Instrument-related temperature biases and their impact on estimation of the long-term ocean temperature variability

Viktor Gouretski

Alfred-Wegener-Institute for Polar and Marine Research, Bremerhaven, Germany

---

Corresponding author address:

Viktor Gouretski, Alfred-Wegener-Institute for Polar and Marine Research,

Bussestr. 24, 24570 Bremerhaven, Germany

E-mail: viktor.gouretski@awi.de

ABSTRACT

WOD2005 global hydrographic dataset is used to estimate temperature and sample depth biases of mechanical and expendable bathythermographs (XBTs) and profiling floats by comparing temperature profiles with more accurate collocated bottle and CTD data. In agreement with earlier studies, the XBT data were found to have both temperature and depth biases. A simple bias model is suggested and applied to the original data to exclude these biases, with estimates of the depth errors being in agreement with independent XBT-CTD inter-comparison experiments. However, applying only depth corrections to XBT temperature profiles without corrections for a pure temperature bias leads to an even larger disagreement with the reference data. Using only depth-corrected XBT data not only overestimates the overall ocean warming, but also creates artificial temperature variability due to the temperature and depth biases being time-variable. The profiling float temperature biases were found to be on average smaller than 0.05^{o}C, but the more precise estimates require a more extensive set of the CTD reference data. The bottle/CTD dataset was found to be internally consistent within about 0.05^{o}C, but a more careful estimation critically depends on the availability of the respective metadata. Global heat content anomalies based on the composite hydrographic dataset including both temperature and depth corrected bathythermograph profiles do not reveal any progressive warming within the upper 1000 m layer between 1950s and mid 1990s. Only the uppermost layer exhibits a gradual warming since the end of 1950s, in a good agreement with the independent estimates of the global land and sea surface temperatures. The study demonstrates that systematic errors pose a significant problem in identifying long-term heat content variations in the Global Ocean.

1. Introduction

Estimation of the long-term temperature changes in the global ocean, a main heat reservoir of the Earth climate system, has received a growing importance during the last years as a part of the climate change issue. Temperature is the most often measured sea water characteristic, with the measurement accuracy generally believed to be superior to accuracy for other parameters, like salinity, dissolved oxygen or nutrients. Accumulation of temperature observations in the oceanographic database stimulated attempts to quantify temperature changes in the Global Ocean.

First estimates of the global ocean warming were presented by Levitus et al. (2000), where the global ocean warming of 0.31^{o}C for the upper 300 meters were reported, with the range of heat content increase on the order of $20 \cdot 10^{22}$ J between 1950s and 1990s. A basin scale warming at intermediate depth between the 1950s and the 1990s was reported for the Southern Ocean by Gille (2002), who compared historical hydrographic data with temperatures measured by the free drifting floats during the 1990s, revealing a 0.17^{o}C warming of the Southern Ocean at depths between 700 and 1000 meters. Using additional temperature profiles Levitus et al. (2005) (hereafter LAB2005) obtained new estimates of a progressive warming of the Global Ocean with an increase of the heat content of 14.5×10^{22}J between 1957 and 1997 for the upper 3 km layer. Lyman et al. (2006) presented estimates of global upper-ocean heat content anomaly from 1993 through 2005 and found a cooling event between 2003 and 2005, with a net loss of 3.2 ($\pm$1.1) x 10^{22} J of heat. However, the reported cooling was later found to be a data artefact, as instrumentation flaws for some Argo floats have been discovered (Nature, 2007).

Growing interest to the problem of the global warming requires a more careful quality assessment of the sparse, inhomogeneous and irregular oceanographic data. Instruments for measuring temperature and other sea water parameters have undergone significant changes since the beginning of the routine observations of the World Ocean, with mechanical instruments being supplanted by more precise electronic devices. However, this change in ocean instrumentation was gradual, so that since 1960s the global oceanographic data base is characterized by a mixture of both mechanical and electronic instruments. In contrary to the sea surface temperature data, where temperature biases were identified and corrections suggested (Folland and Parker, 1995), biases in the subsurface hydrographic data are much less documented.

First estimates of systematic errors data for large hydrographic datasets were obtained by Johnson et al. (2001) and Gouretski and Koltermann (2001). Using a set of cruises, property offsets were calculated in the deep water (below about 2 km) for cross-over areas. However, in both studies only salinity, dissolved oxygen and nutrient inter-cruise offsets have been quantified, with temperature data assumed to be error-free.

The problem of systematic errors in the temperature data and their possible effect on global heat content estimates was first raised by Gouretski& Koltermann (2007). They found the Expendable Bathythermograph (XBT) data to have a significant positive temperature bias, an artefact which obviously leads to an overestimation of the ocean warming. This temperature bias has also exaggerated the amplitude of the decadal scale heat content anomaly variations, creating an artificial positive heat content anomaly between 1973 and 1982. Detection of biases both in XBT and profiling float data shows convincingly that any “blind” use of data from different instrument types or even one of the same type without a proper check for systematic errors contains a major risk for global climatologic estimates.

In this study we aim to assess the effect of systematic errors on the computation of temperature and heat content anomalies. Instrument specific systematic temperature offsets are calculated relative to more accurate “classical” hydrographic data (bottle data plus Conductivity/Temperature/Depth (CTD) data), used as a reference. A method for correcting bathythermograph temperature profiles is suggested and new estimates of temperature and heat content anomalies, based on the corrected data, are presented for the time period from 1950 through 2003.

2. Data

a. Data sources

The last update of the NODC hydrographic data collection –World Ocean Atlas 2005 (WOA2005) (Locarnini et al., 2006) **-** provides the main temperature dataset for this study. It was complemented by a complete set of the North Atlantic profiling float (Argo) data as of the end 2006. Observed level temperature profiles were interpolated on a set of 22 levels between the surface and 1000m (the selected levels are identical to those used in the WOCE Hydrographic Climatology (Gouretski and Koltermann, 2004)). Quality control of the data included crude range check (according to temperature ranges given in Boyer and Levitus, 1994) and a statistical check in 1x2-degree squares. As (1) the amount of available data decreases rapidly with depth and (2) because the most abundant data types (Mechanical (MBT) and Expendable (XBT) bathythermographs) are concentrated in the upper 1000 meters only this layer was selected for the analysis (pure surface temperature observations were not included). The subsurface temperature data in this analysis span the time period between 1947 and 2005.

b. Instrument types

- A variety of instruments have been used to measure subsurface temperature in the ocean, with instruments evolving from the exclusive use of mechanical devices (before mid-1960s) to electronic sensors and sampling instruments (since mid-1960s)**.** The majority of the subsurface temperature data in the WOA2005 dataset was obtained by means of five instrument types: 1) reversing thermometers attached to hydrographic bottles (Nansen type bathometers) 2) Mechanical bathythermographs (MBT); 3) Expandable bathythermographs (XBT), 4) Conductivity-Temperature-Depth (CTD) devices, and 5) profiling floats. Systematic errors of the temperature data from these instrument types are investigated in this study.

1) NANSEN BOTTLES

Nansen bottles were the historically first devices designed for taking deep water samples and measuring temperature at depth. The first detailed description of the method and its accuracy was given by Wüst (1932), based on the results of the German oceanographic expedition in the South Atlantic in 1925-27. A description of the working procedure on Nansen-Bottle stations implemented on the ships of the Woods Hole Oceanographic Institution since 1950s is given by B. Warren (2008). During the procedure known as a hydrographic cast Nansen bottles were attached to the wire and lowered in to the water. The temperature at the sampling depth was recorded by means of reversing mercury thermometers attached to the bottles, with up to 18-19 bottles usually employed during a single cast (Warren, 2008). Pairs of pressure protected and unprotected thermometers were used for the determination of sampling depths using the difference between the temperatures of the two thermometers. Since using such paired thermometers enabled sufficient accuracy only at depths greater than about 200 meters, sampling depths of shallow casts were usually calculated from the amount of the wire paid out and the angle of the wire from the vertical at the height of the hydrographic winch. Such calculations of the sampling depth were often subject to considerable errors as the real shape of the wire with attached bottles under specific local conditions (wind, currents) remained unknown. Though the use of unprotected thermometers was introduced in the beginning of the 20th century (Perlewitz, 1908), it was not until the mid 1930-th that the thermometric sample depth determination became general practice (Iselin, 1936).

2) CONDUCTIVITY-TEMPERATURE DEPTH (CTD)

Observations by means of Nansen bottles have been gradually replaced by a *CTD* - an electronic sampler and profiler for continuous measurements of temperature and salinity (conductivity) from oceanographic ships. The CTD is now a standard oceanographic instrument. Introduced in the middle-1960s CTD have almost entirely replaced Nansen bottles only during 1980s. For instance, in Woods Hole Oceanographic Institution Nansen bottles were entirely supplanted by the CTDs in 1981 (Warren, 2008). Personal communication with colleagues from the United Kingdom, France, Australia and the USA allows crudely define the transition period from Nansen casts to CTDs as 1980-85. However, in some countries which contributed substantially to the oceanographic databases (for instance, the former Soviet Union), this replacement did not obviously took place until 1990s (according to author's personal experience a numerous oceanographic fleet of the Hydrometeorological Service of the former Soviet Union still used Nansen bottles in the end of 1980s).

