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% of the total OHC increase between 1920-1944 and 2010-2025.

1 Introduction

Recognition of the crucial role of the global ocean for the Earth climate system and experimental evidence of 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). Most OHC time series based on subsurface hydrographic observations begin in the late 1950s, as the IGY 1957-1958 observational campaign for the first time enabled quasi-global coverage of the ocean with hydrographic observations.

It should be noted that a significant number of temperature and salinity profiles have been accumulated over the time period preceding the IGY. The "Challenger" expedition 1873-1876 marks the beginning of observational physical oceanography (Thomson and Murray, 1885; Roemmich et al., 2012). Guided by a group of scientists, the expedition produced a large set of deep temperature profiles with the accuracy and vertical resolution needed to quantitatively analyze secular changes in 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 its 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 the 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 that 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, MBT temperature measurements are much less accurate than Nansen casts, and the maximum instrument depth is limited to 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 exist for the time period before the 1950s. Zanna et al. (2018) produced OHC time series since 1871 based on global reconstructions, linking OHC to observed sea surface temperatures (Rayner et al., 2003) using Green's functions derived from data assimilation experiments (Wunsch and Heimbach, 2007). Fu et al. (2025) provided an 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. Using this dataset, we assess changes in ocean heat content in the upper 2000 m layer over a 150-year period from the 1870s to the 2020s.

2 Data

For this analysis, we use only Nansen cast temperature profiles from 1873 to 2025, focusing on the period before 1970. The World Ocean Database (WOD) (Locarnini et al., 2023) (updated 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 metadata exists on whether vertical interpolation was applied. Apart from the WOD archive, we used four temperature profile datasets digitized from 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 to estimate 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 a drop in the number of observations at all levels. For all time periods, the number of data rapidly decreases with depth. Before the 1920s, shallow profiles prevailed, 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.

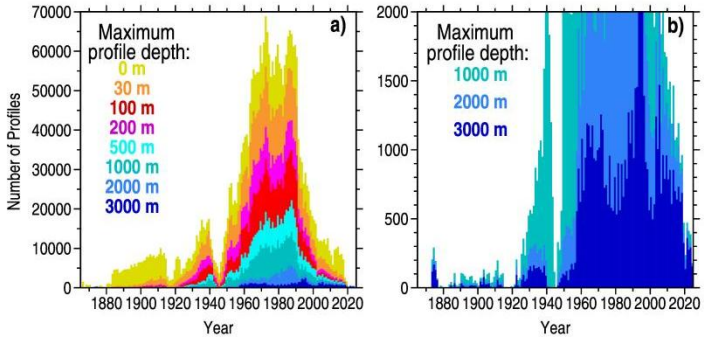

Figure 1. Yearly number of hydrographic profiles used in the study: a) all profiles; b) profiles deeper than 1000 m.

Fig. 2 shows the spatial distribution of profiles for selected time periods and for several maximum profile depths. The two-decade period between 1950 and 1970, which includes the International Geophysical Year campaign (Chapman, 1959), is characterized by quasi-global coverage, even for 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 accomplishment 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 prevail (Fig. 3a), with only 10%, 5%, and 2.5% of all casts 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 that can be attached to the wire. Typically, not more than 10-12 bottles were used (Gould et al., 2013). Moroshkin (1951) reported 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 neighboring bottles typically increases with depth from one or several tenths of a meter near the ocean surface to 500 m or more in the deep part of the profile (**Fig. 3d**).

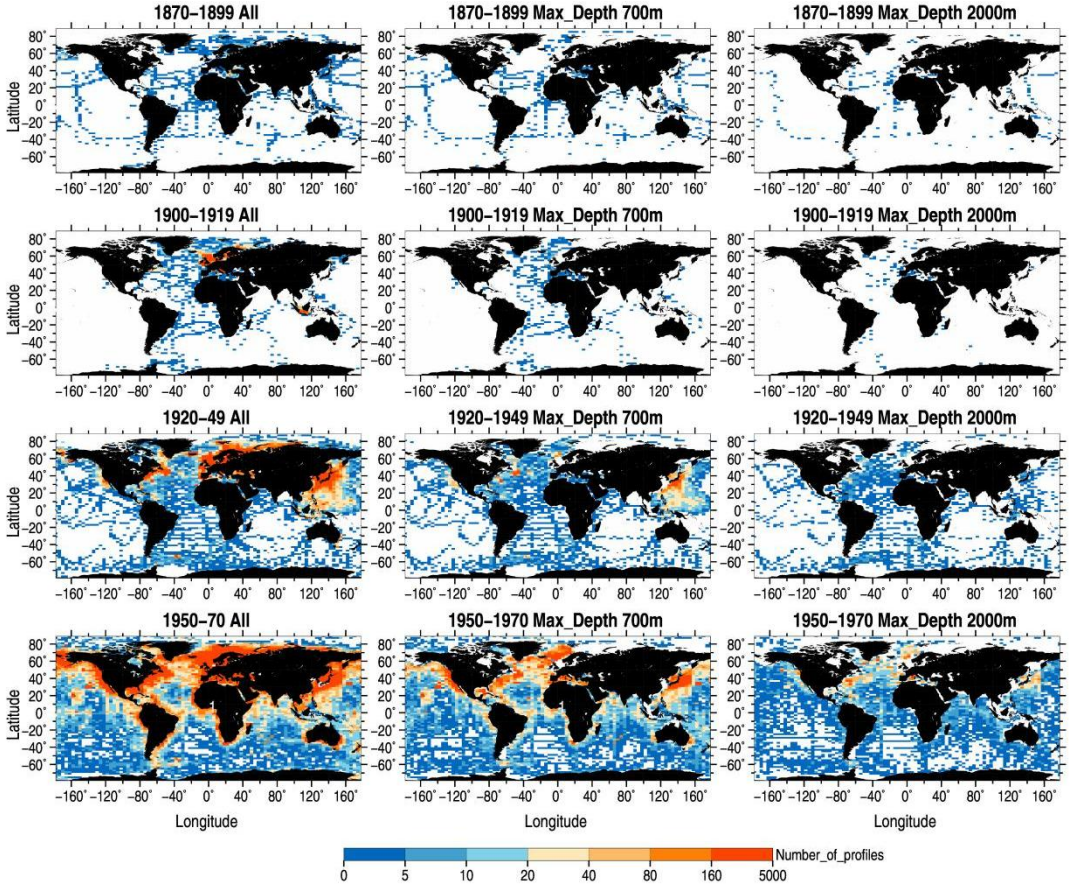

Figure 2. Number of profiles within 2° × 4° Latitude/Longitude boxes for several time-periods and profile maximum depths.

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).

