On the Consistency of the Bottle and CTD Profile Data

VIKTOR GOURETSKI,[a,b] LIJING CHENG,[a,b] AND TIM BOYER[c]

[a] International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[b] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China
[c] National Oceanic and Atmospheric Administration/National Centers for Environmental Information, Silver Spring, Maryland

(Manuscript received 3 February 2022, in final form 8 August 2022)

ABSTRACT: Nansen bottle casts served as the main oceanographic instrumentation type for more than a century since the establishing of the technique in the late 1890s. Between the end of the 1960s and the end of the 1990s Nansen cast technique has been gradually replaced by electronic sensor profilers (CTD). Both instrumentation types are considered as the most accurate among other oceanographic instruments and are often used as the unbiased reference. We conducted a comprehensive investigation of the consistency of the temperature data from Nansen casts and CTD profilers analyzing the quasi-collocated bottle and CTD data between the 1960s and the 1990s when both instrumentation types overlap. We found that Nansen casts tend to overestimate the sample depth with reversing mercury-in-glass thermometer temperatures being on average slightly lower compared to CTD data. Respectively, depth and temperature corrections are provided. Further, we estimated the ocean heat content changes between 1955 and 1990 using (along with all other instrumentation types) corrected and uncorrected Nansen cast data. These calculations show that for the upper 2 km layer the global average warming trend for this time period increases from 0.20 ± 0.05 W m^{-2} for the uncorrected data to 0.28 ± 0.06 W m^{-2} for the corrected data at the 90% confidence level. Finally, we suggest that the Nansen bottle cast profiles be put into a separate instrumentation group within the World Ocean Database.

KEYWORDS: Climate change; Temperature; Data quality control; Databases

1. Introduction

An unbiased ocean in situ dataset is vital for a wide range of oceanographic and climate applications. The World Ocean Database 2018 (WOD) (Boyer et al. 2018)—the largest publicly available, uniformly formatted, and quality-controlled historic ocean profile dataset—consists of data from more than a dozen different oceanographic instruments, each having specific uncertainties and biases. The first efforts to construct global ocean heat content (OHC) time series from ocean profile data revealed the impact of the systematic errors in the data from expendable bathythermographs (XBT) (Gouretski and Koltermann 2007), which introduced artificial variability in the OHC time series. Respectively, a number of correction schemes for XBT data have been developed (Gouretski and Reseghetti 2010; Cowley et al. 2013; Cheng et al. 2018). The data from MBTs were also found to be biased, with the respective corrections suggested by Gouretski and Reseghetti (2010) and by Gouretski and Cheng (2020).

Nansen/Niskin bottles with reversing thermometers (hereafter referred to simply as "bottles") and ship-based conductivity–temperature–depth (CTD) profile data represent two of the most important constituents of the WOD. Due to the high precision of thermometers and pressure/depth measurements compared to MBTs and XBTs and because of the possibility for precruise and after-cruise calibration, bottle and ship-based CTD data can be considered as a reference subset of hydrographic profile data superior in quality to other instrumentation types. However, the consistency of the bottle and CTD data was rarely studied.

At their inception, bottle casts consisted of series of bottles placed on the wire with a messenger (concentric brass weight) placed on the cable below each bottle. The bottles were designed by F. Nansen and collaborators in 1894 (Helland-Hansen and Nansen 1909). The messenger slides down the wire until it reaches the next bottle tripping a trigger to release the upper end of the bottle from the wire. The bottle falls over and release the next messenger. For deep stations several casts were needed to provide the necessary vertical resolution. Thermometer frames were attached to each bottle holding the mercury-in-glass pressure-protected and unprotected thermometers (Negretti and Zambra 1874). Sampling depth was derived from the difference in temperature between protected and unprotected thermometers. When a cast was not equipped with both protected and unprotected thermometers, the sampling depth was estimated from the length of the wire put out and the angle of the wire to the vertical at the deck height. The original design of the Nansen bottle was generally replaced by the Niskin bottle in the late 1960s. Niskin bottles were placed on a line in the same way as Nansen bottles. In case of CTD profilers Niskin are frequently mounted around a circular rosette sampler metal frame to obtain water samples concurrent with the CTD sensor measurements. Rossette frames have the capacity to hold as many as 36 bottles. The CTD rosettes had reversing thermometers on selected bottles to calibrate the CTD temperatures. Throughout the paper, the terms "bottle data" or "bottle cast" encompass temperature data from reversing thermometers attached to Nansen or Niskin bottles. Data obtained by the reversing thermometers attached to the bottles of the CTD rosette are related to the CTD instrumentation group.

Corresponding authors: Viktor Gouretski, viktor.gouretski@posteo.de; Lijing Cheng, chenglij@mail.iap.ac.cn

DOI: 10.1175/JTECH-D-22-0004.1

The technique of working with bottle casts has been standardized and described in a number of manuals. In some cases, manuals provide working instructions valid for all oceanographic vessels of a country. An example is the hydrographic manual for the ships of the former USSR (State Oceanographic Institute 1977). In other cases, the manuals are specific for the ships belonging to a certain organization like the U.S. Navy (Naval Hydrographic Office 1968). Though bottle casts represent one of the earliest standardized oceanographic techniques, the variation in quality of bottle data is striking (Saxton 1964). A detailed review of carrying out bottle stations from late 1950s through the 1970s is given by Warren (2008).

The first electronic profilers and samplers for oceanographic studies were developed in the 1950s (Hamon 1955; Hamon and Brown 1958), and the earliest profile data of this type available in the WOD refer to the year 1964. Commercial production of CTDs began in 1964 (Brown 1974).The manufacturing of electronic salinity–temperature–depth (STD) and later CTD profilers started in several countries (United States, Germany, France, former Soviet Union). The gradual process of supersession of classical bottle cast technique by electronic sensor profilers took several decades. Unfortunately, the metadata needed to unambiguously attribute the instrumentation type are often missing in the oceanographic data held in digital form in archives, and so missing in the WOD. Cruise reports and other documentation detailing the exact instrumentation used by particular institutions, projects, ships, and investigators are scattered across institutional repositories and libraries around the world. Often, this information can most readily be gathered from the investigators themselves. Many of those who were on cruise during the period in which CTDs gradually supplanted bottle casts as the main source of ship-based high-quality temperature (and salinity) data have retired. Therefore, a query was started among oceanographers from different countries to gather institutional knowledge from scientific papers, institutional repositories, and the memory of investigators to find out the time of the earliest CTD implementation, and to document the specific instrumentation used on historic cruises.

The first profilers manufactured in the United States appeared in mid-1960s, but reversing thermometers attached to bottles were still in use to supplement CTD work until 1981 at Woods Hole Oceanographic Institution (Warren 2008). Independently from the U.S. manufacturers, the electronic profiler of the type "bathysonde" was developed at the Institute for Applied Physics in Kiel, Germany, with first trials of the profiler taking place in 1961, and the routine measurements started in 1964 during the Indian Ocean cruise of the R/V *Meteor II* (W. Zenk 2020, personal communication). Similarly, the United Kingdom started using CTDs (e.g., STD profilers) in the late 1960s, with the MEDOC 69 experiment providing one of the first routine CTD implementation examples. In contrast, the Royal Navy ships and weather ships (Stations I and J) stuck with bottle casts for considerably longer (J. Gould 2020, personal communication). The first implementation of CTD profilers on Argentine ships probably goes back to "Islas Orcadas" cruise 11 in late 1976 (A. Piola 2020, personal communication). The French oceanographers started to work with a CTD profiler on or around the time of the "Jean Charcot" cruise during the MEDOC 70 experiment in 1970 (M. Fieux 2020, personal communication). The first profilers on Norwegian ships were implemented in early 1970s, with different ships getting digital instruments at different time (H. Sagen 2020, personal communication). At least three types of electronic profilers were developed in the former USSR. The "ISTOK" profilers were constructed in early 1960s at the Hydrophysical Institute in Sevastopol, former Soviet Union, and were installed at several ships operated by this institute. The ISTOK-3 modification was often used on the USSR fishery reconnaissance vessels (I. Skvorets 2020, personal communication). The large fleet of oceanographic vessels operated by the Hydrometeorological Service of the Soviet Union was equipped with the profilers of the type "Zond-Batometer." However, these types of profilers served rather as an additional type of instrumentation, with bottle casts remaining the main technique to conduct deep-water observations [personal experience of the first author, who participated on cruises with the R/V *Akademik Korolev* (1974) and the R/V *Professor Vieze* (1987–88)]. The profilers of the type "AIST" were used at the research ships operated by the Institute of Oceanology of the Academy of Sciences. In spite of the existence of several types of profilers developed in the former Soviet Union, bottle casts remained the main technique used by the majority of the Soviet ships. Also, in other countries the supersession of bottle casts by CTD profilers took place much later. For instance, the first Icelandic ship was equipped with the CTD profiler in May 1989 (M. Danielsen 2020, personal communication).

This study aims to investigate and quantify the quality of data collected by historical bottle casts, which is the major subsurface temperature data source (and sole deep ocean > 250 m source) before mid-1960s. The manuscript is organized as follows. The data used in the study are described in section 2, followed by the description of the method to calculate temperature offsets for quasi-collocated profiles in section 3. Details of the offset calculations for different profile groups are presented in section 4. Possible error sources in bottle data are discussed in section 5. Based on the comparison with the high-resolution CTD profiles the depth and temperature correction scheme for historic Nansen cast data is developed and presented in section 6 followed by the comparison with the available results from the thermometric depth estimation method in section 7. The total set of Nansen bottle casts is then divided into two groups with the data from the standard and nonstandard levels, respectively, to further investigate the sources for the detected temperature offset (section 7). The calculation of the bottle cast depth bias was repeated based on the salinity data to demonstrate a qualitative agreement between the overall depth corrections based on temperature and salinity data (section 8). Finally, the impact of suggested corrections on the estimates global ocean warming trend is investigated (section 9).

2. Data

a. Bottle and CTD data

WOD was used as the data source for this study. The time period of interest is 1964–99 because the first CTD data in the

TABLE 1. Number of temperature profiles in different categories and groups for the time period 1964–2000. Percentage values are given in parentheses.

Profile category/group description	Category or group name	Number of profiles
Ocean station data (OSD) category within the WOD	OSD	1 951 380
Conductivity–temperature–depth (CTD) category within the WOD	CTDH	599 024
Groups of CTD profiles within OSD category		
Low-resolution CTD paired with high-resolution CTD (code13)	CTDLP	58 750 (1.37%)
Low-resolution CTD closely collocated with high-resolution CTD, except CTDLP	CTDLC	26 839 (1.38%)
Low-resolution CTD, except CTDLP and CTDLC	CTDL	203 312 (10.42%)
Likely low-resolution CTD, except CTDLP, CTDLC, and CTDL	LCTDL	221 043 (11.33%)
Groups of bottle profiles within OSD category		
Nansen bottle profiles	BOT	30 987 (1.59%)
Likely bottle profiles	LBOT	1 865 791 (95.61%)
Very likely bottle profiles	VLBOT	1 644 748 (84.29%)

WOD relates to the year 1964, and the subsequent analysis implies that the bottle cast technique was essentially replaced by the CTD instruments by the 1990s for temperature measurements. By the 1990s, when bottle cast temperature was recorded at all, it was usually to calibrate the CTD temperature sensor. The two data types—bottle and CTD—represent essentially different observational techniques. Whereas during a bottle cast temperature is measured by mercury-in-glass thermometers, electronic sensors are used for CTD profilers. The resistance to electrical current in a platinum thermistor is a measure of the temperature of the water. The pressure at sample depth is also directly measured by CTDs, whereas sample depth for Nansen or Niskin bottles attached to the wire was estimated from the angle and length of the wire put out or calculated from the difference in temperature readings from the paired pressure-protected and pressure-unprotected thermometers. In case when no paired thermometers were used the errors in sample depth could be quite large, because the shape of the wire under the surface is essentially unknown and depends on many generally unknown factors, like wind, water currents, and parameters of the ship hull.