During the CTD cast the data are sent up through a conducting cable for digital recording and pre-processing in a deck unit**.** A Rosette sample is normally used for taking water samples. An important difference to the older Nansen cast procedure is that the bottles of the Rosette sampler can be closed at arbitrary depth, which is measured by a precise pressure sensor.

3) MECHANICAL BATHYTHERMOGRAPH (MBT)

The MBTs were introduced in oceanographic practice in the beginning of the World War II and were used for determining ocean temperatures down to approximately 250 meter depth. The MBT consists of a thermal element attached to a stylus which scribes a trace across a gold-plated or smoked slide. On the ships of the former Soviet Union gelatine-covered slides were used. A capillary copper tube filled with liquid xylene serves as a temperature sensor, that expands and contracts with temperature changes causing the stylus to move. As a pressure sensor a spring loaded piston enclosed in a flexible envelope made of brass bellows is used, with the piston head moving the slide as the pressure changes. The instrument is lowered to a depth and then winched up again with temperature values read off the trace using a special reader with a depth-temperature grid calibrated for each instrument. According to Casciano (1967), over 5000 MBTs were in use alone in the United States in middle 1960s. Only a few evaluations of MBT data quality and reliability exist (Casciano, 1967; Dinkel and Stawnichy, 1973; Stewart, 1963).

4) EXPENDABLE BATHYTHERMOGRAPH (XBT)

First XBTs appeared in 1966 and almost completely replaced MBTs till the beginning of 1990s. The XBT contains a precision thermistor located in the nose of the free falling probe, and a copper wire is used to communicate the temperature measured by the thermistor during the drop back to the ship. Temperature data are recorded and displayed in real time as the probe falls. Probe depth Z_{XBT} is determined from the fall rate equation:

$$Z_{XBT} = at + bt^2 , \qquad (1)$$

where t – is elapsed time since the moment the probe hits the water surface, and coefficients a and b are provided by the manufacturer (Sippican corporation). There are several types of XBT probes: the T-4 and T-6 models are designed for measuring temperature down to the nominal depth of 460m, the T-7 and Deep-Blue (DB) models have the maximum nominal depth of 760 m and the T-5 modification can reach the nominal depth of 1830 meters. Manufacturer provides two different fall rate equations: for the shallow probes (T-4, T-6, T-7 and Deep Blue) and for the deep modofocationT-5.

5) PROFILING FLOATS

Whereas the first autonomous profiling floats were launched in mid-1990s, their extensive implementation started within the Argo program in 1999, with the total number of floats currently amounting to about 3000, whereas a smaller amount of profiling float data was obtained also between 1996 and 1998. These floats drift freely with the ocean current at depth and rise to the surface at pre-defined time intervals. The floats are equipped with a CTD which measures temperature and conductivity while rising to the surface from a parking depth of about 2000m (Argo Project office, 2006). Unlike the ship-based CTD the laboratory calibration of the profiling float sensors is rarely possible as floats retrieval is difficult and expensive. Instead, the performance control of the float CTD is done by comparing float data with local climatological temperature and salinity data. Wong et al. (2003) note a good quality of temperature and pressure measurements, in contrary to salinity measurements which may be affected by a significant sensor drift. However, series problems with temperature measurements have been recently reported (Nature, 2007).

c. Precision of temperature and sample depth measurements

It is important to note that the global subsurface temperature dataset is inhomogeneous, with five main data types differing considerably in precision of both temperature and sample depth. Typical temperature and sample depth errors for different instrument types are summarized in the Table 1. For instrument types under consideration, calibration checks are known to be regularly conducted only on deep water reversing thermometers and CTD sensors. Temperature errors of deep sea reversing thermometers (Nansen casts) and sensor drift errors (CTD) can be eliminated and preliminary data corrected after the post-cruise calibrations are done. In case of XBTs each measuring unit – the XBT probe – is lost after completing the profile, so that correction of the registered temperatures is only possible if collocated independent accurate data are available. Accuracies for temperature and sample depths based on manufacturer specifications or available literature are listed in the Table 1, which renders the CTD data by far the best accuracy among the instrument types considered both in temperature and in depth. Nansen cast bottle data, being relatively accurate in temperature, are often subject to large sample depth errors, especially in the case when no unprotected thermometers were used during the cast.

The assessment of sample depth errors is crucial for the detection of long-term temperature variability in the ocean, since sample depth error is translated into temperature error in case of a (typically) non-zero vertical temperature gradient. To assess the magnitude of possible error the global vertical temperature gradients were calculated, using temperature gridded data from the WOCE Climatology (Gouretski and Koltermann, 2004). For a given systematic depth error of only 1 meter typical equivalent temperature error may exceed 0.2^{o}C above 200m in the regions with the strong thermocline (Fig. 1). Using average values of the vertical temperature gradient and sample depth precisions according to instruments' specifications (Table 1) equivalent temperature errors have been calculated. A total possible absolute temperature error (the sum of the pure temperature error and equivalent temperature error due to the sample depth uncertainty) is shown in Fig. 2, with CTD accuracy superior to all other instrument types throughout the whole depth range.

3. Method of the temperature bias estimation

Unlike random errors systematic errors can be identified and estimated only by comparing observations against independent standards. In this study blended CTD and Nansen bottle data were used as a reference for the time period between 1947 and 2000. As the amount of CTD/Bottle data in the database reduces significantly after about 2000, a blend of profiling float and CTD data were used as a reference for XBT data after 2000. The first step of the analysis was to construct the so called super observations – space and time averaged temperature values. Super observations were constructed separately for each instrument type. The area of the global ocean was subdivided into 1-degree zones, and each zone in turn was subdivided into equal-size squares, with the longitude side of squares changing with the zone latitude to keep the area of the geographical squares equal to that of the squares in the near-equatorial 1-degree zones. Temperature observations within each box were averaged to produce box-averaged values $\langle T_x \rangle$. The time bin for the averaging was selected to be one month.

In the following, the difference between the box-averaged instrument specific temperature and the box-averaged reference (CTD/Bottle) temperature defines the box-averaged bias $\langle \delta \rangle$ for the specific instrument type:

$$\langle T_x \rangle - \langle T_{CTD} \rangle = \langle \delta \rangle + \varepsilon \,, \tag{2}$$

where the random uncertainty ε is mostly due to the within-box synoptic variability, since reference profiles and profiles for the specific instrument type do not exactly overlap in space and time. To estimate mean box averaged uncertainty ε we calculated within-box standard errors for CTD and Bottle data, using observed level profile dataset from the WGHC climatology (Gouretski and Koltermann, 2004). Global average of these standard errors (Fig. 3) gives an approximately linearly decreasing mean ε value from about 0.35^{o}C in the near surface layers to less than 0.1^{o}C below 900 m.

Finally, at each standard level for each month the global mean systematic error $\{\langle \delta \rangle\}$ was estimated as the average $\{\ldots\}$ over all overlapping boxes, allowing averaging out the random eddy noise:

$$\{\langle \delta \rangle\} = \{\langle T_x \rangle\} - \{\langle T_{CTD} \rangle\}. \tag{3}$$

The number of boxes overlapping each month with reference boxes (Fig. 4) varies considerably between the instrument types, with the best overlapping being for mechanical bathythermographs and for the shallow XBT modifications T-4 and T-6. There is a reduction in the number of overlapping boxes for most of the data types due to a rapid decrease of the CTD data percentage in the data base after about 1993. With a strong geographical bias typical for all instrument types (sampling mostly in the Northern Hemisphere) considerable seasonal variations are observed in the number of overlapping boxes for all instrument types, with a typical 2 to 3-fold reduction from summer to winter. Assuming uncorrelated errors ε one can expect typically a 5- to 20-fold reduction of the globally averaged uncertainty $\{\varepsilon\}$ for monthly averaged offset values compared to the mean box uncertainty ε as shown in Fig. 3. The spatial distribution of the overlapping boxes is sparse and inhomogeneous, with the highest data concentration in the North Atlantic and North Pacific oceans. Fig. 5 gives an example of the mean temperature bias relative to reference CTD/Bottle data at 100 meter level for two selected pentads 1970-74 and 1985-89.

