated by the data from the World Ocean Circulation Experiment WOCE) a global warming pattern is revealed for the layer 0-500m, including the Arctic Ocean, with the exception for the central Indian Ocean and the south-western part of the tropical western Pacific Ocean (**Fig. 14d**). The prevailing warming trend remains for the deeper layers. However, the North Atlantic Ocean (1000-2000 m) and the South Indian Ocean (500-1000m) (**Fig. 14 i, n**) both feature strong cooling. The progressing global-scale warming is further revealed for the latest time period between 1990-2009 and 2010-2025 (**Fig. 14 e, j, o)** with the strongest deep warming (1000-2000m) found in the North-West and South Atlantic Ocean.

8.4 Basin-wide and global vertical profiles of temperature anomalies

Box-averaged temperature anomalies are used to calculate vertical profiles of average anomalies for distinct oceans and for the entire global ocean. As already indicated by the spatial anomaly difference pattern (**Fig. 14**) the most profound temperature changes took place in the Atlantic Ocean (**Fig. 15a, b**), with the Pacific Ocean showing the smallest secular temperature changes (**Fig. 15f**). For the global ocean the continuous warming between the 1870s and the 2020s is seen throughout the entire water column, except for a close to zero warming between the 1945-1969 and 1970-1989 time periods within the upper 500 m (**Fig. 15g, h**). Since the 1920s the upper 50 m layer warmed by about 0.65 °C globally, whereas the global temperature increases at 1000 m and 2000 m level amounts to 0.1°C and 0.04 °C respectively.

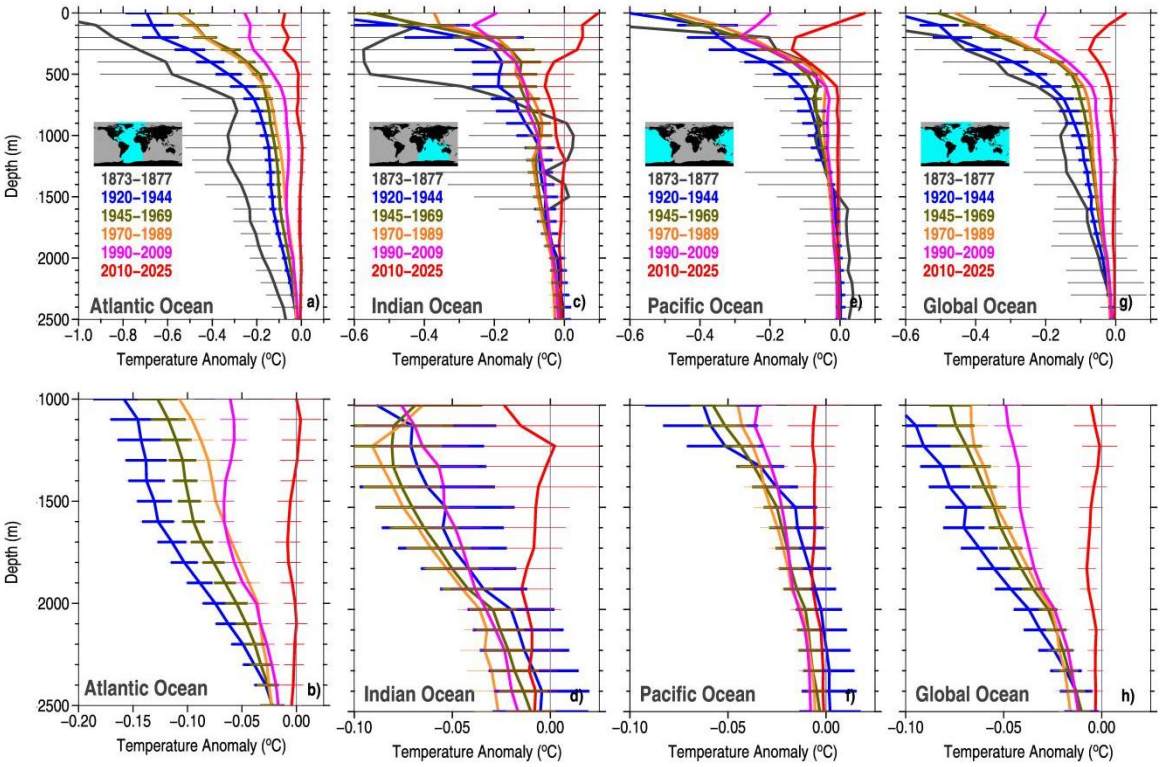


Figure 15. a, c, e, g) Temperature anomaly vertical profiles for the three main oceans and for the entire global ocean and for selected six time periods with standard error bars; b, d, f, h) same but for the depth range below 1000 m. Note the different temperature scale for the Atlantic Ocean.