Bottle cast technique introduced in the end of 1890s remained in use for about a century before electronic profilers almost completely replaced them by the end of the twentieth century.

Bottle casts (Nansen or Niskin bottles) reside in the ocean station data (OSD) category within the WOD (Table 1). Only a small portion of the OSD profiles obtained after 1963 can be unambiguously attributed to bottle cast profiles. There are 30 987 casts from the former Soviet Union for which the WOD temperature instrument is designated "bottle." These profiles are from the bottle group (BOT) and represent solely Nansen cast profiles because Niskin bottles were not used on the ships of the former USSR.

Bottle casts usually gather more oceanographic information than simply temperature and salinity. Chemical and biological constituents of seawater are measured from bottle samples. Often, investigators report pressure, temperature, and salinity from the CTD package lowered with the bottle rosette only at the level at which the bottles were tripped to provide the physical state of the water from which other measurements were extracted. The electronic temperature sensor of the CTD records at a frequency at or higher than 1 Hz. Depending on the speed at which the CTD is lowered, CTDs can record a considerable amount of temperature information in a short vertical distance. The CTD category in the WOD endeavors to include CTDs with a measurement every 5 m or less in some portion of the vertical profile (with allowances for shallow casts, deep casts with lower frequency at deep depths). In this manner the CTD dataset differs from the OSD dataset in that it is a high-resolution dataset (CTDH). CTD data received in tandem with bottle casts at only bottle cast depths are low-resolution CTDs (CTDL). The OSD dataset also includes a considerable amount of historical CTD data which have been received at selected standard depths due to digital conventions which were designed when file size was a critical factor in the exchange of data. These data are included as CTDL data in the OSD dataset. Unfortunately, a considerable fraction of the cast from the OSD dataset are missing metadata preventing unambiguous attribution to CTD or bottle instrumentation. A portion of the CTDL data in the OSD dataset have a CTDH version in the CTD dataset. These "high-resolution pairs" are cross identified in both OSD and CTD datasets in the WOD. In this way users of the OSD dataset have access to CTD values collected at the same time and depth or pressure that water samples are collected and to maintain an inclusive concise set oceanographic variables for bottle cast data in the OSD dataset while maintaining the high-resolution information necessary for some oceanographic studies and not double weighting a single oceanographic cast.

b. Data quality control

Both the bottle and the CTD temperature profile data were quality controlled, using the automated quality control procedure developed for the compilation of the World Ocean Circulation Experiment–Argo Hydrographic Climatology (Gouretski 2018).

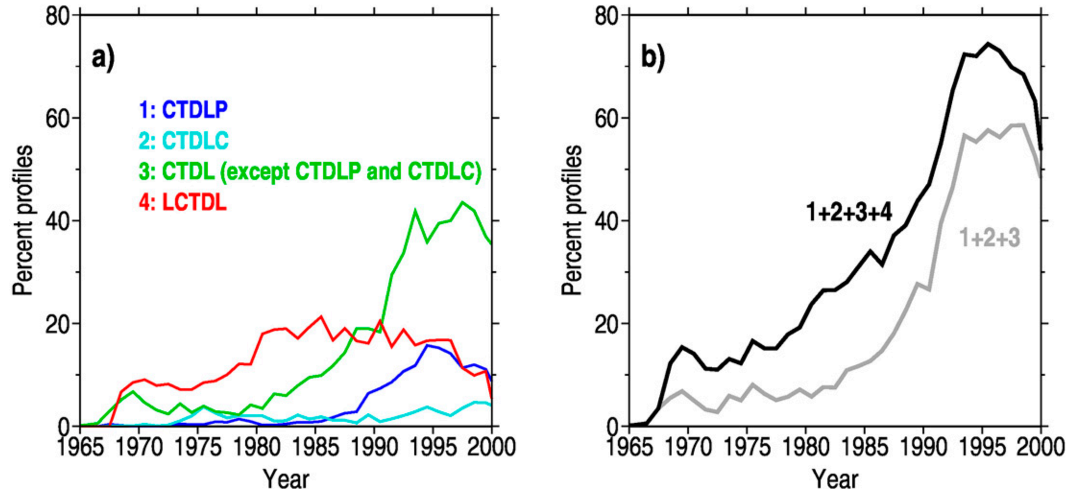

FIG. 1. (a) Yearly percentage of profiles from different groups of low-resolution CTD instrumentation type (see Table 1). (b) As in (a), but for two merged profile groups attributed to low-resolution CTD profiles.

The novelty of this quality control procedure is that the local climatological ranges of temperature at each location and depth are calculated taking the skewness of the distribution into account. This procedure effectively removes true data outliers, while keeping the percentage of falsely identified outliers relatively low. In cases where interpolation of vertical profiles was needed, the Reiniger and Ross (1968) parabolic method was used.

c. *Distinguishing bottle casts from low resolution CTD casts*

Since bottle casts and CTD have principally different methods of sample depth estimation and of measuring temperature it is necessary to distinguish between these two types of observations for the current analysis. We used the following metadata (if available) in WOD profile data files to distinguish between depths and temperatures from bottle casts and from CTDL casts in the OSD group: 1) temperature instrument [variable specific second header code 5 (V5)], 2) casts with concurrent CTDH [second header 13 (S13)], 3) probe type [second header 29 (S29)] and 4) bottle sampler type [second header 97 (S97)]. In the following sections we distinguish between several groups of CTDL casts (Table 1), with the yearly percentage of profiles for each group from the total yearly number of OSD dataset casts given in Fig. 1.

1) LOW-RESOLUTION CTD PROFILES WITH HIGH RESOLUTION PAIRS (CTDLP AND CTDLC GROUPS) AND LIKELY BOTTLE PROFILES (LBOT GROUP)

V5 identifies 58 749 temperature profiles as CTDL casts between 1964 and 2000 with a CTDH counterpart (S13 designation; CTDLP). Collocation checks between the OSD dataset and CTD dataset reveal 26 839 OSD casts closely collocated with CTDH counterparts with no S13 designation (high-resolution pair; CTDLC). These collocated OSD–CTDH pairs are not designated with S13 for a variety of reasons. Most often, CTDLC come from different sources than the CTDH and may not be easily identified with the correct CTDH other than position–date collocation. This lack of identification may be due to missing (or erroneous) information on platform (ship from which rosette was dropped) in one or the other of the datasets or differences in position and date/time which leave uncertain the actual cast among several to which an S13 should be assigned. These differences in data and metadata can be resolved in many cases with sufficient scrutiny. For this work, the 47 503 CTDLC are not used for analysis out of an abundance of caution given the uncertainty in matching datasets.

After exclusion of CTDLP and CTDLC there remained 224 868 depth–temperature profiles in the CTDL group of the OSD dataset (Table 1). The rest of 1 865 791 OSD depth–temperature profiles were attributed to the "likely bottle" (LBOT) group.

2) LIKELY LOW-RESOLUTION CTD PROFILES (LCTDL GROUP) AND VERY LIKELY BOTTLE PROFILES (VLBOT GROUP)

Many of the LBOT group do not include the necessary metadata to unequivocally designate the depth and temperature data as coming from a reversing thermometer (bottle casts).

Three additional tests were applied to the LBOT casts to identify likely sensor data (LCTDL).

Test 1: CTDH casts were grouped according to the ship name. For each ship name the date (year/month/day) of the earliest cast was found. We assume that for each ship after the first implementation of the CTD instrumentation, all subsequent depth/temperature profiles were also taken by means of CTD instruments. This assumption might not hold in each case because scientific teams often used their own CTD equipment which did not necessarily remain on the ship after cruise completion. The ships of the former Soviet Union were excluded from this test. For these ships knowledge on the data of the earliest CTD implementation does not provide useful information, because bottle casts remained the main instrumentation type and electronic profilers were often used in parallel with the classical bottle casts.

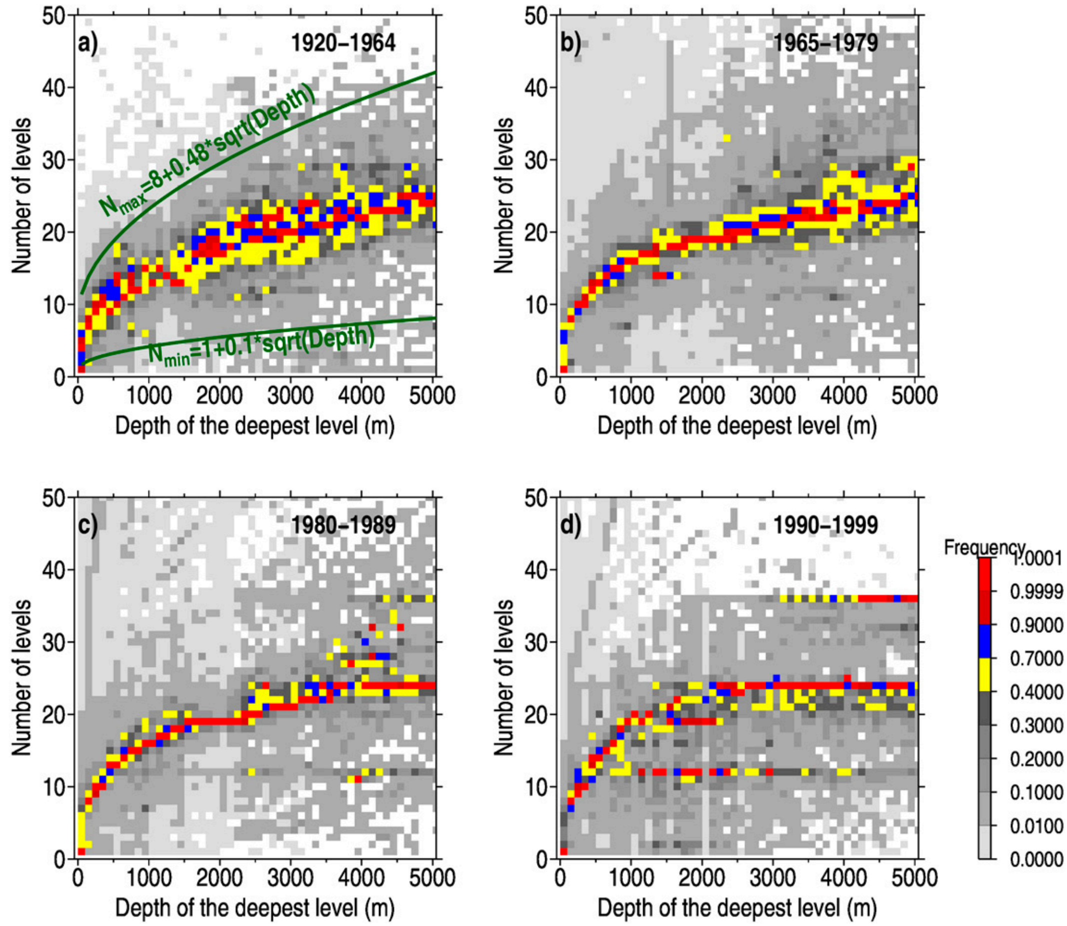

FIG. 2. (a) Histograms of the number of reported levels vs profile depth for OSD instrumentation type for the pre-CTD time period 1920–64. Region between the green lines includes about 96% of all profiles. (b)–(d) As in (a), but for selected time periods after 1964.

Test 2: Number of reported (observed) levels also provide additional information on the instrumentation type. Figure 2 shows frequency histograms for the number of reported levels versus the deepest sampled depth (Z_{last}) for selected time periods. Generally, the number of observed levels increases with the cast depth, with the number of levels being typically less than 40 even for the deepest casts. This reflects the limitation on the bottle cast vertical resolution imposed by equipment (wire, rosette, etc.), as the typical maximum number of bottles rarely exceeded 15, so that deep stations required several casts to provide the desirable full-depth vertical resolution. According to Fig. 2, casts with atypically large number of reported levels are very rare before 1964, when CTD (STD) sensors were introduced, but appear more frequently in the following years. We explain this partly through the presence of CTDL casts in the OSD group. Casts with the number of levels exceeding $N = 10 + 0.48\sqrt{Z_{last}}$ were designated LCTDL (Fig. 2a).