4. XBT biases

Though the XBT temperature data dominate in the upper 700 meters since the 1970s, these instruments are the most error prone oceanic observing system with about 15% of XBTs suffering instrument malfunctions before reaching 250 m (McPhaden et al., 1998). Two main problems specific to the XBT data are reported in the literature: 1) inadequacy of the manufacturer's XBT fall-rate equation and (2) pure temperature biases. In the presence of a (typically) non-zero vertical temperature gradient pure temperature biases are coupled with equivalent temperature errors due to sample depth errors. Following sources of errors in XBT data were reported in a new study of the XBT data quality by Reseghetti et al. (2007): a) XBT probe mass variations, b) differences in the launching height, c) start-up transients due to a different time response of the recording system, d) pure temperature bias of the probe thermistor, e)inadequacy of the fall-rate equation coefficients.

a. XBT-CTD inter-comparison experiments

A considerable number of XBT versus CTD inter-comparison experiments have been conducted since mid-1970s in different regions of the World Oceans in order to check the validity of the manufacturer's fall rate equation coefficients (Table 2). All these independent studies of the XBT fall velocity agree in that the XBT depths based on a manufacturer fall-rate equation coefficients *underestimate* the true sample depths for XBT types T-4, T-6, and T-7 (probes fall faster than stated by he manufacturer). Only for the Japanese analogues of the T-5 type a slower fall rate was reported (Kizu et al., 2005). Hanawa et al. (1994) (here on HRBSS) and Hanawa et al. (1995) made the most detailed study of the XBT fall-rate velocity for the instrument types T-4, T-6 and T-7, based on the XBT-CTD inter-comparison experiments in different regions of the World Ocean between 1985 and 1992. These regional inter-comparisons resulted in a new fall-rate equation, which was recommended as a substitution of the fall rate equation provided by the manufacturer. It should be noted that HRBSS method of calculation depth corrections is *independent* on XBT temperatures and their study does not address the problem of possible systematic temperature errors.

In spite of the recommendation not to use this new equation when archiving the XBT data, part of the XBT data reside in the WOA2005 collection with the new fall-rate equation applied. To produce a consistent original XBT data set we recalculated sample depth using the manufacturer fall-rate equation for those XBT profiles, where HRBSS corrections have been already applied. Respectively, a similar XBT dataset, but with HRBSS corrections applied to all XBT profiles was produced to investigate the effect of these corrections.

For both T-4/T-6 and T-7/DB probe types comparison with collocated CTD/Bottle box-averaged temperatures shows a temperature bias, changing with depth from positive values within the upper 150-200 to slightly negative values in the deeper layers, suggesting systematic offset due to the fall rate parameters deviating from those specified by the manufacturer (Fig. 6). However, the magnitude of the bias is strongly time-dependent, so that the XBT data from 1970s are characterized by a positive bias almost for the whole upper 400/700-meter layer. The magnitude of the positive bias within the upper 100 m is typically in the range 0.1-0.3^{o}C. Application of HRBSS depth corrections deteriorates the agreement with the reference data. With corrections applied to all XBT profiles the warm bias extends over the whole 400 m (T-4/T-6 probes) or 700 m (T-7/DB probes) layer, resulting in about 0.1^{o}C increase of the depth-averaged warm bias.

Though systematic temperature errors in XBT data were known to exist, and the method of calculation of depth errors was respectively proposed to be independent from possible temperature errors (Hanawa et al., 1994,1995), *pure temperature* errors in XBT data were less investigated and were found to vary considerably, depending on the cruise, probe type and acquisition system. However, the majority of available inter-comparison studies indicate a prevailing warm bias (see Table 1 in Gouretski and Koltermann, 2007). Already in an early analysis of the XBT depth errors Seaver and Kuleshov (1982) noted that even a small positive temperature bias, when presented as a depth error, becomes greatly magnified when the ambient vertical temperature gradient is small. Correcting their XBT profiles from the Sargasso Sea area for a small positive temperature bias, they arrived on a more reasonable XBT depth error. The above considerations are true for the depth corrections suggested by McDowell (1978) which are based on the data from the same region with pronounced thermostads.

b. XBT bias model

To explain the observed total XBT temperature bias along with the fact that HRBSS corrections alone do not improve the agreement between the XBT and CTD data, we suggest a simple bias model, which takes account for both a pure temperature bias and an equivalent temperature error in the XBT data due to the presence of the vertical temperature gradient. At each level z for each monthly bin the total box-averaged temperature bias is decomposed as

$$\langle\delta(z)\rangle = \langle\Delta\rangle - \langle\gamma(z)\rangle \cdot \zeta(z) + \varepsilon , \qquad (4)$$

where $\langle\Delta\rangle$ is a pure temperature bias, $\langle\gamma(z)\rangle$ is the box-averaged vertical temperature gradient and $\zeta(z)$ is the XBT depth correction, with the second term representing the equivalent temperature error. As the depth correction is zero at the surface, the pure temperature bias is equal to the surface temperature bias: $\langle\Delta\rangle = \langle\delta(0)\rangle$. Since the box-averaged values of vertical gradients γ and temperature biases δ are subject to noise due to the synoptic variability, averaging {…} over a sufficient number of boxes is required, resulting in the formula for the XBT depth correction at an arbitrary level:

$$\zeta(z) \approx \{\langle\delta(z)\rangle - \langle\delta(0)\rangle\} \cdot \{\langle\gamma(z)\rangle\}^{-1} . \tag{5}$$

Here the assumption is made that the depth correction ζ does not change between the boxes within a respective one-month time bin.

The performance of the bias model is illustrated in Fig. 7 for all XBT types. The model gives a good representation of the total temperature biases with residuals being typically an order of magnitude smaller compared with total biases. Below about 50-100 m (in accordance with independent inter-comparison experiments) all types are characterized by slower velocities (positive corrections) than given by the fall rate equation with manufacturer coefficients. Negative corrections in the upper layer are qualitatively in agreement with inter-comparison results of 103 XBT profiles analysed by Seaver and Kuleshov (1982). They found negative depth corrections above about 150 m for T-7 probes and in the upper ~120 m for T-5 probes. The explanation to the change of the fall rate with depth is that normally the probe velocity upon entry is greater than the calibration velocity, but it slows with depth as the water viscosity increases as the probe enters deeper and colder layers.

Seaver and Kuleshov (1982) investigated theoretically the dependence of the XBT probe velocity on the water temperature and noted that the temperature change from 25 to 10°C increases the kinematic viscosity by 42 %. We calculated mean temperatures for the same time/spatial bins as used for the bias calculations. Comparison of the depth corrections with temperature deviations from the time mean (Fig. 7) implies a certain correlation only for the T-4/T-6 and T-5 types. Slower T-4/T-6 fall velocities between 1979 and 1988 coincide with on average colder water temperatures, whereas slower fall velocities for the T-5 probes correspond approximately to colder water temperatures between 1985 and 1990.

Using our temperature bias model (formula (5)), we calculated depth corrections separately for T-4/T-6 and T-7/DeepBlue probes for two cases: 1) with account for both the pure temperature bias ‹δ(0)› and the bias due to sample depth error and 2) with zero temperature bias equivalent to the account for depth-errors only (Fig. 8). To compare our results with independent XBT-CTD inter-comparisons depth corrections were calculated for each inter-comparison experiment, using values of the coefficients *a* and *b* in the fall rate equation available in the literature (Table 2). Among the total of 33 inter-comparison experiments for T-4/T-6 and T-7/DB probe types HRBSS give fall rate equation coefficients for 13 inter-comparisons, which were conducted in different regions of the World Ocean (the recommended new fall rate coefficients represent values averaged over all regional inter-comparisons). As follows from the Fig. 8 a satisfactory agreement between our values depth corrections and values from independent inter-comparisons is achieved only if the pure temperature bias is taken into account (Fig.8, left panels).

Depth dependence of differences between depth corrections calculated using equation (5) and corrections from independent inter-comparisons is presented in Fig. 9. The agreement is better for the T-4/T-6 types, with differences ranging from -9 to 4 meters. On average, depth corrections based on our bias model tend to be smaller for T-4/T-6 types and larger in case of T-7/DeepBlue probes. Our model confirms the validity of the new HRBSS fall rate equation coefficients: the average difference between the depth corrections based on the bias model and inter-comparison experiments is from -5 to 1 meters for the T-4 and T-6 probe types and within -2 to +9 meters for T-7 probes (respective depth-averaged differences are -1.8m for T-4/T-6 types and +2.7 for T-7/DeepBlue types).

5. MBT biases

Before the introduction of XBTs mechanical bathythermographs were the most abundant data type in the upper 200-meter layer. The number of MBT profiles available in the WOA2005 database reduces considerably from 1990 to 1991, with a resulting abrupt reduction in the mean monthly number of boxes overlapping with CTD/Bottle data from about 300 to less than 30 (Fig. 4). Therefore, temperature bias was not estimated after 1990. Similar to XBT, MBT data are characterized by a positive temperature bias relative to the bottle and CTD data (Fig. 10). A gradual decrease of the temperature bias is observed between 1950s and 1990s.

