

Time, Probe Type, and Temperature Variable Bias Corrections to Historical Expendable Bathythermograph Observations

LIJING CHENG

International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, Beijing, China

JIANG ZHU

International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

REBECCA COWLEY

CSIRO Wealth from Oceans Flagship, and Centre for Australian Weather and Climate Research, Hobart, Tasmania, Australia

TIM BOYER

NOAA/National Oceanographic Data Center, Silver Spring, Maryland

SUSAN WIJFFELS

CSIRO Wealth from Oceans Flagship, and Centre for Australian Weather and Climate Research, Hobart, Tasmania, Australia

(Manuscript received 8 September 2013, in final form 17 February 2014)

ABSTRACT

Systematic biases in historical expendable bathythermograph (XBT) data are examined using two datasets: 4151 XBT–CTD side-by-side pairs from 1967 to 2011 and 218 653 global-scale XBT–CTD pairs (within one month and 1°) extracted from the *World Ocean Database 2009 (WOD09)* from 1966 to 2010. Using the side-by-side dataset, it was found that both the pure thermal bias and the XBT fall rate (from which the depth of observation is calculated) increase with water temperature. Correlations between the terminal velocity A and deceleration B terms of the fall-rate equation (FRE) and between A and the offset from the surface terms are obtained, with A as the dominant term in XBT fall-rate behavior. To quantify the time variation of the XBT fall-rate and pure temperature biases, global-scale XBT–CTD pairs are used. Based on the results from the two datasets, a new correction scheme for historical XBT data is proposed for nine independent probe-type groups. The scheme includes corrections for both temperature and depth records, which are all variable with calendar year, water temperature, and probe type. The results confirm those found in previous studies: a slowing in fall rate during the 1970s and 2000s and the large pure thermal biases during 1970–85. The performance of nine different correction schemes is compared. After the proposed corrections are applied to the XBT data in the *WOD09* dataset, global ocean heat content from 1967 to 2010 is reestimated.

1. Introduction

The ocean stores ~90% of the world's anthropogenic global warming energy (Church et al. 2011), so detection

of changes in ocean storage content is a key method of monitoring the global energy budget (Domingues et al. 2008; Levitus et al. 2012). Great efforts have been made to reconstruct the changes in ocean heat content (OHC) (Domingues et al. 2008; Ishii and Kimoto 2009; Levitus et al. 2009; Palmer and Haines 2009). All studies show a robust global ocean warming in the past 60 yr, but large differences exist between individual estimates (Lyman et al. 2010). Evidence indicates that many of the uncertainties originate from biases in the historical

Corresponding author address: Jiang Zhu, Institute of Atmospheric Physics, Chinese Academy of Sciences, P.O. Box 9804, Beijing 100029, China.
E-mail: jzhu@mail.iap.ac.cn; chenglij@mail.iap.ac.cn; rebecca.cowley@csiro.au; tim.boyer@noaa.gov; susan.wijffels@csiro.au

expendable bathythermograph (XBT) dataset, which dominates the past 50 yr of ocean in situ observations (Lyman et al. 2010).

Biases in XBT data have been identified since the early 1970s (Flierl and Robinson 1977). The sources of these biases originate from depth errors and pure temperature biases (Gouretski and Reseghetti 2010; Cowley et al. 2013). There are no pressure sensors in XBT probes; instead, time is converted to depth by an empirical fall-rate equation (FRE): $D = At - Bt^2$, where the values of coefficients A and B are specified by manufacturers and t is the elapsed time from when the probe hits the sea surface. The usually adopted form of FRE does not accurately describe the XBT fall rate (Cheng et al. 2011; Cowley et al. 2013; Hamon et al. 2012; Kizu and Hanawa 2002b), and there is evidence that the fall rate has changed over time (Cowley et al. 2013; DiNezio and Goni 2011; Good 2011; Gouretski 2012; Gouretski and Reseghetti 2010; Hamon et al. 2012; Ishii and Kimoto 2009; Levitus et al. 2009). Many authors calculated different fall-rate coefficients (although based only on a specific and small dataset of XBT profiles), and sometimes an additional offset term was included in the fall-rate equation.

Pure temperature biases are independent of fall rate and possibly result from thermistor errors or recording system biases (Gouretski and Reseghetti 2010; Cowley et al. 2013). The sensitivity of the XBT thermistor is $\pm 0.01^\circ\text{C}$, but the overall accuracy of an XBT recording system is $\pm 0.2^\circ\text{C}$ (provided by the manufacturer). Experimental tests both in the ocean and laboratory have indicated that an XBT system (recorder plus cables plus XBT probe) usually measures temperatures warmer than reality, indicating a positive bias (Anderson 1980; Gouretski and Reseghetti 2010; Kizu and Hanawa 2002a; Reseghetti et al. 2007; Wright and Szabados 1989), but the causes of this bias are still unknown.

Many field tests have tried to determine the size of these biases by comparing XBT profiles with collocated conductivity–temperature–depth (CTD) profiles, which are listed in detail in Gouretski and Reseghetti (2010). The results show a high cruise-to-cruise and probe-to-probe variability, which complicates their application to the global XBT dataset. Different causes of biases in XBTs are reported in literature and can be broadly classified as follows:

- mechanical (probe type, its weight and dimensions, shape, nose roughness, manufacturer, year) (Green 1984; Seaver and Kuleshov 1982)
- external (launch height, atmospheric conditions, water conditions, maybe ship speed) (Gouretski and Reseghetti 2010; Hamon et al. 2012; Kizu et al. 2005; Reseghetti et al. 2007; Reverdin et al. 2009)

- electrical (recording system, circuitry, ground connection, thermistor, wire)

Several authors suggest a decrease in fall rate in cooler waters because of increased viscosity (Gouretski and Reseghetti 2010; Kizu et al. 2008; Thadathil et al. 2002), but this effect has not been fully quantified, so it is not included in most of global correction schemes. Quantification of the change in fall rate with water temperature will be important for corrections in the Southern Hemisphere and high-latitude regions. Differences in fall rates between manufacturers are also important to quantify (Kizu et al. 2011).

In 1995, Hanawa et al. (1995, hereinafter H95) recalculated the fall rates for some XBT types based on 285 XBT–CTD pairs, and these corrections have been adopted as the fall rate by manufacturers and the Intergovernmental Oceanographic Commission (IOC) and the Integrated Global Ocean Services System (IGOSS) since then. Subsequently, Gouretski and Koltermann (2007) provided the first evidence of the time-varying feature of XBT biases, showing that H95's correction is not applicable to all time periods. Adding to the confusion, since 1995 XBT metadata have contained a mix of manufacturer FRE and H95 FRE and in some cases, no indication at all of which FRE has been used.

Attempts to remove biases from the XBT dataset for use in climate studies have been made in the last 5 years. Pseudoreference datasets have been used in some cases, where each XBT profile and its reference profile are located within a specified spatial and temporal distance, usually a grid box in space and time (Gouretski and Reseghetti 2010; Hamon et al. 2012; Ishii and Kimoto 2009; Wijffels et al. 2008). These datasets are briefly summarized below:

- Wijffels et al. (2008, hereinafter W08) propose a depth-independent but time-varying depth correction factor and separate probes into two groups: shallow and deep.
- Ishii and Kimoto (2009, hereinafter IK09) also assume that all XBT biases are due to depth error, and they model depth error as a linear equation as function of elapsed time. Their corrections are also time varying.
- Levitus et al. (2009, hereinafter L09) examine the total XBT bias, which is time and depth varying, but they do not separate the probe types.
- Gouretski and Reseghetti (2010, hereinafter GR10) separate the depth error from the pure thermal bias and suggest both are time varying. Additionally, depth error is assumed to be depth varying.
- Good (2011, hereinafter GD11) uses a different technique to estimate the depth error by comparing XBT

TABLE 1. Main characteristics of different XBT probe types (1 kt = 0.51 m s⁻¹).

	Maximum depth(m)	Ship speed (kt)	Acquisition time (s)	A (m s ⁻¹) Sippican FRE	B (m s ⁻²) Sippican FRE
T4	460	30	70.5	6.472	0.00216
T6	460	15	70.5	6.472	0.00216
T7	760	15	118.3	6.472	0.00216
DB	760	20	118.3	6.472	0.00216
T5	1830	6	290.5	6.828	0.00182
T10	200	10	32.1	6.301	0.00216

depth with ocean depth measured by the General Bathymetric Chart of the Oceans (GEBCO). The results show XBT depth biases are time varying.

- [Gouretski \(2012, hereinafter G12\)](#) extends the method proposed in [GD11](#) and models a linear depth error with depth. A similar depth error history to previous studies is found and a time-varying pure thermal bias is presented.
- [Hamon et al. \(2012, hereinafter H12\)](#) model the depth error using a linear equation with an offset and includes corrections for the pure thermal bias. Their corrections are separated into 0–200-m-averaged high and low temperature groups and deep and shallow groups, rather than by probe type/manufacturer.

Comparisons among these corrections show a robust time variation in the depth errors and the pure thermal biases. However, the uncertainties among the corrections are as large as natural intradecadal variation ([Lyman et al. 2010](#)).

Despite the extensive amount of work to define the biases in XBTs, how to obtain a reliable correction scheme for the historical XBT dataset is still debated because of a lack of understanding about why the biases exist in XBT data. In a recent study by [Cowley et al. \(2013, hereinafter CW13\)](#), a large number of side-by-side XBT–CTD comparisons (more than 4000 pairs) from the past 50 years were collected to provide a new and accurate insight into the biases. They report that the pure thermal bias is likely temperature and time varying but independent of depth, and that depth error is depth and time varying and possibly dependent on temperature. In [CW13](#), an independent method ([Cheng et al. 2011](#)) is applied to side-by-side data to calculate fall rate coefficients and pure thermal bias, implying that they are all variable with time (hereafter [CW13-CH](#)). However, since the XBT–CTD pairs used in [CW13](#) were mostly collected by researchers under ideal conditions, are the results of these ~4000 data representative of the ~2 200 000 XBT profiles globally?

In this study, we try to answer this question by examining the XBT bias in two XBT–CTD comparison datasets: the side-by-side XBT–CTD pairs (from [CW13](#)) and a global-scale XBT–CTD dataset. We introduce the two datasets and the method in [section 2](#).

The results derived from the two datasets are presented in [sections 3](#) and [4](#). In [section 5](#), the correction scheme for historical XBT data is proposed. In [section 6](#), the performances of our corrections are tested and compared with other independent corrections. In [section 7](#), global ocean heat content is reestimated after our corrections to the historical XBT data in the *World Ocean Database 2009 (WOD09)* dataset. The conclusions are presented in [section 8](#).

2. Data and methods

a. Introduction to XBT probe types

The current official XBT manufacturers are Sippican Corporation (now Lockheed Martin Corporation, hereinafter Sippican) in the United States and TSK (Tsurumi Seiki, hereafter TSK) in Japan under a Sippican license. In the 1990s, a small amount of XBT probes was independently manufactured by Sparton of Canada, using their own models but they were very similar to Sippican XBTs. Detailed XBT–CTD comparison studies indicate that the two manufacturers' XBTs (Sippican and TSK) have slightly different designs, which lead to different fall rates ([CW13](#); [Kizu et al. 2011](#)). Studies based on global-scale data similarly show a difference in the depth error and the pure thermal bias between the Sippican and TSK types ([Ishii and Kimoto 2009](#)). Furthermore, differences in bias history exist among probe types from the same manufacturer.

The XBT is a relatively easy-to-use, robust, and accurate sensor that measures the temperature of a water column, designed to be dropped from a moving ship. Unfortunately, it has no pressure sensor, so its depth is estimated by a fall-rate equation provided by Sippican, the inventor of the XBT, with experimentally estimated fall-rate coefficients valid in the World Ocean. There are different XBT types, defined by their nominal maximum depth (200 m, T10; 460 m, T4–T6; 760 m, T7–DB; and 1830 m, T5). Three other types (1000 m, FD; 460 m, T11; and 2000 m, T12) were manufactured, but their presence in the historical archives are negligible. [Table 1](#) listed the main characteristics of the different XBT

types (maximum depth, maximum allowed ship speed, maximum acquisition time, and fall-rate equation coefficients). Sometimes, there are two versions of XBT probe types for the same depth rating, one for low speed and one for faster ships. Sippican stated that T4–T6–T7–DB have the same characteristics in water, so that they have the same coefficients for the fall-rate equation. T10, FD, and T5 have different coefficients. Operational experience shows that an XBT is able to reach up to about 20% deeper than the nominal maximum depths: ~250 m for T10, ~550 m for T4–T6, ~900 m for T7–DB, ~1200 m for FD, and ~2000 m for T5 without any evidence of errors. On the other hand, wire breaks, insulation penetrations, and leakage defects often shorten the maximum depth of the profile. The thermistor samples at ~0.1s, nearly coincident with the thermistor time constant. Since the early 1970s, it has also been verified in field tests comparing XBT records with more accurate and expensive instruments (usually CTD) that the uncertainties in XBT data were beyond the values stated by the manufacturer ($\pm 0.2^\circ\text{C}$ in temperature and 2% or 5 m in depth, whichever is greater), whereas the stated instrumental resolution is 0.01°C in temperature and about 0.65 m in depth.

From a technical point of view, XBT–CTD comparison tests do not represent typical operational launching conditions. During an XBT–CTD test, the ship is required to be motionless, so wake and air turbulence due to hull motion and water turbulence in the near surface are absent. Most XBT–CTD pairs in the side-by-side dataset were performed from a stationary ship, and the height of the launch platform and test calibration details are frequently not included in reports/papers describing the comparison. These details could provide more insight into the reason for the offset term.

b. Side-by-side XBT–CTD pairs

In the side-by-side XBT–CTD pairs dataset, each XBT profile is collocated in space and time with a CTD profile. This data source includes historical XBT–CTD experiments and data from the *WOD09*. The dataset analyzed in this paper is the same as that used in *CW13*. The dataset contains 4115 pairs, including 2096 high-resolution pairs (with a depth interval of less than 1 m for XBT, and less than 2 m for CTD), and 2019 low-resolution pairs ($>1\text{-m}$ depth intervals in XBT and $>2\text{-m}$ depth intervals in CTD or bottle). This side-by-side dataset is available via the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Data Access Portal (doi:10.4225/08/52AE99A4663B1).

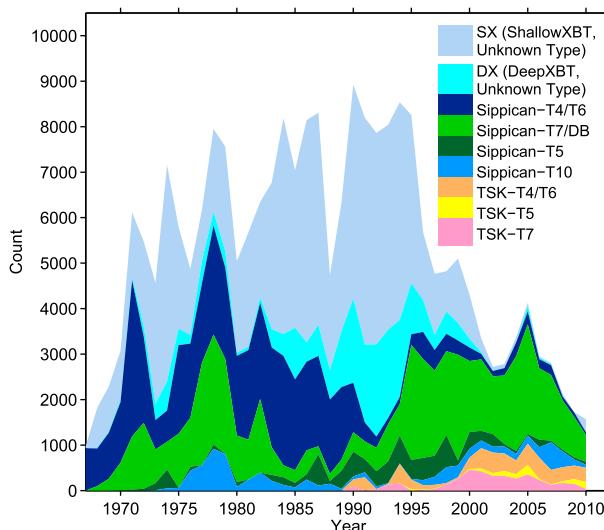


FIG. 1. Temporal distribution of XBT profiles of different probe types in the global-scale dataset. Curves of different groups are stacked.

c. Global-scale XBT–CTD pairs

Global-scale XBT–CTD pairs were sourced from *WOD09* (Boyer et al. 2009) using the following criteria:

- The locations of the XBT and reference CTD profiles are within 1° .
- The times of the XBT and CTD profiles are within 30 days.
- There are 218 653 individual XBT profiles. Each XBT profile is used only once and a reference CTD profile is constructed from the CTD/bottle data within the distance/time range specified above. The method used to obtain the reference CTD profile is:
- For each XBT profile, n CTD profiles are located within the distance/time range.
- The n CTD profiles are interpolated to standard depths: from 1 to 800 m with a 5-m interval. These CTD profiles are denoted as CT_1, CT_2, \dots, CT_n . The distances between each CTD profile and the XBT profile are calculated and denoted as D_1, D_2, \dots, D_n .
- The CTD-buddy profile is calculated by weighting all of the CTD profiles according to their distance from the XBT:

$$CT_b = (CT_1/D_1 + CT_2/D_2 + \dots + CT_n/D_n) / (1/D_1 + 1/D_2 + \dots + 1/D_n),$$

where CT_b is the resultant CTD reference profile.

In the global-scale pairs dataset, pre-1990, shallow XBTs (T4–T6–T10) make up nearly half of the total

TABLE 2. Number of XBTs split by probe type in the WOD09 global pairs dataset.

Profile	Sippican					TSK					Overall total	
	G1-T7-DB	G3-T4-T6	G5-T10	G6-T5	Total	G7-T4-T6	G8-T5	G9-T7-DB	Total	G4-SX		G2-DX
Count	50 001	48 569	8 198	3 586	110 354	5 342	812	3 724	9 878	82 678	15 743	218 653

dataset (Fig. 1; Table 2). T7s (deep XBT) make up the other half. XBTs with unknown type dominate during 1983–95. Deep XBTs (T5–T7–DB–FD) are the main part of the dataset after 1992. TSK XBTs become more common in the dataset after 1988; the amount of TSK XBTs is consistent in time from 1995 to 2010.

According to the manufacturers, the T4 and T6 probes are structurally the same (the same applies to T7 and DB), the only difference being the length of the wire on the canister spool. There is no significant difference in the fall-rate error or the pure thermal bias between the T7 and DB or the T4 and T6 probe types (in CW13). However, there is a statistically significant difference between T7–DB and T4–T6 (CW13), the TSK and Sippican probes, T10, and other probe types. The unknown type makes up the majority of the dataset and will have a different behavior because of the mix of probe types it contains. Keeping these differences in mind, XBT profiles are collected into nine groups: G1 includes Sippican T7–DB; G2 contains unknown-type profiles with a maximum depth deeper than 550 m post-1985 (540 m pre-1985), denoted as DX; G3 includes Sippican T4–T6; G4 includes all of the unknown-type profiles with a maximum depth shallower than 540 m pre-1985 (550 m post-1985), denoted as SX; G5 includes Sippican T10; G6 includes Sippican T5, since the terminal depth is as deep as 1800 m, which is distinct from the other probes; G7 includes the T4–T6 probes manufactured by TSK; G8 includes the TSK T5; and G9 includes the TSK T7–DB.

d. Method to calculate fall-rate coefficients and pure thermal bias

The method proposed in Cheng et al. (2011, herein-after CH11) is used in this paper to calculate the fall-rate coefficients from the side-by-side dataset and the global-scale dataset. The pure thermal bias is determined after the calculated fall-rate coefficients are applied. The method is also used in CW13 as an independent method to estimate the fall-rate coefficients and the pure thermal biases in the side-by-side dataset. We summarize the main steps of this method:

- The fall-rate equation suggested by the manufacturers is $D = At - Bt^2$, which is revised to $D = At - Bt^2 - \text{offset}$, where A and B are the terminal fall rate and deceleration term, respectively; offset “is a correction describing the up-to-now unpredictable phenomena occurring

mainly in the near-surface layer, at the start up, with a deviation from the manufacturer’s equation” (CH11, p. 247).

- The individual fall-rate equation is obtained by minimizing the standard deviation of the temperature differences between the XBT and CTD profiles. This process helps to isolate the depth error from the pure thermal bias—that is because the standard deviation of the temperature difference is zero when a XBT temperature curve is parallel to the reference CTD profile.
- The residuals after correcting an XBT profile by using the individual fall-rate equation are regarded as the pure thermal bias.
- The final A , B , and offset coefficients for the whole XBT dataset are obtained by calculating the weighted median of these coefficients from each individual XBT–CTD comparison, and the weights are the depths of each individual XBT profile. This weighting gives more importance to deep profiles.