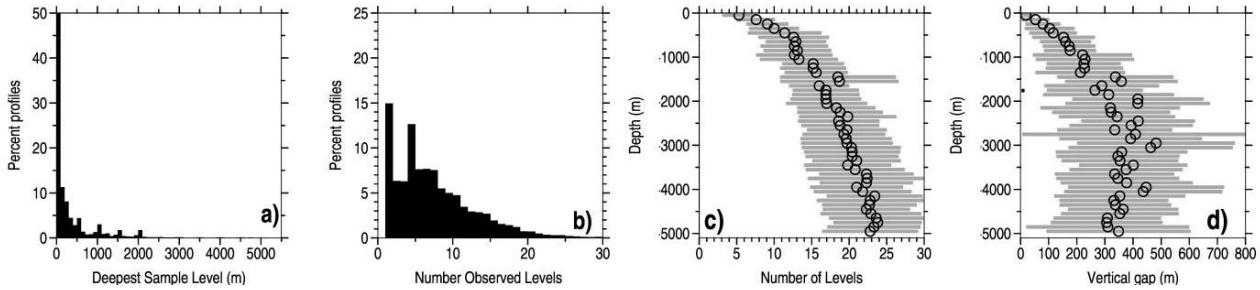

Figure 3. a) Deepest sample level histogram; b) number of observed levels histogram; c) mean number of observed levels versus cast depth; d) mean vertical gap size versus depth.

3 Data quality control

Imperfections and instrument malfunctions, especially in the early period, errors in reading temperature values, lack of proper thermometer calibration, and errors in sample depth lead to both random and systematic errors in temperature. The data used in the study underwent three quality-control steps: automated statistical 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 in 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 imperfection of the early observational technique.

3.2 Manual quality control

Temperature profiles were subjected to manual quality control. The data have been screened for remaining outliers in each 5° × 20° Latitude/Longitude box (**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 that passed the automated quality control as a background reference. An increased percentage of manually identified outliers is clearly seen below 1000-1500 m, 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 that passed both the automated and manual quality control are shown in **Fig. S3** for selected boxes. Overall, manual quality control identified 0.17% of all temperature observations as outliers, with 0.85% of profiles affected by the manual check. We note that, despite improved accuracy and reliability of modern oceanographic instrumentation, manual (expert) control and data evaluation remain a necessary step in the preliminary analysis of hydrographic data (Swift, 2010).

4 Vertical interpolation of the observed level data

The choice of 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 because quality control can exclude some observed levels. As indicated by 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**).

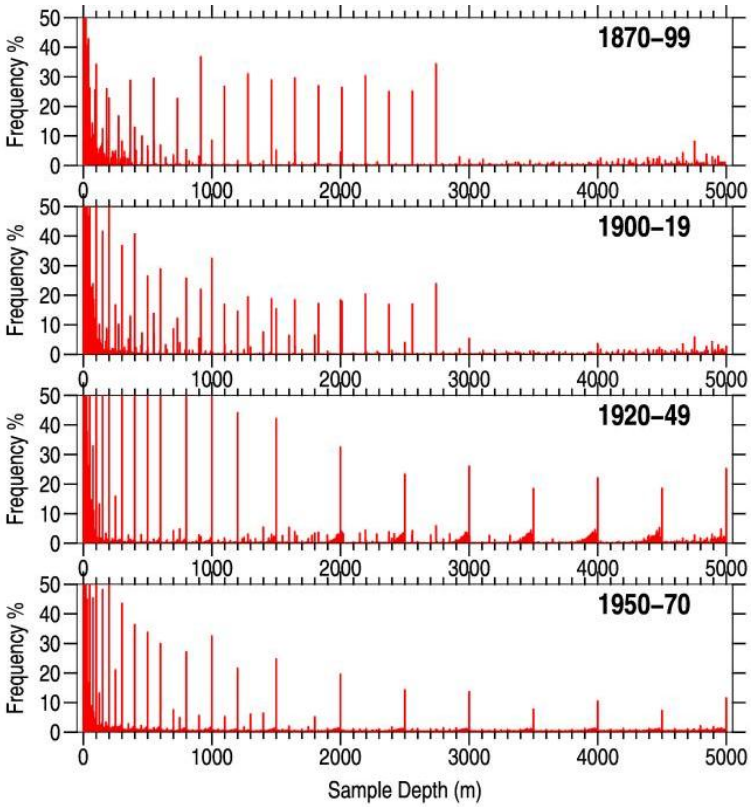

Figure 4. Sample depth frequency histograms for OSD temperature profiles for selected time periods.

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 a distance-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.

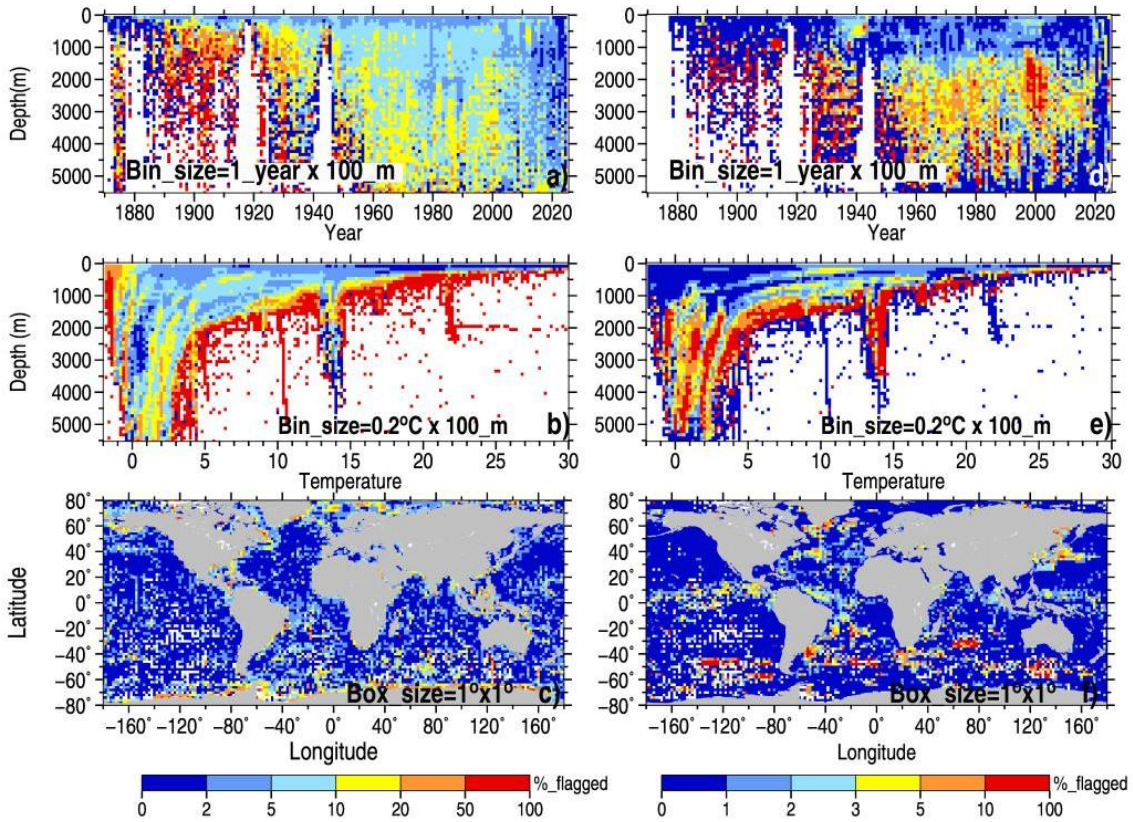

Figure 5. Outlier percentage for OSD temperature profiles: a-c) automated statistical quality control; d-f) manual quality control. Outlier percentage is shown for depth/year bins (a, d), depth/temperature bins (b, e), 1° × 1° latitude/longitude boxes (c, f).