Both temperature and depth biases of the MBT profiles were investigated by Stewart (1963), Casciano (1967), and Dinkel and Stawnychy (1973). It was found, that accuracies of $\pm$ 0.1^{o}F (0.056^{o}C) for temperature and $\pm$ 1% for depth as specified by the manufacturer were practically never achieved because of mishandling and abuse in field work, long intervals between calibrations, and instrument errors due to hysteresis, response time and changes in set (Stewart, 1963). In all three studies listed above MBTs were tested in a special test facility allowing for changes in pressure and temperature. Unfortunately, these test results (Table 3) are based on a very small number of MBTs. However, all tests indicate positive temperature and depth biases. As the total MBT temperature bias results from the pure temperature bias and sample depth error, we applied the same bias model (equation (5)) to the MBT data. Comparison of modelled MBT depth corrections with independent results (Fig.11**)** shows qualitative agreement above 150 m (modelled corrections slightly negative), but disagreement at deeper levels with modelled corrections being as high as 4 meters at 250 m depth. It should be noted however that depth errors determined by Dinkel and Stawnychy (1973) are based on tests of only four bathythermographs, with one of which returned from field service. Casciano (1967) results might be more reliable as they are supported by the calibration checks on 144 new and rebuilt MBTs. Unfortunately, depth and temperature errors from these calibrations are presented in the form of histograms, so that no vertical profile of the depth or temperature error is available for comparison with our results. It is also not clear which depth range is most appropriate for the comparison with histogram derived mean bias values. However, the agreement of our depth corrections with Casciano (1967) estimates is better than 1,5 meter within the upper 150 meters (Fig. 11). Finally, we note, that the mechanical bathythermographs are characterized by smaller and less time-variable biases compared to the expendable bathythermographs. This renders them a useful source of historical data for the upper 250 m layer.

6. Profiling floats biases

Autonomous profiling floats represent the newest instrument type within the oceanographic data holdings. The floats are equipped with a CTD, but unlike the ship-based CTD no laboratory sensor calibration is possible as the floats are lost after completion of their expected mean lifespan of about 4 years. The Argo temperatures are accurate to $\pm 0.005^{o}$C and depths to ± 5 meters. Comparison with CTD data indicates generally small biases of different sign (typically less than 0.05^{o}C), thus reflecting a generally good float performance (Fig.12). Largest positive biases are found within the upper 150-200 meters between 1996 and 2000, probably indicating some problems during the initial development phase. Unfortunately, the number of CTD data available for comparison seems to be insufficient, with the total number of CTD profiles in the WOA2005 decreasing rapidly after 1990s. Therefore, these inter-comparison results should be considered as preliminary, until more CTD profiles become available.

7. Discrepancies between the CTD and Nansen cast data

In this study a composite subset of both bottle and CTD data was used as a reference with the assumption that these two data types are by far the most accurate of the whole WOA2005 hydrographic data collection. There are two categories of the CTD and bottle data in the

WOA2005: a) Ocean Station Data (OSD) data type, which includes bottle (Nansen and Niskin) and low resolution CTD stations and b) high resolution CTD stations. Though the temperature accuracy of the thermistors (CTD) and reversing thermometers (Nansen casts) are comparable (Table 1), the accuracy of the sample depth determination by Nansen casts is considerably lower and may be affected by systematic biases. Temperature super-observation were constructed separately for the bottle casts and CTD casts respectively, with 24-month running mean differences (Bottle-CTD) shown in Fig. 13. According to our calculations, the absolute depth-averaged offset is generally less than 0.05°C, with bottle cast data being colder in the upper 100 m layer since about 1973. Assuming that only large-scale bottle-CTD difference patterns are reliably reproduced by the analysis, we note a transition from positive to negative Bottle-CTD temperature differences after 1980. One possible explanation of this change is the transition from using Nansen casts to the CTD rosette bottles, both residing within the WOA2005 OSD data. As noted by Warren (2008), the Woods Hole Oceanographic Institution fully changed form Nansen cast technique to CTD devices in 1981. The time period of this instrumentation change is confirmed by a number of personal communications. According to Y.-H. Park the transition from Nansen casts to CTDs occurred around 1980 on French oceanographic ships. T. Byrne reports that by 1985 CTD data was collected on virtually all physical oceanographic voyages of the National Facility in Australia, which agrees with the estimate by L. Tally for the National oceanographic facilities in the USA. B. King estimates the transition time from Nansen casts to CTDs in the United Kingdom to be around early 1980s. According to A. Mantyla the Nansen casts were used on the CalCOFI ships (California Cooperative Oceanic Fisheries Investigations) until 1987. In spite of detectable offsets between the CTD and Nansen cast data, these offsets are much smaller compared with those for all kinds of bathythermographs, so that the use of Nansen and CTD profiles as a reference is obviously justified.

8. Effect of instrument temperature biases on global temperature anomaly calculations

Monthly box-averaged temperatures were used as input data to estimate global temperature and heat content anomalies within the upper 1000-meter layer. We choose a 25 year reference period (1971-1995) for the anomaly calculations. Monthly reference fields were produced for the upper 400 meter layer and annual reference climatology were used between 400 and 1000 m. The choice of the reference period duration was a trade off between our aim to provide both a sufficient data coverage and to make the reference period as short as possible. Only bottle and CTD profiles were used in order to minimize possible bias effects. Optimal interpolation was implemented to produce the reference temperature fields, using the decorrelation radius of 555 km. No interpolation was performed for the grid-nodes with less than three box-averaged temperature values within the decorrelation radius. Because of insufficient sampling, reference temperatures (especially monthly mean values above 400 m) are not available for all grid-nodes. Finally, at each standard level monthly box-anomalies were computed and global time series were produced by integrating box-anomalies spatially.

In the absence of systematic errors, calculation of global temperature anomalies based on different instrument types would produce essentially the same results, provided the time-spatial sampling patterns were similar for each data type. In reality, global anomalies, calculated for each instrument type separately (Fig. 14), exhibit significant differences due to both instrument specific temporal/spatial sampling and instrument related temperature and/or depth systematic errors. In this study we do not attempt to estimate errors due to inadequate sampling with the primary focus on the impact of the instrument specific biases. According to Gouretski and Koltermann (2007), systematic errors may pose a significant problem in identifying long-term heat content variations in the Global Ocean.

To demonstrate the impact of biases on anomaly calculations, global temperature anomalies were first calculated separately for each instrument type (Fig. 14). In the upper-most layers anomaly time series for different instrument types show the decadal scale variability similar to that exhibited by the bottle/CTD data, with discrepancies in the form of a relatively time-invariant instrument specific temperature offset**.** In contrary, virtually no agreement between the bottle/CTD and bathythermograph anomaly time series is observed in the deeper layers. The most pronounced difference is the absence of the "warm decade" (~1970-80) in the bottle/CTD time series which is pronounced in all types of XBT data (most clearly seen for the 400 m level, Fig. 14). As in the earlier study (Gouretski and Koltermann, 2007), we explain this feature as an artefact due to the time-variable temperature bias in the XBT data, described in details in the previous sections (see Fig.6). Correcting only XBT sample depths by means of the new HRBSS fall rate coefficients deteriorates the agreement with the reference data similar to the offset calculations. Since the spatial distribution of both bathythermograph and bottle/CTD profiles does not change considerably within the upper 500-700 m, this artefact can not be attributed to the change of sampling patterns with depth.

In order to further investigate the effect of systematic errors, global temperature anomalies were calculated using the composite temperature dataset for the following three cases: 1) original data, 2) depth-corrected XBT data according to HRBSS and 3) depth and temperature corrected MBT and XBT data according to this study. In the last case the MBT and XBT original profile data were corrected for temperature and depth bias using the bias model described above. To correct observed temperature profiles, time-dependent pure temperature bias and depth corrections (calculated using formula (5)) were implemented. Fig. 15 illustrates a strong impact of data inconsistencies on the estimates of the global temperature anomalies. Depth- and temperature-corrected data do not evidence a warming trend of the upper 1000-meter layer until mid-1990s. Only after about 1995-96 a rapid temperature increase is detected by all kinds of instruments (Fig. 14).

In addition to temperature anomalies time series of the global heat content anomaly (GHCA) were computed for selected layers (Fig. 16). Whereas linear trends computed for the time period 1948-2004 indicate an increase of the GHCA, the slope of the linear regression depends critically on the dataset used. Application of the Hanawa et al. (1994) depth corrections leads to an approximately two-fold increase of the regression slope, compared to the original data and the case, when both depth and temperature are corrected for bathythermograph profiles. Linear trends, calculated for the time period 1948-1997, indicate an insignificant global *cooling*, when original or temperature- and depth-corrected bathythermograph data are used. On the contrary, using of only depth-corrected XBT data still results in a positive slope ($0.1572 \cdot 10^{22}$ J/year for 1948-2004 versus $0.1212 \cdot 10^{22}$ J/year for 1948-1996). Linear trend model gives obviously an insufficient representation of the secular heat content changes in the Global Ocean, considering the shortness of the time series and the high GHCA variability. We use linear trends here in order to illustrate the strong dependence of linear trend parameters on instrument biases.