3. Results from the side-by-side dataset

a. Temperature’s influence on pure thermal bias

CW13 found a small but increasing trend of the pure thermal bias with column-averaged temperature by using side-by-side data. Though the relationship between the water temperature and the pure thermal bias is robust and is suggested by several independent studies (Gouretski and Reseghetti 2010; Reseghetti et al. 2007; Reverdin et al. 2009), CW13 finds that the temperature-varying thermal error plays a minor role in determining the time variation of the biases. Similar to CW13, we compared the ΔT at each 1-m interpolated depth with the CTD temperature at that depth, high-resolution T7–DB data within 1993–2005, and all of the T4–T6 data within 1982–93 are used according to Fig. 14 in CW13. The relationship (the function of “Robustfit” in MATLAB is used to fit a linear model) is described as

$$\begin{aligned} \text{T7 - DB: } dT_{\text{temp}} &= 0.00140(\pm 0.0005)\text{Temp}_{\text{column}} \\ &+ 0.0139(\pm 0.0056)^{\circ}\text{C}, \end{aligned} \quad (1)$$

$$\begin{aligned} \text{T7 - T6: } dT_{\text{temp}} &= 0.00167(\pm 0.0014)\text{Temp}_{\text{column}} \\ &+ 0.0115(\pm 0.0202)^{\circ}\text{C}, \end{aligned} \quad (2)$$

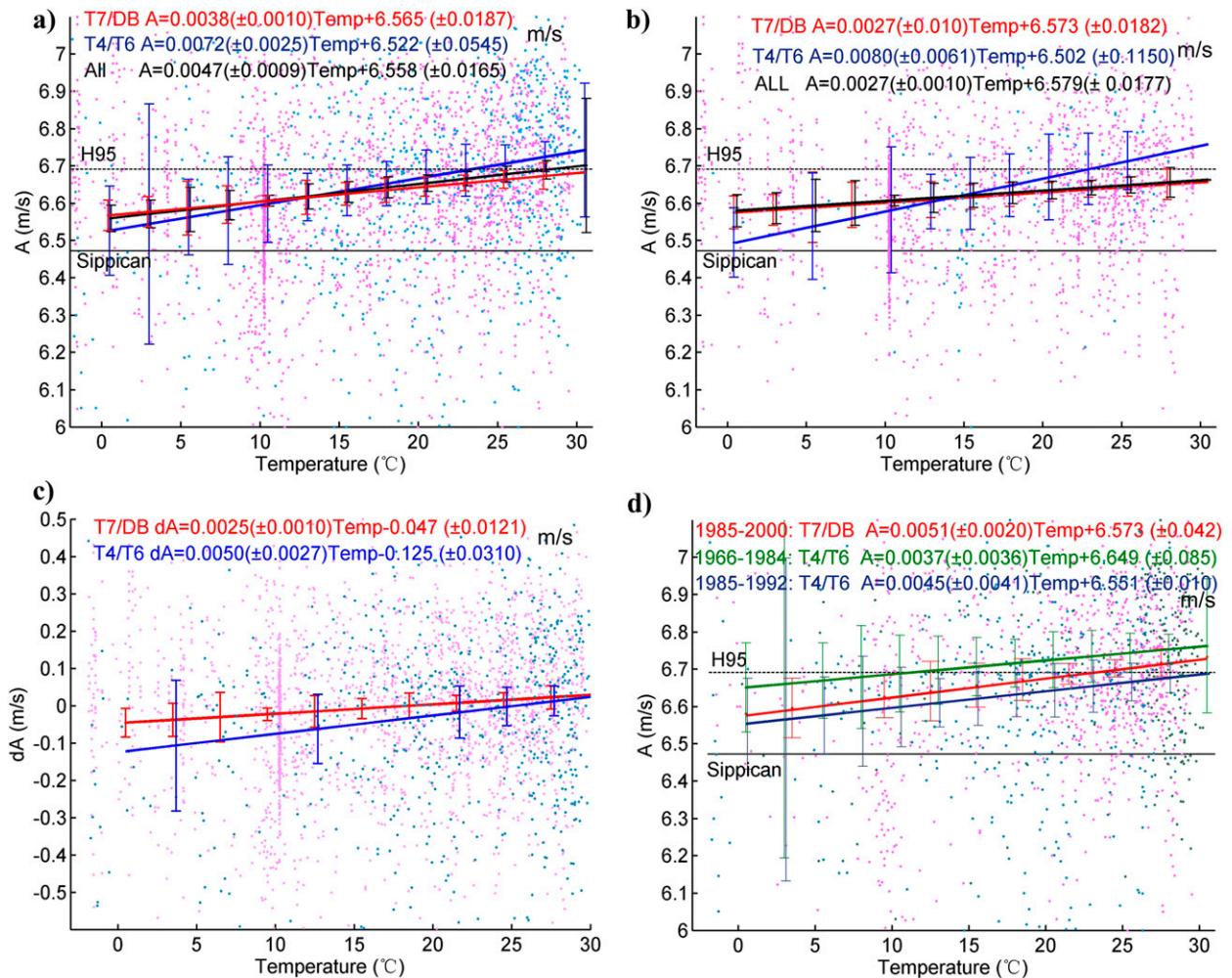


FIG. 2. Average fall rate coefficient A calculated over 0–100 m for T4–T6 in light blue dots and T7–DB in pink dots. Linear regression is obtained by using Robustfit: blue line for T4–T6, red line for T7–DB, and black lines for the whole dataset. The error bars are $2\times$ the standard error. Obtained by using (a) the whole dataset, (b) high resolution data, (c) dA after removing time variation, and (d) data from period 1985–2000 for T7–DB, and 1966–84 and 1985–92 for T4–T6.

where $\text{Temp}_{\text{column}}$ is the column-averaged temperature over all depths and the 95% confidence intervals are in parentheses. The trend for T7–DB is more robust than that for T4–T6, which suggests near-90% uncertainties. The statistical p value of the trend is 0.11×10^{-10} for T4–T6 and 0.21×10^{-11} for T7–DB, resulting in a rejection of null hypothesis (that there is no significant trend between the pure thermal bias and temperature) at the 99% confidence level. The results are in general agreement with the results from XBT calibrations in a temperature bath, quoted in Reseghetti et al. (2007), Reverdin et al. (2009), GR10, and CW13. This temperature dependence had been ascribed to both temperature-resistance characteristics of the XBT thermistor and the wire length (Reseghetti et al. 2007; Seaver and Kuleshov 1982).

b. Temperature's influence on depth error

Temperature's influence on the XBT fall rate has been documented in several papers (Gouretski and Reseghetti 2010; Kizu et al. 2005; Thadathil et al. 2002), suggesting an increasing fall rate with water temperature due to decreasing water viscosity. However, an explicit relationship has not been described with the exclusion of Kizu et al. (2008) even for expendable CTD (XCTD) probes. In this paper, we try to quantify the temperature's influences on the terminal value of the fall rate, A , which is derived by fitting coefficient A over the 0–100-m integrated water temperature. The 0–100-m depths are chosen instead of terminal depth or sea surface temperature because (i) the terminal depths of XBTs in the side-by-side pairs are not exactly the same, varying

from 100 to 1000 m, depending on probe type and maximum depth attained. Using column-averaged temperature may give results dominated by the variation of the terminal depth. (ii) The terminal velocity is most likely affected by the upper-ocean (i.e., 100 m) temperature rather than the deeper ocean temperature. (iii) Terminal velocity (described by coefficient A) is reached within 0–100 m.

The relationship between temperature and the averaged coefficient A calculated over 0–100 m for the whole side-by-side dataset (Fig. 2a) shows a slightly increasing trend for both T7–DB (2220 high-resolution pairs used) and T4–T6 (684 high-resolution pairs) with a slope of $0.0038 (\pm 0.0010)$ and $0.0072 (\pm 0.0025) \text{ m s}^{-1} (\text{C}^\circ)^{-1}$, respectively, but the high-resolution data (Fig. 2b) have slightly different slopes: $0.0027 (\pm 0.0010) \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T7–DB (1569 high-resolution pairs used) and $0.0080 (\pm 0.0061) \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T4–T6 (only 89 high-resolution pairs, unfortunately). The difference is so evident probably because we have removed data from the early period: most of the early period has low-resolution data and the remaining high-resolution data show a slope more representative of recent years.

Since the coefficient A has shown temporal variability, to fully separate the time variation from changes in A with temperature, we subtract the yearly averaged A (in Tables 3–5 in CW13) from each individual A . The relationship of the residual (dA) and temperature is presented in Fig. 2c, showing an increasing trend with temperature with slopes of $0.0025 (\pm 0.0010)$ and $0.0050 (\pm 0.0027) \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T7–DB and T4–T6, respectively, $\sim 30\%$ smaller than that in Fig. 2a. In this scenario, we assume that the time variation has been totally removed.

Using data within a period of almost constant fall rate and containing a large amount of data from high latitudes (CW13) (1985–2000 for T7–DB, 1985–92 for T4–T6) (Fig. 2d), coefficient A increases with temperature, with a slope of $0.0051 (\pm 0.0020) \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T7–DB, $\sim 20\%$ more than that in Fig. 2a; and $0.0045 (\pm 0.0041) \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T4–T6, similar to Fig. 2a. A smaller slope [$0.0037 \pm 0.0036 \text{ m s}^{-1} (\text{C}^\circ)^{-1}$] is found when 1966–84 data are used for T4–T6.

Based on the four tests shown above, the fall-rate coefficient A increases with temperature from 0.0025 to $0.0051 \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T7–DB and from 0.005 to $0.008 \text{ m s}^{-1} (\text{C}^\circ)^{-1}$ for T4–T6 when different subsets of the side-by-side dataset are used.

In this paper, we use the relationship in Fig. 2c, where we remove the time variation in A as a first guess of the A –temperature relationship. In short, the obtained relationship for T4–T6 and T7–DB is, with 95% confidence interval, as follows:

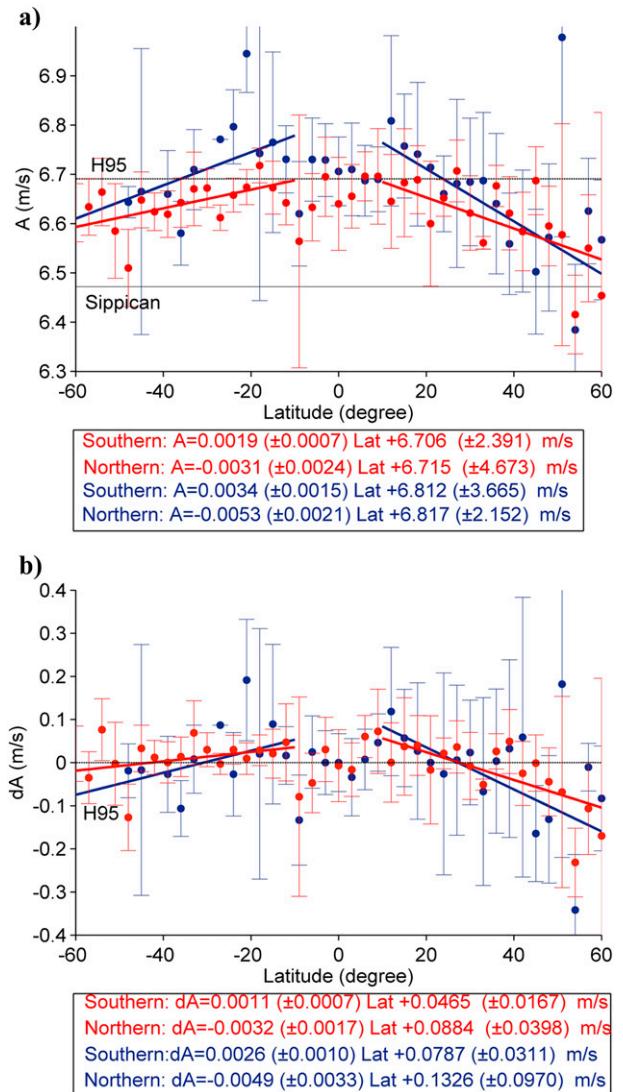


FIG. 3. (a) Fall rate coefficient A as function of latitude for T4–T6 in blue and T7–DB in red. Colored lines are the results of Robustfit of the terminal velocity (A) with latitude from -60° to -10° in the Southern Hemisphere and from 60° to 10° in the Northern Hemisphere. Error bars are $2\times$ standard errors. (b) Zonal-averaged dA when time variation of the coefficient A is removed.

$$\text{T4 - T6: } dA_{\text{temp}} = 0.0050(\pm 0.0027)\text{Temp}_{100} \quad (\text{m s}^{-1}), \tag{3}$$

$$\text{T7 - DB: } dA_{\text{temp}} = 0.0025(\pm 0.0010)\text{Temp}_{100} \quad (\text{m s}^{-1}). \tag{4}$$

The statistical p value of the trend is 0.246×10^{-3} for T4–T6 and 0.58×10^{-5} for T7–DB, resulting in a rejection of the null hypothesis (there is no significant trend between dA_{temp} and Temp_{100}), suggesting the two trends are robust. The different temperature dependence of the

T4–T6 and T7–DB probes are possibly due to the small sample size and the low resolution of the T4–T6 sample. Physically, the two types of probes are identical when submerged in water, but they have different weight distributions: for T7–DB there is more wire in the upper part of the probe and less weight for the zinc nose.

Since a relationship between fall rate and water temperature is apparent, the fall rate may be variable with latitude as well. Despite some large uncertainties at some latitudes, the terminal fall rate appears to be lower at high latitudes and higher in the subtropics in both hemispheres (Fig. 3a). However, in the tropics (from -10° to $+10^\circ$), the terminal fall rate seems to be constant, approximating the H95 FRE for both T7–DB and T4–T6. From 60° to 10° S and from 60° to 10° N, coefficient A increases from high to low latitudes for both T7–DB and T4–T6.

In Fig. 3b, the temporal variability of A is removed. When we compare the four zonal slopes of dA after removing the time variation with the slope of A (in Fig. 3a), the slope of the fall rate is reduced by 4%–40%. The largest reduction exists for T7–DB data in the Southern Hemisphere, where the slope is reduced from $0.0019 (\pm 0.0007)$ to $0.0011 (\pm 0.0007)^\circ\text{C} (\text{C}^\circ)^{-1}$, around 40%. This test indicates that the time variation feature of A may only partly reduce the zonal variation, since the removal of the time variation only partly reduces the zonal variation of the total bias.

c. Correlations between fall-rate coefficients

In this section we examine the possible correlations among the three coefficients in our proposed fall-rate equation. We assume A to be the main term and investigate representing the B and offset terms as a function of A and investigate whether they are linearly dependent on the speed A . Using this approach, the speed A is supposed to be able to allow us to estimate both B , representing the variation of the probe mass and ambient conditions, and offset, linked to the strength of the initial transient time and the deployment conditions.

The correlations between coefficients B and A are obtained by fitting the data using Robustfit in MATLAB for T4°C T6, T7°C DB, and TSK (all types from the side-by-side dataset) separately. We looked for a linear relationship between A and B (such as $B = a_1A + a_2$) and we found (Fig. 4a) the following:

$$\begin{aligned} \text{T4 - T6: } B &= 0.0069(\pm 0.0007)A \\ &\quad - 0.0435(\pm 0.0045) \quad (\text{m s}^{-2}), \quad (5) \end{aligned}$$

$$\begin{aligned} \text{T7 - DB: } B &= 0.0070(\pm 0.0002)A \\ &\quad - 0.0440(\pm 0.0016) \quad (\text{m s}^{-2}), \quad (6) \end{aligned}$$

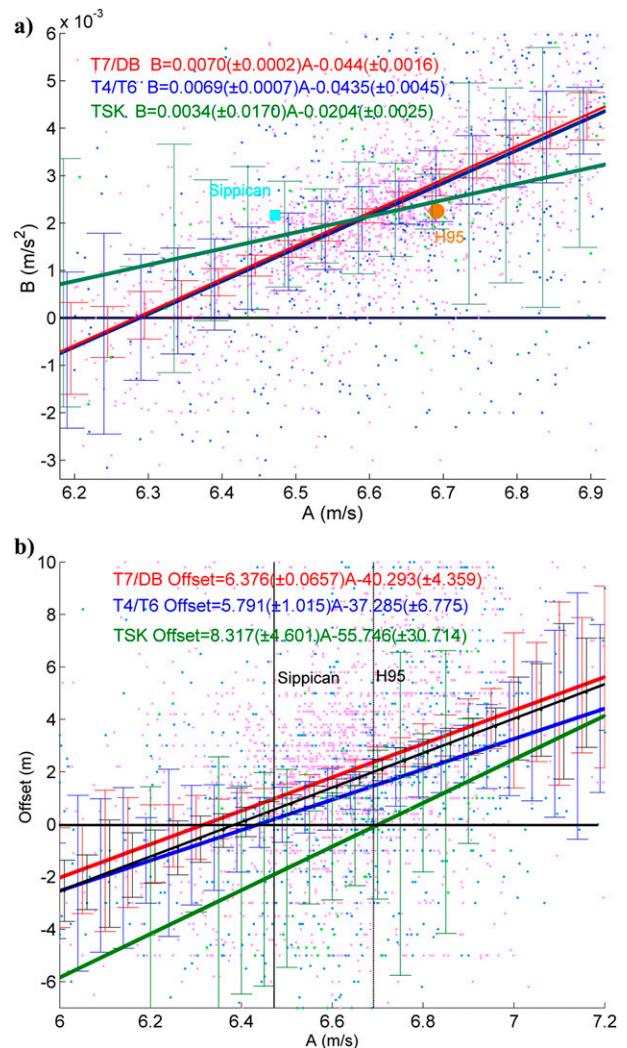


FIG. 4. (a) Correlation between A (terminal fall rate) and B (falling deceleration) for T4–T6 (blue) and for T7–DB (red), TSK (green), and all data (black). (b) Correlation between fall-rate coefficient A and surface offset for T4–T6 (blue), T7–DB (red), TSK (green). Colored dots are the individual coefficients for T7–DB (pink), T4–T6 (light blue), and TSK (green). Correlation is obtained using Robustfit with a 95% confidence interval and error bars are $2\times$ the standard error.

$$\begin{aligned} \text{TSK: } B &= 0.0034(\pm 0.0017)A \\ &\quad - 0.0204(\pm 0.0025) \quad (\text{m s}^{-2}), \quad (7) \end{aligned}$$

where the 95% confidence interval is given in parentheses. The statistical p values for the three trends are 0.274×10^{-7} , 0.386×10^{-11} , 0.47×10^{-2} for T4–T6, T7–DB, and TSK, respectively, indicating the null hypothesis (there is no significant trend) is rejected. The results for T7–DB have a smaller range of uncertainties (smaller p value) than for T4–T6 because of a larger number of T7–DB profiles. The correlations

for TSK are completely different from the Sippican probes, and all the error bars overlap, indicating that the statistical difference of the results for TSK is less robust.

Then, we looked for a linear correlation (offset = $b_1A + b_2$) for the T7-DB, T4-T6, and TSK probe types. The results are (with the 95% confidence interval) (Fig. 4b):

$$\begin{aligned} \text{T4 - T6: offset} &= 5.791(\pm 1.015)A \\ &\quad - 37.285(\pm 6.775) \text{ (m)}, \end{aligned} \quad (8)$$

$$\begin{aligned} \text{T7 - DB: offset} &= 6.376(\pm 0.657)A \\ &\quad - 40.293(\pm 4.359) \text{ (m)}, \end{aligned} \quad (9)$$

$$\begin{aligned} \text{TSK: offset} &= 8.317(\pm 4.601)A \\ &\quad - 55.746(\pm 30.714) \text{ (m)}. \end{aligned} \quad (10)$$

The statistical p values for the three trends are 0.345×10^{-7} , 0.380×10^{-11} , 0.0101 for T4-T6, T7-DB and TSK respectively, indicating the rejection of the null hypothesis test. The correlation for TSK is different from T4-T6 and T7-DB.