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/m^3 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. Throughout 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.

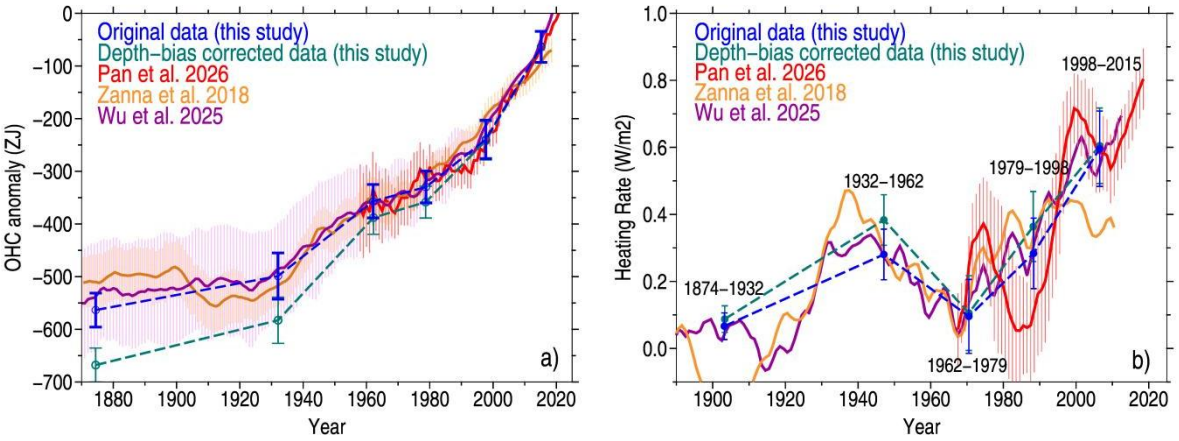


Figure 16. a) Global OHC anomalies for selected time periods for original data (blue) and depth-bias corrected data (cyan), OHC anomaly time series from Pan et al. (2026) (red), Zanna et al. (2018) orange), and Wu et al. (2025) (magenta) are shown for comparison; b) heating rate for five time periods: 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).

9 Discussion and conclusions

The global archive of temperature profiles (Locarnini et al., 2023) was used in a number of studies where the evolution of the thermal state of the global ocean was estimated (Levitus et al., 2005; Gouretski and Koltermann, 2007; Cheng and Zhu, 2016; Pan et al., 2026). However, the entire potential of the historical data archive has not been used as these analyses were based on the data obtained after 1940s. This study takes in focus the oldest available subsurface temperature observations, starting with the first subsurface temperature measurements for scientific purposes during the voyage of the British ship "Challenger" in the 1870s. Since the old data are less accurate and often provide only a crude vertical resolution a reliable interpolation method is of importance. We applied the anomaly method by Frankignoul (1981) to the historical profiles to reduce the interpolation error while interpolating over big vertical gaps. All historical profiles have undergone a three-step quality control, which includes the automated statistical quality control procedure described by Gouretski (2018), the manual control, and the automated control of temperature anomalies relative to the base-line climatology. The updated WOCE/Argo global hydrographic climatology (Gouretski, 2018) is used, which is based on the high-quality temperature profiles for the time period 1990-2025.

Because of the general data paucity, we calculated mean temperature anomalies in 5° × 5° latitude/longitude boxes which are used for the estimation of basin-averaged temperature anomalies for 5- to 25-year-long time periods. Our results confirm a progressive warming of the subsurface ocean since 19th century, which is seen both globally and for the three main oceans in all layers above 2000 m. The estimated OHC change since 1950s agrees well with the recent results by Pan et al. (2026), which are based on a much larger dataset and apply a temperature anomaly mapping method (Cheng and Zhu, 2016).

We show that the beginning of the ocean warming can be reliably traced back to the 1920s and likely even to the end of the 19th century. Available historical data suggest a strong warming of the global ocean before the International

Geophysical Year 1957-1958. We find that between 1920-1944 and 1945-1969 the global OHC increased by 142.2 ± 48.7 (194.9 ± 54.2) ZJ. These values amount to 32.7 ± 6.8 (37.5 ± 3.8) % of the total OHC increase between 1920s and 2020s for the original and depth corrected temperature profiles respectively.