Test 3: Due to the limited number of bottles vertical resolution of bottle casts changes with increased depth of cast. Spacing between bottle trips is also greater at deeper levels, where the water column becomes more homogeneous. For CTD profilers the vertical sampling typically remains the same throughout the water column. For each bottle cast with at least 10 reported levels the mean spacing and the standard deviation of the spacing between the bottles were calculated in order to identify profiles with constant vertical spacing. Profiles with constant vertical spacing were designated LCTDL. All casts which do not belong to CTDLP, CTDLC, CTDL, and LCTDL groups compose the very likely bottle profile group (VLBOT group).

d. *High-resolution CTD profiles*

The CTD dataset within the World Ocean Database includes 599 024 casts with a vertical resolution of 1–2 dbar between 1964 and 2000. The metadata in most cases make it possible to identify the manufacturer and the type of the CTD profiler. The high-resolution CTD profiles are represented by the CTDH group (see Table 1). The high-resolution profiles are assumed to be superior in accuracy over the bottle profiles

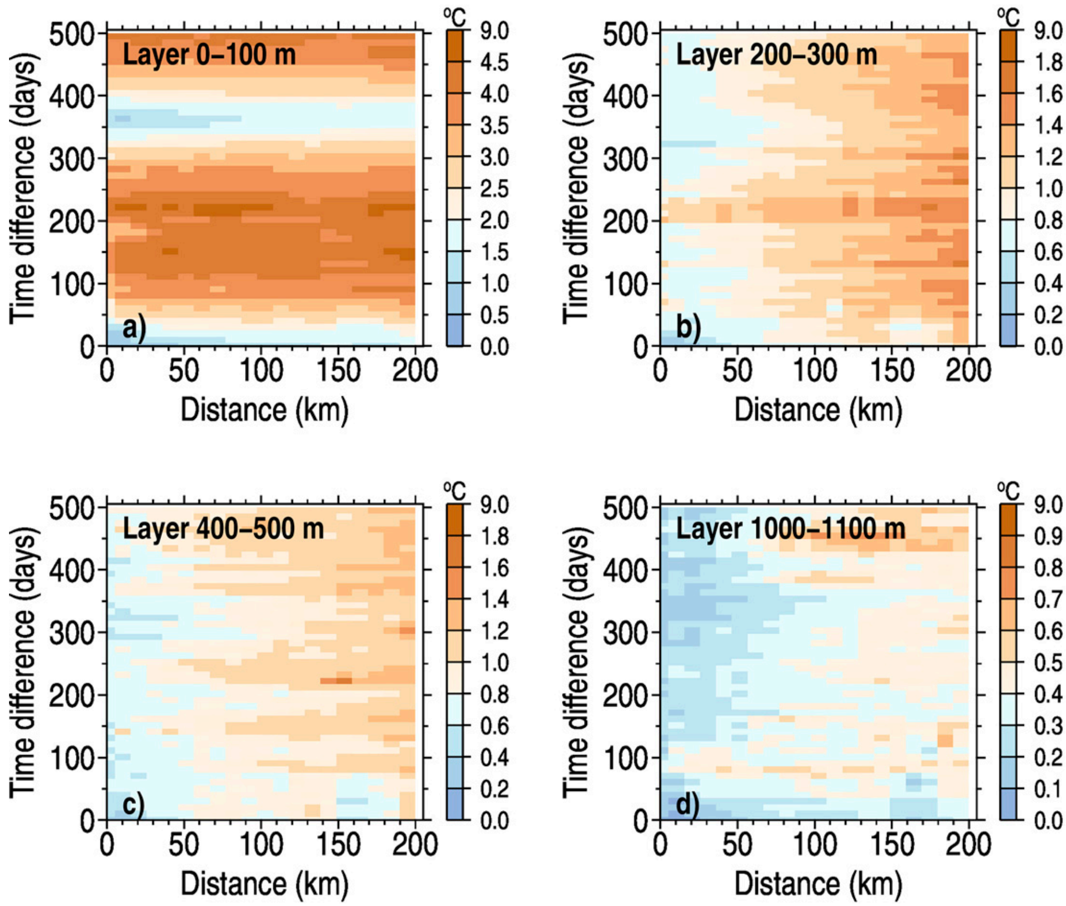

FIG. 3. Mean absolute offset vs time and spatial separation between quasi-collocated profiles for selected layers: (a) 0–100; (b) 200–300; (c) 400–500; (d) 1000–1100 m.

and serve as the reference when analyzing collocated data in this study.

3. Calculation of the offset between quasi-collocated bottle and CTD temperature profiles

To estimate the offset between bottle and CTD temperature profiles (temperature values TB and TC, respectively) the method applied by Gouretski and Cheng (2020) to derive temperature bias for MBTs was used. The essentials of the method are as follows. For each TB, all TC situated within the radius R and obtained within τ days from the occupation date of the bottle cast were selected. For each observed bottle temperature differences δ_i= TB − TC_i (i = 1, … , N, where N is the number of TC in the collocation bubble) were calculated. In case of more than one TC, the weighted average of all T differences is taken: $\Delta = \Sigma\delta_i w_i/\Sigma w_i$, where weights $w_i = \exp[(\Delta r)^2/R^2 + (\Delta t)^2/\tau^2]$, where Δr is the distance between TB and TC and Δt is the time difference. This weighted average represents the estimate of the temperature offset between the TB and the collocated TC. The TB − TC differences are calculated at the observed bottle profile depths, so that only the CTDH profiles need to be vertically interpolated. Below we explain the reasons why CTDL temperature profiles are not used as a reference along with the high-resolution temperature profiles.

To decide on the size of the influence bubble we calculated absolute median TB − TC temperature offsets using the collocated pairs for 10 km × 5 days spatial–time bins for three layers: 0–100, 200–300, 400–500, and 1000–1100 m (Fig. 3). Except for the uppermost 100 m layer with pronounced seasonal temperature variability the choice of the spatial size of the collocation bubble has greater impact on the median absolute offset compared to the variations in the temporal separation between the profiles.

The calculated individual offsets are vertically interpolated at regularly spaced 10 m levels in order to obtain the average TB − TC differences (offsets) for individual years. Due to the general paucity of the data, a massive averaging of individual offsets is needed in order to effectively reduce noise due to temporal and spatial temperature variability. We combine individual offsets for each 10 m level within the 7-yr time window centered at each calendar year between 1971 and 1990, and the yearly values of the temperature offsets are then

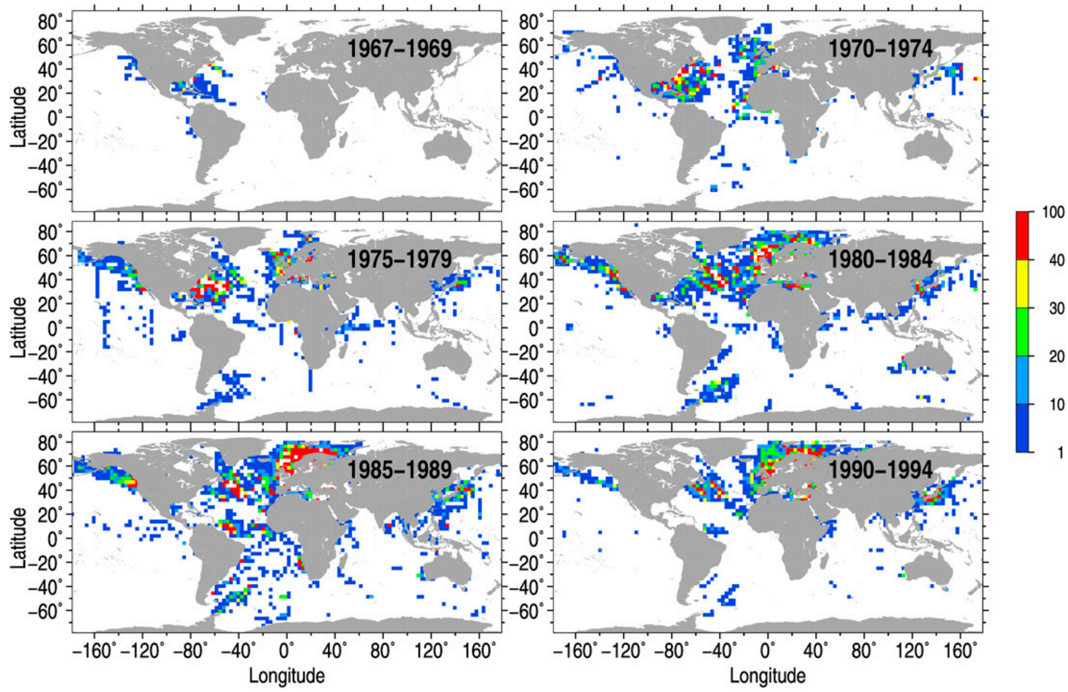

FIG. 4. The number of the collocated bottle and CTD profiles in 3° × 3° latitude–longitude boxes for selected pentades.

calculated as the median of all individual offsets within the time window.

To test the impact of the collocation bubble size on the temperature offset pattern, experiments were conducted for 16 different bubbles, with the spatial size between 50 and 200 km and the maximum time difference between the collocated profiles within 15 and 60 days. Changing the bubble size within the ranges indicated above does not lead to qualitative changes in the offset pattern, with the offset standard deviation from the mean over 16 runs being typically lower than 0.02°–0.04°C. Based on these tests we selected the collocation bubble size of $R = 150$ km and the maximum time difference between the observations $\tau = 45$ days. For these size of the collocation bubble this yields a total of 119 744 collocated pairs of VLBOT and CTDH profiles are available (the CTD profiles might contribute multiple time to the collocated pairs because of overlapping bubbles). The spatial distribution of the collocated pairs for selected pentades (Fig. 4) shows that the majority of collocations is situated in the more frequently sampled regions of the Northern Hemisphere, especially along the eastern U.S. coast and in the northeastern part of the Atlantic Ocean. The much less numerous collocated pairs during the end of 1960s are concentrated almost solely near the North America, indicating the initial implementation of CTD profilers on the U.S. ships. Before the 1980s the majority of the collocated profiles are found in the subtropical and moderate regions of the World Ocean, whereas during the subsequent years the majority of the collocated pairs are in the subpolar and polar regions of the Northern Hemisphere.

4. Temperature offsets between different groups of data

We calculated temperature offsets between five different profile groups and the CTDH data (Table 2) with Fig. 5 showing the respective offset patterns versus depth and year. The overall offset pattern is very similar for VLBOT and LBOT temperature profiles, with the latter group having 22% more collocated profiles. In contrast, the subsets of CTDL temperature profiles exhibit different offset patterns. The paired low- and high-resolution CTD profiles demonstrate the best agreement with the absolute offsets below 200 m typically less than 0.02°C. The offsets for the CTDLC and CTDH temperature profiles is similarly low below about 400 m, with higher offsets at shallower levels. The offsets for the CTDL temperature profiles are predominantly negative above 600–800 m levels. The worst agreement with the CTDH temperature profiles is observed for the LCTDL temperature profiles. Based on the

TABLE 2. Instrumentation groups and the respective number of profiles collocated with the high-resolution CTD data. The names of the instrumentation groups are explained in Table 1.