Updates of the global surface temperature time series are now available from several research groups. To bring our results in perspective with these independent studies, we compared our surface temperature anomalies with time series produced by the Climatic Research Unit (CRU) of the University of East Anglia (http://www.cru.uea.ac.uk)). Whereas our time series (Fig. 17) refers only to the sea surface temperature, the CRU time series are based on both ocean and land components. The CRU anomaly time series are computed by averaging areally 5° x 5° grid-box temperature anomalies using the HadCRUT3v dataset**.** These box anomalies are based on a much larger set of land-based meteorological and marine sea surface temperature observations compared to the WOA2005 dataset. In spite of differences in the original data and methods there is a good agreement between both types of curves since the end of 1950s, with all time series exhibiting similar decadal-scale variability. The comparison also indicate a time-lag of a few years between the CRU time series (based both on land and sea data) and our ocean time series. Before the International Geophysical Year 1957-1958 an offset of 0.2-0.3°C is observed, though the course of the curves is very similar. This difference may be an indication of a not-excluded temperature bias or/and of an inadequate spatial sampling. The latter explanation seems plausible as the hydrographic sampling had a strong geographical bias to a few regions in the Northern Hemisphere before the completion of the hydrographic program of the International Geophysical Year 1957-1958. The last argument is supported by the fact, that the agreement of our anomaly time series before 1957 with the CRU Northern Hemisphere curve is better than with the CRU global curve.

9. Concluding comments

In this study we analyzed systematic errors in the subsurface temperature data and their impact on temperature and heat content anomaly estimates for the Global Ocean. The systematic temperature errors were identified through the comparison of the instrument specific data with the collocated bottle and CTD data, which were served as a reference. In agreement with the earlier results by Gouretski and Koltermann (2007), the mechanical and expendable bathythermograph data were found to be highly biased, with temperature and sample biases to be time variable.

A simple bias model, unifying the pure temperature bias and the bias due to sample depth error, was suggested and proved against independent XBT-CTD inter-comparison experiments. Our results confirmed the validity of the Hanawa et al. (1994) new fall rate equation. However, application of this equation without a proper account for a pure temperature bias leads to an even larger disagreement (in temperature) between the collocated CTD and XBT profiles. Temperature and depth biases were determined separately for all types of XBT probes. The fall velocity for three main XBT types (T-4/T-6, T-7/DeepBlue and T-5) was found to be different, so that two different fall-rate equations are obviously needed for the most commonly used T-4/T-6 and T-7/DeepBlue modifications.

Temperature offsets relative to the collocated CTD stations were also determined for profiling floats. The largest disagreement was found in the upper 150-200 m layer for the floats launched before the beginning of the Argo program. As the number of CTD profiles obtained after 2000 is relatively small in the WOA2005 dataset, further inter-comparison studies are needed to obtain more reliable estimates of profiling float temperature biases.

This study was focused on the instrument specific biases and their possible impact on the estimation of secular changes in the World Ocean. Our analysis has demonstrated, that usage of only depth-corrected XBT data leads to the overestimation of the heat content increase in the Global Ocean which was reported by Levitus et al. (2005). No global warming trend is detected within the upper 1000 m layer between 1948 and 1996 if all types of bathythermograph data are corrected both for temperature and sample depth biases. Only since mid-1990s a rapid temperature increase is detected by all kinds of instruments. Clearly, further refinement of the method of bias estimation is needed in order to arrive a higher degree of consistency within the available subsurface temperature dataset. This further work is critically dependent on the availability of the metadata, which would allow identification of possible error sources.

Acknowledgements

Thanks S. Danilov, M. Schröder and A. Sy for their helpful comments. Discussions with F. Reseghetti with respect to XBT issues are appreciated. I also wish to thank S. Levitus for sending me the archive papers related to the assessment of the MBT performance. I am grateful to B. King, A. Mantyla, Y.-H. Park, L. Talley and T. Byrne for their personal communications with respect to the changes in hydrographic instrumentation.

References

Argo Project Office (2006): Argo, an array of profiling floats observing the ocean in real time. Prepared by the Argo Project Office. Printed by Japan Agency for Marine-Earth Science and Technology, 8 pp.

Boyed, J.D. and R.S. Linzell (1993), The Temperature and Depth Accuracy of Sippican T-5 XBTs, *J. Atmos. Oc. Tech.*, 9, 470-483.

Boyer, T.P. and S. Levitus (1994): Quality control and processing of historical temperature, salinity and oxygen data. NOAA Technical Report NESDIS 81, 65,

Casciano, D.L. (1973): Mechanical Bathythermograph. Geo-Marine Technology, December/January 1966/67, p.19-21.

Dinkel, C.R. and M. Stawnychy (1973): Reliability Study of Mechanical Bathythermograph. Marine Technology Soc. Journal, vol. 7(3),p.41-47.

Folland, C.K. and D.E. Parker (1995): Correction of instrumental biases in historical sea surface temperature data, Q. J. R. Meteorol. Soc., 121, 319-367.

Folland, C.K., N.A. Rayner, S.J. Brown, T.M. Smith, S.S.P. Shen, D.E. Parker, L. Macadam,P.D. Jones, R.N. Jones, N. Nichols, and D.M.H. Sexton (2001), Global temperature change and its uncertainties since 1961, *Geophys. Res. Lett.*, 28, 2621-2624.´

Gille, S. (2002): Warming of the Southern Ocean Since 1950s. Science, 295, p.1275-1277.

Gould W.J. (1991): RRS Charles Darwin Cruise 50, IOMSL cruise report, 221, 41 pp.

Gouretski, V.V. and K. Jancke (2001): Systematic errors as the cause for an apparent deepwater property variability: global analysis of the WOCE and historical hydrographic data. *Prog. Oceanogr.*, 48, 337-402.

Gouretski, V.V. and K.P. Koltermann (2004): WOCE Global Hydrographic Climatology, Berichte des Bundesamt für Seeschiffahrt und Hydrographie, Nr.35/2004, 52 pp., 2 CDROMs.

Gouretski, V.V. and K.P. Koltermann (2007): How much is the Ocean really warming? Geophysical Research Letters, 34, L01610, doi:10.1029/2006GL027834

Green, A.W. (1984): Bulk dynamics of the expendable bathythermograph (XBT). Deep Sea Research, 31, 415-426.

Hallock, Z.R. and W.J. Teague (2002): The fall rate of the T-7 XBT, *J. Atmos. Oc. Tech.*, 9,470-483.

Hanawa K. and Y. Yoshikawa (1991): Re-examination of depth error in XBT data. Journal of Atmospheric and Oceanic Technology, 8, 422-429.

Hanawa, K., P. Rual, R. Bailey, A. Sy and M. Szabados (1994): Calculation of new depth equations for expendable bathythermographs using a temperature-error-free method

(Application to Sippican/TSK T-7, T-6 and T-4 XBTs), UNESCO technical papers in marine sciences, 67, 46 pp.

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 Research, 42, No.8, p.1423-1451.

Henin C. (1989): Experimental error of the expendable bathythermograph in the western tropical Pacific. Paper presented at the first session of the TOGA XBT ad hoc Panel of Experts, Noumea, New Caledonia, June 1989, pp. 9 (unpublished manuscript).

Heinmiller, R.H., C.C. Ebbesmeyer, B.A. Taft, D.B. Olson, and O.P. Nikitin (1983): Systematic Errors in Expendable Bathythermograph, J. Phys. Oceanogr., 12, 592-600.

Iselin, C.O'D. (1936): A study of the circulation of the western North Atlantic. Papers in Physical Oceanography and Meteorology, 4, 4, 101p.

Johnson, G.C., P.E. Robbins and G.E. Hufford (2001): Systematic adjustments of Hydrographic Sections for Internal Consistency, Journal of Atmospheric and Oceanic Technology, 18, p.1234-1244.

Kizu, S. and K. Hanawa (2002): Recorder-dependant Temperature error of Expendable Bathythermograph, *J. Oceanogr.*, 58, 469-476.

Kizu, S. H. Yoritaka and K. Hanawa (2005): A new Fall-rate Equation for T-5 Expendable Bathythermograph (XBT) by TSK, *J. Oceanogr.*, 61,115-121.

Levitus, S., J. Antonov, T. Boyer and C. Stephens (2000): Warming of the World Ocean, *Science.*, 287, 22252229, DOI:10.1126/science.287.5461.2225.

Levitus, S., J. Antonov and T. Boyer (2005): Warming of the World Ocean, 1955-2003. *Geophys. Res. Lett.*, 32, L12602, doi:10.1029/2005GL023112.

Locarnini, R. A., A. V. Mishonov, J. I. Antonov, T. P. Boyer, and H. E. Garcia (2006): *World Ocean Atlas 2005, Volume 1: Temperature.* S. Levitus, Ed. NOAA Atlas NESDIS 61, U.S. Government Printing Office, Washington, D.C., 182 pp.

McDowell, S. (1977): A note on XBT accuracy. POLYMODE News, No. 29, 1-4.

McDowell (1978): A cautionary note on T-5 XBTs. POLYMODE News, No. 58, WHOI, Woods Hole, MA (unpublished manuscript).