10 Data availability

The quality control procedure described above was applied to the OSD profiles between 1873 and 2025 from the World Ocean Database and to the additional digitized datasets described in the paper. The resulting dataset comprises observed level data with quality flags and data interpolated to 10-meter levels. The data are in NetCDF format with metadata. The complete dataset (Gouretski et al., 2026) can be accessed freely at https://zenodo.org/records/22764102 and http://doi.org/10.12157/IOCAS.20260520.001.

Author contributions.

VG and LC – conceptualization, methodology, writing; JD – software, editing, data formatting.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgments. We thank the National Large Scientific and Technological Infrastructure "Earth System Numerical Simulation Facility" (https://cstr.cn/31134.02.EL) for the technical support. We are thankful to Yuying Pan, Laure Zanna and Quran Wu for providing their OHC time series. Rebecca Cowley's help with publication references on manual quality control is appreciated. Our special thanks go to Franco Reseghetti, Peter Koltermann and to the two anonymous reviewers, for their detailed comments and suggestions.

Financial support. This study is supported by National Natural Science Foundation of China (Grant no. 42261134536), National Key R&D Program of China (Grant No. 2023YFF0806500), the International Partnership Program of the Chinese Academy of Sciences (Grant No. 060GJHZ2024064MI), Asian Cooperation Fund, the New Cornerstone Science Foundation through the XPLORER PRIZE.

References

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.

Austin, J. F. and McConnel, A.: James Six F. R. S.: Two hundred years of the Six's self-registering thermometer, Notes Rec. R. Soc. Lond., 35, 49—65, https://doi.org/10.1098/rsnr.1980.0004, 1980.

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., and Weatherall, P.: Global bathymetry and elevation data at 30 arcsec resolution: SRTM30_PLUS, Mar. Geod., 32, 355—371, https://doi.org/10.1080/01490410903297766, 2009.

Borishanski, L. S.: Manual for the performance of work on standard hydrological sections and marine hydrological surveys, Gidrometeoizdat, Leningrad, 284 pp., https://search.rsl.ru/ru/record/01005762178 (last access: 8 October 2026), 1954 (in Russian).

Boyce, F. M.: A brief study of the protected deep-sea reversing thermometer, Manuscript Report Series (Oceanographic and Limnological), No. 227, 17 pp., https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/40648199.pdf (last access: 8 October 2026), 1966.

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, Journal of Climate, 29, 4817-4842, DOI: 10.1175/JCLI-D-125-0801.1

Brennecke, W.: Die ozeanographischen Arbeiten der Deutschen Antarktischen Expedition 1911–1912, Archiv der Deutschen Seewarte, XXXIX, No. 1, Hamburg, https://digitale-bibliothek.bsh.de/viewer/api/v1/records/141316/tei/en/ (last access: 8 October 2026), 1921.

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.

Chapman, S.: The International Geophysical Year, 1957–58, Nature, 175, 55–56, https://doi.org/10.1038/175055a0, 1959.

Cheng, L. and Zhu, J.: Benefits of CMIP5 multimodel ensemble in reconstructing historical ocean subsurface temperature variation, J. Climate, 29, 5393–5415, https://doi.org/10.1175/JCLI-D-15-0730.1, 2016.

Cheng, L. J. and Coauthors: Record high temperatures in the ocean in 2024, Adv. Atmos. Sci., 42, 1092–1109, https://doi.org/10.1007/s00376-025-4541-3, 2025.

Frankignoul, C.: On the depth interpolation of repeated hydrographic data, J. Geophys. Res., 86, 2001–2004, https://doi.org/10.1029/JC086iC03p02001, 1981.

Garcia, H. E., Boyer, T. P., Baranova, O. K., Locarnini, R. A., Mishonov, A. V., Grodsky, A., Paver, C. R., Weathers, K. W., Smolyar, I. V., Reagan, J. R., Seidov, D., and Zweng, M. M.: World Ocean Atlas 2018: Product Documentation, A. Mishonov, Technical Editor, https://doi.org/10.25923/ng6j-ey81, 2019.

Gebbie, G. and Huybers, P.: The little ice age and 20th-century deep Pacific cooling, Science, 363, 70–74, https://doi.org/10.1126/science.aar8413, 2019.