Why are the coefficients correlated? The three coefficients are physically independent of each other, because they are used to independently model the XBT falling motion. However, the results show they are significantly correlated with each other. We discuss the possible reasons for these correlations in appendix A and the impact of the offset term on the A - B correlation in appendix B. A recent study based on a numerical simulation of the XBT probe entry indicates that the velocity term is probably dominant, and it seems to influence both the deceleration and offset terms (Abraham et al. 2014). By using the correlations, we make an assumption about the XBT bias: the correlations are an intrinsic feature of the fall-rate equation. The correlations shown above are used as the first guess to apply to the global-scale dataset.

d. Test of the correlations using the side-by-side dataset

To test whether the XBT FRE is dominated by term A and if a new version of FRE in terms of A is possible, we first check the side-by-side data using the specific A coefficients and then substitute A into the B and offset terms in Eqs. (5)–(7) and (8)–(10), respectively. After this step, the residual thermal bias is compared with the total thermal bias and the pure thermal bias (calculated as described in CW13) in Fig. 5. Biases for both the T4-T6 and T7-DB probe types are reduced to an amount comparable to the pure thermal bias, suggesting that the correlation works well. The correlation method is not as effective during the

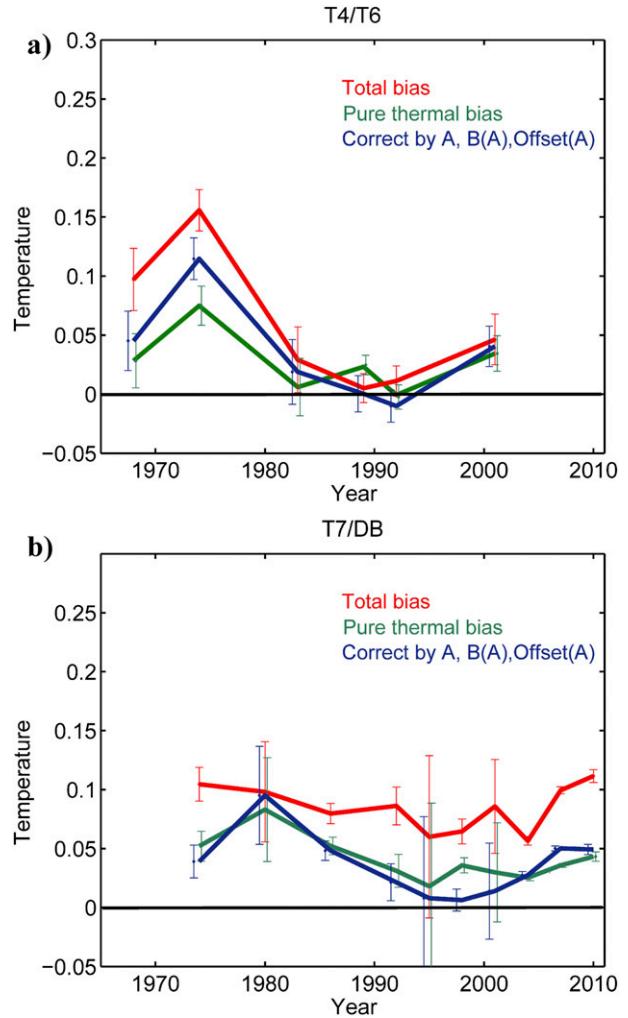


FIG. 5. Time-varying total thermal bias in the side-by-side dataset as a function of time (red); pure thermal bias (depth error removed, green). Thermal bias after corrections using the individual A of each pair, along with B and offset, which are calculated from the $B(A)$ and offset(A) correlations (blue). Error bars are $2 \times$ the standard errors: (a) T4-T6 and (b) T7-DB.

years around 1975 for T4-T6 when the correlation method reduces the thermal bias by about half compared to the pure thermal bias. Since the number of pre-1980 pairs in the side-by-side dataset is small, large uncertainties are expected. These results reasonably suggest that A is the dominant coefficient in determining XBT depth. Therefore, we apply these correlations to the global-scale analyses, trying to correct depth in a similar way.

Because of the small size of the side-by-side dataset, the spatial and temporal distribution is limited and the errors obtained using the correlation method are large in the early time period. To try to reduce these errors and to get more information on the XBT bias history, we examine the biases in the global-scale XBT dataset.

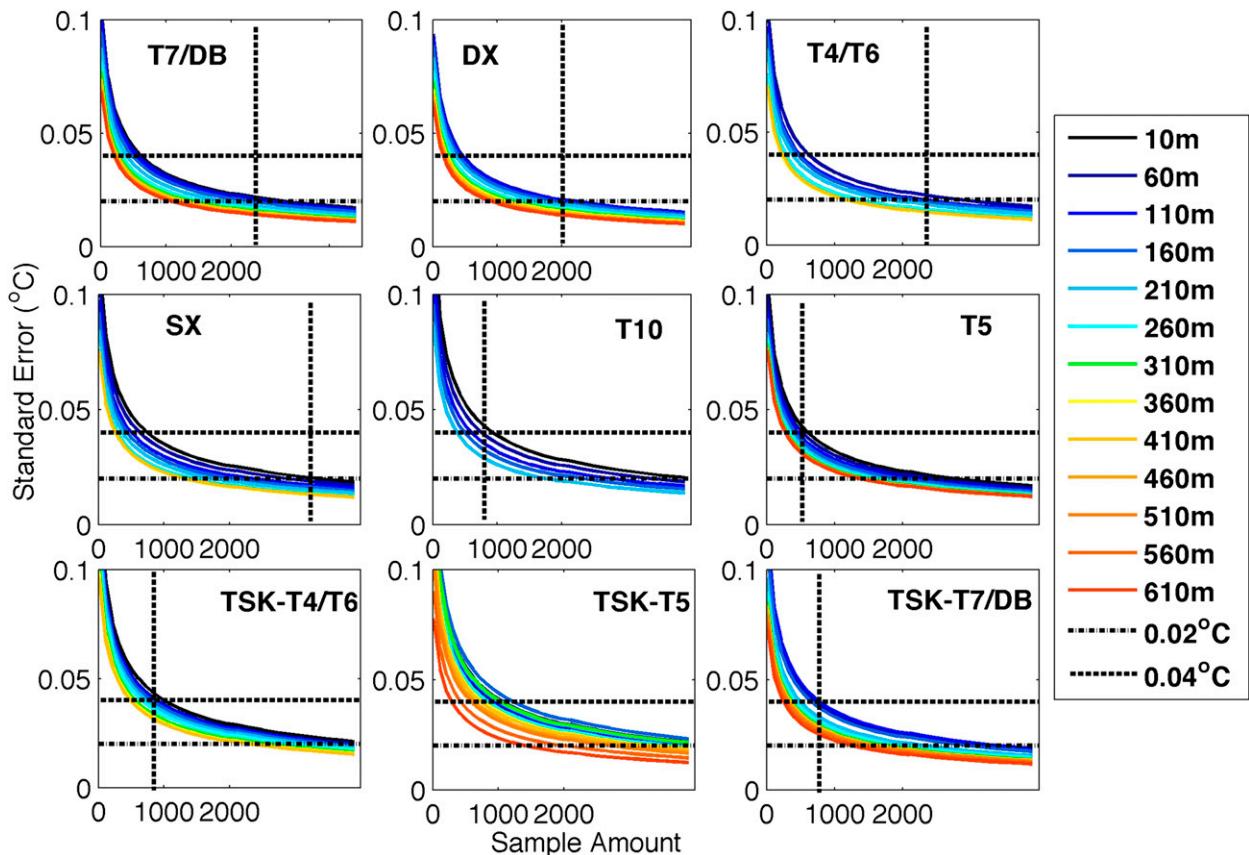


FIG. 6. Bootstrap tests. The standard error as a function of sampling sizes from 50 to 3800 is presented for the nine groups, respectively. Colors represent the standard errors at depths from 10 to 610 m in increments of 50 m. Vertical dashed line marks how many pairs are used in this study in each time period; TSK-T5 is not presented because its data amounts are too small. The standard errors of 0.02°C and 0.04°C are presented as horizontal dashed lines.

4. Results from global-scale pairs

a. Simplified FRE to model the XBT depth in the global-scale pairs dataset

Since the global-scale pairs are not true XBT-CTD comparisons, to achieve computational efficiency and to avoid uncertainties introduced by using four variables (three fall-rate coefficients and the pure thermal bias), the fall-rate equation is simplified from the equations in section 3c to

$$D = At - (a_1A + a_2)t^2 - (b_1A + b_2). \quad (11)$$

For the global-scale dataset, Eq. (11) is used to model the XBT depth (D). The values for the constants a_1 , a_2 , b_1 , and b_2 are given in section 3c for each probe type and coefficient A is calculated for each XBT-CTD pair in the global-scale dataset.

b. Bootstrap analyses to determine the length of each temporal bin

A bootstrap test is needed to determine the length of each temporal bin to obtain statistically significant

results, since the global-scale pairs are not evenly distributed over time and small numbers of pairs in some time periods may lead to spurious results.

The depth of each XBT profile is calculated using Eq. (11), leaving only the pure thermal bias (individual corrections). Next, a bootstrap analysis is conducted on the temperature differences after individual corrections. For each group, we randomly select n pairs (from 50 to 3750 in increments of 100) at a specific depth (from 5 m to the terminal depth in increments of 50 m). The median of the temperature differences is then calculated. This procedure is repeated 200 times and 200 medians are obtained. The standard error of these medians is calculated at each depth for each n and each group (Fig. 6).

The pure thermal biases have larger uncertainties in the upper ocean for all the groups, where the water properties change with space and time much more than in the deeper ocean. For G1-T7-DB, G2-DX, G3-T4-T6, and G4-SX, the standard errors for all depths, except 10 m, decrease to 0.02°C (2 times the XBT thermistor sensitivity) when the number of pairs is more

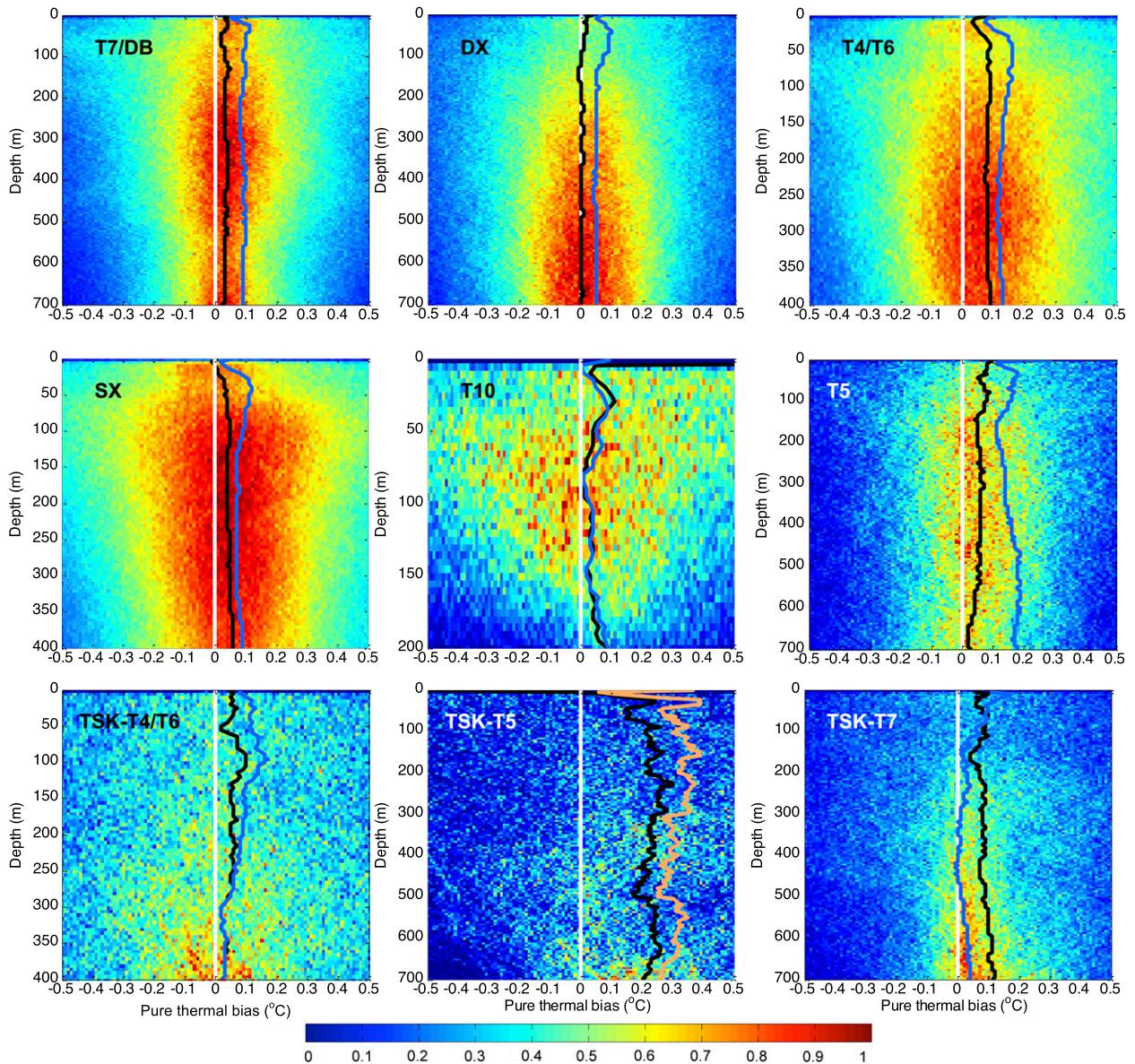


FIG. 7. Frequencies of pure thermal biases after individual corrections for the nine groups. Temperature differences are first counted in each 0.01°C (x axis) and 5-m -depth (y axis) grid box and then all of the frequencies are normalized by the maximum frequency. Medians of the pure thermal bias at depths before (blue solid line, with exception of TSK-T5 in orange solid) and after (black solid line) corrections are shown.

than a threshold of $\sim(2200, 2000, 2200, \text{ and } 3000, \text{ respectively})$. There are larger errors for G5-T10 than the other Sippican groups, and the standard errors for all depths (except 10 m) decrease to 0.04°C (4 times the XBT thermistor sensitivity) when there are more than 800 pairs. The standard errors for T5 (except 10 m) are less than 0.04°C when there are more than 700 pairs. For the TSK pairs, 900 and 850 pairs are needed for TSK-T4-T6 and G9-TSK-T7-DB, respectively, to decrease the standard errors to 0.04°C . However, for G8-TSK-T5,

there are only ~ 800 pairs in total, so we do not specify a threshold for this group.

Based on these results, we select time windows to diagnose the time variation of the depth and the pure thermal bias. Initially, 5-yr bins centered on each year are used from 1967 to 2010. Most XBTs delivered from Sippican-TSK are probably used in about 2 years, so a 5-yr window could easily include numerous distinct batches, with distinct fall rates. Additionally, in each 5-yr bin, if the number of pairs is less than the threshold shown above, then the length of the bin is extended until

the number of pairs is equal or greater than the threshold. The maximum length of each window is 10 years. We used 7-yr bins for the TSK-T5 group, owing to the limited amount of data.

c. Results after individual corrections

After the application of the individual corrections to the XBT profiles, the depth error in each profile is removed, and the residuals are regarded as pure thermal bias. CW13 found the pure thermal bias is generally consistent with depth. This property is used to test our method of depth correction. Before individual corrections, all Sippican XBTs have depth-varying temperature biases (Fig. 7). After correction of the depth error, the depth variation is reduced for all groups, leaving a more depth-independent thermal bias as found in CW13. The medians of the pure thermal biases are approximately 0.030°, 0.00°, 0.08°, 0.04°, 0.04°, 0.06°, 0.05°, 0.31°, and 0.09°C from G1 to G9, respectively. Positive bias is found in all probe types over the past 45 yr. TSK probes seem to have larger pure thermal biases than Sippican probes, and T4–T6 probes have a warmer bias than T7–DB for both manufacturers.

d. Time variable pure thermal bias

As suggested by previous studies, the pure thermal bias is time variable and contributes to the XBT bias history. In this study, the residuals after the individual corrections to the depth error are regarded as pure thermal bias. The yearly averaged pure thermal bias is not always consistent over depth (Fig. 8, left panels). Studies by GR10 and G12 using global-scale data also show pure thermal biases varying with depth. The source of the variations in thermal bias with depth found in the global-scale studies could be the background mesoscale and seasonal errors associated with the large spatial and temporal distances between the XBT data and the reference CTD data.

The residuals are averaged over depth to obtain a depth-averaged bias (Fig. 8, right panels), compared with results from GR10, G12, H12, CW13-CH, CW13, and L09. Generally, L09 shows much larger temperature corrections for each group than the other schemes, since they encapsulate depth-related biases as well as pure thermal bias.

The pure thermal biases for G1-T7-DB show a bias shift near 1984 from a larger bias (~0.15°C) to a much smaller bias, similar to the results from the side-by-side data (CW13/CW13-CH).

G2-DX has a similar time evolution as T7-DB, but with an earlier bias shift near 1980. This result differs significantly from H12 and G12, who find a near-consistent pure thermal bias over time, but match well with GR10.

The T4–T6 shows a much later decrease of pure thermal bias (~1985) compared with the side-by-side data result in CW13 (~1975). CW13 suggests that the change from strip charts to digital recording systems may be responsible for the thermal bias shift during early 1980s. It is possible this shift happens earlier in the side-by-side data (where scientific organizations and navies were testing new digital equipment) than in the global-scale data.

The G4-SX matches well with G12-SX pre-1985 and is generally consistent with other schemes (except L09) since 1995. A larger pure thermal bias is detected from 1985 to 1995 for SX with very larger uncertainties, due to a small data amount at this period. Both G4-SX and G2-DX show an earlier shift compared with T7-DB and T4–T6, suggesting a mix of other probes or FREs in the SX and DX groups. Significant differences appear between the DX and SX groups, so we recommend not merging the two datasets into a single group.

For Sippican T10, the pure thermal bias shifts from a slightly negative bias pre-1990 to a consistent warm bias after 1990. The bias reaches a maximum near 1995–2000. G12's result for T10 is well within the error bars of our results.

Sippican T5 probes show a larger bias pre-1980 (~0.25°C) than later (~0.05°C), values consistently larger than the other Sippican probes. For TSK-T4–T6 and TSK-T5, the pure thermal bias again shows a shift from a larger warm bias to a smaller bias in the 1990s. The TSK-T7 probes show an increasing bias history over time from the late 1980s to the 2000s.