4.1 Case of uneven distribution of observed levels over depth

The straightforward application of the RR method to observed level profiles was found to result in unrealistically large excursions on the interpolated temperature profile for a not-negligible number of profiles. Such cases typically occur when the observed levels are distributed extremely unevenly so that neighboring vertical gaps differ by more than an order of magnitude. Such profiles occur, for instance, when the entire temperature profile at a hydrographic station is composed of two or more hydrographic casts, each covering a part of the entire water column. When such profiles are merged together in the final station table, the deepest observed depth of the shallower cast might be very close to the shallowest observed level of the deeper cast. In other cases, small gaps may result from including interpolated levels along with observed levels in the resulting station table, or from large gaps in profiles due to outlier rejection. Since the interpolating polynomial must pass through each point of the original temperature profile, unrealistic overshoots appear on the interpolated profile. **Fig. 6** shows several examples of such profiles. In some cases, the overshooting amounts to several degrees. To avoid the overshooting, the original profiles are checked for pairs of levels that are very close to each other before interpolation. Such pairs are then replaced by one level referred to as the middle point, and temperature is taken as the average of the temperatures at the two levels. Another option is to use a simple linear interpolation in such cases.

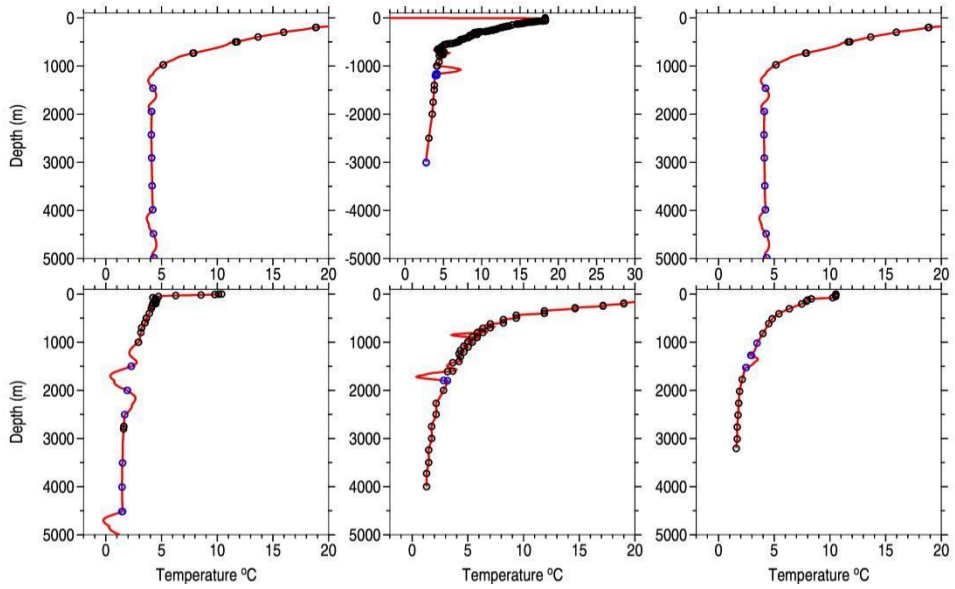

Figure 6. Examples of temperature profiles with uneven distribution of observed levels resulting in unrealistic overshooting of interpolation polynomials in RR method: observed levels (black circles), observed levels with a small distance to a neighbor level (blue circles), interpolated profile (red).

4.2 Dependence of interpolation error on distance between observed levels.

Interpolating non-linear temperature profiles over large gaps can lead to significant temperature errors. 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-square 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-square 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. **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.

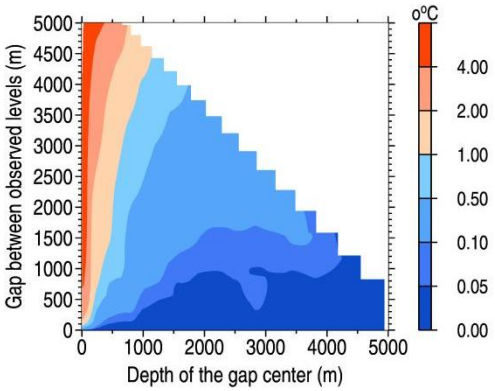

Figure 7. Global root mean squared interpolation error versus depth and the vertical gap size. Error refers to the center of the gap.

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

As 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.

5. Estimation of the Nansen cast depth bias

Only under ideal conditions does the wire with attached Nansen bottles take a vertical position over the entire wire length. Because currents and wind cause ship drift during occupation of the hydrographic station, the wire deviates from the vertical, so the length of wire paid out is greater than the actual depth at which the temperature was measured. Brennecke (1921) suggested using 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 than the protected one, and the difference depends on hydrostatic pressure, which can be converted to 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).

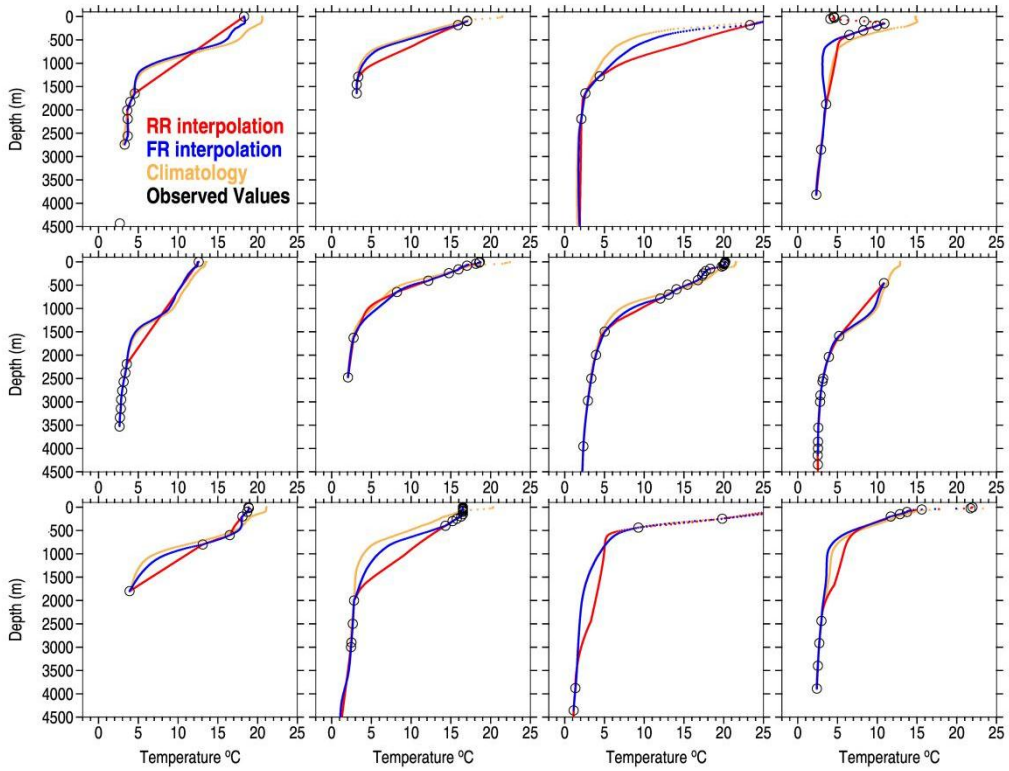

Figure 8. Comparison of RR (red) and FR (blue) interpolation methods for 16 arbitrarily selected temperature profiles with large vertical gaps. Observed level values are marked with black circles, WAGHC climatology temperature profiles are shown in green.