McPhaden, M.J., A.J. Busalacchi, R. Cheney, J.R. Donguy, K.S. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G.T. Mitchum, P.P. Niiler, J. Picaut, R.W. Reynolds, N. Smith and K. Takeuchi (1998): The tropical ocean-global atmosphere observing system: a decade of progress, *J. Geophys. Res.*, 103(C7), 14,169-14,240.

Nature (2007) Artefacts in ocean data hide rising temperatures, 447, 3, p.8.

Perlewitz, P. (1908): Ozeanographische Versuche und Beobachtungen an Bord S.M.S. “Möwe” und S.M.S. “Zeiten” im Sommer 1907. Annalen der Hydrographie und Maritimen Meteorologie XXXVI, Jahrgang (1908), Heft 1, 1-5.

Reseghetti, F., M. Borghini, and G.M.G. Manzella (2007): Factors affecting the quality of XBT data – results of analysis on profiles from the Western Mediterranean Sea. Ocean Sci., 3, 59-75.

Siever, G. A., and S. Kuleshov (1982): Experimental and analytical error of the expendable bathythermograph, *J. Phys. Oceanogr.*, 12, 592-600.

Singer J.J. (1990): On the error observed in electronically digitized T-7 XBT data. Journal of Atmospheric and Oceanic Technology, 7, 603-611.

Stewart, R. L. (1963): Test and Evaluation of the Mechanical Bathythermograph. Unpublished manuscript. Marine Sciences department, U.S. Naval Oceanographic Office, Washington 25, D.C., 33 p

Sy A. und J.Ulrich (1990): North Atlantic ship-of-opportunity XBT programme – 1989 data report. Wissenschaftlich- Technische Berichte aus dem Deutschen Hydrographischen Institut, 1990-2, pp.89.

Warren, B.A. (2008): Nansen-bottle stations at the Woods Hole Oceanographic Institution, Deep-Sea Res., in press

Wüst, G., G. Bennecke and H.H.F. Meyer (1932): Ozeanographische Methoden und

Instrumente. In: Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs- und Vermessungsschiff „Meteor“ 1925-1927, Band IV, Erster Teil.

Table 1. Temperature and depth accuracies for different instrument types

Instrument	Temperature Precision, °C	Sample depth precision	Profile Percentage of total (0-100m)/(0-1000m)	Year of introduction
Nansen casts	~0.01	<= 6%FS(0-200m) <=1,5%FS(>200m)	9.5 / 10.6	1897
CTD	0.001	0.015%FS	18.0 / 20.8	1967
MBT	0.2	>1%Z	33.5 / 00.0	1940
XBT	0.1	2% Z	35.9 / 65.2	1966
Profiling floats	0.002	0.015%FS	3.2 / 6.6	1996

FS – full scale, Z - depth

Table 2. Summary of the fall rate equation coefficients obtained from the XBT-CTD inter-comparisons

No.	Year	Multiplicative correction	a	b	XBT type	Number of profiles	Reference or/and data source	Note
1	1976.416	1.0356 *)	-	-	T4	12	Mantyla (pers. comm., 1976)	Comparison with Nansen casts
2	1986.000	1.0500	6.796	227	T4	35	Henin, 1989	
3	1989.070	1.0571	6.810	281	T4	41	Hanawa et al., 1994	
4	1989.457	0.9567	6.553	138	T6	9	Hanawa and Yoshikawa, 1991	
5	1990.720	1.0095	6.566	164	T4	17	Hanawa et al., 1994	
6	1991.540	1.0259	6.656	194	T4	80	Hanawa et al., 1994	
7	1992.000	1.0403	6.708	266	T4	52	Hanawa et al., 1994	
8	1992.125	0.9923	6.514	61	T4	15	Hanawa et al., 1994	
9	2003.900	1.0152	6.570	220	T4 T6	55	Resegetti et al., 2007	
10	1976.500	1.0200 *)	6.601	220	T7	47	Saever and Kuleshov, 1982	Data from McDowell, 1977
11	1977.960	1.0121 *)	-	-	T7	103	Seaver and Kuleshov, 1982	Account made for 0.025°C T-bias Depth-corrections from Fig. 4
12	1978.154	0.9890	6.450	131	T7	139	Green, 1984	
13	1978.154	0.9950	6.440	215	T7	139	Heinmiller et al., 1983	0 to 325 m
14	1978.154	1.0440	6.757	226	T7	139	Heinmiller et al., 1983	326 to 760 m
15	1985.833	1.0435 *)	-	-	T7	14	Singer, 1990	
16	1985.960	1.0441	6.741	253	T7	12	Hanawa and Yoshikawa, 1991	
17	1985.960	1.0376	6.715	225	T7	15	Hanawa Yoritaka, 1987	
18	1986.125	1.0419 *)	-	-	T7	11	Singer, 1990	
19	1987.125	1.0261	6.652	203	T7	7	Hanawa and Yoshikawa, 1991	
20	1987.125	1.0259	6.652	201	T7	9	Hanawa et al., 1994	

21	1987.540	1.0301	6.666	224	T7	13	Hanawa et al., 1994	
22	1987.708	1.0352	6.941	413	T7	8	Hanawa and Yoshikawa, 1991	
23	1988.000	1.0414	6.723	253	T7	37	Hanawa et al., 1994	
24	1988.500	1.0408	6.674	329	T7	22	Hanawa et al., 1994; (data from Yoshida et al., 1993)	
25	1989.476	1.0075	6.562	148	T7	10	Hanawa and Yoshikawa, 1991	
26	1989.500	1.0200	6.601	220	T7	11	Sy and Ulrich, 1990	
27	1989.750	1.0402	6.747	200	T7	12	Hanawa et al., 1994	
28	1989.900	1.0676	6.854	324	T7	25	Hanawa et al., 1994	
29	1990.020	1.0324	6.680	226	T7	21	Hanawa et al., 1994	
30	1990.080	1.0311	6.672	225	T7	24	Hanawa et al., 1994	
31	1990.375	1.0514	6.798	238	T7	144	Hallock and Teague, 1992	
32	1990.500	1.0500	6.796	227	T7	29	Gould, 1991	
33	1991.700	1.0101	6.561	179	T7	12	Hanawa and al., 1994	
34	2004.100	1.0393	6.720	60	T7 DB	55	Resegetti et al., 2007	
35	1977.500	1.0117 *)	-	-	T5	?	Seaver and Kuleshov, 1982 (data from McDowell, 1978)	Account made for 0.025°C T-bias; depth-corrections from Seaver and Kuleshov (1982), Fig. 6
36	1991.500	0.9806	6.705	162	T5	34	Boyd and Linzell, 1993	
37	1996.990	0.9590	6.541	187	T5	300	Kizu and Hanawa, 2005	

*) Multiplicative corrections were calculated using absolute depth corrections below 100 m

Table 3. Calibration results for mechanical bathythermographs

Authors	Year	Number of MBTs	Mean T-bias (°C)	Mean depth bias (m)
Stewart (1963)	1963	1	0.36	6.0
Casciano (1967)	1966	29	0.08	0.4
Dinkel and Stawnychy (1973)	1972	12	0.21	3.6

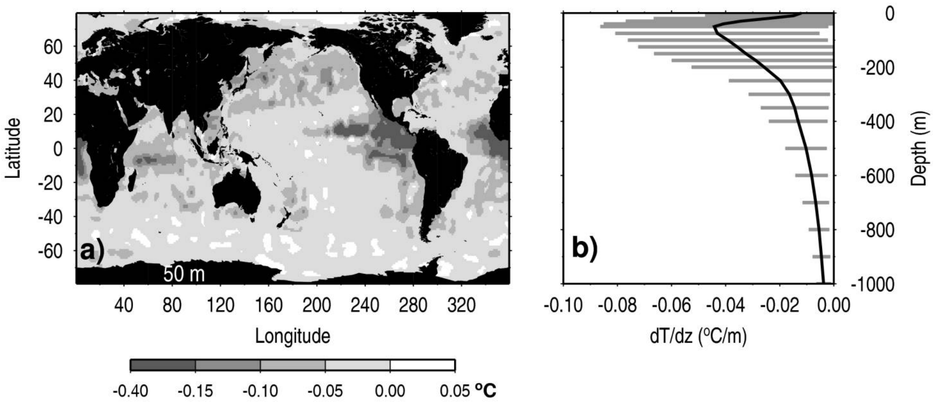

Fig.1. Vertical temperature gradient based on the WOCE Global Hydrographic Climatology (Gouretski and Koltermann, 2004): (a) 200 m level; (b) global-averaged.