Although some consistencies in pure thermal bias are found between global-scale data in this study and the side-by-side data in CW13 (such as T7-DB, SX), the discrepancies are significant between T4–T6 and DX. For example, during 1970–90, the pure thermal bias experiences a shift, but the shift time is different for the two datasets. This discrepancy indicates that the corrections derived from the side-by-side dataset may not be adequate to represent the bias history in the global XBT dataset.

e. Time variable depth error

In this section, we examine the time variation in the XBT fall rate. Using the scheme to calculate the depth from correlations between A and the other coefficients (described in section 3c), the depth of an XBT profile is fully characterized by the fall-rate coefficient A . The time variation of A is obtained by calculating the maximum depth-weighted median of A from the global-scale pairs in each year bin (Fig. 9, left panels) and comparing the results from the side-by-side data. The results show again the existence of the time-varying

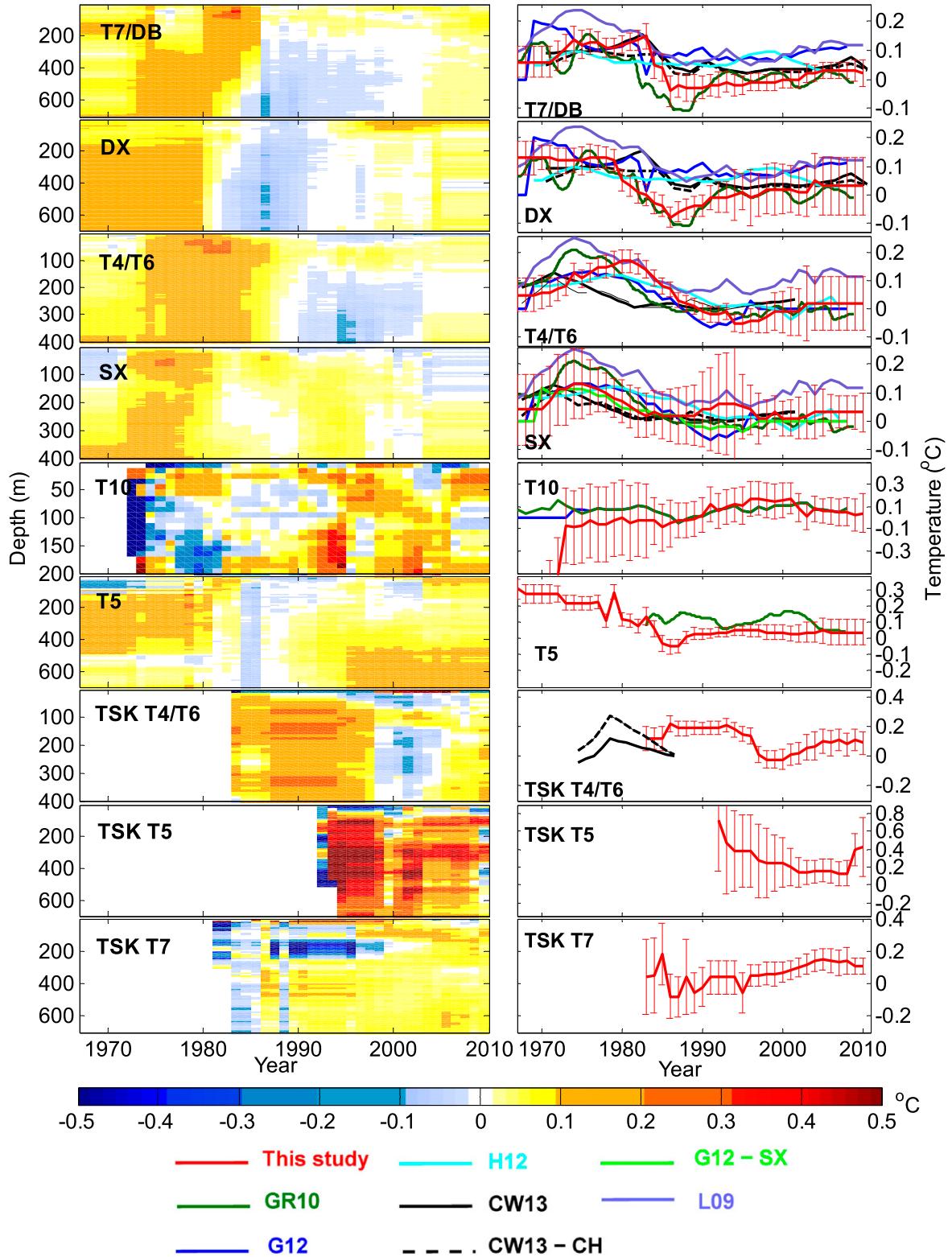


FIG. 8. (left) Pure thermal biases as function of depth and calendar year for the nine different XBT datasets. (right) Depth-averaged pure thermal bias is presented in red. Error bars are $2\times$ standard errors. Pure thermal biases obtained from different independent studies are also included: CW13, CW13-CH, G12, GR10, H12, L09, and G12-SX.

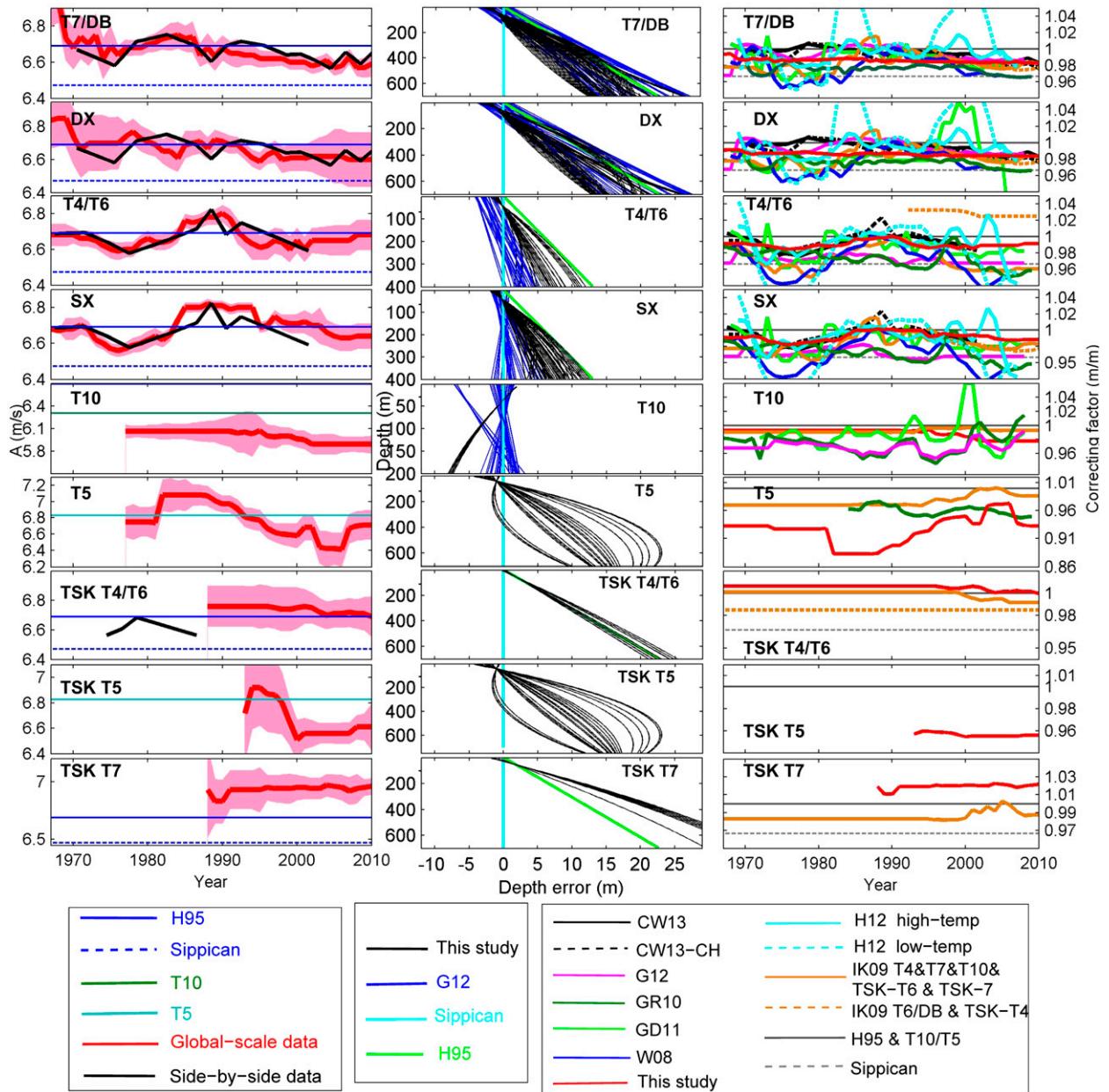


FIG. 9. (left) Fall rate coefficient A as function of calendar year for the nine different XBT datasets (red) with $2\times$ the standard error (shading); and coefficient A from side-by-side data (black). (middle) Depth error at each year as function of depth (black) for the nine groups, compared with the depth error diagnosed by G12 (blue) and H95 (green) for comparison; the manufacturer depth is regarded as the reference with zero depth error (cyan). (right) Depth-averaged correction factor from independent studies, compared with the results of this study (red).

depth error (DiNezio and Goni 2011; Gouretski and Koltermann 2007).

There is a similar evolution for G1-T7-DB, G2-DX, G3-T4-T6, and SX groups. The value of A reaches a peak near 1984 for T7-DB and DX, but near 1989 for T4-T6 and SX. These results from the global-scale data agree well with those from the side-by-side data, suggesting that the side-by-side dataset seems to be able to

represent the time variation of the depth error in the global XBT dataset. On the other hand, G5-T10 shows a decreasing fall rate since 1997. The value of A for G6-T5 has a peak around 1985 and it then decreases. For TSK-T4-T6 and T7, the A values are all slightly larger than in the H95 FRE and show a smaller time variation. Recent Sippican T7-DB and DX probes after 2000 have a fall-rate equation closer to Kizu et al.

(2011)'s results than to the manufacturer's or H95's fall-rate equation.

The corresponding depth errors as a function of depth (Fig. 9, middle column) lie between the H95 FRE in green and the manufacturer FRE in cyan. For G1-T7-DB and G2-DX, the depth errors diagnosed in this study generally agree with those presented in G12. A contradiction is found in the Sippican-T10 group between G12 and this study. The errors in the upper 100 m are negative in this study compared with both the manufacturer's and H95's FREs.

We then compared the results based on our correction scheme with those proposed by the following independent studies: H95, W08, IK09, GR10, GD11, H12, G12, CW13, and CW13-CH. First, it is apparent that there are differences among these studies:

- the corrections are depth constant (W08, GD11) or depth variable (H95, IK09, GR10, G12, H12, CW13, and CW13-CH), and here depth variable is defined as a depth error (or temperature bias) that depends on depth
- pure thermal bias corrections are included (GR10, G12, H12, CW13, and CW13-CH) or not (H95, W08, IK09, and GD11), and only L09 presents thermal bias correction
- the corrections are proposed separately for different probe types (IK09, GD11, G12, CW13, CW13-CH), or according to the depth, namely, shallow or deep XBTs (W08, GR10, and H12)

To account for these differences, the depth corrections are converted into a depth correction factor (Fig. 9, right column) defined by $CF = D_{\text{cor}}/D_{\text{uncor}}$, where D_{cor} is the XBT depth after correction and D_{uncor} is the XBT depth in the global-scale dataset. The D_{uncor} is calculated from the H95 equation with the exception of T5 and T10, which use the original manufacturer's fall rate (see Table 1). The correction scheme of W08 is time varying but constant over depth. The CF for IK09, CW13, CW13-CH, GD11, and this study is calculated using $CF = \text{mean}[D_{\text{cor}}(t)/D_{\text{uncor}}(t)]$, where t is the elapsed time from 1 s to approximately the maximum acquisition time (40s for T10, 80s for T4-T6-SX-TSK-T4-T6, 140s for T7-DB-DX-TSK-T7, 350s for T5-TSK-T5) with a bin of 1 s. The $D_{\text{cor}}(t)$ is calculated according to each scheme's corrections. For GR10, GD11, and G12, CF is calculated from $CF = \text{mean}[D_{\text{cor}}(d)/D_{\text{uncor}}(d)]$, where d is the depth calculated from the H95 FRE (or a specific manufacturer's coefficients for T5-T10) from 1 m to the terminal depth (250, 520, and 890 m for T10, T4-T6-SX-TSK-T4, T7-DB-DX-T5-TSK-T5-T7, respectively) with a bin of 1 m. The $D_{\text{cor}}(d)$ is the corrected XBT depth calculated using each scheme. It should be noted that the correction factor is only an

approximation of the depth error and assumes the depth error is depth independent. However, it is still worthwhile comparing these correction schemes in this manner to provide insight into the depth error history.

Generally, all of the correction factors for G1-G4 show a slightly increasing fall rate with time during 1975-85, and a decreasing fall rate after 1990, where the corrections mostly lie between H95 and the manufacturer's FRE. The depth-averaged corrections for T7-DB and DX for all of the correction schemes except H12 and GD11 agree quite well from 1987 to 2010. All of the schemes share a similar history, but our results show a much weaker time variation compared with other studies.

The schemes of W08, IK09, and GD11 have a larger time variation of depth corrections for T7-DB, DX, T4-T6, and SX, especially before 1990, which may be due to the assumption of zero pure thermal bias, leading to the overestimation of depth errors. Even separating the depth error from the pure thermal bias, H12 still shows a dramatic time variation of depth error, which is larger than schemes that also separate the two error sources (G12, GR10, CW13, and this study).

The depth correction factor for T10 has different histories in the five independent studies of GR10, G12, GD11, IK09, and this study. We found a slightly decreasing correction factor for T10 from the 1990s to the 2000s. For T5, all three schemes (GR10, IK09, and this study) show a peak within 2000-05 but differ for pre-1995. IK09 and this study suggest an increasing correction factor from 1980 to 2005. For TSK probes, our corrections are stable over time and generally agree with IK09 for TSK-T4-T6.

In all, the difference in depth correction factors between the global-scale dataset (this study) and the side-by-side dataset (CW13, CW13-CH) are small (within one standard error bar as in Fig. 9, left column), suggesting that the depth error is consistent over different datasets. The small number of pairs in the CW13 side-by-side dataset means comparisons with the global-scale trends for the T10, T5 and TSK groups cannot be made.

f. The influence of temperature on depth error and pure thermal bias

From the analyses of the side-by-side dataset, both the fall rate and the pure thermal bias increase with water temperature. Therefore, two questions are proposed:

- Is the time variation of the fall rate and the pure thermal bias caused by the different temperature of seawater sampled by the XBT system during XBT-CTD comparison tests?
- How does water temperature contribute to the historical XBT bias?

The average temperature of the water column sampled by XBTs during the comparison tests during the early era (pre-1980), based on the XBT data distribution only, is up to 10°C warmer than that in the latter time period (post-2000). How much does the temperature contribute to the time variation of the XBT bias? To answer this question, we apply two temperature-related corrections to the historical dataset. The A coefficient was modified as described in Eqs. (3) and (4), and the temperatures were modified following Eqs. (1) and (2). After these steps, we found a reduced magnitude of biases (with the exception of TSK-T4-T6-T10) and a slightly reduced time variation (with the exception of TSK-T7) (Fig. 10).

The regional variation of bias may be strongly influenced by water temperature, since the upper-ocean temperature has more than a 20°C difference between the tropical and high-latitude regions. At six depths (100, 200, 300, 400, 500, and 600 m), the zonal gradient of the bias is reduced significantly when the temperature-related corrections are applied (Fig. 11). However, the residual of the biases after the temperature-related corrections (blue) is larger at the lower latitudes, since the XBT was mainly deployed at low latitudes in the early era, when data show a larger warm bias.

5. A time, probe type, and temperature variable bias correction

In section 3, we determined the relationship among the three fall-rate coefficients, and the relationships between the water temperature and the fall rate/pure thermal bias. Based on these results, in section 4, we simplified the fall-rate equation and diagnosed the time evolution of the fall rate and the pure thermal bias using the global-scale dataset. The influence of temperature on the fall rate and the pure thermal bias was also examined. Using these results, a new correction scheme for historical XBT data is proposed, by using the results from the global-scale dataset. Global-scale data, instead of side-by-side data, are used because (i) the side-by-side dataset contains insufficient data to derive corrections for probe types other than T4-T6 and T7-DB; (ii) although we found the fall-rate coefficient for the two datasets had a similar time variation, the pure thermal bias was considerably different (such as in the T4-T6 and SX groups). Since the global pairs database is more representative of the global historical archive, we chose to use the depth error and thermal bias calculated from this dataset to get our final correction scheme.

a. Corrections for pure thermal bias

We assume the pure thermal bias is determined by both time and water temperature. For an XBT profile in

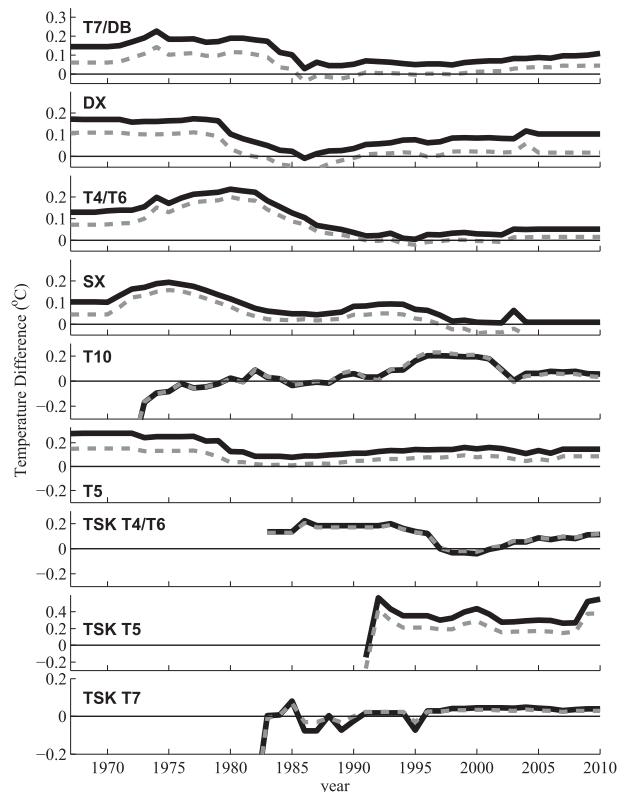


FIG. 10. Yearly averaged temperature difference (XBT – CTD) before (black solid) and after (gray dashed) temperature-related corrections for the nine different XBT datasets.

a specific year (denoted as “year”), the correction to the pure thermal bias is determined by

$$dT = dT_{\text{year}} + dT_{\text{temp}},$$

where dT_{temp} is the temperature-dependent part of the thermal bias, which is specific to each XBT probe type. Initially, the column-averaged temperature for the XBT is calculated from the XBT profile. The term dT_{temp} is calculated using the relationship between the pure thermal bias and the column-averaged temperature described in Eqs. (1) and (2). For the global-scale dataset, the yearly averaged values of dT_{temp} are listed in Table 3.

The term dT_{year} represents the time variable part of the pure thermal bias. It is calculated by subtracting the yearly averaged pure thermal bias (dT_{temp} , Table 3) from the total pure thermal bias (red lines in Fig. 8, right panels). The values of dT_{year} in each year for each group based on the global-scale dataset are presented in Table 3.

By summing dT_{year} and dT_{temp} , the correction to the temperature measurements (dT) is obtained.

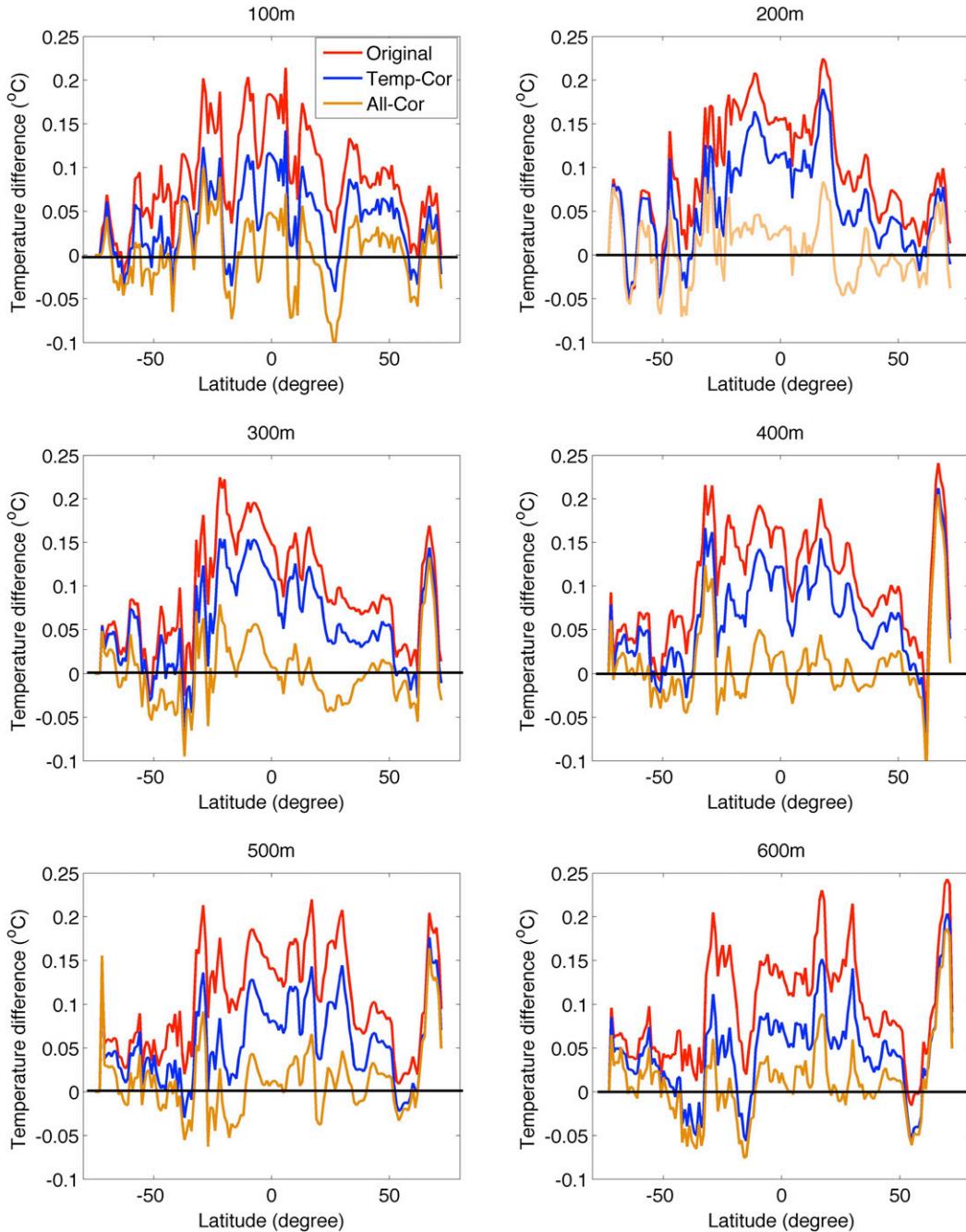


FIG. 11. Zonally averaged temperature difference (XBT – CTD) without any corrections (red), after temperature-related corrections (blue) and after temperature-related and time-dependent corrections (orange). Temperature difference is shown at depths of 100, 200, 300, 400, 500, and 600 m.

b. Correcting the depth error

We assume that the depth error is affected by calendar year and 0–100-m-averaged water temperature. The depth in each XBT profile is first recalculated to the H95 FRE (except for the T5 and T10 groups), then corrected

using the terms dA_{year} and dA_{temp} , which represent the time and temperature variable corrections, respectively.