Bottle profile group	Number of collocated pairs
VLBOT	119 744
LBOT	145 764
CTDL	38 710
CTDLC	21 684
CTDLP	38 850
LCTDL	89 738

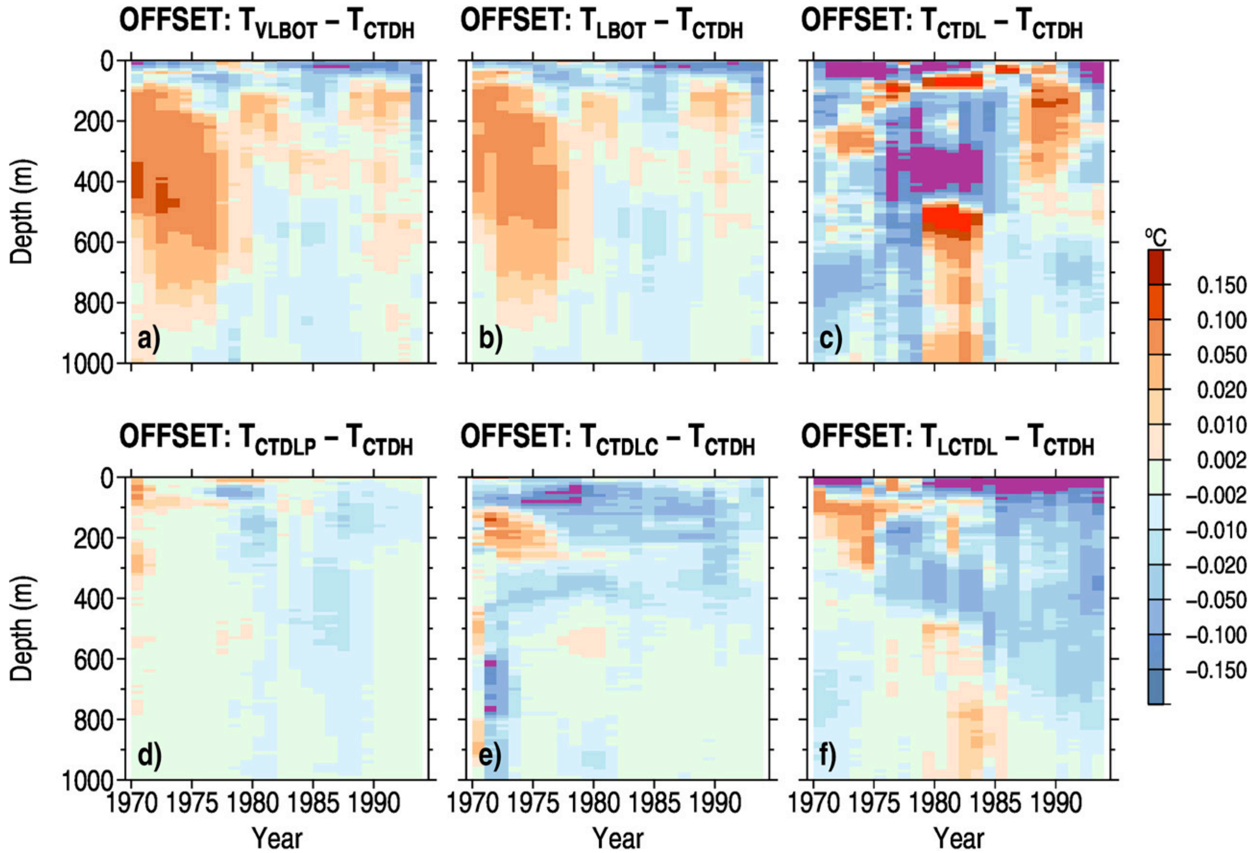

FIG. 5. Temperature offset relative to the high-resolution CTD data for the following groups of the OSD profiles: (a) VLBOT; (b) LBOT; (c) CTDL; (d) CTDLP; (e) CTDLC; (f) LCTDL (see Table 1).

above results and with the aim to keep the reference data as homogeneous as possible we do not use the CTDL temperature profiles for the derivation of biases in bottle data and rely on the comparison between the VLBOT and CTDH groups.

Figure 6 provides several statistics for the VLBOT and CTDH collocated pairs. According to Fig. 6a, the percentage of VLBOT collocated with CTDH changes over the time period in question, increasing from a few percent in 1960s to about 18% in 1990. The collocated profile pairs are unevenly distributed between the countries. According to Fig. 6b, more than 50% of all VLBOT between 1973 and 1992 come from the ships of the former Soviet Union, Russia, and the Ukraine. We treat VLBOT from these three countries as one and the same group due to similar standards and oceanographic

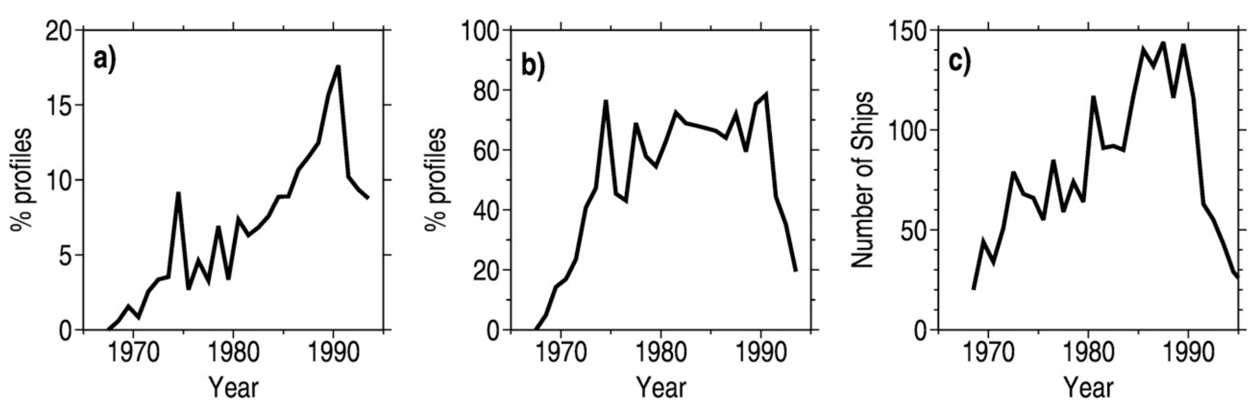

FIG. 6. (a) Percentage of bottle profiles (VLBOT group) collocated with high-resolution CTD profiles. (b) Percentage of VLBOT profiles from Russian and Ukrainian ships among all collocated profiles. (c) Yearly number of ships that contributed with VLBOT profiles collocated with high-resolution CTD profiles.

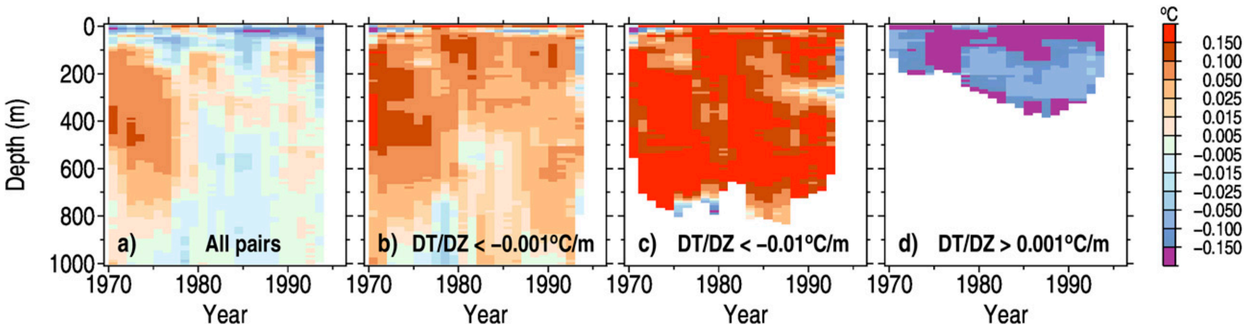

FIG. 7. Temperature offset of bottle data relative to high-resolution CTD data for different ranges of the vertical temperature gradient: (a) all collocated VLBOT–CTDH pairs; (b) $dT/dz < -0.001°C\ m^{-1}$; (c) $dT/dz < -0.01°C\ m^{-1}$; (d) $dT/dz > 0.001°C\ m^{-1}$.

practices used on the ships of these countries. In case the ships from this group differed systematically in their maneuverability from ships of other countries, it could have systematic impact on wire angle and shape, but the available metadata are not sufficient to check this hypothesis. Finally, we note, that the number of distinct ships contributing to the collocated dataset varies significantly over time, similarly to the percentage of the collocated profiles (Fig. 6c). Between 1984 and 1991, more than 100 different ships contributed each year to the collocated dataset.

5. Temperature and depth errors in bottle cast data

Basically two kinds of errors influence the accuracy of the bottle cast data: 1) errors related to the estimation of the bottle sampled depths and 2) errors in measuring temperature by means of reversing mercury-in-glass thermometers.

a. Depth error

One of the motivations for the current study was to check whether bottle casts exhibit a systematic depth error because of the difficulties with estimation of bottle depth. The possible depth error has no expression in temperature in case of a vertically homogenized water column. However, thick thermostads are relatively rare in the upper ocean and occur mostly in the high latitude regions where vertical convection in winter effectively homogenize the water column. Much more common in the upper ocean is the case of a nonzero vertical temperature gradient. In this case the error in depth translates into the respective error in temperature, with temperature error magnitude being proportional to the magnitude of the vertical temperature gradient.

When paired thermometers were not used on a hydrographic cast the length of the wire put out assumed to be representative of the sample depth. However, the position of the wire usually deviates from the vertical, so that the actually sampled depths might be on average smaller than the respective length of the wire put out. It should be noted, that even if the wire is straight and vertical, the measurement of wire out is not very accurate and could conceivably be biased. To investigate the existence of a systematic depth error in bottle cast data, we calculated yearly VLBOT − CTDH temperature offsets for several ranges of the vertical temperature gradient. The results of the calculations presented in Fig. 7 indicate that bottle casts tend to overestimate the sample depth, with larger biases corresponding to the stronger vertical temperature gradients. In the regions where temperature decreases with depth (predominant situation in the World Ocean) the offset is positive, whereas in the regions with temperature inversion the offset is negative. Both cases imply *depth overestimation* by bottle casts.

The overall relationship between the VLBOT − CTDH depth offset and the vertical temperature gradient is illustrated by Figs. 8a and 8b, where the median offsets in gradient-depth bins are shown. For all depth levels offsets are positive for negative gradients and are negative in case of temperature inversion.

b. Reversing thermometer errors

The histogram in Fig. 8b reveals a weak negative VLBOT − CTDH temperature offset of −0.02° to −0.05°C for the vertical temperature gradient range 0.000° to $-0.001°C\ m^{-1}$ for all levels above 1000 m. In the case of a very homogeneous water column the calculated temperature offset cannot be explained by the uncertainty in the bottle depth estimation.

The reversing thermometers used on historical bottle casts are prone to a number of errors. Sverdrup et al. (1944) elaborate on the following error sources: 1) errors of reading, 2) correction errors arising due to the difference between the in situ temperature and temperature during thermometer reading, 3) corrections errors arising from limit of accuracy test, 4) correction errors arising from change of thermometer zero point after the manufacture of the thermometer, and 5) errors of the breaking-off device. With respect to the second correction type in the above list it should be noted that the proper relaxation time is necessary before reading the reversing thermometers. De Visser and Hodgson (1983) conducted a number of simulation experiments to investigate the effect of premature reading of reversing thermometers. They found that maximum errors due to premature reading occur during the first 20 min, However, for the high precision measurements, this time may be insufficient to achieve thermal equilibrium. The magnitude of the error due to premature reading depends on the absolute value of the difference between time constants for the main and auxiliary thermometers. According to the simulations the error magnitudes on the order of 0.01°C are possible even for relaxation

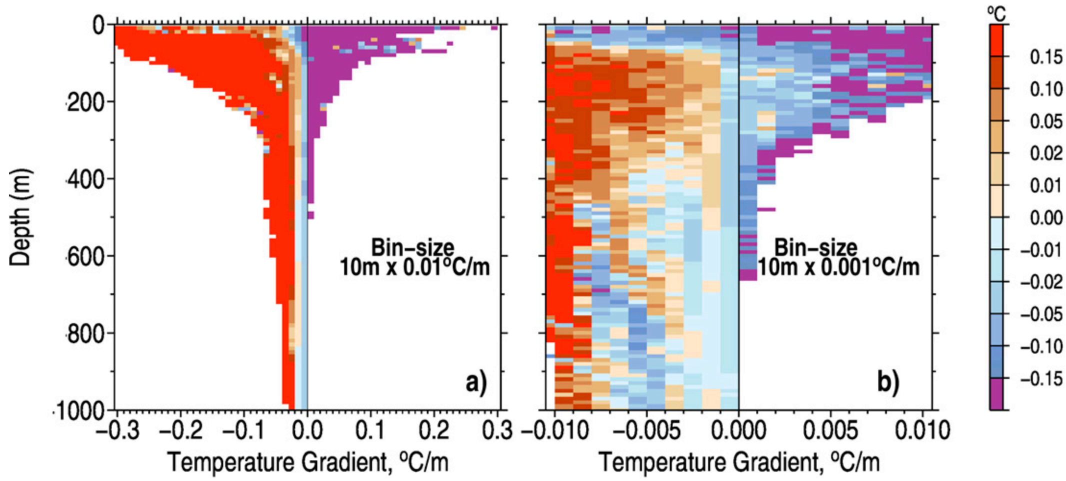

FIG. 8. (a) Median VLBOT − CTDH temperature offset in depth–gradient bins (bin size: 10 m × 0.01°C). (b) As in (a), but for the bin size 10 m × 0.001°C.

time greater than 30 min. Unfortunately, the oceanographic databases do not possess metadata regarding the thermometer time constants to apply respective corrections. Our estimate of the overall thermal bias between 1971 and 1990 and for the layer 200–800 m is −0.02°C, i.e., well within the envelope provided by the simulations by De Visser and Hodgson (1983).