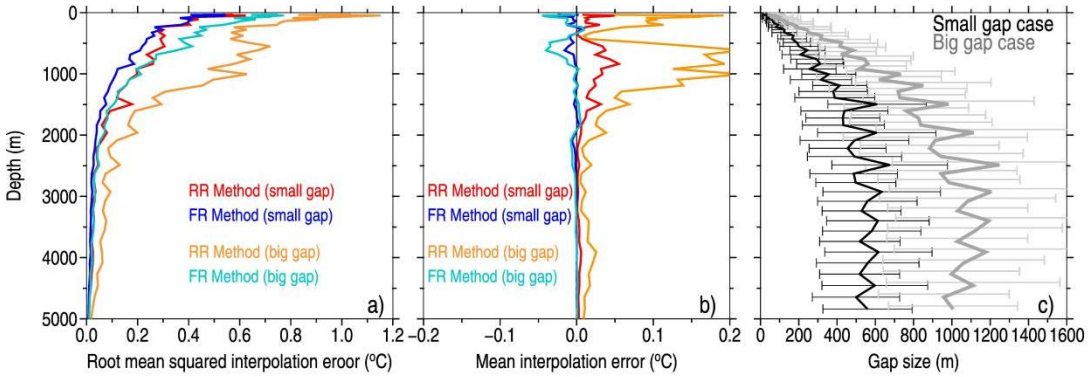

Figure 9. Global root mean squared (a) and global mean (b) error for interpolated temperature values obtained by means of RR method (red and orange) and FR method (blue and cyan). Red and blue profiles correspond to the small gap case, orange and cyan correspond to the big gap case. Explanation Is given in the text. 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. For a substantial number of Nansen casts, the reported sample depths 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 casts 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 proposed another method for estimating depth bias based on analysis of the sample-depth histogram. Here we apply this method to derive depth corrections for old Nansen casts obtained before 1970. As illustrated in Fig. 4, the disproportionately large number of observations comes from a limited number of so-called standard levels chosen as target depth levels for sampling. The non-standard depth levels can be linked to profiles where the thermometric method was used or where corrections were applied to account for wire deviation from the vertical.

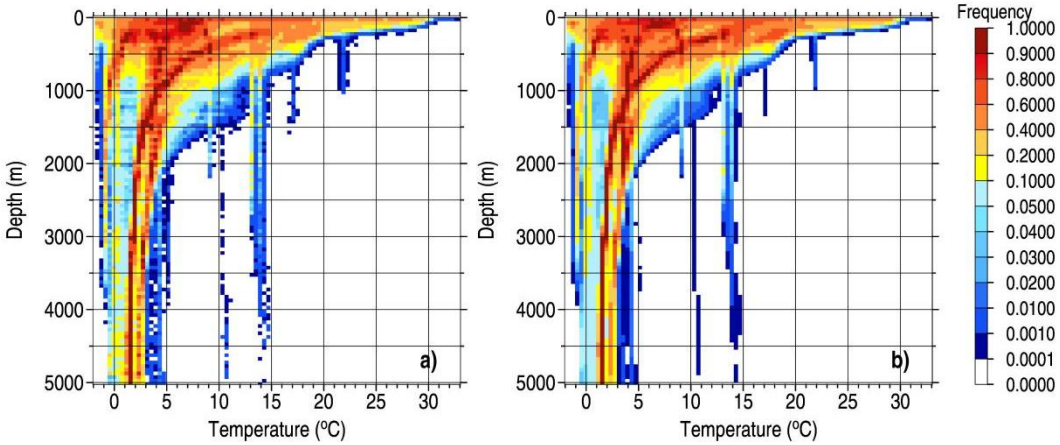

Figure 10. Normalized temperature-depth histograms for validated observed level data (a) and for interpolated data (b).

Fig. 11a, b show frequency distributions of sampled depths around standard levels. We obtained these histograms after removing the standard levels and then smoothing with a Gaussian kernel (the frequencies at standard depths are not zero due to the Gaussian smoothing). After some experimenting, we applied a Gaussian kernel width of 20 m to prevent over-smoothing of the histograms. 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)**. Using these histograms, the ratio R = (Z_{max})/(Z_{stand}) was calculated for each standard level at or below 50 m (**Fig.11c**). Levels above 50 m were not used as they mostly fall within the upper mixed layer or a seasonal thermocline. Corrections are also given in **Table A1**. As shown in Fig. 11c, reported sample depths slightly overestimate the actual depths reached. We note that Nansen cast bottle depths are reported in round meters, limiting 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 to 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 goal 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 baseline 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 the climatology temperature field (**Fig. S6),** which varies in the range 2014-2015 and 2003-2006 above and below the Argo maximum depth, respectively.

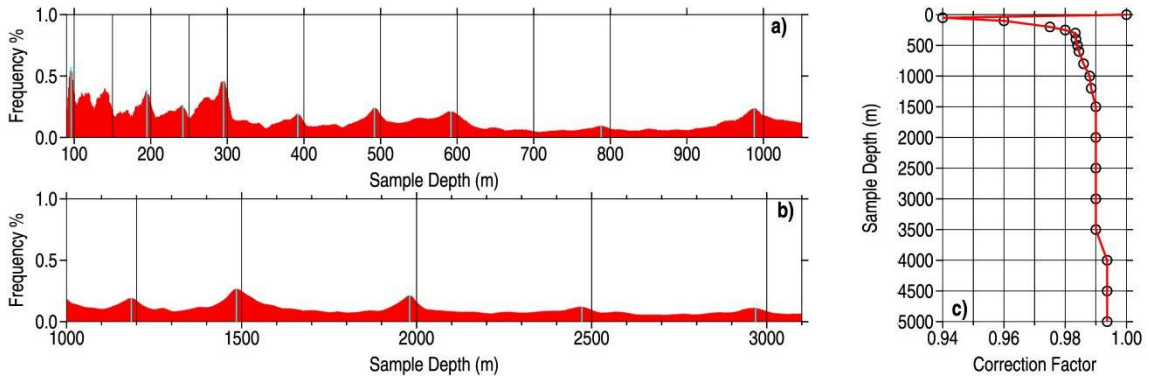

Figure 11. a, b) Smoothed sample depth histograms with deleted standard levels. Black vertical lines correspond to the standard depth; cyan bars denote maximum frequency. Differences between the standard depth and the depth of the respective local maximum frequency represent the depth bias; c) depth correction factor versus depth.