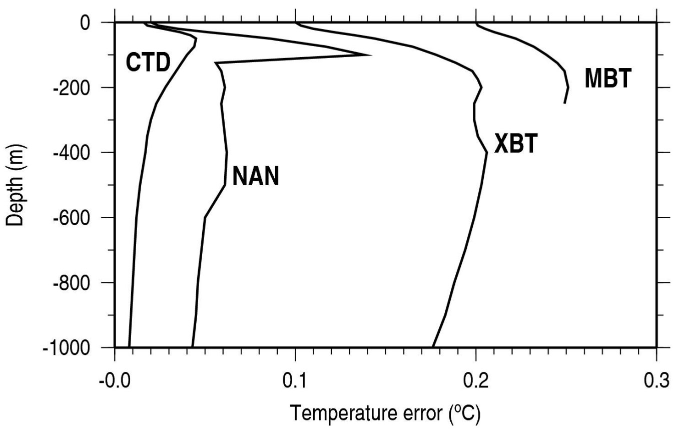

Fig.2. Estimates of the total absolute temperature error (pure temperature error plus equivalent temperature error due to the sample depth error) based on instrument type accuracy specifications.

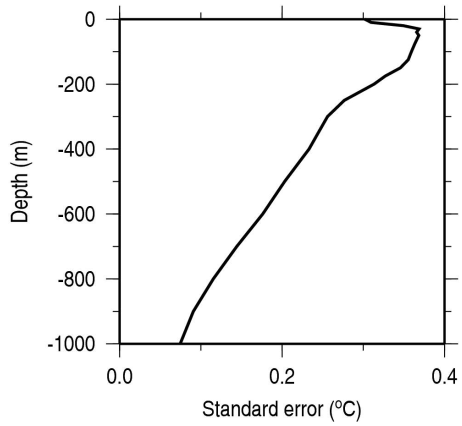

Fig.3. Global averaged standard error of the 111x111 km box-averaged temperatures based on the CTD and bottle data from the WOCE Hydrographic Climatology (Gouretski and Koltermann, 2004).

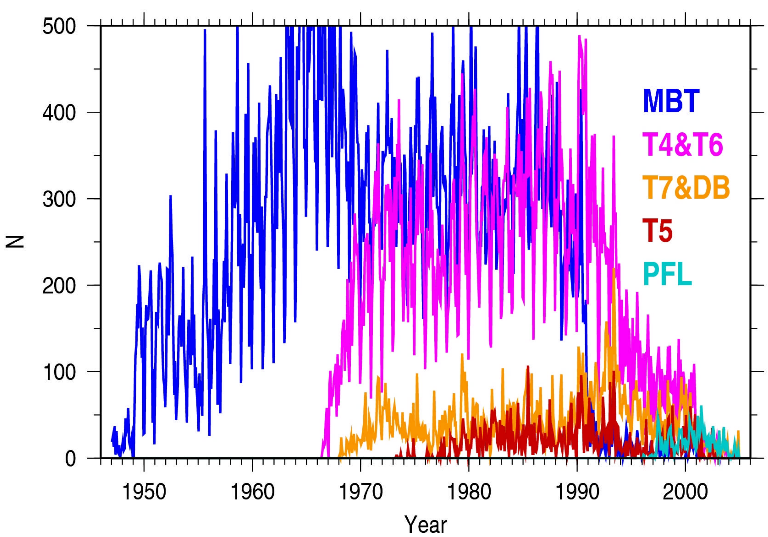

Fig.4. Number of 111x111 km boxes overlapping with reference bottle and CTD data for different instrument types.

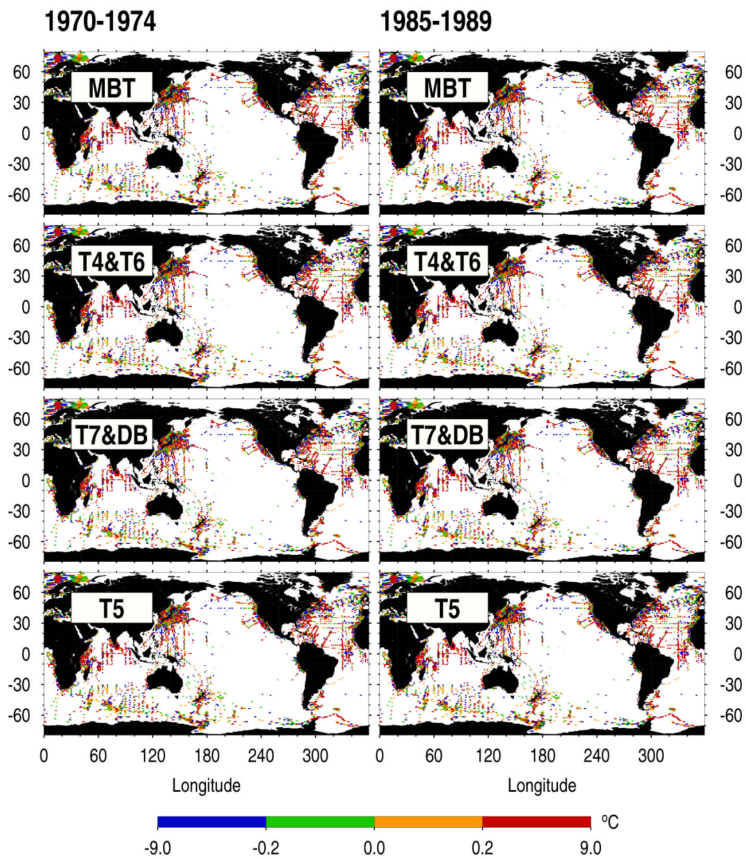

Fig.5. Average MBT and XBT temperature biases for two selected 5-year periods at 100 m.

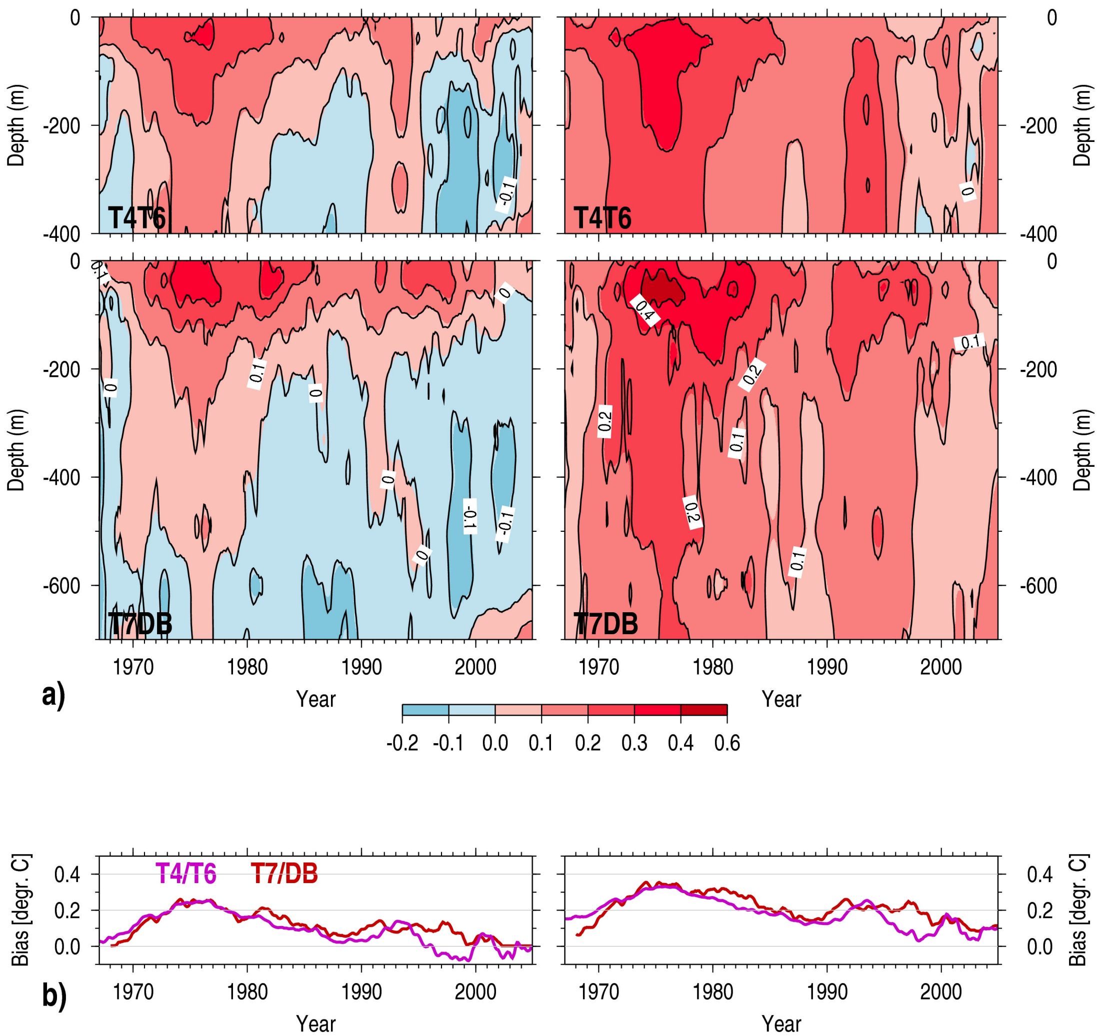

Fig.6 Temperature biases of XBT probe types T-4/T-6 and T-7/DeepBlue: a) time evolution with depth; b) depth-averaged biases. Left panels correspond to original data, right panels correspond to depth-corrected data according to Hanawa et al. (1994) fall-rate equation.