6. Depth and temperature correction scheme for historical bottle data

Similar to the previous studies of the biases in MBT and XBT data (Gouretski and Reseghetti 2010; Cheng et al. 2014; Gouretski and Cheng 2020), we consider here the bias model, which takes into account both the depth and the thermal bias in bottle data. The CTD data are taken as the bias free reference. Compared to the bathythermograph bias studies, far fewer collocated pairs are available for bottle and CTD data. For instance, for the VLBOT − CTDH data comparison 119 744 collocated profiles pairs are available between 1967 and 1999 (Table 2). After a number of experiments we finally selected a 7-yr-wide time window centered at each calendar year. The choice of the time window width is a trade-off between the smoothness of the bias pattern and the intention to minimize the time window width.

For each collocated pair the vertical interpolation of temperature and VLBOT − CTDH temperature difference profiles on 10 m levels is performed. These interpolated collocated profiles provide the input for the bias correction procedure. First, the thermal bias is subtracted from the original temperatures of the bottle casts. Second, the original depths of each bottle cast are changed by multiplying by the depth correction factor, which is increased iteratively by the increment of 0.0001 in the range 0.7–1.3. After each iteration the median VLBOT − CTDH offset over all pairs at each level is calculated. The value of the depth correction factor, which results in the smallest residual median offset, corresponds to the optimal depth correction. Figure 9 shows the optimal depth correction along with the original and residual offsets and the offset reduction. The offset reduction for each level and year is determined as the difference between the absolute values of the original and the residual offset. The mean reduction factor R is defined as $R = \Sigma\,|B_{\mathrm{original}}(i, k)|/\Sigma\,|b_{\mathrm{residual}}(i, k)|$, where year $i = 1971, \ldots,$ 1993, and the 10 m depth level $k = 5, \ldots, 101$ [the uppermost 40 m layer ($k < 5$) is not used]. In case of the application of the yearly corrections, the substantial reduction of the original temperature offset is achieved, with the average reduction factor equal to 2.71.

The derived optimal yearly depth corrections exhibit variations from year to year. At least part of these variations reflect the limitations of the data basis (wrong attribution to the data type due to the lack of metadata, insufficient amount of the collocated pairs) rather than the real changes in biases related to the changes in measurement techniques, since the Nansen cast method remained essentially unchanged over almost a century. We also calculated *the overall depth corrections* as median of all yearly values (Figs. 9a,b, bold red curve). The overall depth corrections are given in Table A1 in the appendix. Within the upper 1000 m the overall median depth correction does not exceed 10 m. The depth corrections imply depth overestimation by bottle casts persistent over the whole 20-yr intercomparison period. A few exceptions can be seen for the uppermost 50 m layer and for the layer below 850 m. In the uppermost layer calculations of the depth corrections are affected by the strong time and spatial variability whereas in the deepest layer the smaller number of the available collocated data might have impact on calculations. Figures 9g and 9h reflect changes in the geographical distribution of the available collocated pairs over time. Before 1978 the collocated pairs are available mostly in the warmer and typically stronger stratified regions (see also maps in Fig. 4). The total bottle temperature bias pattern (Fig. 9d) broadly correlates with the depth–time distribution of the vertical temperature gradient (Fig. 9g) and

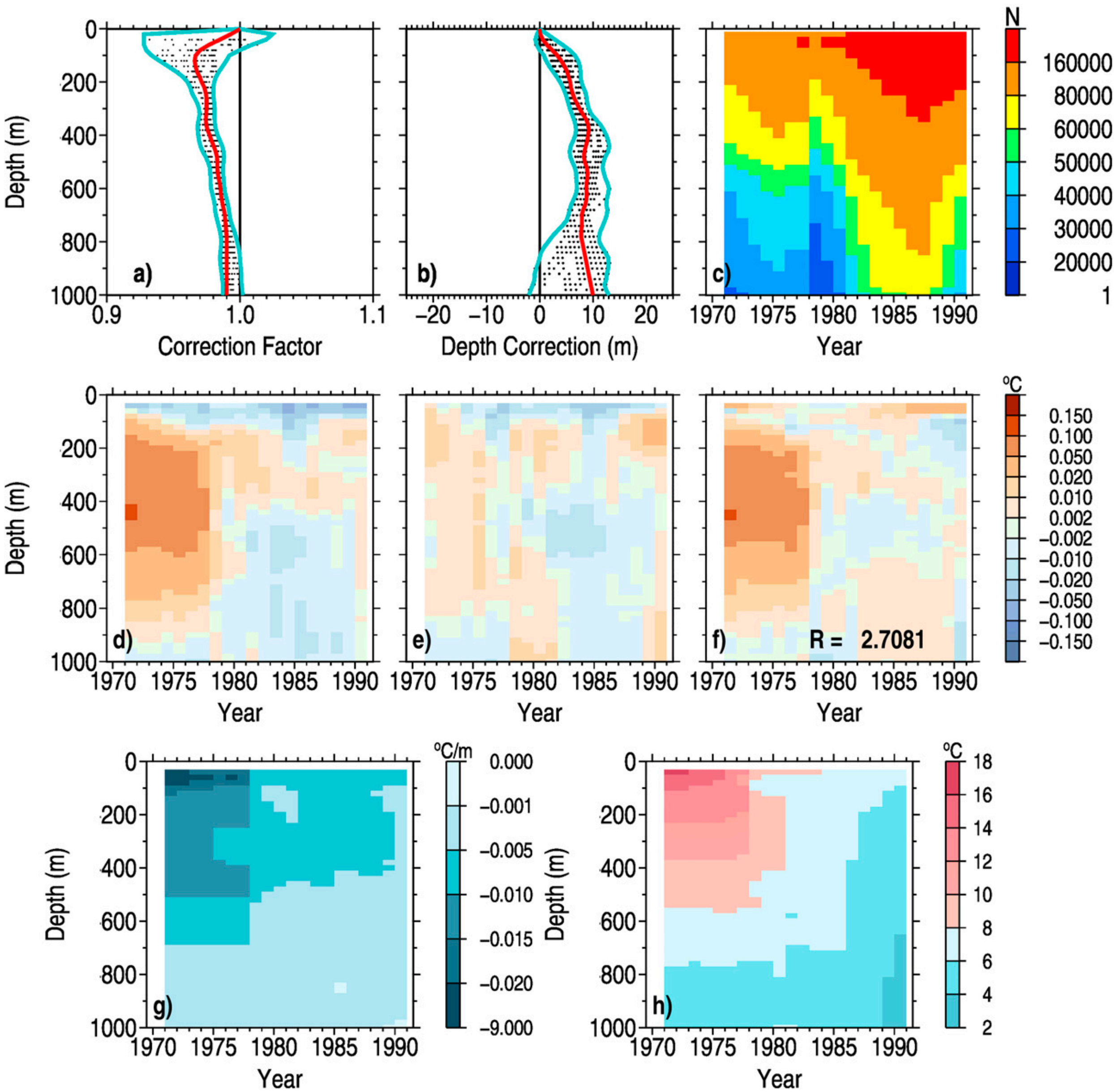

FIG. 9. (a) The optimal yearly depth correction factor (dots), the overall depth correction factor for the years 1970–93 (red curve), the minimum and maximum values of the depth correction factor (cyan curves); (b) as in (a), but for the optimal depth correction; (c) number of collocated pairs available within the 7-yr time window; (d) original temperature offset; (e) residual temperature offset; (f) offset reduction and the overall offset reduction factor *R*; (g) vertical temperature gradient; (h) temperature [(g) and (h) are based on bottle thermometer data].

temperature (Fig. 9h). The general depth overestimation by Nansen casts leads to a stronger positive total bias for the years with stronger vertical gradient.

7. Standard and nonstandard level data

a. *Bottle cast vertical sampling*

Our calculations imply that bottle cast data on average exhibit both thermal and depth biases when compared with the collocated reference CTDH data. As suggested in the section 5b, the thermal bias may appear due to the premature reading of the reversing thermometers, which have not reached thermal equilibrium after being drawn on board. The most obvious reason for the depth bias is explained through the imperfectness of the sample depth estimation. Bottles with thermometer frames were distributed along the whole length of the wire in order to take samples (and measure temperature) close to some predefined target depths. When paired thermometers were not used the only parameters used to estimate sample depths were the length of the

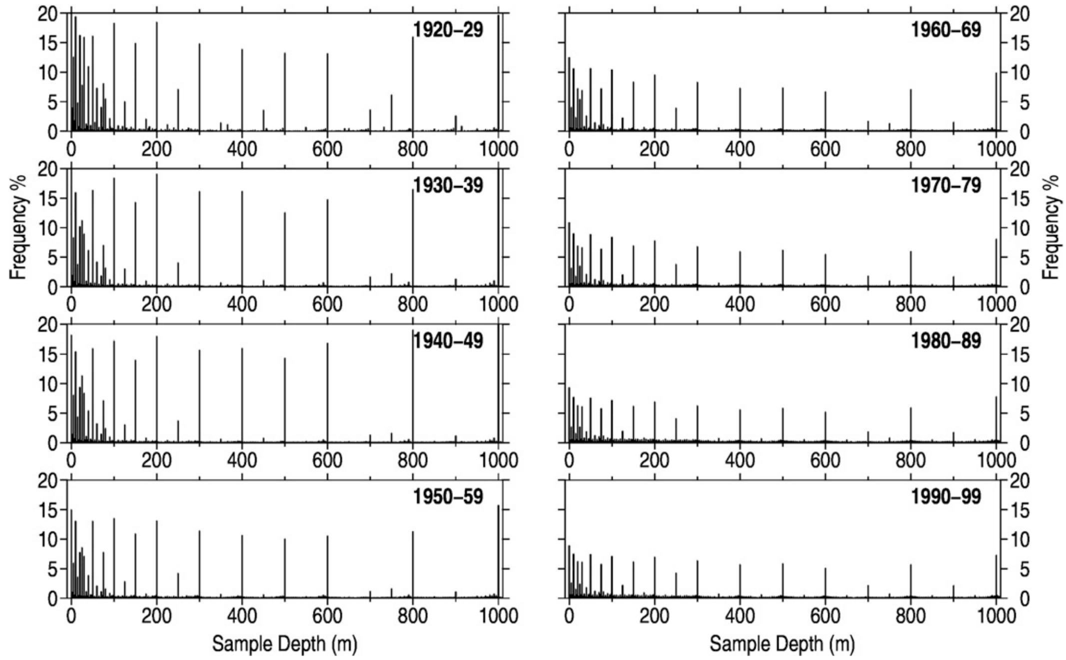

FIG. 10. Sample depth histograms for the OSD temperature profiles for the decades between 1900 and 1999.

wire put out and the angle of the wire to the vertical at the height of ship deck.