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 at 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.

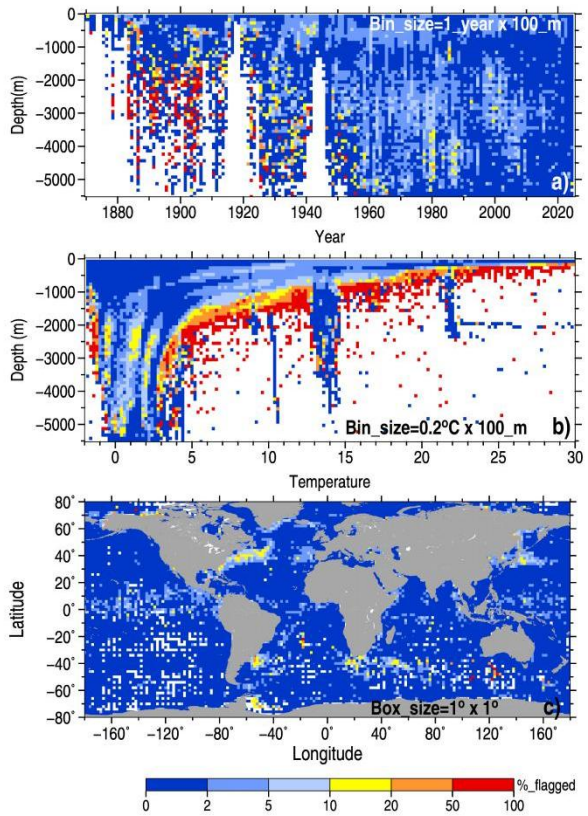

Figure 12. Outlier percentage for temperature anomalies: a) percentage in depth/year bins, b) depth/temperature bins, c) 1° × 1° Latitude/Longitude boxes.

7 Adjustment of the oldest temperature profiles

Although the first subsurface temperature measurements in the WOD database date to cruises from the end of the 18th century, the voyage of "Challenger" (1873-1876) represents the first large-scale oceanographic expedition with

temperature measurements conducted by scientists. Comparing the "Challenger" temperature data with Argo-era temperatures revealed profound warming of the global ocean along the expedition's entire track (Roemmich et al., 2012). Almost simultaneously with the "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). Neither cruise had 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 the larger size, better vertical resolution, and wider geographical scope of the "Challenger" data, along with the fact that the measurements and their analysis were 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 the "Challenger" - "Gazelle" comparison, profiles in all three oceans are used (**Fig. 13b**). Distribution of point temperature anomalies over depth (**Fig. 13d, 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.

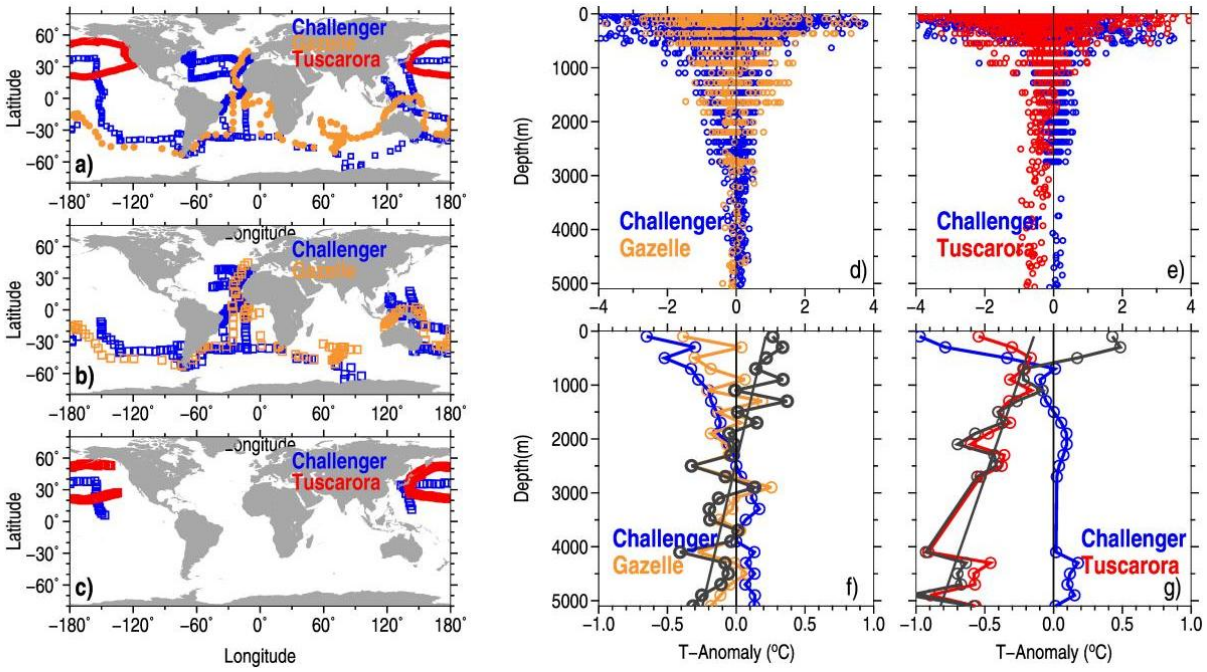

Figure 13. a) Station locations for "Challenger", "Gazelle", and "Tuscarora" expeditions; b) location of "Challenger" and "Gazelle" stations used for data inter-comparisons; c) same as b) but for "Challenger" and "Tuscarora" data; d) comparison of temperature anomalies for "Challenger" and "Gazelle" 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 the 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 5-25 year long time periods, rather than for distinct calendar years. The earliest time spans the quasi-global oceanographic survey conducted by "Challenger", "Gazelle", and "Tuscarora". The selected time periods are: 1873-1876, 1920-1944, 1945-1969, 1970-1989, 1990-2009, 2010-2025. For the calculation of temperature anomalies, we use profiles interpolated on 10-meter levels and described in Section 4. For each time period, mean anomalies are calculated for 5° × 5° Latitude/Longitude boxes.

8.2 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 the patchiness of the basin-scale temperature anomaly pattern due to remaining biases and mesoscale noise. However, despite a coarse sampling within each layer, a progressive warming between the 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.