In detail, XBT depth (D) is recalculated as

$$D = At - Bt^2 - \text{offset} \quad \text{with}$$

$$A = A_{\text{XBT}} + dA_{\text{year}} + dA_{\text{temp}},$$

TABLE 3. Terms dT_{temp} and dT_{year} for the nine different XBT datasets centered on each year from 1966 to 2010. The two values are added together to obtain dT .

	G1-T7-DB		G2-DX		G3-T4-T6		G4-SX		G5-T10		G6-T5		G7-TSK-T4-T6		G8-TSK-T5		G9-TSK-T7	
	dT_{year}	dT_{temp}	dT_{year}	dT_{temp}	dT_{year}	dT_{temp}	dT_{year}	dT_{temp}										
1966	0.0311	0.0288	0.0755	0.0344	0.0192	0.0307	-0.0038	0.0437			0.2801	0.0298						
1967	0.0214	0.0385	0.0955	0.0344	0.0170	0.0329	-0.0031	0.0430			0.2801	0.0298						
1968	0.0208	0.0391	0.0955	0.0344	0.0123	0.0376	-0.0026	0.0425			0.2501	0.0298						
1969	0.0208	0.0391	0.1024	0.0275	0.0069	0.0430	-0.0024	0.0423			0.2501	0.0298						
1970	0.0213	0.0386	0.1022	0.0277	0.0169	0.0430	-0.0009	0.0408			0.2501	0.0298						
1971	0.0217	0.0382	0.1022	0.0277	0.0171	0.0428	0.0331	0.0368			0.2501	0.0298						
1972	0.0425	0.0374	0.0924	0.0275	0.0267	0.0432	0.0775	0.0324	-0.1068	0.0367	0.2485	0.0314						
1973	0.0732	0.0367	0.0910	0.0289	0.0373	0.0426	0.0800	0.0299	-0.1173	0.0272	0.1889	0.0310						
1974	0.0992	0.0357	0.0905	0.0294	0.0954	0.0345	0.1014	0.0285	-0.0968	0.0267	0.1889	0.0310						
1975	0.0834	0.0365	0.0899	0.0300	0.0767	0.0332	0.1017	0.0282	-0.0555	0.0254	0.1889	0.0310						
1976	0.0932	0.0367	0.0995	0.0304	0.1008	0.0291	0.0928	0.0271	-0.0855	0.0254	0.1991	0.0308						
1977	0.0938	0.0361	0.1001	0.0298	0.1214	0.0285	0.0826	0.0273	-0.0854	0.0253	0.0793	0.0306						
1978	0.0750	0.0349	0.0993	0.0306	0.1217	0.0282	0.0737	0.0262	-0.0449	0.0248	0.2588	0.0311						
1979	0.0744	0.0355	0.0882	0.0317	0.1323	0.0276	0.0536	0.0263	-0.0243	0.0242	0.0874	0.0325						
1980	0.0889	0.0310	0.0156	0.0343	0.1425	0.0274	0.0427	0.0272	-0.0749	0.0248	0.0772	0.0327						
1981	0.1006	0.0293	-0.0055	0.0354	0.1404	0.0295	0.0224	0.0275	-0.0441	0.0240	0.0524	0.0275						
1982	0.1064	0.0285	-0.0213	0.0312	0.1301	0.0298	0.0040	0.0259	-0.0542	0.0241	0.1124	0.0275						
1983	0.1212	0.0287	-0.0313	0.0315	0.1113	0.0286	-0.0052	0.0251	-0.1139	0.0238	0.0319	0.0280	0.1635	0.0264				
1984	0.0215	0.0284	-0.0616	0.0315	0.0810	0.0289	0.0043	0.0256	-0.0741	0.0240	-0.0778	0.0277	0.1635	0.0264				
1985	0.0106	0.0293	-0.0708	0.0307	0.0519	0.0280	-0.0057	0.0256	-0.0541	0.0240	-0.0779	0.0278	0.1635	0.0264				
1986	-0.0695	0.0294	-0.1096	0.0295	0.0427	0.0272	-0.0058	0.0257	-0.0542	0.0241	-0.0174	0.0273	0.1635	0.0264				
1987	-0.0494	0.0293	-0.0874	0.0273	0.0029	0.0270	-0.0070	0.0269	-0.0344	0.0243	0.0022	0.0277	0.1615	0.0284				
1988	-0.0591	0.0290	-0.0766	0.0265	-0.0075	0.0274	-0.0073	0.0272	0.0056	0.0243	0.0028	0.0271	0.1621	0.0278	0.7057	0.0328		
1989	-0.0594	0.0293	-0.0662	0.0261	-0.0173	0.0270	-0.0071	0.0270	0.0956	0.0243	0.0028	0.0271	0.1606	0.0293	0.7057	0.0242	0.0085	0.0314
1990	-0.0584	0.0283	-0.0665	0.0264	-0.0477	0.0276	0.0133	0.0266	0.0352	0.0247	0.0123	0.0276	0.1808	0.0291	0.4433	0.0266	0.0085	0.0314
1991	-0.0481	0.0280	-0.0373	0.0272	-0.0478	0.0277	0.0329	0.0270	0.0750	0.0249	0.0112	0.0287	0.1507	0.0292	0.3541	0.0258	0.0085	0.0314
1992	-0.0482	0.0281	-0.0386	0.0285	-0.0377	0.0276	0.0328	0.0271	0.0954	0.0245	0.0211	0.0288	0.1208	0.0291	0.3541	0.0258	-0.0893	0.0292
1993	-0.0392	0.0291	-0.0390	0.0289	-0.0780	0.0279	0.0327	0.0272	0.0818	0.0281	0.0210	0.0289	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
1994	-0.0490	0.0289	-0.0293	0.0292	-0.0783	0.0282	0.0227	0.0272	0.1413	0.0286	0.0211	0.0288	0.1208	0.0291	0.3541	0.0258	0.0207	0.0292
1995	-0.0489	0.0288	-0.0193	0.0292	-0.0681	0.0280	0.0118	0.0281	0.1412	0.0287	0.0211	0.0288	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
1996	-0.0388	0.0287	-0.0384	0.0283	-0.0681	0.0280	0.0118	0.0281	0.1312	0.0287	0.0211	0.0288	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
1997	-0.0386	0.0285	-0.0262	0.0261	-0.0681	0.0280	-0.0172	0.0271	0.1412	0.0287	0.0211	0.0288	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
1998	-0.0380	0.0279	-0.0042	0.0241	-0.0579	0.0278	-0.0172	0.0271	0.1312	0.0287	0.0116	0.0283	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
1999	-0.0378	0.0277	-0.0040	0.0239	-0.0379	0.0278	0.0047	0.0252	0.1213	0.0286	0.0116	0.0283	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
2000	-0.0273	0.0272	-0.0039	0.0238	-0.0374	0.0273	-0.0161	0.0260	0.1312	0.0287	0.0115	0.0284	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
2001	-0.0272	0.0271	-0.0040	0.0239	-0.0374	0.0273	-0.0154	0.0253	0.1413	0.0286	0.0006	0.0293	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
2002	-0.0275	0.0274	-0.0140	0.0239	-0.0364	0.0263	-0.0183	0.0282	0.1413	0.0286	0.0004	0.0295	0.1104	0.0295	0.3541	0.0258	0.0207	0.0292
2003	-0.0055	0.0254	-0.0189	0.0288	-0.0052	0.0251	-0.0073	0.0272	-0.0114	0.0239	0.0100	0.0299	0.0348	0.0251	0.1184	0.0215	0.0899	0.0300
2004	-0.0050	0.0249	0.0211	0.0288	-0.0048	0.0247	0.0007	0.0292	0.0549	0.0250	0.0101	0.0298	0.0450	0.0249	0.1284	0.0215	0.1091	0.0308
2005	0.0050	0.0249	-0.0004	0.0303	-0.0055	0.0254	0.0007	0.0292	0.0280	0.0319	0.0201	0.0298	0.0754	0.0245	0.1375	0.0224	0.119	0.0309

TABLE 3. (Continued)

	G1-T7-DB		G2-DX		G3-T4-T6		G4-SX		G5-T10		G6-T5		G7-TSK-T4-T6		G8-TSK-T5		G9-TSK-T7	
	dT_{year}	dT_{temp}																
2006	0.0050	0.0249	-0.0004	0.0303	-0.0052	0.0251	0.0007	0.0292	0.0297	0.0302	0.0095	0.0304	0.0654	0.0245	0.1275	0.0224	0.109	0.0309
2007	0.0050	0.0249	-0.0006	0.0305	-0.0063	0.0262	0.0007	0.0292	0.0197	0.0302	0.0095	0.0304	0.0854	0.0245	0.1083	0.0216	0.0993	0.0306
2008	0.0032	0.0267	-0.0007	0.0306	-0.0065	0.0264	0.0004	0.0295	0.0196	0.0303	0.0095	0.0304	0.0550	0.0249	0.1084	0.0215	0.1089	0.0310
2009	0.0032	0.0267	-0.0007	0.0306	-0.0047	0.0246	0.0002	0.0297	-0.0103	0.0302	0.0094	0.0305	0.0847	0.0252	0.3747	0.0252	0.0797	0.0302
2010	-0.0074	0.0273	-0.0008	0.0307	-0.0034	0.0233	0.0002	0.0297	0.0210	0.0189	0.0094	0.0305	0.0644	0.0255	0.3948	0.0251	0.0809	0.0290

where A_{XBT} is the H95 fall rate (6.691 m s^{-1}) for all the probe types with the exception of T10 and T5-TSK-T5, where $A_{\text{XBT}} = 6.301 \text{ m s}^{-1}$ for T10 and 6.828 m s^{-1} for T5-TSK-T5. The coefficients B and offset are calculated from the correlations $B = a_1A + a_2$ and $\text{offset} = b_1A + b_2$, where the coefficients a_1, a_2 are given in Eqs. (5)–(7). Coefficients b_1, b_2 are constant with time and presented in Eqs. (8)–(10).

To correct the global-scale data, the relationships obtained from the side-by-side dataset are used. XBT profiles with types T7-DB, DX, and T5 use the relationship obtained for T7-DB; profiles of type T4-T6, SX, and T10 use the relationship obtained for T4-T6; and all of the TSK profiles use the relationship obtained using the results of the TSK pairs.

The term dA_{temp} is the part of the A coefficient influenced by water temperature. For each XBT profile, the 0–100-m-averaged temperature is obtained by using the XBT temperature records (Temp_{100}). Next, the A -temperature relationship described in Eqs. (3) and (4) is used to calculate dA_{temp} (Table 4).

The term dA_{year} is the time variable part of the XBT fall rate. This term is obtained by subtracting the temperature varying fall rate (dA_{temp}) from the total time-varying term A_{total} (as shown with red lines in the left panels of Fig. 9): $dA_{\text{year}} = A_{\text{total}} - A_{\text{XBT}} - A_{\text{temp}}$. The yearly average values of dA_{year} are shown in Table 4.

It should be noted that Tables 3 and 4 list the corrections applied to the WOD09 dataset, where the temperature-related parts (dT_{temp} and dA_{temp}) are the yearly averaged values. This can be used to calculate historical ocean heat content. However, if one needs a more careful correction to an individual XBT profile, the temperature-related part should be calculated for this individual XBT profile, and then the time variable part (dT_{year} and dA_{year} in Tables 3 and 4) should be added to this temperature-related part to get the final correction.

6. Evaluation of the correction scheme

a. Temperature differences between XBT and CTD in the global-scale dataset

In Fig. 12, temperature differences between XBT and CTD in the global-scale dataset before and after XBT bias corrections (as in section 5) are presented at each year—depth for nine groups. Before correction, the XBT data for each group have a time-varying warm bias using the H95 FRE (or the Sippican FRE for T10–T5). After depth errors are removed from the global-scale dataset, reduced warm biases still exist and are time variable. After correction for both the fall-rate error and

TABLE 4. Temperature variable and time variable parts of A (dA_{temp} and dA_{year}) for the nine different XBT datasets in each year. The two values are added to the H95 fall rate (6.691 m s^{-1}) for G1, G2, G3, G4, G6, G9; and the T5 fall rate (6.828 m s^{-1}) for G6 and G8, the T10 fall rate (6.301 m s^{-1}) for G5 together to obtain a total fall rate coefficient A used to calculate XBT depth (yearly A is shown in Fig. 9, left panels).

	G1-T7-DB		G2-DX		G3-T4-T6		G4-SX		G5-T10		G6-T5		G7-TSK-T4-T6		G8-TSK-T5		G9-TSK-T7	
	dA_{year}	dA_{temp}	dA_{year}	dA_{temp}	dA_{year}	dA_{temp}	dA_{year}	dA_{temp}										
1966	0.1633	0.0556	0.0973	0.0516	-0.1211	0.1100	-0.1242	0.1181			-0.1726	0.0545						
1967	0.1702	0.0587	0.0973	0.0516	-0.1255	0.1144	-0.1242	0.1131			-0.1726	0.0545						
1968	0.1851	0.0638	0.1073	0.0516	-0.1312	0.1201	-0.1323	0.1112			-0.1726	0.0545						
1969	-0.0238	0.0627	0.1138	0.0451	-0.1322	0.1211	-0.1205	0.1094			-0.1726	0.0545						
1970	-0.0736	0.0625	0.0033	0.0456	-0.1323	0.1212	-0.1187	0.1076			-0.1726	0.0545						
1971	-0.0126	0.0615	-0.0767	0.0456	-0.1516	0.1205	-0.0979	0.1068			-0.1713	0.0532						
1972	-0.0512	0.0601	-0.0363	0.0452	-0.1523	0.1212	-0.1145	0.1034			-0.1714	0.0533						
1973	-0.0007	0.0596	-0.0430	0.0519	-0.1582	0.1171	-0.1704	0.0993	-0.3585	0.1174	-0.1714	0.0533						
1974	-0.1285	0.0574	-0.0443	0.0532	-0.1779	0.1068	-0.1894	0.0983	-0.2904	0.0493	-0.1314	0.0533						
1975	-0.0579	0.0568	-0.0447	0.0536	-0.1830	0.0919	-0.2185	0.0974	-0.2905	0.0494	-0.1313	0.0532						
1976	-0.1052	0.0541	-0.0450	0.0539	-0.1745	0.0734	-0.2241	0.0930	-0.2885	0.0474	-0.1313	0.0532						
1977	-0.0674	0.0563	-0.0047	0.0536	-0.1649	0.0738	-0.1896	0.0685	-0.2880	0.0469	-0.1313	0.0532						
1978	-0.0670	0.0559	0.0241	0.0548	-0.1737	0.0826	-0.1615	0.0604	-0.2877	0.0466	-0.1313	0.0532						
1979	-0.0676	0.0565	-0.0181	0.0570	-0.1354	0.0843	-0.1517	0.0606	-0.2859	0.0448	-0.1313	0.0532						
1980	-0.0455	0.0544	-0.0408	0.0597	-0.1180	0.0869	-0.1385	0.0674	-0.2855	0.0444	-0.1313	0.0532						
1981	-0.0386	0.0575	-0.0617	0.0606	-0.1508	0.0997	-0.1262	0.0751	-0.2889	0.0478	-0.1414	0.0533						
1982	-0.0134	0.0523	-0.0991	0.0580	-0.1486	0.0975	-0.1387	0.0676	-0.2899	0.0488	0.2173	0.0346						
1983	-0.0012	0.0501	-0.0991	0.0580	-0.1309	0.0898	-0.1083	0.0672	-0.2952	0.0541	0.2171	0.0348						
1984	0.0088	0.0501	-0.1385	0.0574	-0.1025	0.0914	-0.0284	0.0673	-0.2896	0.0485	0.2135	0.0384						
1985	-0.0161	0.0550	-0.0565	0.0554	-0.0232	0.0921	0.0426	0.0663	-0.2867	0.0456	0.2167	0.0352						
1986	-0.0051	0.0540	-0.0416	0.0505	-0.0281	0.0870	0.0443	0.0646	-0.2869	0.0458	0.2167	0.0352						
1987	-0.1051	0.0540	-0.0172	0.0361	-0.0122	0.0811	0.0425	0.0664	-0.2869	0.0458	0.2148	0.0371						
1988	-0.0646	0.0535	-0.0491	0.0280	0.0005	0.0884	0.0602	0.0687	-0.2878	0.0467	0.2099	0.0320						
1989	-0.1152	0.0541	0.0025	0.0264	0.0052	0.0837	0.0480	0.0709	-0.2884	0.0473	0.1499	0.0320						
1990	-0.0595	0.0484	-0.0095	0.0284	0.0339	0.0750	0.0460	0.0729	-0.2882	0.0471	0.1106	0.0313						
1991	-0.0821	0.0410	-0.0347	0.0336	0.0064	0.0725	0.0281	0.0708	-0.2882	0.0471	0.1106	0.0313						
1992	-0.0824	0.0413	-0.0404	0.0393	-0.0664	0.0753	0.0351	0.0738	-0.3092	0.0481	0.0477	0.0342						
1993	-0.0966	0.0455	-0.0826	0.0415	-0.1029	0.0718	0.0391	0.0698	-0.2985	0.0474	-0.0590	0.0409						
1994	-0.0928	0.0417	-0.1265	0.0454	-0.1390	0.0779	0.0422	0.0667	-0.3169	0.0458	-0.0784	0.0403						
1995	-0.1103	0.0392	-0.1467	0.0456	-0.1586	0.0775	-0.0682	0.0671	-0.2845	0.0534	-0.0964	0.0383						
1996	-0.1082	0.0371	-0.0961	0.0350	-0.1472	0.0761	-0.0634	0.0723	-0.4092	0.1081	-0.1940	0.0359						
1997	-0.1060	0.0349	-0.0736	0.0325	-0.1164	0.0753	-0.0107	0.0696	-0.4190	0.1079	-0.2429	0.0348						
1998	-0.1017	0.0306	-0.1235	0.0324	-0.1279	0.0768	-0.0479	0.0768	-0.4090	0.1079	-0.2605	0.0324						
1999	-0.1234	0.0323	-0.1131	0.0320	-0.1666	0.0755	-0.0581	0.0770	-0.4055	0.1044	-0.2702	0.0321						
2000	-0.1216	0.0305	-0.1130	0.0319	-0.1089	0.0678	-0.0621	0.0810	-0.4462	0.1051	-0.2800	0.0319						
2001	-0.1198	0.0287	-0.1130	0.0319	-0.1589	0.0678	-0.0509	0.0798	-0.4300	0.0789	-0.1812	0.0331						
2002	-0.1381	0.0270	-0.1030	0.0319	-0.1054	0.0643	-0.0757	0.0846	-0.4728	0.0617	-0.1816	0.0335						
2003	-0.1270	0.0259	-0.1213	0.0502	-0.1006	0.0595	-0.0565	0.0654	-0.4535	0.0424	-0.4228	0.0347						

TABLE 4. (Continued)

	G1-T7-DB		G2-DX		G3-T4-T6		G4-SX		G5-T10		G6-T5		G7-TSK-T4-T6		G8-TSK-T5		G9-TSK-T7	
	dA_{year}	dA_{temp}	dA_{year}	dA_{temp}	dA_{year}	dA_{temp}	dA_{year}	dA_{temp}										
2004	-0.1052	0.0241	-0.1213	0.0502	-0.0941	0.0530	-0.1043	0.0732	-0.4575	0.0464	-0.4419	0.0338	-0.0833	0.0922	-0.3078	0.0397	0.2261	0.0528
2005	-0.1453	0.0242	-0.1274	0.0563	-0.1009	0.0598	-0.1243	0.0732	-0.4811	0.0700	-0.4419	0.0338	-0.0724	0.0913	-0.3083	0.0402	0.2156	0.0533
2006	-0.1353	0.0242	-0.1274	0.0563	-0.0967	0.0556	-0.1348	0.0737	-0.4820	0.0709	-0.4543	0.0362	-0.0732	0.0921	-0.3078	0.0397	0.2055	0.0534
2007	-0.0868	0.0257	-0.1447	0.0536	-0.1050	0.0639	-0.1348	0.0737	-0.4817	0.0706	-0.1943	0.0362	-0.0658	0.0947	-0.3041	0.0360	0.1759	0.0530
2008	-0.1468	0.0257	-0.1450	0.0539	-0.0755	0.0644	-0.1509	0.0998	-0.4822	0.0711	-0.1657	0.0376	-0.0767	0.0956	-0.2530	0.0349	0.1950	0.0539
2009	-0.1486	0.0275	-0.1448	0.0537	-0.0712	0.0601	-0.1634	0.1123	-0.4821	0.0710	-0.1557	0.0376	-0.0776	0.0965	-0.2651	0.0470	0.2051	0.0538
2010	-0.1313	0.0302	-0.1454	0.0543	-0.0549	0.0438	-0.1634	0.1123	-0.4305	0.0194	-0.1549	0.0368	-0.0989	0.0978	-0.2651	0.0470	0.2159	0.0530

the pure thermal bias, the temperature differences are reduced further and mostly eliminated, although there remain some small errors.

b. Quantification of correction performance

As suggested by Tim Boyer in the 2010 XBT Bias and Fall Rate Workshop in Hamburg, Germany, the corrections on XBT data should (i) be able to successfully meet a performance criterion, such as being able to correct 95% of all depths to within 0.025°C versus CTD in statistical tests; (ii) be relatively easy to implement, fully documented, and extendable into the future; and (iii) cover all XBT types, all observation depths, and all time periods (T. Boyer 2010, meeting presentation). Recent correction schemes for XBTs tend to satisfy criteria (ii). However, a consistent validation scheme has not been agreed upon. Existing schemes are validated mainly (i) by calculating the global ocean heat content history with and without the corrected XBT data or (ii) by comparing the correction scheme with the other schemes, to look for similarities between schemes. In this section, we propose a method of measuring correction performance.