Thermometric depth measurement, e.g., the simultaneous use of paired pressure-protected and pressure-unprotected reversing thermometers attached to the same bottle, was established in the oceanographic practice since the 1920s, when several important oceanographic expeditions have been conducted (Wüst 1933; Deacon 1937; Sverdrup et al. 1944). However, details on the implementation of this method on thousands of oceanographic cruises which contributed to the WOD archive remained essentially unknown. Such parameters as wire angle and wire length, and positions of the bottles with paired thermometers were usually retained only in desk log journals, but not in the final expedition data reports. The analysis of the thermometrically measured depths showed that the wire angle typically gets smaller with depth, so that tables for estimating sample depth were developed (Kirejev 1939). In spite of the missing data on wire length and angle in hydrographic archives, it is known that 1) paired thermometers were used as a rule at the limited number of sampled levels, and that 2) the paired thermometers usually were not used at levels shallower than 100–400 m. The strategy for the bottle casts was to take samples and measure temperature at a number of prespecified depths, e.g., at so-called standard levels. Some institutions or countries followed the recommendations of the respective hydrographic manuals to select these standard levels. Typically, the spacing between the bottles (and thermometers) increases with depth in order to optimally describe the vertical structure with the limited number of sampled levels. Figure 10 shows sample level depth (z) frequency histograms for the OSD temperature profiles for selected decadal time periods. For each sample depth z_k, the frequency $f(z_k)$ is defined as $f(z_k)=100N(z_k)/N(z \geq z_k)$, where $N(z_k)$ is the number of observations at level z_k and $N(z \geq z_k)$ is the total number of observations at or below z_k.

According to the figure, the set of the most frequently chosen standard levels remained approximately the same since the beginning of the twentieth century. For the time period in question, e.g., between mid-1960s and mid-1990s, about 15 "main" standard levels between the surface and 1000 m depth can be identified. These levels have the relative frequency of greater than 5%.

Unfortunately, for the overwhelming majority of the bottle casts it is unknown whether the depths of the standard levels represent the target sample levels, the actually sampled levels, or levels for which temperature and other parameters were obtained by the vertical interpolation on standard levels. For instance, for the bottle casts of the former Soviet Union the final table of the hydrological and hydrochemical observations TGM-3M reported temperature and other parameter values on both the observed and standard levels (State Oceanographic Institute 1977).

b. *Standard level depth bias derived from the sample depth histograms and from the thermometric depth estimation method*

We used sample depth histograms to highlight the possible standard level depth bias. Figure 11a shows the histogram (red) for the time period 1920–70. To produce this histogram, 15 "main" standard levels (see section 7a) between the surface and 1000 m depth have been excluded. Finally, the histogram was smoothed with a running mean five-point filter. The peaks of this nonstandard level histogram are close to the

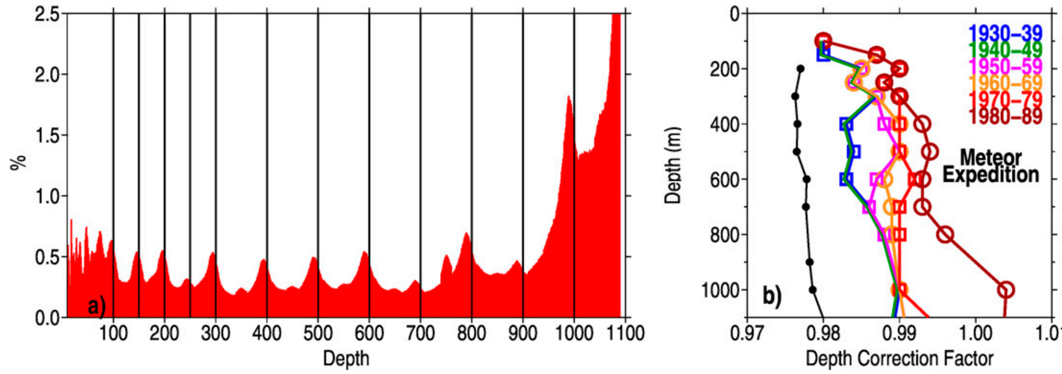

FIG. 11. (a) Sample depth histogram for the years 1920–70 with "main" standard levels (black lines) excluded. (b) Depth correction factor derived from level histograms for several decades, with the correction factor for the *Meteor* 1925–27 expedition shown in black.

respective standard levels (shown by the vertical black lines) being systematically shifted to shallower depths. We assume that the depths of the peaks correspond to the mean sample depths which were actually reached during the bottle casts with thermometrically measured depths. We further assume that sample depth targeting procedure (e.g., placing of the bottles on the wire) was the same regardless of whether the paired thermometers had been used or not. Under these assumptions the difference between the target (standard) depth and the depth of the nearest histogram peak can be interpreted as the systematic difference between the target and actually sampled depths and may be used to calculate the respective depth correction factors shown for selected decades in Fig. 11b. The calculated depth correction factors are in a broad agreement with the optimal depth correction indicating the predominant sample depth overestimation. As noted above, the thermometric method of sample depth estimation usually was not applied for bottle casts shallower than 100–200 m. The hydrographic manuals for different countries and agencies recommend different depth levels below which the paired pressure-protected and unprotected reversing thermometers should be used. Unfortunately, the WOD metadata do not provide information on which bottles were equipped with paired thermometers. However, the details on thermometric depth measurements are available in published historical cruise reports. The in-depth evaluation of the thermometric depth estimation method is given in the reports of the German Atlantic Expedition (GAE) 1925–27 (Wüst et al. 1932; Wüst 1933). The expedition reports provide data on the length of the wire put out (L) and on the actual sample depth (D), so that the correction factor $K = D/L$ can be directly calculated for the positions of the bottles equipped with paired thermometers (the paired thermometers were used below 200 m). Mean depth correction factor obtained from 1002 thermometric depth measurements during the GAE is shown in Fig. 11b. Differences between the wire length and the actual sample depth for the GAE are available for different values of wire angle to the vertical at deck height; therefore, the weighted mean correction values are shown. It should be noted that the GAE depth corrections characterize bottle depth overestimation specific to the particular ship and expedition, with the depth corrections being larger compared to the average corrections derived from the level depth histograms.

c. Bottle–CTD temperature offsets for the collocated data at standard and nonstandard sample levels

Only a small fraction of all OSD profiles (less than 5%) report data only at "main" standard levels (see sections 7b and 7c), and the majority of profiles reports data also at other depth levels. To split all OSD profiles into the standard and nonstandard level groups, the extended set of standard levels was defined as follows: all levels round to 5 m between the surface and 100 m depth, all levels round to 10 m between 100 and 200 m, and all levels round to 50 m between 200 and 1000 m. Using this extended set of 46 standard levels, we find that the yearly percentage of VLBOT profiles with data only at these standard levels decreases from almost 100% during the year 1917 to about 37% for the year 1999 (Fig. 12). The approximately 20% reduction of the standard level profile percentage is observed after the mid-1940s indicating a general change in the vertical sampling strategy or/and a more frequent use of paired thermometers. The CTDL group is characterized by the

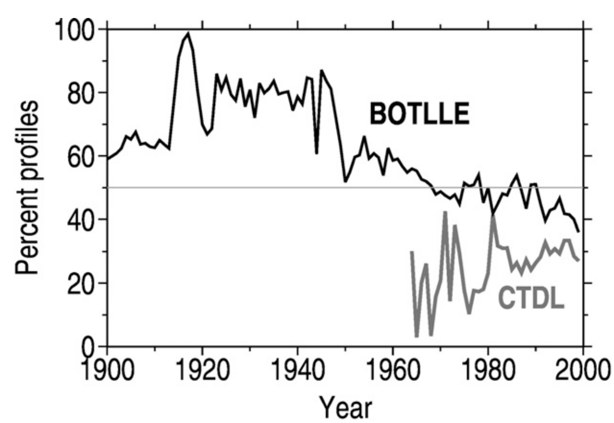

FIG. 12. Percentage of profiles from the standard level group for bottle (black) and CTDL (gray) data.

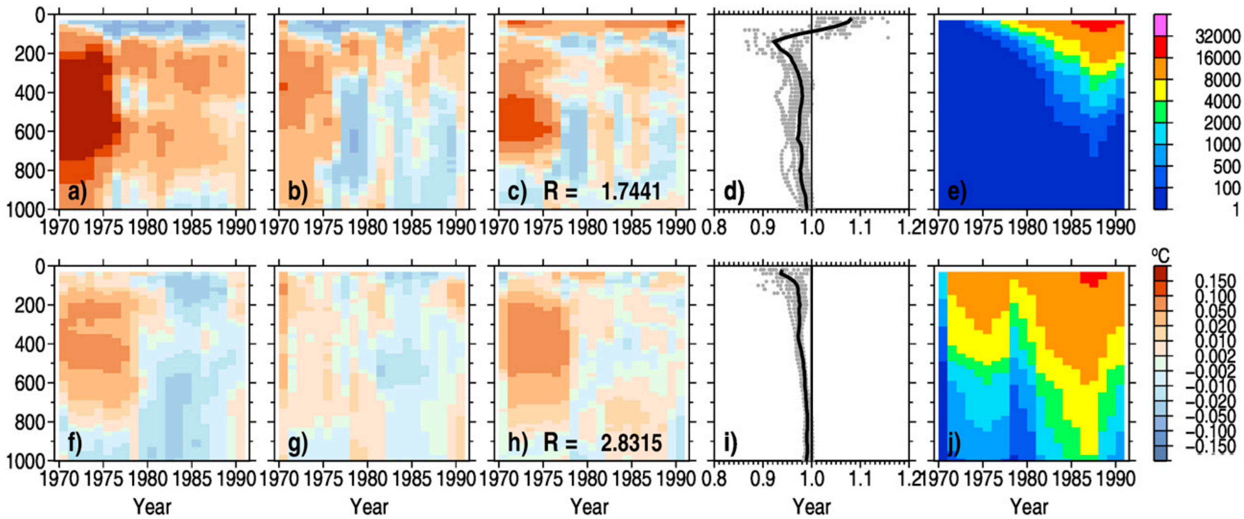

FIG. 13. (a) Bottle–CTDH temperature offset for the standard level group of bottle profiles; (b) residual offset; (c) offset reduction and the overall reduction factor R; (d) optimal yearly (gray dots) and overall (black curve) depth correction factor; (e) number of collocated pairs; (f)–(j) as in (a)–(e), but for the nonstandard level group.

approximately 10%–40% lower percentage of standard level profiles compared to bottle profiles, implying that the vertical sampling in the CTDL group is less strongly linked to sampling at standard levels.

We calculated VLBOT–CTDH temperature offsets for standard and nonstandard level groups separately (Fig. 13). Standard level profiles are characterized by a stronger positive temperature bias compared to the nonstandard level data. Respectively, the derived optimal depth corrections suggest that the standard level profiles exhibit a stronger tendency to overestimate the actual sample depth below about 100 m. However, unlike for the nonstandard level profiles, the sign of the derived optimal depth correction changes around 100 m level, indicating depth underestimation for the upper 100 m layer. To check the issue, we calculated depth corrections for several groups of the standard level casts: shallow (<100 m) casts, deep (>100 m) casts, casts north of 40°N, casts south of 40°N. In all cases the calculations resulted in positive depth corrections (depth underestimation) within the upper 100 m layer.

As pointed by one of the reviewers, the depth underestimation above 100 m for the standard level group might be due to the tendency to set the meter wheel to zero (which corresponds to the water–air interface) closer to the point on the wire where the uppermost bottle remains under the surface during the maximum downward displacements of the sea surface corresponding to wave troughs. This would contribute a uniform offset for all bottles of the cast which would be overwhelmed by the growing depth overestimation at depths beyond a threshold.