Gould, J., Sloyan, B., and Visbeck, M.: Chapter 3 – In situ ocean observations: a brief history, present status, and future directions, International Geophysics, 103, 59–81, https://doi.org/10.1016/B978-0-12-391851-2.00003-9, 2013.

Gouretski, V.: World Ocean Circulation Experiment – Argo Global Hydrographic Climatology, Ocean Sci., 14, 1127–1146, https://doi.org/10.5194/os-14-1127-2018, 2018.

Gouretski, V. and Cheng, L.: Correction for systematic errors in the global dataset of temperature profiles from mechanical bathythermographs, J. Atmos. Ocean. Tech., 37, 841–855, https://doi.org/10.1175/JTECH-D-19-0205.1, 2020.

Gouretski, V. and Koltermann, K. P.: How much is the ocean really warming?, Geophys. Res. Lett., 34, L01610, https://doi.org/10.1029/2006GL027834, 2007.

Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global ocean database, Deep-Sea Res. Pt. I, 57, 812–834, https://doi.org/10.1016/j.dsr.2010.03.011, 2010.

633 Gouretski, V., Jungclaus, J. H., and Haack, H.: Revisiting the Meteor 1925–1927 hydrographic dataset reveals
634 centennial full-depth changes in the Atlantic Ocean, Geophys. Res. Lett., 40, 2236–2241,
635 https://doi.org/10.1002/grl.50503, 2012.
636 Gouretski, V., Cheng, L., and Boyer, T.: On the consistency of the bottle and CTD profile data, J. Atmos. Ocean. Tech.,
637 39, 1869–1887, https://doi.org/10.1175/JTECH-D-22-0004.1, 2022.
638 Gouretski, V., Cheng, L., Du, J., Xing, X., Chai, F., and Tan, Z.: A consistent ocean oxygen profile dataset with new
639 quality control and bias assessment, Earth Syst. Sci. Data, 16, 5503–5530, https://doi.org/10.5194/essd-16-5503-
640 2024, 2024.
641 Gouretski, V., Cheng, L., and Du, J.: IAP Historical Nansen Cast Temperature Profile Dataset [data set], Marine
642 Science Data Center, Chinese Academy of Sciences, https://msdc.qdio.ac.cn (last access: 8 October 2026),
643 https://doi.org/10.12157/IOCAS.20260520.001, 2026.
644 Helland-Hansen, B. and Nansen, F.: The Norwegian Sea: its physical oceanography based upon the Norwegian
645 researches 1900–1904, Vol. II, Det Mallingske Bogtrykkeri, 390 pp., https://imr.brage.unit.no/imr-
646 xmlui/handle/11250/114874 (last access: 8 October 2026), 1909.
647 Hydrographisches Amt der Admiralität: Die Forschungsreise S.M.S. “Gazelle” in den Jahren 1874 bis 1876, II. Theil,
648 Physik und Chemie, https://epic.awi.de (last access: 8 October 2026), 1888.
649 Levitus, S., Antonov, J., and Boyer, T.: Warming of the world ocean, 1955–2003, Geophys. Res. Lett., 32, L02604,
650 https://doi.org/10.1029/2004GL021592, 2005.
651 Köhl, A.: Evaluating the GECCO3 1948-2018 ocean synthesis – a configuration for intializing the MPI-ESM climate
652 model, Q J R Meteorol Soc., 146, 2250–2273, https://doi.org/10.1002/qj.3790, 2020.
653 Li, Y., Church, J. A., McDougall, T. J., and Baker, P. M.: Sensitivity of observationally based estimates of ocean heat
654 content and thermal expansion to vertical interpolation schemes, Geophys. Res. Lett., 49, e2022GL101079,
655 https://doi.org/10.1029/2022GL101079, 2022.
656 Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Reagan, J. R., Boyer, T. P., Seidov, D., Wang, Z., Garcia, H. E.,
657 Bouchard, C., Cross, S. L., Paver, C. R., and Dukhovskoy, D.: World Ocean Atlas 2023, Volume 1: Temperature,
658 A. Mishonov, Technical Editor, NOAA Atlas NESDIS 89, https://doi.org/10.25923/54bh-1613, 2024.
659 Moroshkin, K. V.: Experience of work with domestic unprotected thermometers, Collected Works of the Institute of
660 Oceanology, Academy of Sciences of the USSR, 5, 78–92, https://search.rsl.ru/ru/record/01006019350 (last
661 access: 8 October 2026), 1951 (in Russian).
662 Nansen, F.: The Norwegian North Polar Expedition, 1893–1896, Scientific Results, edited by F. Nansen, Vol. VI,
663 London, 659 pp., https://archive.org/details/cu31924012205765 (last access: 8 October 2026), 1905.
664 Negretti, H. and Zambra, J. W.: On new deep-sea thermometer, Proc. R. Soc., 22, 238–241,
665 https://doi.org/10.1098/rspl.1873.0034, 1874.
666 Pan, Y. and Coauthors: Ocean heat content sets another record in 2025, Adv. Atmos. Sci.,
667 https://doi.org/10.1007/s00376-026-5876-0, 2026.
668 Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan,
669 A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth
670 century, J. Geophys. Res., 108, 4407, https://doi.org/10.1029/2002JD002670, 2003.
671 Reiniger, R. F. and Ross, C. K.: A method of interpolation with application to oceanographic data, Deep-Sea Res., 15,
672 185–193, https://doi.org/10.1016/0011-7471(68)90040-5, 1968.