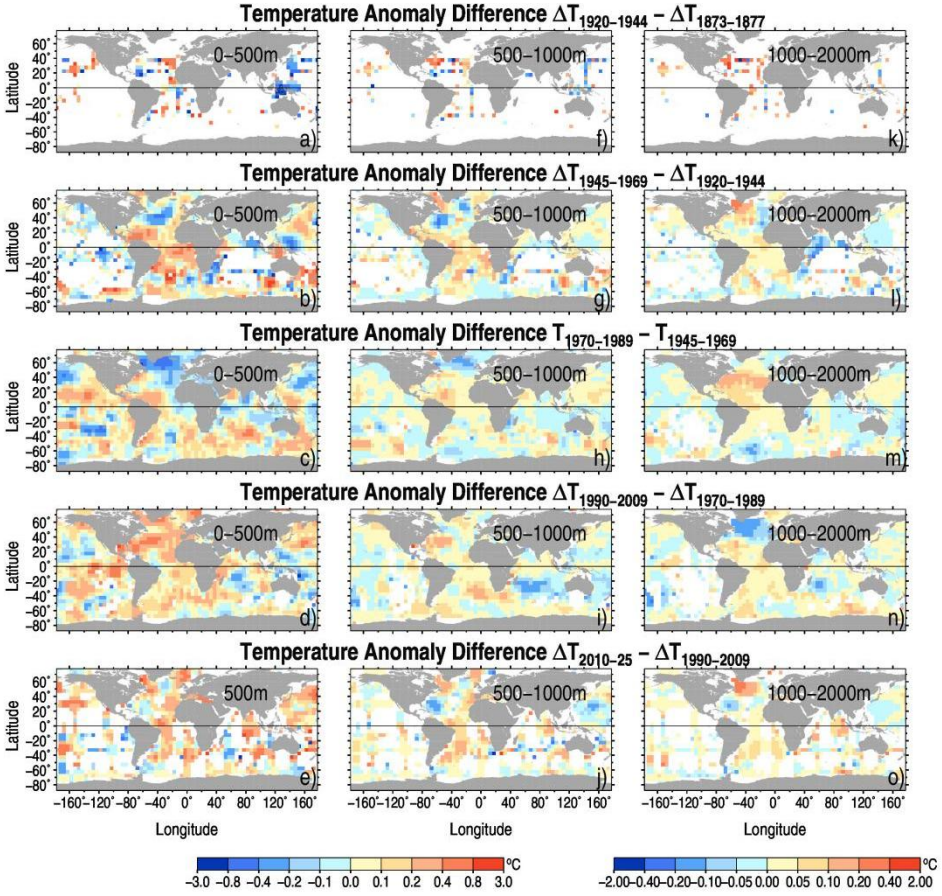

Figure 14. Layer-averaged temperature anomaly differences between selected time periods in 5° × 5° latitude/longitude boxes: 0-500 m (a-e), 500-1000 m (f-j), 1000-2000 m (k-o). Note a different color scale for the layer 1000-2000 m.

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 0-500 m, including the Arctic Ocean, with the exception of 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 (500-1000 m) (**Fig. 14i, n**) both feature strong cooling. The latest period (1990-2009 and 2010-2025) shows continued global warming (Fig. 14 e, j, o), with the strongest deep warming (1000-2000 m) in the North-West and South Atlantic Ocean.

8.3 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 oC globally, whereas the global temperature increase at 1000 m and 2000 m levels amounts to 0.1 ^{o}C and 0.04 ^{o}C respectively.

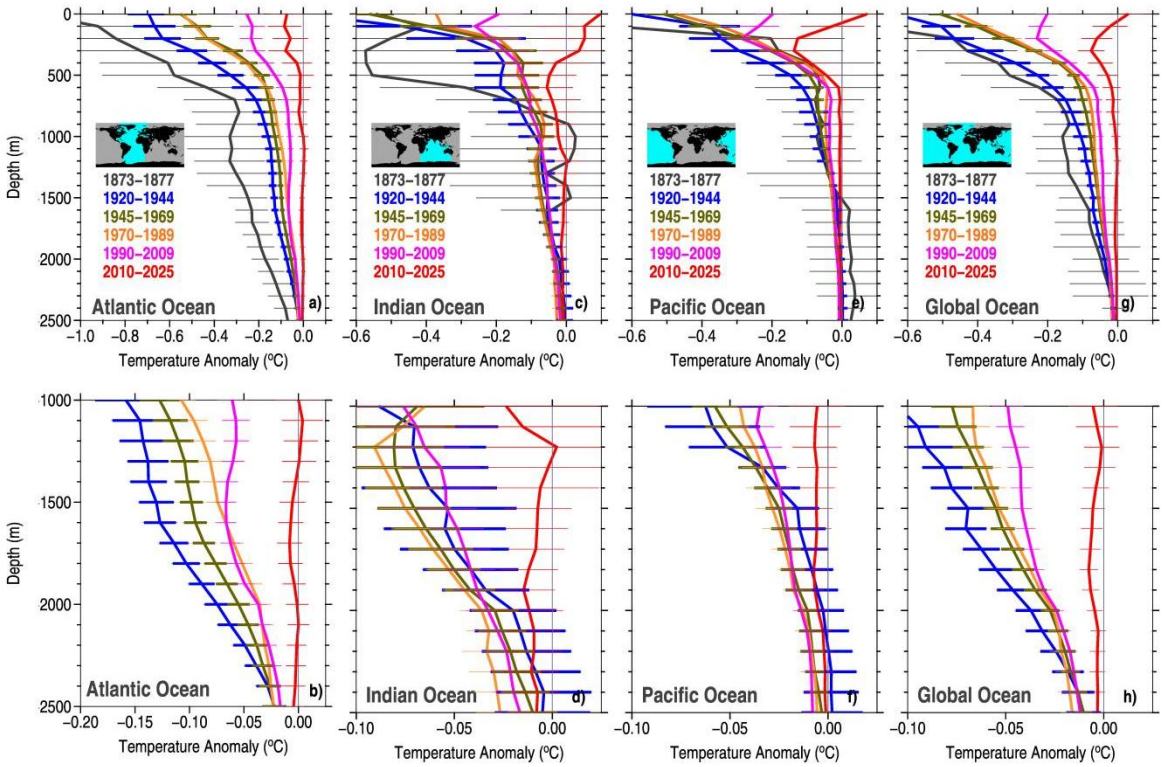

Figure 15. a, c, e, g) Temperature anomaly vertical profiles for the three main oceans and for the entire global ocean and for selected six time periods with standard error bars; b, d, f, h) same but for the depth range below 1000 m. Note the different temperature scale for the Atlantic Ocean.

8.4 Error estimation

The box-averaged temperature anomalies in each 100 m layer are calculated using all observed anomaly profiles within the respective 5° × 5° 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 ± 0.04 °C. Errors associated with the mesoscale variability are much larger. According to **Fig. S7**, the standard deviation of temperature anomalies within a 2° × 4° 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 ° × 5 ° 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 autocovariances 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**.

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 parameter values were used for the OHC calculations: 3991.87 J/kg °C as the seawater specific heat capacity, 1028 kg/m3 as the average seawater density, and 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 the Atlantic, Indian, and Pacific oceans, with weights equal to individual ocean areas.

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 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 1961, global ocean heat content increased by 142.4 ± 19.7 ZJ, compared with 435.7 ± 22.1 ZJ between 1932 and 2015, accounting for 32.7 ± 6.8% of the total OHC increase since 1932. Our calculations further suggest that warming likely also occurred between the 1870s and the 1920s, amounting to 64.6 ± 85.9 ZJ if original sample depths not corrected for possible depth bias are used. However, this value is statistically insignificant because of large errors from imperfections in the earliest data and insufficient spatial sampling between 1873 and 1877.