When paired thermometers are not used and no adjustments are applied for the wire angle at the deck height the depth of each bottle of a cast is given by the length of the wire paid out between the sea surface and the respective bottle. The thermometrically measured bottle depth therefore cannot be deeper than the length of the wire (assuming the error in wire length is negligible), e.g., the derived optimum depth correction factor should be less than one. We note that the derived optimal depth correction factor (Fig. 9a) depends on the magnitude of the thermal bias, which is equal to −0.02°C in these calculations. Using a stronger negative bias of about −0.3°C would result in depth correction factor values less than one. However, we are unaware of literature sources reporting such high temperature offsets for the reversing thermometers.

8. Derivation of the bottle data depth bias from the analysis of salinity collocated profiles

Though temperature observations are the focus of this study, the majority of the OSD profiles report data both for temperature and salinity, so that salinity (and any other variable) can potentially be used to estimate the depth bias using the analysis of collocated profiles. Salinity offsets between bottle and CTDH data show qualitatively similar patterns as for temperature for different ranges of the salinity vertical gradient: the stronger the gradient, the higher the offset magnitude. For negative salinity gradients, positive offset values prevail (Fig. 14b), whereas for the regions with salinity increasing with depth the calculated VLBOT − CTDH salinity offsets are negative (Fig. 14c), similar to the case of temperature inversion (cf. with Figs. 7a–d).

The application of the depth correction scheme (see section 6) to the salinity collocated dataset was not successful. The method does not work in the areas with very weak vertical gradients, where measurement uncertainties lead to false estimates of the vertical gradients. According to Fig. 14d, the ratio of the typical vertical gradient to the measurement uncertainty (~0.02°C for temperature and 0.02 on the PSS-78 scale for salinity) is an order of magnitude smaller for salinity compared to

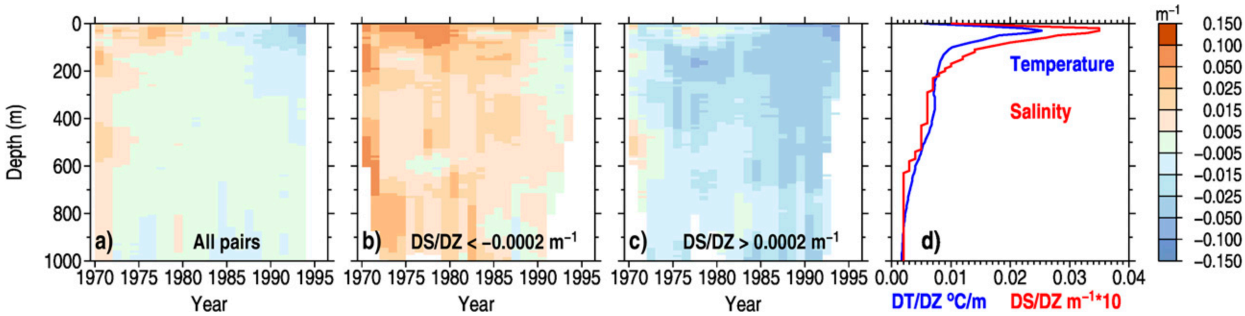

FIG. 14. Bottle–CTDH salinity offset: (a) all collocated pairs; (b) for $dS/dz < -0.0002\ \mathrm{m}^{-1}$; (c) for $dS/dz > 0.0002\ \mathrm{m}^{-1}$; (d) overall median vertical gradient for temperature (blue) and salinity (red) for the collocated bottle and CTDH data.

temperature. Therefore, we recommend using depth corrections for bottle data based on the temperature collocated dataset.

9. Impact of suggested corrections on the ocean heat content estimation

The heat energy which accumulates in the Earth climate system due to the progressing global warming is known to accumulate in the global oceans, which store more than 90% of all excessive heat energy (IPCC 2021). The OHC time series is well documented since the global implementation of the Argo program (Roemmich et al. 2015; Cheng et al. 2019; Meyssignac et al. 2019), with Argo profiling floats providing the temporal and spatial sampling within the upper 2000 m which is adequate to resolve year-to-year changes of the OHC. The estimation of the OHC for the earlier years is less accurate due to the inhomogeneous sampling and systematic errors, found both in the mechanical bathythermograph (MBT) (Gouretski and Cheng 2020) and XBT data (Gouretski and Koltermann 2007; Cheng et al. 2016), which were the main contributors to the upper ocean observing system from the 1950s to the early 2000s. Prior to the introduction of the CTD bottle casts represent the only deep reaching oceanographic instrumentation making the proper account for possible biases in bottle data especially important for the estimation of the OHC.

To illustrate the possible impact of the unaccounted biases in bottle data, we calculated the OHC anomaly time series between 1950 and 2000 for the layer 0–2000 m (Fig. 15) using the gap filling method by Cheng et al. (2017). The calculations are based on the WOD temperature profile data from all instrument types including bottle data. Two curves in Fig. 15a correspond to the noncorrected (blue) and corrected (black) bottle data, with profile data from other instruments being identical in both cases. For the corrected OHC time series we subtracted the overall thermal bias and applied the overall median depth correction factor for bottle profiles (see Fig. 9a and Table A1). To characterize the possible spread in OHC estimates due to the uncertainty in depth corrections (shaded area in Figs. 15a,b) we calculated OHC time series using minimum and maximum values of the depth correction factor, respectively (Fig. 9a). The depth correction factor was calculated only to the depth of 1000 m due to the paucity of the collocated data at deeper levels. The value of the correction factor at 1000 m was used for all levels below 1000 m. We note that below 800 m a larger spread of yearly depth correction factor estimates is observed as the consequence of the reduced number of available collocated profile pairs, with the maximum depth correction factor slightly exceeding one below 900 m level.

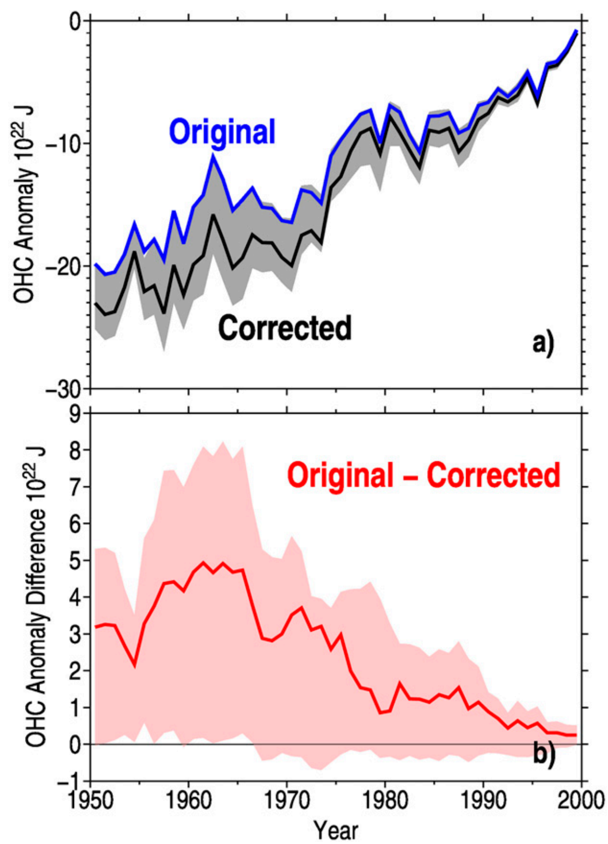

FIG. 15. (a) Yearly ocean heat content anomaly for the corrected (black) and uncorrected (blue) bottle data between 1950 and 2000. (b) Yearly difference between the original and corrected ocean heat content anomaly estimates shown in (a). Shading corresponds to the OHC estimate spread between the maximum and minimum value of the depth correction factor.

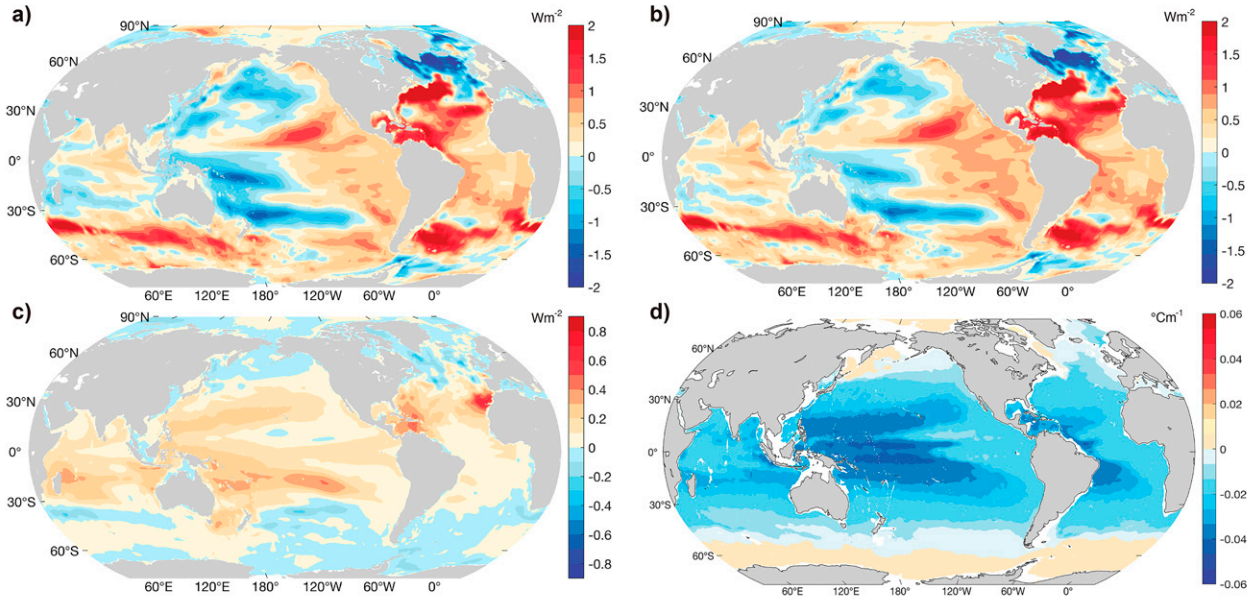

FIG. 16. Ocean heat content anomaly trend for the years 1955–90 based on WOD temperature profile data with (a) original and (b) corrected bottle data. (c) Difference between (a) and (b). (d) Mean vertical temperature gradient for the layer 100–800 m based on the WAGHC climatology (Gouretski 2018).

Because of the gradual shift from Nansen cast to CTD instrumentation the corrected and uncorrected time series become essentially identical after mid-1990s. We have calculated the OHC warming trend for the layer 0–2000 m between 1955 and 1990 using corrected and uncorrected bottle data. Applying corrections to bottle data leads to the increase of the global ocean warming trend from 0.20 ± 0.05 to 0.28 ± 0.06 W m^{-2} for the 90% confidence level.

Figures 16a and 16b show spatial distribution of the OHC trend for the layer 0–2000 m between 1955 and 1990 for uncorrected and corrected bottle data. Overall, the regions where warming trend has increased due to the application of corrections prevail over the regions where warming trend becomes weaker for corrected data (Fig. 16c). Since the magnitude of the total temperature bias depends on the thermal stratification, we produced the map of the mean vertical gradient for the layer 100–800 m (Fig. 16d) based on the WOCE–Argo Global Hydrographic Climatology (Gouretski 2018). Corrections have the most significant impact in the tropical belt, the western tropical Pacific exhibiting the largest increase. This is in agreement with the strong regional thermocline which makes the correction impact especially pronounced.

10. Efforts to improve data and metadata for better quantification of bias

Metadata identifying instrumentation (CTD or bottle), country, institute, and project associated with measurements, and measurement method (wire out versus reversing thermometer) as well as quality information about the measurements are imperative for better quantification of bottle versus CTD temperature offset. However, as shown here, this information is often not available in the WOD or in the original archived data. The International Quality Controlled Oceanographic Database (IQuOD) (Cowley et al. 2021) is an international group dedicated to providing a internationally agreed upon standardized high quality ocean temperature profile dataset for calculation of ocean heat content (IQuOD Team 2018). One aspect of the IQuOD is to provide the most complete set of metadata for ocean profile data as possible (Palmer et al. 2018). This can be achieved in the case of bottle and CTD data through scrutiny of cruise reports and scientific papers based on cruise data to augment the metadata available in the WOD.