Roemmich, D., Gould, W. J., and Gilson, J.: 135 years of global ocean warming between the Challenger expedition and the Argo Programme, Nat. Clim. Change, 2, 425–428, https://doi.org/10.1038/nclimate1461, 2012.

Spilhaus, A. F.: A bathythermograph, J. Mar. Res., 1, 95–100, https://elischolar.library.yale.edu/journal_of_marine_research/522 (last access: 8 October 2026), 1938.

State Oceanographic Institute: Manual for hydrological works in oceans and seas, 2nd edn., Gidrometeoizdat, Leningrad, 725 pp., https://mgmtmo.ru/edumat/hydrology/Guidelines_hydrol_work_oceans_seas_77.pdf (last access: 8 October 2026), 1977 (in Russian).

Swift, J. H.: Reference-quality water sample data: notes on acquisition, record keeping, and evaluation, The GO-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidelines, IOCCP Report No. 14, ICPO Publication Series No. 134, Version 1, http://www.go-ship.org/Manual/Swift_DataEval.pdf (last access: 8 October 2026), 2010.

Thomson, W. and Murray, J.: The voyage of H.M.S. Challenger 1873–1876, Narrative, Vol. 1, First Part, Ch. III, Johnson Reprint Corporation, http://archimer.ifremer.fr/doc/00000/4751/ (last access: 8 October 2026), 1885.

Thomson, C. W. and Murray, J.: Report on the scientific results of the voyage of H.M.S. Challenger during the years 1873–76, Physics and Chemistry, Vol. II, Her Majesty Stationery Office, London, https://www.biodiversitylibrary.org/item/30429#page/1/mode/1up (last access: 8 October 2026), 1889.

U.S. Hydrographic Office: Deep-sea soundings in the North Pacific Ocean obtained in the United States Steamer Tuscarora, Commander George E. Belknap, Government Printing Office, Washington, D.C., https://www.nodc.noaa.gov/OC5/PACIFIC2009/HTML/books.html (last access: 8 October 2026), 1874.

U.S. Naval Oceanographic Office: Instruction manual for obtaining oceanographic data, U.S. Government Printing Office, Washington, D.C., 232 pp., https://epic.awi.de/id/eprint/32525/1/manual_oce-data.pdf (last access: 8 October 2026), 1968.

Wu, Q., Gregory, J. M., Zanna, L., and Khatiwala, S.: Time-varying global energy budget since 1880 from a reconstruction of ocean warming, Proc. Natl. Acad. Sci. USA, 122, e2408839122, https://doi.org/10.1073/pnas.2408839122, 2025.

Wunsch, C. and Heimbach, P.: Practical global oceanic state estimation, Physica D, 230, 197–208, https://doi.org/10.1016/j.physd.2006.09.040, 2007.

Wüst, G., Böhnecke, G., and Meyer, H. F.: Ozeanographische Methoden und Instrumente, in: Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs- und Vermessungsschiff “Meteor” 1925–1927, Band IV, Erster Teil, Walter de Gruyter & Co., Berlin, https://ci.nii.ac.jp/ncid/BA82926827 (last access: 8 October 2026), 1932.

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., and Heimbach, P.: Global reconstruction of historical ocean heat. storage and transport, Proc. Natl. Acad. Sci. USA, 116, 1126–1131, https://doi.org/10.1073/pnas.1808838115, 2019.

Appendix A: Corrections for depth bias

Table A1. Depth bias corrections 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.