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 reconstructing full-depth temperature changes using sea-surface temperature anomalies 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 the 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 rounded 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 increase the warming rate because they shift observed depths upward, cooling temperature profiles under the prevailing negative temperature gradient. Applying depth corrections increases the total energy gain between 1930 and 2015 by 16.0 ± 7.2%, from 435.7 ± 22.1 ZJ to 505.2 ± 22.2 ZJ. The respective OHC change between 1962 and 2015 increases by 8.5 ± 7.8 % from 293.4 ± 16.5 ZJ to 317.7 ± 16.5 ZJ. The applied depth correction factor (**Fig. 11c**) varies from 0.94 at 50 m to 0.99 at 1500 m, resulting in different regional impacts as the vertical temperature gradient 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).

Fig. 16b shows the average heating rate across the six analyzed time periods. The heating rate is estimated as the difference between two adjacent time-mean global OHC anomalies divided by the time span between them. Our calculations indicate a heating rate of about 0.07 ± 0.07 W m^{-2} before the 1930s. During the 1920s and 1960s, the heating rate increases to 0.31 ± 0.03 W m^{-2} and is very close to the values for the 1970-1989 and 1990-2009 periods. Consistent with Pan et al. (2026), the fastest OHC growth of 0.66 ± 0.04 W m^{-2} occurs in the latest period, 1990-2009 to 2010-2025. 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 the 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. Like our results, both studies suggest a heating-rate maximum between 1935 and 1950.

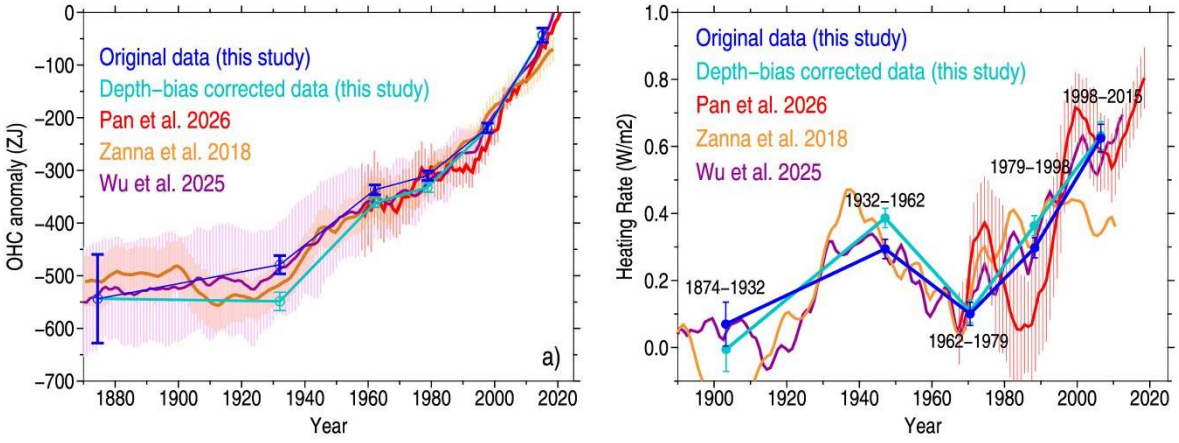

Figure 16. a) Global OHC anomalies for selected time periods for original data (blue) and depth-bias corrected data (cyan), OHC anomaly time series from Pan et al. (2026) (red), Zanna et al. (2018) orange), and Wu et al. (2025) (magenta) are shown for comparison; b) heating rate for five time periods: 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 show standard error (this study); 95% confidence interval (Pan et al., 2025); one standard deviation (Zanna et al., 2018); and one standard deviation from 73 ensemble members (Wu et al., 2025).

We did not apply depth corrections for the earliest data from "Challenger", "Gazelle", and "Tuscarora" because different instrumentation was used on these cruises (a hemp rope instead of a steel one, no bottles attached to the wire). For this reason, the heating rate between the first two time periods is close to zero in the case of the depth-corrected profiles (**Fig. 16b**, blue line). However, two arguments are in favor of a positive depth bias also in the oldest hydrographic profiles: a) unavoidable deviation of the sounding line from the vertical under real meteorological conditions; b) a significantly larger elasticity of the hemp rope compared to the steel wire, with the Young's modulus for steel being four times higher than that for hemp. Allowing for possible positive depth bias in the earliest "Challenger", "Gazelle", and "Tuscarora" data would likely result in a positive heating rate between the 1870s and 1920s.

9 Discussion and conclusions

The global archive of temperature profiles (Locarnini et al., 2023) has been used in many studies to estimate the evolution of the global ocean's thermal state (Levitus et al., 2005; Gouretski and Koltermann, 2007; Cheng and Zhu, 2016; Pan et al., 2026). However, these analyses have not used the full potential of the historical data archive, as they rely on data collected after the 1940s. This study focuses on 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 interpolation error when interpolating over large vertical gaps. All historical profiles underwent a three-step quality control process, including the automated statistical quality control procedure described by Gouretski (2018), manual control, and automated control of temperature anomalies relative to the baseline climatology. We use the updated WOCE/Argo global hydrographic climatology (Gouretski, 2018), based on high-quality temperature profiles for 1990-2025.

Because of the general data paucity, we calculated mean temperature anomalies in 5° × 5° latitude/longitude boxes and used them to estimate basin-averaged temperature anomalies over 5- to 25-year periods. Our results confirm progressive warming of the subsurface ocean since the 19th century, seen both globally and in the three main oceans in all layers above 2000 m. The estimated OHC change since the 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. We find that between 1920-1944 and 1945-1969 the global OHC increased by 142.4 ± 19.7 ZJ. This value amounts to 32.7 ± 6.8% of the total OHC increase between the 1920s and 2020s.

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

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 (1980) 35 (1): 49–65, 1980.

Belknap, G. E.: Deep Sea Soundings in the North Pacific Ocean obtained in the United States Steamer Tuscarora, Kessinger Publishing, 90p., 1874.

Becker, J. J., Sandwell, D. T., F. Smith,W. H., 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, Marine Geodesy, 32, 4, 355-371, 2009.

Borishanski, L.S.: Manual for the performance of work on standard hydrological sections and marine hydrological surveys, Leningrad, Gydrometeoizdat, 284 p., 1954 (in Russian).

Boyce, F. M.: A brief study of the protected deep-Sea reversing thermometer, Manuscript Report Series (Oceanographic and Limnological), No. 227, 17p., 1966

Brennecke, W.: Die ozeanographischen Arbeiten der Deutschen Antarktischen Expedition 1911-1912. Archiv der Deutschen Seewarte, 1921, XXXIX No 1, Hamburg, 1921.

Chapman, S.:The International Geophysical Year, 1957-58, Nature, **175**, 55–56, 1959.

Cheng L. and Zhu, J.: Benefits of CMIP5 Multimodel Ensemble in Reconstructing Historical Ocean Subsurface Temperature Variation, Journal of Climate, 29, 5393 – 5415. DOI: 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, C3, 2001-2004, 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., Zweng, M. M.: World Ocean Atlas (2018): Product Documentation. A. Mishonov, Technical Editor, 2019.