At first, we assume that the temperature differences between XBT and CTD data in the global-scale dataset have a normal distribution with a mean (or median) of zero if there is no systematic bias. The distributions of the temperature differences between XBT and CTD ($T_{xbt} - T_{ctd}$) before and after correction are inserted in 10-m bins from a depth of 5 m to the terminal depth of each profile. Then, the frequencies of ($T_{xbt} - T_{ctd}$) differences are normalized by the maximum frequency (Fig. 13). Before this procedure, the ($T_{xbt} - T_{ctd}$) differences are positively biased in all nine groups. The mean (cross) and the median (circle) of the temperature differences are not the same, implying asymmetry in the distributions. All of the correction schemes are used to correct the XBT profiles in all groups. If there is not a correction on a specific type for a correction scheme (such as T5 in GD11), then the corrections are made according to the corrections for unknown types (such as UNK in GD11). If there is no correction available for unknown types, then corrections are made according to their shallow or deep groups. CW13/CW13-CH is not included in this comparison for TSK groups because CW13 provides corrections on TSK-T6 before 1986, but we found a very small amount of TSK data before 1986 in the global-scale dataset.

After correction, we highlight three aspects:

- (i) Mean and median. For all data, the scheme proposed in this study has reduced the total biases and the residuals are distributed around zero for all groups. Correction schemes based on

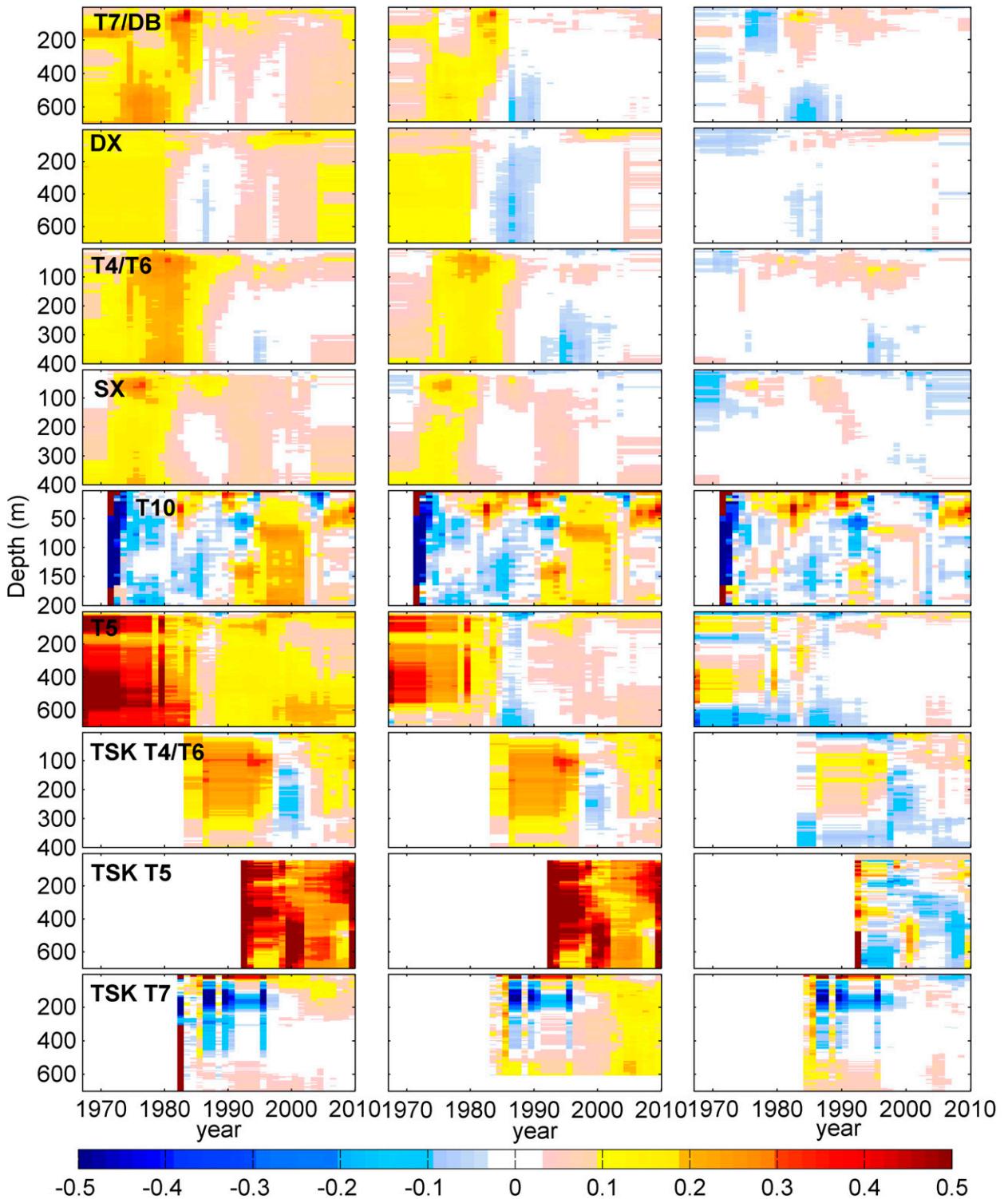


FIG. 12. Temperature differences ($T_{\text{xbt}} - T_{\text{ctd}}$) in the global-scale data for the nine groups (left) before corrections, (middle) after correcting for depth error, and (right) after correcting both the depth error and pure thermal bias.

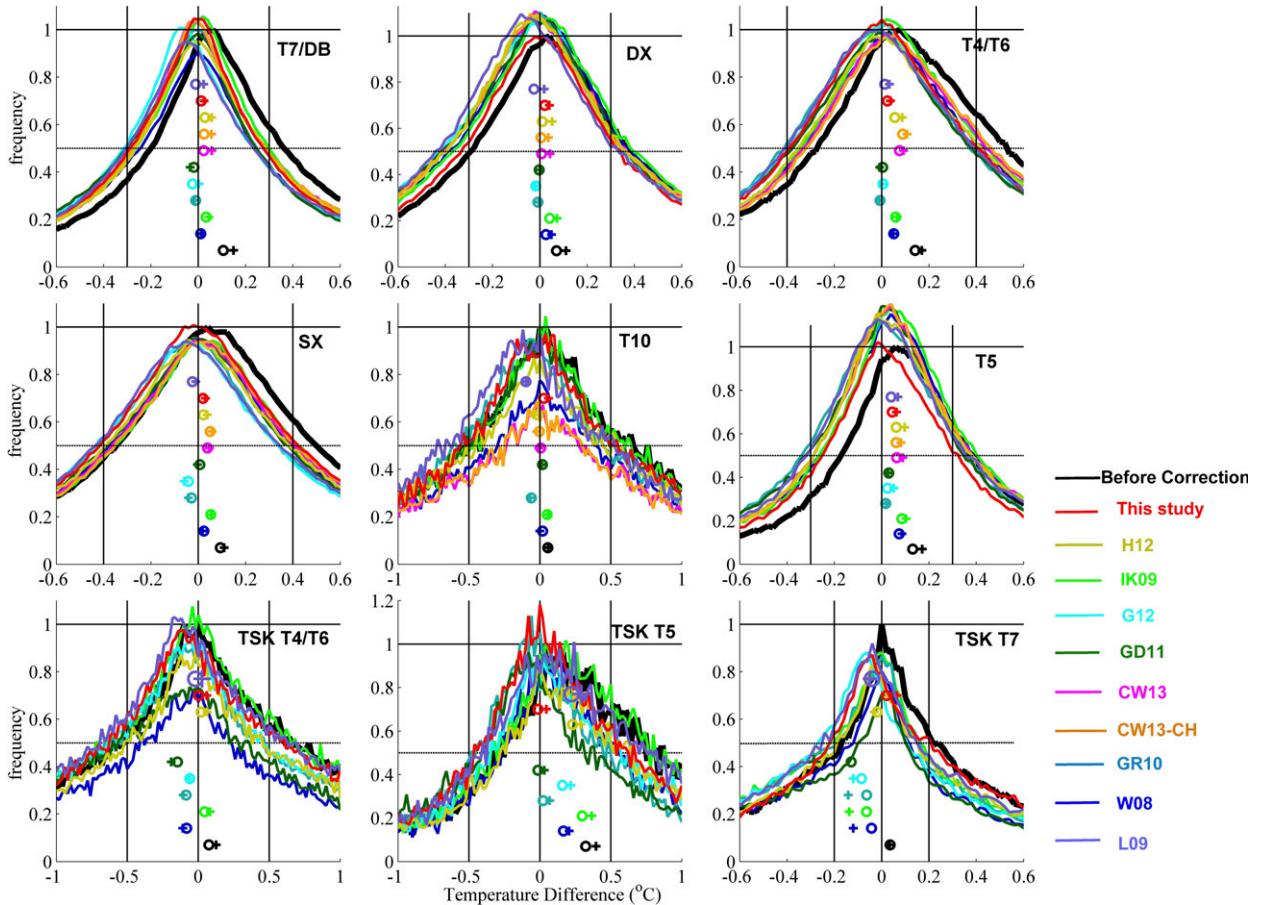


FIG. 13. Distribution of temperature differences ($T_{x_{bt}} - T_{ctd}$) for the nine groups before (black) and after corrections from different correction schemes—IK09, W08, G12, GD11, H12, CW13, CW13-CH, GR10, L09, and this study—in different colors. Mean and median of the distributions are shown as crosses and circles, respectively.

the side-by-side dataset (CW13 and CW13-CH) underestimates the biases in T4/T6 and SX. IK09 underestimates the bias for all groups (except TSK-T7). Maybe this is the reason why the historical OHC time series based on IK09 always shows a weaker long-term trend.

The L09 median is negative for most probe types, suggesting an overcorrection, but positive (undercorrected) for T4-T6, T5-TSK-T5. Almost for all groups and correction schemes (except TSK-T7), the mean is distributed at the right side of (warmer than) the median, indicating asymmetric residual distributions. None of the schemes correct the TSK groups well with the exception of this study (and H12 for TSK-T4-T6 and TSK-T7). L09 looks to perform similarly to H12 for TSK-T4-T7 and TSK-T7

(ii) Symmetry. Using the 0.5 frequency line (dashed line) as a reference point, the uncorrected data are always biased to the positive. The correction scheme proposed in this study reduces the asymmetry.

(iii) Peak. The results of this study have the peak near zero for all groups with the exception of TSK T4-T6 and TSK T7. The position of the peak varies for each correction scheme.

These results give an insight into the distribution of the biases. To quantify the performance of the correction schemes, more detailed statistics are required. We suggest that after correction, the distribution of the ($T_{x_{bt}} - T_{ctd}$) differences should meet the following criteria:

- (i) The ($T_{x_{bt}} - T_{ctd}$) differences should be minimized. The median of the ($T_{x_{bt}} - T_{ctd}$) difference is the first quantity used to assess the correction performance, denoted as the median.
- (ii) The distribution of the ($T_{x_{bt}} - T_{ctd}$) differences should be symmetrical around zero (i.e., it has a normal distribution), and its skewness (calculated by the routine Skewness in MATLAB), is defined as the third standardized moment of a random variable X : For a sample of n values, the sample skewness is

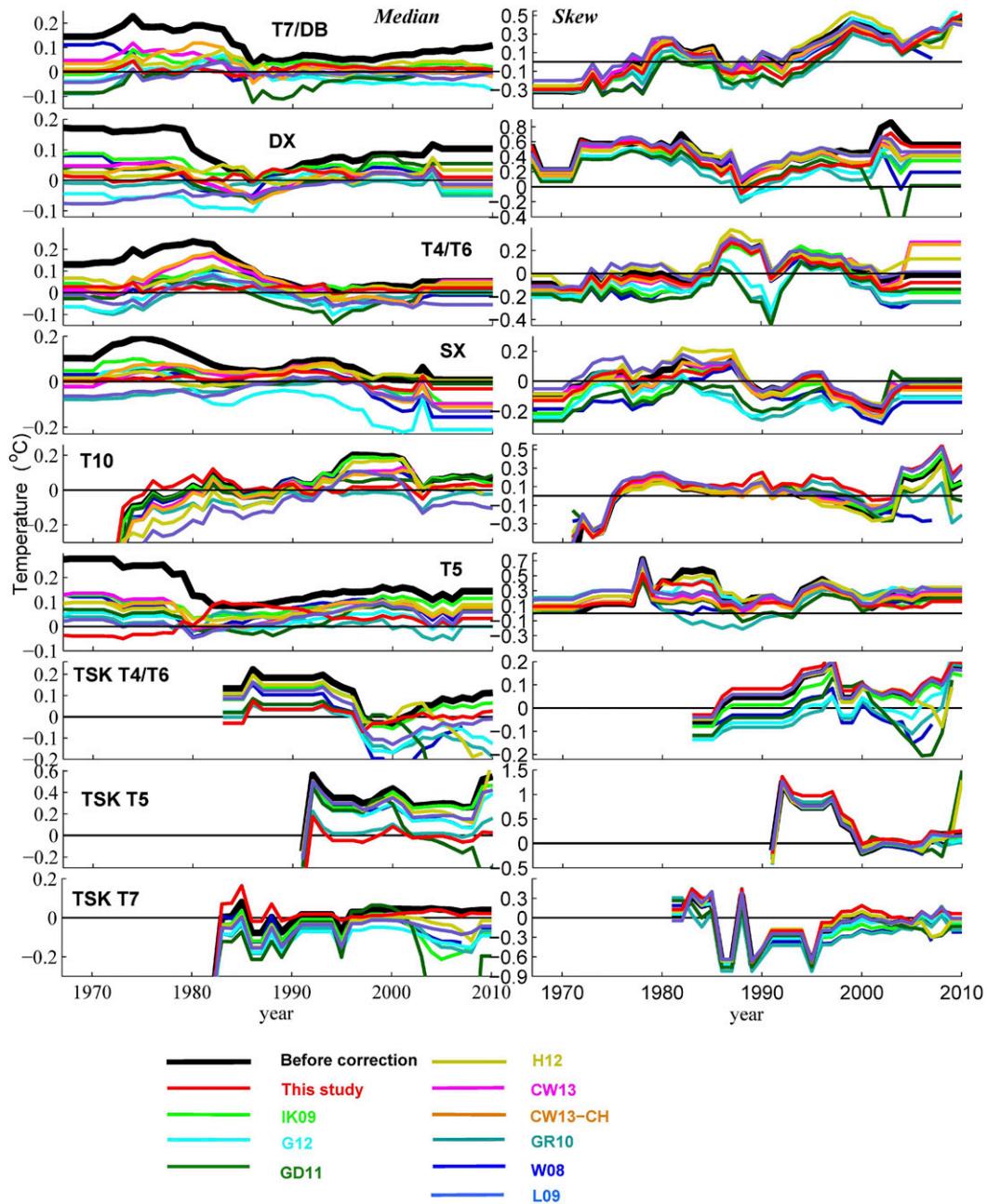


FIG. 14. Comparison of correction performance among different schemes. (left) Median of the temperature differences between XBT and CTD after correction as a function of calendar year. (right) Skewness of the distribution of temperature differences. Comparison is taken for the nine different XBT datasets in each row.

$$\text{skewness} = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}},$$

where \bar{x} is the sample mean, m_3 is the sample third central moment, and m_2 is the sample variance.

These two indices are assessed before and after applying the correction schemes to the XBT profiles in the global-scale dataset. They are calculated using the available $(T_{\text{xbt}} - T_{\text{ctd}})$ differences in 10-m bins from 5 m to the terminal depth in each time window. After a comparison of the median of the $(T_{\text{xbt}} - T_{\text{ctd}})$ differences before and after corrections, Fig. 14 shows the following:

- The biases are all significantly reduced for the T7–DB, DX, T4–T6, and SX groups for all correction schemes.
- For G3–T4–T6 and G4–SX, the correction schemes of CW13/CW13–CH underestimate the total bias from 1975 to 1985, which is due to the much earlier shift of the pure thermal bias in CW13 as presented in Fig. 8.
- All of the schemes reduce the bias in T5.
- For T10 and all of the TSK groups, all the previous corrections are weak in correcting the biases, but we underline that most of the correction schemes include no special treatment for the T5 (with the exception of IK09), T10 (excepting IK09, GD11, and G12), and TSK groups (excepting IK09 and G12).

We suggest that these probe types should be separated, since their bias history seems to be different when compared with other groups.

Comparison of the skew of the results shows that the distribution of the $(T_{\text{xbt}} - T_{\text{ctd}})$ differences is not always symmetrical around zero but still positively biased in general. This asymmetry may be caused by the following:

- Incomplete corrections to XBT biases, especially to the depth error.
- Errors due to the large temporal–spatial differences between an XBT and the correspondent CTD, introducing mesoscale noises into the distribution.
- “Leakage” from small insulation penetrations in the copper wire of the XBT, causing a gradual anomalous increase in temperature with depth (Bailey et al. 1994; Heinmiller et al. 1983; Kizu and Hanawa 2002a). This bias exists in a large part of the global XBT dataset; it is difficult to detect using automated quality-control procedures, and leads to a positive skew. Anderson (1980) states that, among the known sources of wrong measurements, there are more effects (mainly electrical effects) producing a warm variation than a cold one. This means that (statistically speaking) small errors that are not easily identified are more likely to result in a warm bias in XBTs.
- Surface thermal inertia caused by large differences in air temperature between the XBT storage area immediately prior to the drop and water temperature, which are also affected by the recording system used (Kizu and Hanawa 2002b). Usually, an XBT can take 3–4 m to equilibrate with the water temperature, but this time is partially dependent on the coupling between the probe and the recording system (up to ~10 m; Kizu and Hanawa 2002a). This transient depends on the thermistor and the acquisition system,

which includes also the computer and the ground connection of the recording system to the hull of the ship. Surface thermal inertia may introduce a cool bias near the sea surface, which may contribute to the negative skew.