11. Conclusions

Our collocation analysis reveals offsets between bottle cast and CTD sensor data both for temperature and salinity. Though the number of collocations is less than in the collocation analyses for XBT and MBT the offset pattern remains rather stable for different choices of the collocation bubble. Assuming the CTD data are superior in quality and precision compared to the bottle cast data, our analysis suggests the existence of the small negative pure thermal bias in historical bottle data, which we explain through the effect of premature reading of the reversing thermometers. Taking account of this pure thermal bias we find a consistent depth bias, with reported depths overestimating the actual depth. The depth overestimation amounts to 1%–3% between 100 and 1000 m. We explain the derived depth bias through the impact of the bottle casts where the thermometric method of the sample depth estimation was not used. We estimate that accounting for biases in Nansen cast data leads to the increase of the

TABLE A1. Overall depth correction factor.

Depth (m)	Minimum depth correction factor	Maximum depth correction factor	Median depth correction factor
0	1.0000	1.0000	1.0000
20	0.9280	1.0243	0.9930
40	0.9280	1.0197	0.9850
60	0.9280	1.0087	0.9767
80	0.9297	0.9997	0.9700
100	0.9337	0.9920	0.9667
120	0.9397	0.9897	0.9660
140	0.9460	0.9870	0.9667
160	0.9520	0.9843	0.9677
180	0.9573	0.9820	0.9697
200	0.9607	0.9807	0.9713
220	0.9627	0.9800	0.9733
240	0.9640	0.9800	0.9743
260	0.9660	0.9803	0.9750
280	0.9680	0.9807	0.9750
300	0.9693	0.9810	0.9750
320	0.9697	0.9810	0.9747
340	0.9690	0.9810	0.9743
360	0.9683	0.9813	0.9747
380	0.9683	0.9820	0.9757
400	0.9687	0.9833	0.9773
420	0.9693	0.9850	0.9787
440	0.9703	0.9867	0.9803
460	0.9720	0.9877	0.9817
480	0.9743	0.9880	0.9827
500	0.9760	0.9877	0.9830
520	0.9773	0.9873	0.9830
540	0.9777	0.9873	0.9833
560	0.9780	0.9877	0.9840
580	0.9780	0.9880	0.9848
600	0.9783	0.9883	0.9854
620	0.9793	0.9890	0.9857
640	0.9803	0.9900	0.9863
660	0.9810	0.9910	0.9870
680	0.9813	0.9917	0.9880
700	0.9820	0.9923	0.9887
720	0.9830	0.9930	0.9890
740	0.9840	0.9943	0.9893
760	0.9850	0.9957	0.9897
780	0.9857	0.9970	0.9900
800	0.9860	0.9980	0.9900
820	0.9857	0.9990	0.9900
840	0.9853	0.9997	0.9900
860	0.9850	1.0000	0.9900
880	0.9853	1.0003	0.9900
900	0.9857	1.0007	0.9900
920	0.9863	1.0010	0.9900
940	0.9867	1.0010	0.9900
960	0.9873	1.0013	0.9900
980	0.9873	1.0017	0.9900
1000	0.9870	1.0020	0.9900

global warming trend between 1955 and 1990 by 36%. Finally, we suggest the bottle cast profiles to be put into the separate instrumentation group within the World Ocean Database. We also note that accounting for depth and temperature biases in the historical bottle data would also require the respective updates of the biases, derived for MBTs and XBTs, because in both cases bottle cast data were used as the reference and were assumed to be bias-free.

Acknowledgments. This study is supported by Strategic Priority Research Program of the Chinese Academy of Sciences (XDB42040402), Youth Innovation Promotion Association, CAS (2020-077), the National Natural Science Foundation of China (42076202, 42122046), and the National Key Scientific and Technological Infrastructure Project "Earth System Science Numerical Simulator Facility" (Earth Lab). The funding from NOAA/Oceanic and Atmospheric Research/Global Ocean Monitoring and Observing is also acknowledged. We are thankful to the staff of the National Center for Environmental Information for preparing the invaluable collection of the hydrographic data which served as a basis for this study. We are thankful to John Gould, Walther Zenk, Alberto Piola, Steve Diggs, Igor Skvorets, Magnus Danielsen, Rebecca Cowley, Michele Fieux, Vladimir Tereschenkov, Helge Sagen, Bruce Huber, Magnus Danielsen, Shoichi Kizu, Peter Koltermann, Viacheslav Maslennikov (deceased), and Sergej Lesenkov, who shared with us their memory and knowledge about the transition from Nansen cast to CTD instrumentation in different countries and institutions. Jim Swift provided useful comments on the preliminary version of the manuscript. We thank two anonymous reviewers whose comments helped to improve the manuscript.

Data availability statement. Temperature and salinity profile data analyzed during the current study reside in the World Ocean Database (Boyer et al. 2018) and are available from the following public domain resource: https://www.ncei.noaa.gov/access/world-ocean-database-select/dbsearch.html. The overall and yearly sample depth correction for the Nansen cast profiles are available from http://www.ocean.iap.ac.cn/pages/dataService/dataService.html.

APPENDIX

Depth Correction Factor

Table A1 provides the overall depth correction factor.

REFERENCES

Boyer, T. P., and Coauthors, 2018: World Ocean Database 2018. NOAA Atlas NESDIS 87, 207 pp.

Brown, N., 1974: A precision CTD microprofiler. *Proc. IEEE Int. Conf. on Engineering in the Ocean Environment*, Newport, RI, IEEE, 270–278, https://doi.org/10.1109/OCEANS.1974.1161443.

Cheng, L., J. Zhu, R. Cowley, T. Boyer, and S. Wijffels, 2014: Time, probe type, and temperature variable bias corrections to historical expendable bathythermograph observations. *J. Atmos. Oceanic Technol.*, **31**, 1793–1825, https://doi.org/10.1175/JTECH-D-13-00197.1.

——, K. E. Trenberth, M. D. Palmer, J. Zhu, and J. P. Abraham, 2016: Observed and simulated full-depth ocean heat-content changes for 1970–2005. *Ocean Sci.*, **12**, 925–935, https://doi.org/10.5194/os-12-925-2016.

——, ——, J. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3**, e1601545, https://doi.org/10.1126/sciadv.1601545.

——, H. Luo, T. Boyer, R. Cowley, J. Abraham, V. Gouretski, F. Reseghetti, and J. Zhu, 2018: How well can we correct systematic errors in historical XBT data? *J. Atmos. Oceanic Technol.*, **35**, 1103–1125, https://doi.org/10.1175/JTECH-D-17-0122.1.

——, J. Abraham, Z. Hausfather, and K. E. Trenberth, 2019: How fast are the oceans warming? *Science*, **363**, 128–129, https://doi.org/10.1126/science.aav7619.

Cowley, R., S. Wijffels, L. Cheng, T. Boyer, and S. Kizu, 2013: Biases in expendable bathythermograph data: A new view based on historical side-by-side comparisons. *J. Atmos. Oceanic Technol.*, **30**, 1195–1225, https://doi.org/10.1175/JTECH-D-12-00127.1.

——, and Coauthors, 2021: International Quality-Controlled Ocean Database (IQuOD) v0.1: The temperature uncertainty specification. *Front. Mar. Sci.*, **8**, 689695, https://doi.org/10.3389/fmars.2021.689695.

Deacon, G. E. R., 1937: *The Hydrology of the Southern Ocean*. Cambridge University Press, 124 pp.

De Visser, A. M., and R. T. Hodgson, 1983: Equilibrium times for reversing thermometers. *J. Geophys. Res.*, **88**, 6077–6080, https://doi.org/10.1029/JC088iC10p06077.

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

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

——, and F. Reseghetti, 2010: 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. I*, **57**, 812–833, https://doi.org/10.1016/j.dsr.2010.03.011.

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

Hamon, B. V., 1955: A temperature-salinity-depth recorder. *ICES J. Mar. Sci.*, **21**, 72–73, https://doi.org/10.1093/icesjms/21.1.72.

——, and N. L. Brown, 1958: A temperature-chlorinity-depth recorder for use at sea. *J. Sci. Instrum.*, **35**, 452–458, https://doi.org/10.1088/0950-7671/35/12/307.

Helland-Hansen, B., and F. Nansen, 1909: *The Norwegian Sea: Its Physical Oceanography Based upon the Norwegian Researches 1900–1904*. Vol. II. Det Mallingske Bogtrykkeri, 390 pp.

IPCC, 2021: *Climate Change 2021: The Physical Science Basis*. Cambridge University Press, 2409 pp.

IQuOD Team, 2018: International Quality Controlled Ocean Database (IQuOD) version 0.1—Aggregated and community quality controlled ocean profile data 1772-present. National Centers for Environmental Information, accessed 26 March 2018, https://doi.org/10.7289/v51r6nsf.

Kirejev, I. A., 1939: Tables for actual sample depth of the series of bathometer (in Russian). *Sev. Morskoj Put*, **14**, 85–99.

Meyssignac, B., and Coauthors, 2019: Measuring global ocean heat content to estimate the Earth energy imbalance. *Front. Mar. Sci.*, **6**, 482, https://doi.org/10.3389/fmars.2019.00432.

Naval Hydrographic Office, 1968: Instruction manual for obtaining oceanographic data. Naval Oceanographic Office Publ. 607, 3rd ed., 232 pp.

Negretti, H., and J. W. Zambra, 1874: II. On a new deep-sea thermometer. *Proc. Roy. Soc. London*, **22**, 238–241, https://doi.org/10.1098/rspl.1873.0034.

Palmer, M. D., T. Boyer, R. Cowley, S. Kizu, F. Reseghetti, T. Suzuki, and A. Thresher, 2018: An algorithm for classifying unknown expendable bathythermograph (XBT) instruments based on existing metadata. *J. Atmos. Oceanic Technol.*, **35**, 429–440, https://doi.org/10.1175/JTECH-D-17-0129.1.

Reiniger, R. F., and C. K. Ross, 1968: A method of interpolation with application to oceanographic data. *Deep-Sea Res. Oceanogr. Abstr.*, **15**, 185–193, https://doi.org/10.1016/0011-7471(68)90040-5.

Roemmich, D., J. Church, J. Gilson, D. Monselesan, P. Sutton, and S. Wijffels, 2015: Unabated planetary warming and its ocean structure since 2006. *Nat. Climate Change*, **5**, 240–245, https://doi.org/10.1038/nclimate2513.

Saxton, R. K., 1964: Oceanographic instruments hydrographic work at sea. Oceanographic Workshop 1964, Lamont Geological Laboratory Tech. Rep. CU-12-64, 3–19.

State Oceanographic Institute, 1977: *Handbook of Hydrological Studies in Oceans and Seas* (in Russian). Israel Program for Scientific Translations, 719 pp.

Sverdrup, H. U., J. A. Fleming, S. L. Seaton, C. C. Ennis, and H. C. Hendrix, 1944: *Observations and Results in Physical Oceanography*. Carnegie Institution of Washington, 156 pp.

Warren, B. A., 2008: Nansen-bottle stations at the Woods Hole Oceanographic Institution. *Deep-Sea Res. I*, **55**, 379–395, https://doi.org/10.1016/j.dsr.2007.10.003.

Wüst, G., 1933: Thermometric measurement of depth. *Int. Hydrogr. Rev.*, **10**, 28–49.

——, G. Böhnecke, and H. H. F. Meyer, 1932: Ozeanographische Methoden und Instrumente. *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs und Vermessungsschiff "Meteor" 1925–1927*, Vol. IV, Verlag von Walter de Gruyter and Co., 298 pp.