Gebbie, G. and Huybers, P.: The little ice age and 20th-century deep Pacific cooling. Science, 363, 70-74, 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. and F. Reseghetti: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. Deep-Sea Research I, Vol. 57(6), pp. 812-834, doi:10.1016/j.dsr.2010.03.011, 2010.

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

Gouretski, V., Jungclaus, J. H., and Haack, H.: Revisiting the Meteor 1925–1927 hydrographic dataset reveals centennial full-depth changes in the Atlantic Ocean, Geoph. Res. Letters, 40, 10, 2236-2241, https://doi.org/10.1002/grl.50503, 2012.

592 Gouretski, V.: World Ocean Circulation Experiment – Argo Global Hydrographic Climatology. Ocean Science, 14,
593 1127-1146, 2018.
594 Gouretski, V. and Cheng, L.: Correction for Systematic Errors in the Global Dataset of Temperature Profiles from
595 Mechanical Bathythermographs, Journal of Atmospheric and Oceanic Technology, 37, 5, 841-855,
596 https://doi.org/10.1175/JTECH-D-19-0205.1, 2020.
597 Gouretski, V., Cheng, L., and Boyer, T. On the Consistency of the Bottle and CTD Profile Data. J. Atmos. Oceanic
598 Technol., 39, 1869–1887, https://doi.org/10.1175/JTECH-D-22-0004.1, 2022.
599 Gouretski, V., Cheng, L., Du, J., Xing, X., Chai, F., and Tan, Z.: A consistent ocean oxygen profile dataset with new
600 quality control and bias assessment, Earth Syst. Sci. Data, 16, 5503–5530, https://doi.org/10.5194/essd-16-5503-
601 2024, 2024.
602 Gouretski, V., Cheng, L., and Du, J.: IAP Historical Nansen Cast Temperature Profile Dataset [data set], Marine
603 Science Data Center, Chinese Academy of Sciences (https://msdc.qdio.ac.cn),
604 doi:10.12157/IOCAS.20260520.001, 2026.
605 Helland-Hansen, B., and Nansen, F.: The Norwegian Sea: Its Physical Oceanography Based upon the Norwegian
606 Researches 1900–1904. Vol. II. Det Mallingske Bogtrykkeri, 390pp., 1909.
607 Hydrographisches Amt der Admiralität: Forschungsreise S.M.S. “Gazelle” in den Jahren 1874 bis 1876, II. Theil,
608 Physik und Chemie, 1888.
609 Levitus, S., Antonov, J., and Boyer, T.: Warming of the world ocean, 1955-2003, Geophysical Research Letters, 32,
610 L02604, doi:10.1029/2004G, 2005.
611 Li, Y., J. A. Church, T. J. McDougall, and P. M. Baker: Sensitivity of Observationally Based Estimates of Ocean Heat
612 Content and Thermal Expansion to Vertical Interpolation Schemes, Geophysical Research Letters, 49,
613 org/10.1029/2022GL101079, 2022.
614 Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Reagan, J. R., Boyer, T. P., Seidov, D.,Wang, Z., Garcia, H. E.,
615 Bouchard, C., Cross, S.L., Paver, C. R., and Dukhovskoy, D.: World Ocean Atlas 2023, Volume 1: Temperature.
616 A. Mishonov Technical Ed. NOAA Atlas NESDIS 89, doi.org/10.25923/54bh-1613, 2024.
617 Moroshkin, K.V.: Experience of work with domestic unprotected thermometers, Collectedworks of the Institute of
618 Oceanology, Academy of Sciences of the USSR, 5, 78, 92, 1951 (in Russian).
619 Nansen, F.: The Norwegian North Polar Expedition, 1893–1896, Scientific Results, edited by Fridtjof Nansen, Volume
620 VI, London, 659 p., 1905.
621 Negretti, H. and Zambra, J. W.: On new deep-sea thermometer. Proc. R. Soc. (1874) 22 (148-155): 238–241, 1874.
622 Pan, Y, and Coauthors: Ocean heat content sets another record in 2025, Adv. Atmos. Sci.,
623 https://doi.org/10.1007/s00376-026-5876-0, 2026.
624 Rayner N. A, Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., Kaplan, A.:
625 Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth
626 century, Journal of Geoph. Res., 108, D14, 4407, DOI: https://doi.org/10.1029/2002JD002670, 2003.
627 Reiniger, R. F. and Ross, C.K.: A method of interpolation with application to oceanographic data, Deep-Sea Res.,
628 15,2,185-193, 1968.
629 Roemmich, D., Gould, W. J., and Gilson, J.: 135 years of global ocean warming between the Challenger expedition and
630 the Argo Programme. Nature Clim Change, **2**, 425–428, https://doi.org/10.1038/nclimate1461, 2012.

Spilhaus, A. F.: A bathythermograph. Journal of Marine Research, 1, (2), https://elischolar.library.yale.edu/journal_of_marine_research/522, 1938.

State Oceanographic Instit: Manual for hydrological works in oceans and seas, second edition, Leningrad, Gidrometeoizdat, 725 p, 1977.

Swift, J. H.: Reference-quality water sample data: Notes on acquisition, record keeping, and evaluation, IPCC Report No. 14, ICPO Publication Series No. 134, Version 1, 2010.

Thomson, W. and Murray, J.: The Voyage of H.M.S. Challenger 1873-1876. Narrative, Vol.1, First Part, Ch. III, ed. Johnson Reprint Corporation, available at http://archimer.ifremer.fr/doc/00000/4751/, 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, 1889.

U.S. Hydrographic Office: Deep-Sea Soundings in the North Pacific Ocean obtained in the United States Steamer Tuscarora, Commander George E. Belknap, Washington D. C., Government Printing Office, 1874.brary.yale.edu/journal_of_marine_research/522, 1938.

U.S. Naval Oceanographic Office: Instruction manual for obtaining oceanographic data, Washington D.C., U.S.Government Printing Office, 232 p., 1968.

Wu, Q., Gregory, G. M., Zanna, L., and Khatwala, S.: Time-varying global energy budget since 1880 from a reconstruction of ocean warming, Earth, Atmospheric, and Planetary Sciences, 122, 20, https://doi.org/10.1073/pnas.2408839122, 2025.

Wunsch, C. and Heimbach, P.: Practical global oceanic state estimation. Physica D, 230, 197-208, doi: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 (ed. von Walter de Gruyter &CO.), Berlin, 1932.

Zanna, L., Khatiwaia, S., Gregory, J.M., Ison, J., and Heimbach, P.: Global reconstruction of historical ocean heat storage and transport, Proceedings of the National Academy of Sciences 116(4), DOI:10.1073/pnas.180883811523, 2019.

Appendix A: Corrections for depth bias

Table A1. Depth bias corrections for historical Nansen casts

Original	Depth Correction Factor
0050	0.9400
0100	0.9600
0200	0.9750
0250	0.9800
0300	0.9833
0400	0.9835
0500	0.9840
0600	0.9845
0800	0.9860
1000	0.9880
1200	0.9885
1500	0.9900
2000	0.9900
2500	0.9900
3000	0.9900
3500	0.9900
4000	0.9937
4500	0.9937
5000	0.9937

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