- Different vertical resolution of the data, which introduces vertical variations in the $(T_{\text{xbt}} - T_{\text{ctd}})$ differences.

Although the contribution of these factors to the distribution cannot be addressed in a single paper, we highlight several properties, which help us to identify the different behaviors of errors for different probe types:

- The T7–DB group has a negative skew pre-1980, while DX shows positive skew at the same time. After 1980, the skew of both groups trends downward from 1980 to 1988 and upward from 1990 to 2000. This indicates that there are more mixed probe types in the DX group before 1980 (perhaps including some T5s, Sparton–XBTs).
- The T4–T6 and SX groups show a very similar trend of skew, with two peaks near 1987 and 1994. This similarity suggests that the majority of the SX group is of the T4–T6 type. But the skew for SX is generally more positive than that for T4–T6 before 1985, implying that maybe some T10 data are included in the SX groups, since T10 data have a consistently positive skew in most years.
- The T5 group has a consistent positive skew from 1966 to 2010, implying that some T5s may be included in the DX group, since the skew of the DX group is always more positive than the T7–DB group.
- The TSK groups show a different skew history compared with the Sippican probes. Studies by Kizu et al. (2011) have shown that Sippican and TSK have different characteristics. Only the thermistor is the same, being produced by the same external manufacturer. Any other component of an XBT probe is independently produced and assembled by two manufacturers. For example, TSK probes seem to be less affected by leakage than Sippican probes, probably because TSK uses a different wire with more robust insulation (Kizu et al. 2011).

c. Evaluation by an independent dataset

An independent dataset (EN3–v2a dataset; Ingleby and Huddleston 2007) is used in this section to further evaluate the performance of various correction schemes. EN3a–v2a is a fully quality-controlled dataset, with data sourced from the *World Ocean Database 2005*, the Global Temperature and Salinity Profile

Programme (GTSP), Argo, and Arctic Synoptic Basin-Wide Observations (ASBO). XBT–CTD pairs are identified using the same strategy proposed in section 2b. All correction schemes shown in the previous sections are used to correct the XBT profiles, with results shown in Fig. 15. All of the correction schemes considerably reduce the total XBT bias. The method presented in this study appears to be the best one, reducing the bias within the range of $-0.04^{\circ}\sim 0.04^{\circ}\text{C}$ at any depth. However, there seems to be a systematical warm bias from 0 to 450 m and a cold bias from 450 to 700 m. Almost all of the correction schemes give a similar depth variable temperature difference between XBT and CTD, but the reason is unknown. L09 does not show the depth variable difference, but it overcorrects equally at each depth. Using this independent dataset, our correction scheme results in a reduction of the systematic XBT bias.

This test gives us the opportunity to test the sensitivity of the strategy used to obtain a reference CTD. Previously, we averaged all of the CTD profiles around an XBT profile together to get the reference CTD buddy; an alternative choice is to use the nearest CTD profile to a given XBT profile as a reference buddy. The results of both choices are shown in Fig. 15, indicating very similar results using the two strategies.

d. Evaluation by checking the temperature dependency of XBT bias

How effective are the various correction schemes at reducing the temperature dependency in the XBT bias? Using the global-scale pairs, we binned the temperature differences between the XBT and CTD buddies at each 10-m depth from 0 to 700 m into 0.5°C water temperature bins (Fig. 16). The median and mean of the temperature differences at each temperature bin are calculated. An increase in XBT bias (median) with temperature is apparent before applying any corrections, with a slope of $0.0151^{\circ} \pm 0.0013^{\circ}\text{C} (^{\circ}\text{C})^{-1}$. This slope is much stronger than that found in the side-by-side dataset. When only applying the temperature-related part of our correction scheme, this temperature dependency is reduced by 22%. When both the temperature-related and the time-related parts of our correction scheme are applied, this temperature dependency is reduced by 34% [$0.0101 \pm 0.0010^{\circ}\text{C} (^{\circ}\text{C})^{-1}$]. When other correction schemes are applied, the temperature dependency is reduced by 27% (GD11), 28% (G12), 15% (CW13), 19% (CW13-CH), 20% (L09), 14% (IK09), 40% (GR10), 7% (H12), and 17% (W08). GR10 and the scheme presented in this study result in a larger reduction of temperature dependency than the other schemes. However, even after bias corrections, temperature dependency can still be

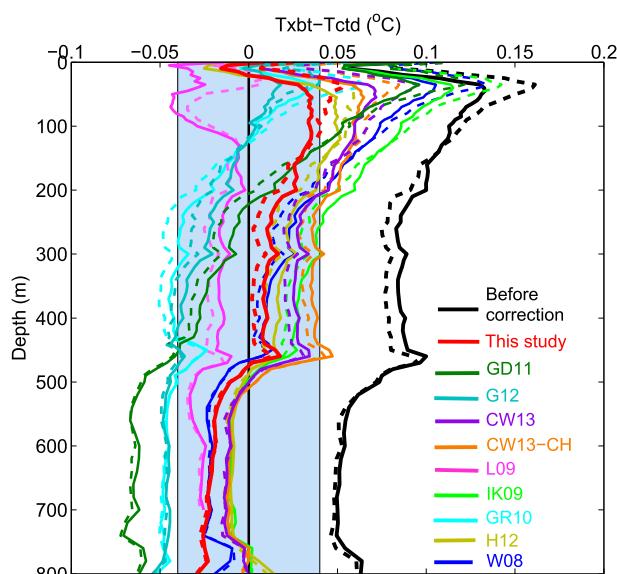


FIG. 15. Temperature difference between XBT and CTD buddies in the EN3-v2a dataset before (black) and after bias corrections by using different correction schemes (colors). CTD buddies are created by using two pairs-detection strategies: one averages all CTD profiles around a XBT profile together (solid) and the other one uses the nearest CTD profile (dashed). The temperature interval from -0.04° to 0.04°C is shaded in light blue.

found, suggesting that there may be some unknown temperature-varying errors.

7. Re-estimation of global ocean heat content

After applying our corrections to the XBT data in the WOD09 dataset, the historical ocean heat content from 1967 to 2010 is reestimated according to the methodology presented in L09. XBT data are from the World Ocean Database as of January 2011, which includes data to the end of December 2010. Quality-control procedures for the XBT data follow the procedures outlined in Boyer and Levitus (1994). XBT data are corrected by using three alternative correction schemes: L09, CW13, and this new scheme. All available data (XBT, MBT, CTD, bottle, moored buoy, glider, undulating CTD, drifting buoy) were used to obtain yearly anomalies and then calculated by subtracting standard-level interpolated values from mean monthly climatology, and averaging over 1° grid boxes.

The yearly anomalies were objectively analyzed to map to the full-ocean grid, following L09. The mean monthly climatologies were calculated from profiling floats only, to remove any possible effect due to XBT bias. Float cycles used were collected from 2004 to 2011, delayed-mode quality controlled when possible,

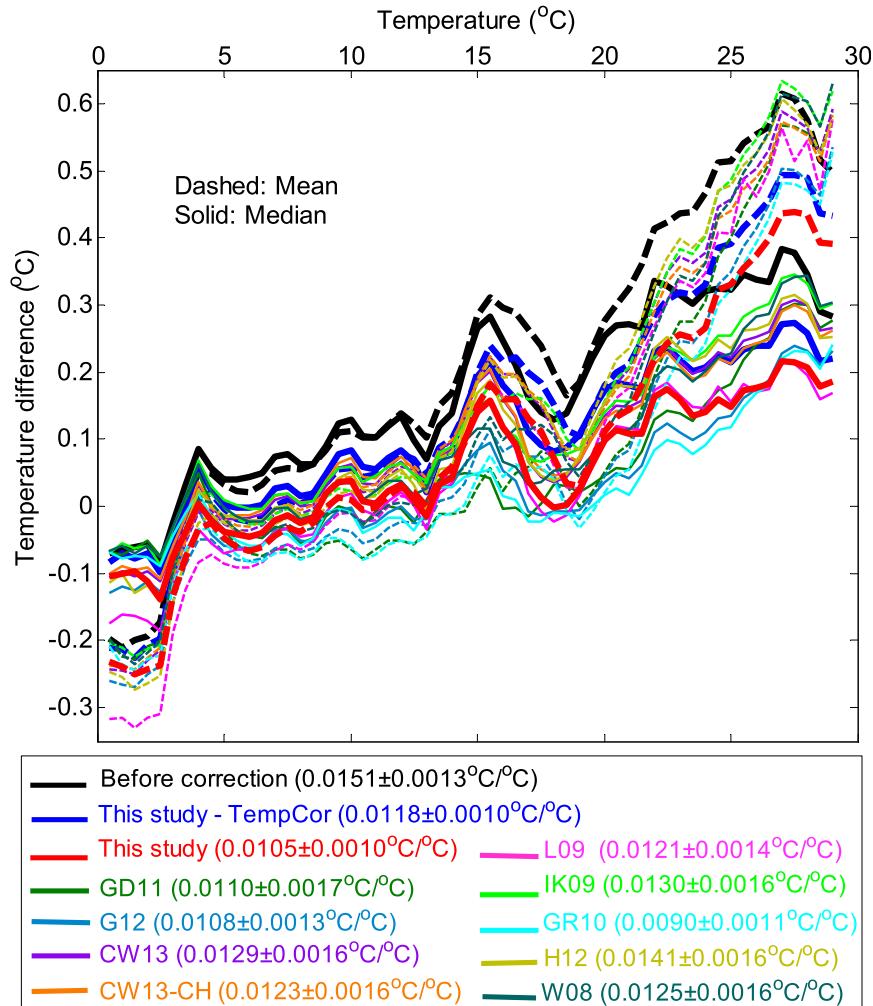


FIG. 16. Temperature differences from the global-scale dataset between XBT and CTD vs water temperature before (black) and after various XBT bias corrections (colors). Temperature differences at each standard depth are binned into 0.5°C bins and then the mean (dashed) and median (solid) of temperature differences at each temperature bin are calculated. The linear trend is attached with a 95% confidence interval.

real-time data otherwise. Temperature anomaly fields were converted to ocean heat content using a volume of water dependent on vertical length halfway between adjacent standard depth levels. Ocean heat content was then summed over all layers, 0–700-m depth, to arrive at the final yearly ocean heat content values.

The OHC under three correction schemes shows consistent structures as in Fig. 17. The warming “hump” during 1970–85 is reduced compared with both raw data and with CW13 corrections. The linear trend of the OHC changes from 1967 to 2010 is about ~ 0.36 ($10^{22} \text{ J yr}^{-1}$) based on our results. The new correction scheme gives a robust confirmation of the so-called global warming of ocean water, due to a better

evaluation of the critical period from 1960 to 1990, when XBT data dominated the ocean observation system.

8. Summary

The systematic biases in historical XBT data were examined by comparing XBT with CTD profiles in two datasets: side-by-side XBT–CTD comparisons and global-scale XBT–CTD pairs extracted from the *WOD09* dataset. The combination of the two datasets provided a new insight into the XBT bias history.

We first suggested in this paper that the structure of the original Sippican FRE can be improved: the addition of an offset term helps to better reproduce the XBT

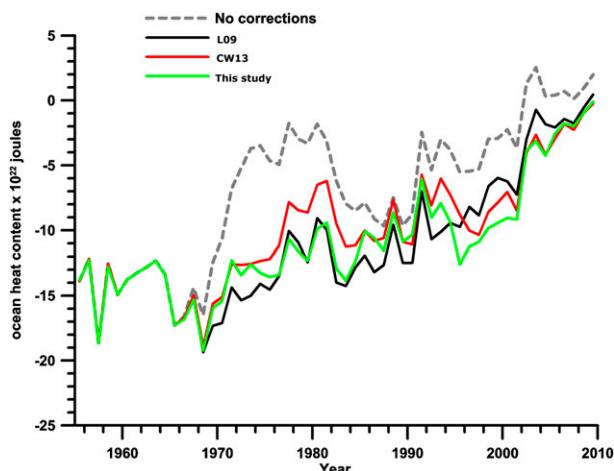


FIG. 17. Re-estimation of historical OHC from 1967 to 2010 after XBT corrections proposed in this study (green), compared with that with XBT raw data (dashed gray), and corrections by L09 (black) and CW13 (red).

probe motion since the beginning. Moreover, we proposed and verified that the three coefficients in this new FRE are correlated with each other. We accordingly hypothesized that the A - B and A -offset correlations are part of the intrinsic nature of the XBT FRE. So, a new approach was presented to correct the XBT depths using only one fall-rate coefficient A , where the other two coefficients B and offset were calculated from A based on their correlation with A . In such an approach, the number of unknown variables needed to describe the XBT motion was reduced from four (A , B , offset, and pure thermal bias) to two (A and pure thermal bias). A test based on side-by-side dataset using this new description indicated the model works well, enabling us to apply the correlations to the global-scale dataset. We also verified that A increases slightly with water temperature in the side-by-side dataset. Then, this relationship was applied to the global-scale dataset due to its influence on the local variation of the measured bias.

By using the global-scale XBT-CTD pairs dataset, we obtained robust indications of a different temporal evolution of the XBT falling motion dependent on probe type. Similar characteristics to previous studies were reproduced: the XBT falls slower than the H95 FRE states during 1975–85 and after 2000. The most likely source of fall-rate variation over time is changes in the probe design by the manufacturer (such as probe weight, shape, dimension, etc.). A time variable pure thermal bias was also obtained; shallow and deep XBTs have a warmer period before 1985 as large as 0.2°C , but T10 and T5 probes have different thermal bias histories compared with other Sippican probes. This could be

partially linked to the use of analog recording systems as well, as noted in the XBT bias literature. In addition, the different influence of even digital recording systems on the quality of the XBT profiles became evident after the field tests described by Wright and Szabados (1989). The TSK-T4 and TSK-T5 groups showed a thermal bias shift near 1998–99, while TSK-T7 showed a consistent thermal bias over time. Comparisons among several independent studies showed a similar bias history for XBT profiles, but all results had a large variation.

Therefore, a new time-, temperature-, and probe-type-dependent correction scheme was proposed to correct the historical XBT data. As validation of this new approach, we corrected the XBT data in the global-scale XBT-CTD dataset: the so calculated thermal biases are significantly reduced when compared to previous results. We checked the performance and the robustness of the new correction scheme via an objective performance test based on two simple statistics: the median of the temperature differences between XBT and CTD, and the skew of their distribution. Ideally, both these values need to be minimized so that any systematic biases are removed. We applied this technique to nine different correction schemes available in the literature. We noted that all schemes improved the median of the thermal differences, but the skew has marginally significant improvements. These metrics provide a new insight into the behavior of the biases of each probe type.

The XBT data in an alternative, fully quality-controlled dataset (EN3-v2a) was corrected by using various correction schemes. The results suggest all of the correction schemes do not perfectly correct the EN3 dataset; the best scheme is the one proposed in this study. After correction, the 0–700-m averaged bias is eliminated, and the bias at each depth is also reduced to $\pm 0.04^{\circ}\text{C}$.

Based on the results presented in this study, XBT biases in both the WOD09 and the EN3-v2a dataset can be reduced by using our proposed correction scheme. Application of this new correction scheme to an estimation of the global ocean heat content was evaluated. This re-estimation of global ocean heat content from 1967 to 2010 results in a robust warming trend.

Acknowledgments. This work is supported by the project “Structures, Variability and Climatic Impacts of Ocean Circulation and Warm Pool in the Tropical Pacific Ocean” of the National Basic Research Program of China (Grant 2012CB417404), National Natural Science Foundation of China (Grant 41075064), and Key Technologies R&D Program of China (Grant 2011BAC03B02).

APPENDIX A

Possible Reasons for Correlations in *A–B* and *A–offset*

The correlations suggested in section 3c indicate that the three fall-rate coefficients are not independent of each other. Therefore, we assume that the correlations are an intrinsic feature of the XBT fall rate, and then we used their correlations in the global-scale dataset. It is impossible now to directly verify this assumption; instead, in this section, we discuss some possible reasons for the correlations.

a. Behavior of the noise?

CH11 simulated an XBT profile by using the manufacturer’s FRE to calculate the depth and sampled the temperature from a CTD profile using H95’s FRE. In this way, the depth error is the only error source. The simulated XBT profile is corrected using both the integrated and H95 methods. Both sets of results show that *A* and *B* are distributed along a line in the *A–B* plane. As discussed by H95, the coefficients *A* and *B* are distributed in a statistical ellipse on the *A–B* plane, and the major axis of the ellipse lies along the line of the least depth error. The noises in the model or data force the resultant coefficients to distribute in an elliptical shape, which is similar to the correlation of the two coefficients we presented in Fig. 4.

As suggested in both analyses, if there is no surface offset, the correlation between *A* and *B* is likely to be related to the behavior of the noise, due to the inaccuracies in the fall-rate model, to improper methods, or noise from the measurements. If something different with respect to the description of the “standard” two-coefficient FRE happens in the near-surface layer (during the first seconds of the XBT motion), then the three coefficients will also be distributed in an ellipsoid following the math in H95, and a correlation still exists.

b. Terminal velocity error?

An offset term in an FRE sometimes appears in the XBT literature. If we exclude early “step” descriptions (Heinmiller et al. 1983; Seaver and Kuleshov 1982), then it is since the 1990s that this option was proposed as a way to improve the description of the first seconds of the XBT motion (Biggs 1992; Boyd and Linzell 1993; Hallock and Teague 1992; Prater 1991; Reseghetti et al. 2007; Singer 1990). The motivation for including this term is to model the depth error in the upper-40 m, accounting for the dynamic adjustment of the probe in the first few seconds in the water. Tests conducted in very shallow waters show that the probe is accelerating in the

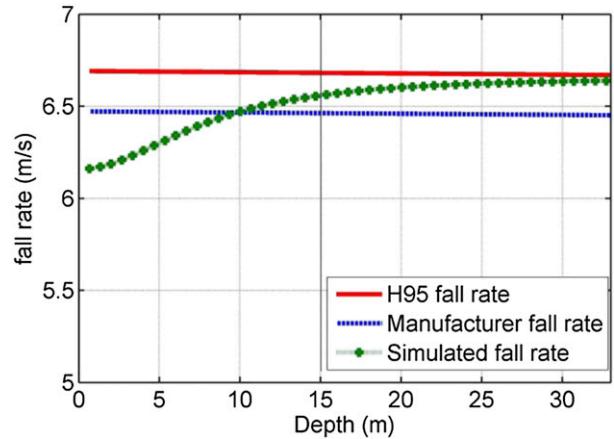


FIG. A1. Fall rate in the first 35 m. Simulated probe velocity in the upper ocean (0–35 m) calculated according to the real-ocean test presented in GR10 (green dotted), H95 FRE (red solid), and manufacturer FRE (blue).

first 3 s (GR10) until it reaches a stable fall rate. Whatever the physical explanation of the surface offset, all of the factors lead to a variable velocity in the surface layer compared with a constant velocity *A*. We now introduce a simple conceptual model to examine the effect of this start-velocity error on the calculation of the fall-rate equation.

Assuming that the actual fall rate of the probe in the first 3 s ($0 s < t < 3 s$) in the water is modeled by $v_1(t) = A_1 - 2B_1t - m/(t^2 + s^2)$, where *t* is the elapsed time after the probe hits the sea surface; v_1 is the probe velocity, which varies with *t*; the term $m/(t^2 + s^2)$ models the velocity in the first seconds, apparent as an acceleration when $m > 0$; and *s* is a constant coefficient modeling the velocity change.

The XBT depth according to the actual fall rate is

$$D_1 = A_1t - B_1t^2 - (m/s) \tan^{-1}(t/s),$$

$$\approx A_1t - B_1t^2 - (m/s)[(t/s)^1/1 - (t/s)^3/3 + (t/s)^5/5 - \dots],$$

where the series expansion is used to simplify the function of $\tan^{-1}(t/s)$. If we summarize the higher-order terms $[(t/s)^3/3 + (t/s)^5/5 - \dots]$ in D_1 and denote the sum as *h*, then the depth D_1 can be simplified as:

$$D_1 = A_1t - B_1t^2 - mt/s^2 - mh/s,$$

$$= (A_1 - m/s^2)t - B_1t^2 - mh/s.$$

The coefficients offset and *A* will be offset = mh/s and $A = A_1 - ms^{-2}$, respectively, which shows that the two coefficients are linked by the terminal velocity error [modeled by $-m/(t^2 + s^2)$]. Therefore, if there is a negative

start-velocity error, then this error will finally lead to a positive offset term and an A -offset correlation when calculating the FRE.

To further clarify this model, we tried to simulate the tests described in GR10, where results of the XBT probes dropped in a shallow area (bottom depth down to about 50 m) are shown. They find that the average probe velocity changes from 6.1–6.2 m s⁻¹ in the first 5.5 m to 6.45–6.6 m s⁻¹ at 47.2 m. In our simulation, the coefficients in fall rate v_1 are set to be $m = 0.8$, $s^2 = 1.5$, $A_1 = A_{H95} = 6.691$ m s⁻¹, and $B_1 = B_{H95} = 0.00225$ m s⁻² to get a similar fall-rate structure in the first 30 m as in GR10. Our simulated fall rate is shown as the green line in Fig. A1. According to the math presented above, we can calculate the depth error compared with H95 at 5 s (~35 m) as

$$\begin{aligned} D_{\text{error}} &= \int_0^4 (v_{H95} - v_1) dt = \int_0^4 (A_{H95} - 2B_{H95}t - v_1) dt \\ &= \int_0^4 m/(t^2 + s^2) dt = (m/s) \tan^{-1}(4/s) \approx 0.83 \text{ m}. \end{aligned}$$

The obtained depth error at 4 s (~30 m) is ~0.83 m, about half of the mean surface offset calculated from the side-by-side data (~1.8 m as in Fig. A2a), a value nearly coincident with the correction obtained in a different way in Reseghetti et al. (2007), where a shift up of about 2.0 m (or 0.3 s) was proposed.

In the real world, the surface velocity will not be exactly represented by $v_1(t) = A_1 - 2B_1t - m/(t^2 + s^2)$; instead, the velocity in the first seconds seems to be a complex function of elapsed time (in a typical standard FRE, the velocity is linearly decelerating with time since the beginning of the XBT motion). However, the start-velocity error will contribute to the A /offset correlation.

c. Start-time error?

The start-time error will occur if the XBT data recording system begins working before the probe hits the sea surface. Real ocean conditions are complex with surface waves, spray, bubbles (both natural and induced by the ship motion), etc.—all of them can lead to a false early recording start. Start time might also be affected by wake produced by the motion of the hull, creating both waves and turbulences in the water. Unfortunately, it is almost unrealistic to drop the XBT probe into unperturbed water without a drop having a significant component of the motion parallel to the sea surface (something like a javelin throw). On the other hand, the thermistor in the XBT probe has a time constant of about 0.10–0.15 s and requires around 0.60 s to detect

a thermal signal (Reseghetti et al. 2007 and references therein). A start-time error near the sea surface (denoted as dt) can contribute to the correlation between A and offset. In detail,

If the manufacturer's FRE is used, then the FRE can be rewritten as $D = A(t - dt) - B(t - dt)^2 = At - Bt^2 - (Adt - 2dt \times t + Bdt^2)$.

We define as the offset term: $\text{offset} = Adt - 2dtBt + Bdt^2 \approx Adt$, which results in an apparent correlation between offset and A .

In the case where surface water sprays induce a fake contact, the acquisition system will start before the probe hits the water's surface, and dt is always positive. If the dt is set to 0.15 s, then the resultant offset will be about 0.9–1 m, also approximately half of the systematical bias we detect in the side-by-side data (~1.8 m). If there is a water spray in the ocean with height of ~0.5 m above the sea surface, and the probe begins to work when splashed with the water at a velocity of 7 m s⁻¹, then there will be a time error of about ~0.072 s. This time error leads to an offset of ~0.5 m.

Although there is no direct observational evidence confirming such an error, it is possible that the start-time error can contribute to the offset term.

APPENDIX B

Impact of the offset Term on A - B correlation

We assumed that the FRE with a positive offset term better describes the actual fall rate of a probe, with an average depth offset of about ~1.8 m. What if the FRE does not contain an offset term? To test the effect of the offset term, we use the FRE without an offset term to correct the side-by-side data using the integrated method. The terminal velocity without the offset term (median $A = 6.53$ m s⁻¹) is smaller than that with offset term ($A = 6.64$ m s⁻¹) by 0.11 m s⁻¹ (Fig. A2a). A comparison of the A - B correlation with and without the offset term (Fig. A2b) shows the A - B correlation has a greater slope in the case of zero offset at the surface. We present some simple math to explain this slope shift:

We suppose that the actual fall-rate equation is $D = At - Bt^2 - C$, where $A = 6.64$ m s⁻¹, $B = 0.00228$ m s⁻², and $C = 1.8$ m are the medians of all of the FRE coefficients with the offset term; see Fig. A2a. The manufacturer FRE is $D_1 = A_1t - B_1t^2$. If one uses this manufacturer's FRE (D_1) to calculate the actual depth (D), the depth error has to be minimized as $D_{\text{error}} = D_1 - D = 0$, where D_{error} is defined as

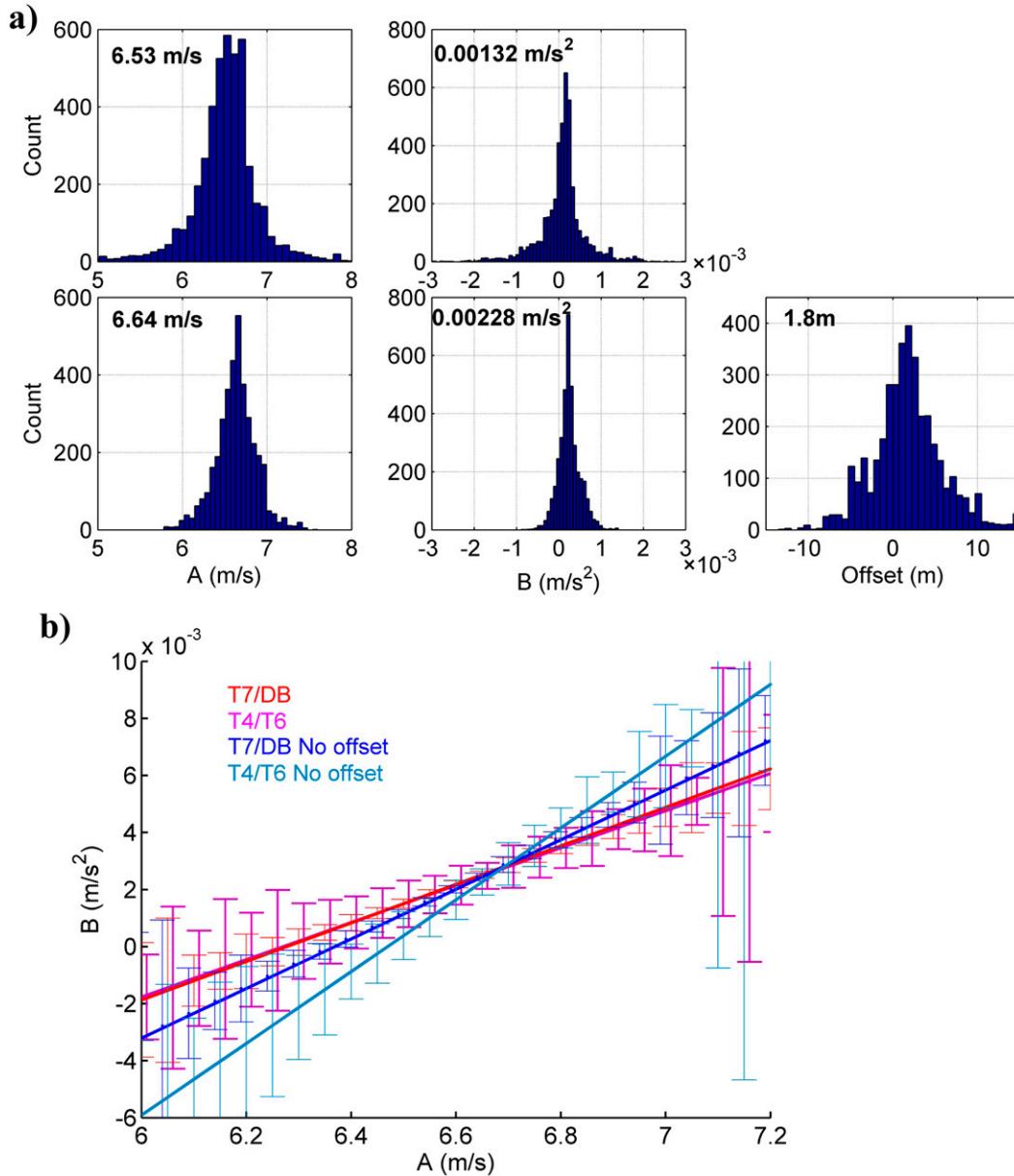


FIG. A2. Comparison of coefficients for manufacturer FRE without offset term and calculated FRE with offset term: (a) Basic statistics of the fall-rate coefficients A and B using the manufacturer FRE form (top) and the coefficients A , B and offset in our FRE (bottom). (b). Comparison of A – B correlations with (red and purple) and without (blue and cyan) the offset term for both T4–T6 and T7–DB probe types. Error bars are two standard errors.

$$D_{\text{error}} = \int_0^T [(A - A_1)t - (B - B_1)t^2 - C] dt$$

$$= (A - A_1)T^2/2 - (B - B_1)T^3/3 - C \times T,$$

where T is the acquisition time (the time when a probe reaches its terminal depth). If we simplify the slope of the A – B correlation to $s = B/A$ and $s_1 = B_1/A_1$ for the two different FRE models D and D_1 , respectively, then the slope $s = 0.00228/6.64 = 0.000343$ with units of

1 (s)^{-1} at $T = 120 \text{ s}$ (the acquisition time of the T7–DB probe at a depth of $\sim 760 \text{ m}$). Then we impose the minimization of $D_{\text{error}} = 0$, which is $(B/s - B_1/s_1)T^2/2 - (B - B_1)T^3/3 - C \times T = 0$. So, s_1 is obtained by the following equation: $s_1 = 1/[B/sB_1 - 2(B - B_1)T/3B_1 - 2C/BT]$.

This solution of s_1 has dependence on B_1 . Since the coefficient B is mainly due to the probe weight loss during its fall, we assume B is independent of the variant choice of FRE. When B_1 has the same value as B ,

$s_1 = 1/(B/sB_1 - 2C/BT)$, is always greater than the slope s ; this explains the slope difference when the offset term is included in Fig. A2b. For T4–T6, T is 80 s, so the slope s_1 for T4–T6 will be greater than s_1 for T7–DB under the same conditions, as shown in Fig. A2b. The resultant $A_1 = B_1/s_1$ will always be smaller than $A = B/s$, again explaining the differences in A (Fig. A2a).

Based on this test, we conclude that if there is no offset term in the FRE, then the surface offset can lead to the stronger A – B correlation and smaller A to compensate for the difference at the surface.

REFERENCES

- Abraham, J., J. Gorman, F. Reseghetti, E. Sparrowa, J. Stark, and T. Shepard, 2014: Modeling and numerical simulation of the forces acting on a sphere during early-water entry. *Ocean Eng.*, **76**, 1–9, doi:10.1016/j.oceaneng.2013.11.015.
- Anderson, E. R., 1980: Expendable bathythermograph (XBT) accuracy studies. Naval Ocean Systems Center Tech. Rep. 550, 201 pp.
- Bailey, R., A. Gronell, H. Phillips, E. Tanner, and G. Meyers, 1994: Quality control cookbook for XBT data. CSIRO Marine Laboratories Rep. 221, 83 pp.
- Biggs, D. C., 1992: Nutrients, plankton, and productivity in a warm-core ring in the western Gulf of Mexico. *J. Geophys. Res.*, **97**, 2143–2154, doi:10.1029/90JC02020.
- Boyd, J. D., and R. S. Linzell, 1993: Evaluation of the Sparton tight-tolerance AXBT. *J. Atmos. Oceanic Technol.*, **10**, 892–899, doi:10.1175/1520-0426(1993)010<0892:EOTSTT>2.0.CO;2.
- Boyer, T., and S. Levitus, 1994: Quality control and processing of historical oceanographic temperature, salinity and oxygen data. NOAA Tech. Rep. NESDIS 81, 64 pp.
- , and Coauthors, 2009: *World Ocean Database 2009*. S. Levitus, Ed., NOAA Atlas NESDIS 66, 216 pp.
- Cheng, L. J., J. Zhu, F. Reseghetti, and Q. P. Liu, 2011: A new method to estimate the systematic biases of expendable bathythermograph. *J. Atmos. Oceanic Technol.*, **28**, 244–265, doi:10.1175/2010JTECH0759.1.
- Church, J. A., and Coauthors, 2011: Revisiting the Earth's sea-level and energy budgets from 1961 to 2008. *Geophys. Res. Lett.*, **38**, L18601, doi:10.1029/2011GL048794.
- 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, doi:10.1175/JTECH-D-12-00127.1.
- DiNezio, P. N., and G. J. Goni, 2011: Direct evidence of a changing fall-rate bias in XBTs manufactured during 1986–2008. *J. Atmos. Oceanic Technol.*, **28**, 1569–1578, doi:10.1175/JTECH-D-11-00017.1.
- Domingues, C. M., J. A. Church, N. J. White, P. J. Gleckler, S. E. Wijffels, P. M. Barker, and J. R. Dunn, 2008: Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, **453**, 1090–1096, doi:10.1038/nature07080.
- Flierl, G. R., and A. R. Robinson, 1977: XBT measurement of thermal gradients in the MODE eddy. *J. Phys. Oceanogr.*, **7**, 300–302, doi:10.1175/1520-0485(1977)007<0300:XMOTGI>2.0.CO;2.
- Good, S. A., 2011: Depth biases in XBT data diagnosed using bathymetry data. *J. Atmos. Oceanic Technol.*, **28**, 287–300, doi:10.1175/2010JTECH0773.1.
- Gouretski, V., 2012: Using GEBCO digital bathymetry to infer depth biases in the XBT data. *Deep-Sea Res. I*, **62**, 40–52, doi:10.1016/j.dsr.2011.12.012.
- , and K. P. Koltermann, 2007: How much is the ocean really warming? *Geophys. Res. Lett.*, **34A**, L01610, doi: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, doi:10.1016/j.dsr.2010.03.011.
- Green, A. W., 1984: Bulk dynamics of the expendable bathythermograph (XBT). *Deep-Sea Res.*, **31A**, 415–426, doi:10.1016/0198-0149(84)90093-1.
- Hallock, Z. R., and W. J. Teague, 1992: The fall rate of the T-7 XBT. *J. Atmos. Oceanic Technol.*, **9**, 470–483, doi:10.1175/1520-0426(1992)009<0470:TFROTT>2.0.CO;2.
- Hamon, M., G. Reverdin, and P.-Y. Le Traon, 2012: Empirical correction of XBT data. *J. Atmos. Oceanic Technol.*, **29**, 960–973, doi:10.1175/JTECH-D-11-00129.1.
- Hanawa, K., P. Rual, R. Bailey, A. Sy, and M. Szabados, 1995: A new depth–time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). *Deep Sea Res. I*, **42**, 1423–1451, doi:10.1016/0967-0637(95)97154-Z.
- Heinmiller, R., C. Ebbesmeyer, B. Taft, T. Olson, and O. Nikitin, 1983: Systematic errors in expendable bathythermograph (XBT) profiles. *Deep-Sea Res.*, **30A**, 1185–1197, doi:10.1016/0198-0149(83)90096-1.
- Ingleby, B., and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles—Historical and real-time data. *J. Mar. Syst.*, **65**, 158–175, doi:10.1016/j.jmarsys.2005.11.019.
- Ishii, M., and M. Kimoto, 2009: Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *J. Oceanogr.*, **65**, 287–299, doi:10.1007/s10872-009-0027-7.
- Kizu, S., and K. Hanawa, 2002a: Recorder-dependent temperature error of expendable bathythermograph. *J. Oceanogr.*, **58**, 469–476, doi:10.1023/A:1021261214950.
- , and —, 2002b: Start-up transient of XBT measurement. *Deep-Sea Res. I*, **49**, 935–940, doi:10.1016/S0967-0637(02)00003-1.
- , S. I. Ito, and T. Watanabe, 2005: Inter-manufacturer difference and temperature dependency of the fall-rate of T-5 expendable bathythermograph. *J. Oceanogr.*, **61**, 905–912, doi:10.1007/s10872-006-0008-z.
- , H. Onishi, T. Suga, K. Hanawa, T. Watanabe, and H. Iwamiya, 2008: Evaluation of the fall rates of the present and developmental XCTDs. *Deep-Sea Res. I*, **55**, 571–586, doi:10.1016/j.dsr.2007.12.011.
- , C. Sukigara, and K. Hanawa, 2011: Comparison of the fall rate and structure of recent T-7 XBT manufactured by Sippican and TSK. *Ocean Sci.*, **7**, 231–244, doi:10.5194/os-7-231-2011.
- Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, doi:10.1029/2008GL037155.
- , and Coauthors, 2012: World Ocean heat content and thermocline sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.*, **39**, L10603, doi:10.1029/2012GL051106.
- Lyman, J. M., S. A. Good, V. V. Gouretski, M. Ishii, G. C. Johnson, M. D. Palmer, D. M. Smith, and J. K. Willis, 2010: Robust warming of the global upper ocean. *Nature*, **465**, 334–337, doi:10.1038/nature09043.

- Palmer, M. D., and K. Haines, 2009: Estimating oceanic heat content change using isotherms. *J. Climate*, **22**, 4953–4969, doi:10.1175/2009JCLI2823.1.
- Prater, M. D., 1991: A method for depth and temperature correction of expendable probes. *J. Atmos. Oceanic Technol.*, **8**, 888–894, doi:10.1175/1520-0426(1991)008<0888:AMFDAT>2.0.CO;2.
- Reseghetti, F., M. Borghini, and G. M. R. Manzella, 2007: Factors affecting the quality of XBT data—Results of analyses on profiles from the western Mediterranean Sea. *Ocean Sci.*, **3**, 59–75, doi:10.5194/os-3-59-2007.
- Reverdin, G., F. Marin, B. Bourles, and P. L’Herminier, 2009: XBT temperature errors during French research cruises (1999–2007). *J. Atmos. Oceanic Technol.*, **26**, 2462–2473, doi:10.1175/2009JTECHO655.1.
- Seaver, G. A., and S. Kuleshov, 1982: Experimental and analytical error of the expendable bathythermograph. *J. Phys. Oceanogr.*, **12**, 592–600, doi:10.1175/1520-0485(1982)012<0592:EAAEOT>2.0.CO;2.
- Singer, J. J., 1990: On the error observed in electronically digitized T-7 XBT data. *J. Atmos. Oceanic Technol.*, **7**, 603–611, doi:10.1175/1520-0426(1990)007<0603:OTEOIE>2.0.CO;2.
- Thadathil, P., A. K. Saran, V. V. Gopalakrishna, P. Vethamony, N. Araligidad, and R. Bailey, 2002: XBT fall rate in waters of extreme temperature: A case study in the Antarctic Ocean. *J. Atmos. Oceanic Technol.*, **19**, 391–396, doi:10.1175/1520-0426-19.3.391.
- Wijffels, S. E., J. Willis, C. M. Domingues, P. Barker, N. J. White, A. Gronell, K. Ridgway, and J. A. Church, 2008: Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise. *J. Climate*, **21**, 5657–5672, doi:10.1175/2008JCLI2290.1.
- Wright, D., and M. Szabados, 1989: Field evaluation of real-time XBT systems. *Diving Safety and Physiology, Ocean Engineering/Technology*, D. Anderson, Ed., Vol. 5, *Oceans ’89: An International Conference Addressing Methods for Understanding the Global Ocean*, Publ. 89CH2780-5, IEEE, 1621–1626, doi:10.1109/OCEANS.1989.587130.