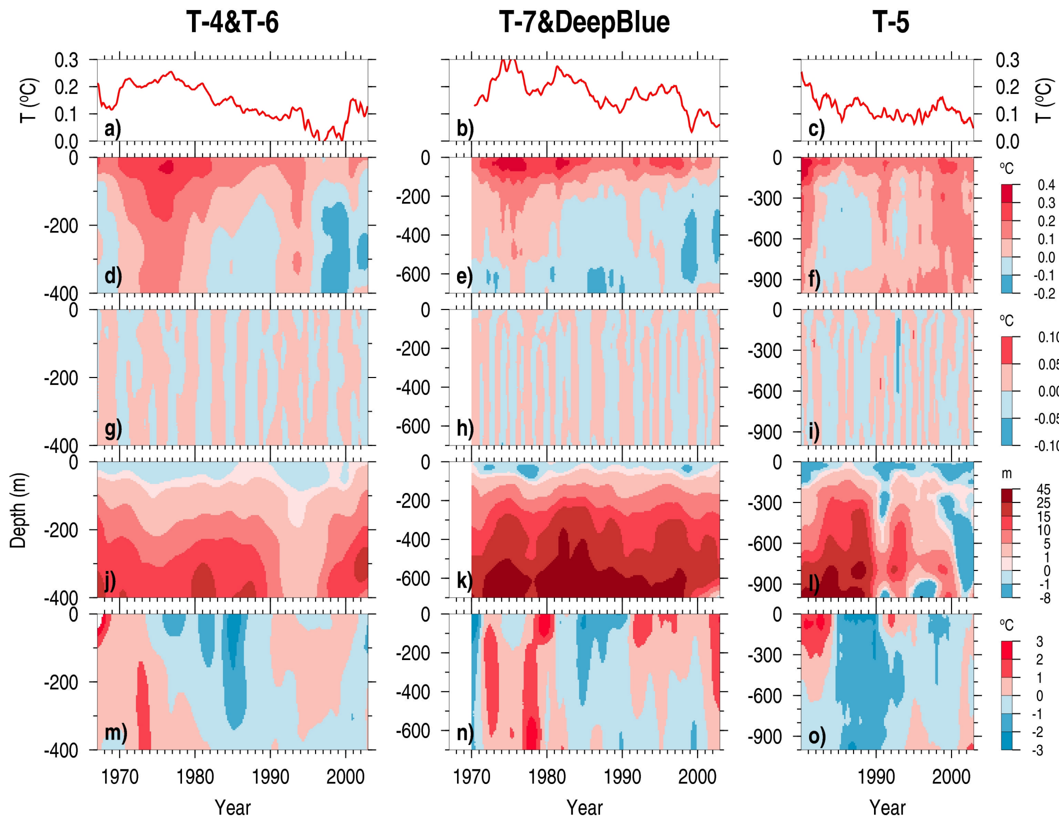

Fig.7. Application of temperature and depth bias model to different XBT probes types: (a)-(c) -observed temperature bias at the surface; (d)-(f) modelled temperature bias at depths; (g)-(h) – residual (observed minus modelled); (j)-(l) depth correction; (m)-(o) water temperature anomaly.

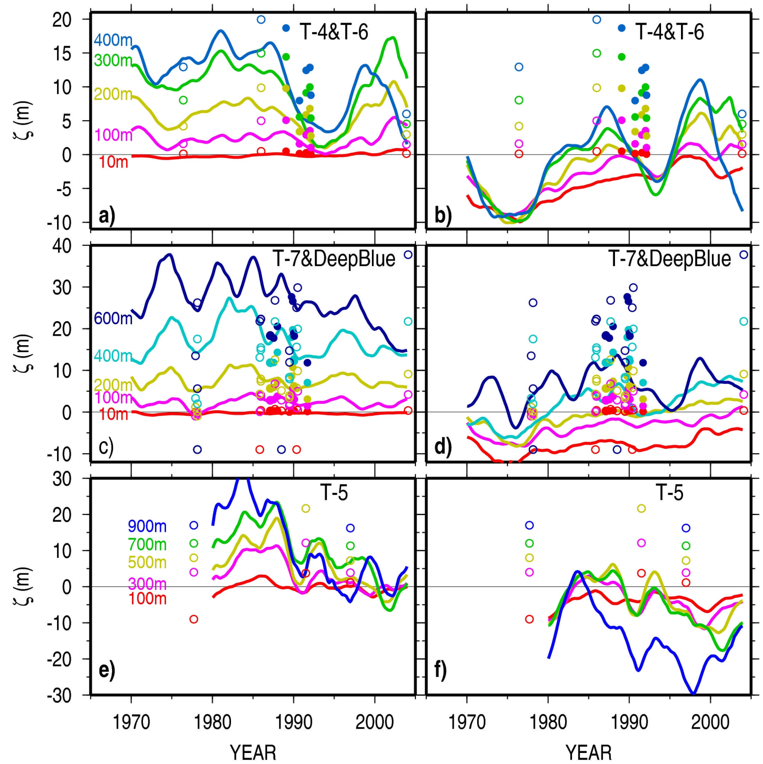

Fig.8. Depth-corrections for different types of XBT probes at selected depths (coloured curves). (a), (c), (e) depth and temperature bias model; (b), (d), (f) depth bias model. Circles show depth corrections determined during independent XBT versus CTD inter-comparisons (Table 2).

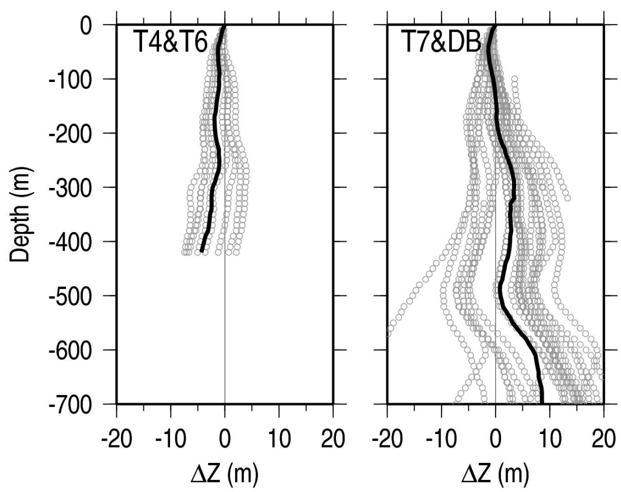

Fig.9. Difference between depth-corrections estimated in this study and depth corrections from independent XBT-CTD inter-comparisons (Table 2) (circles). Black line corresponds to the average difference.

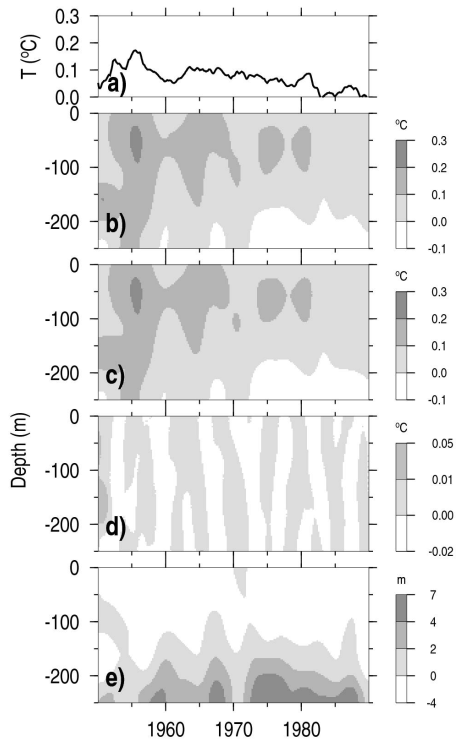

Fig.10. Application of temperature and depth bias model to the mechanical bathythermograph data: (a) observed temperature bias at the surface; (b) observed temperature bias at depths; (c) modelled temperature bias; (d)– residual (observed minus modelled); (e) depth correction.

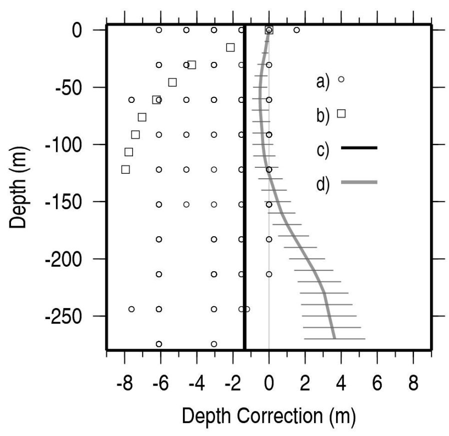

Fig. 11 Comparison of MBT depth corrections (this study, (d)) with independent MBT calibration results by (a) Dinkel and Stawnychy (1973), (b) Stewart (1963) and (c) Casciano (1967).

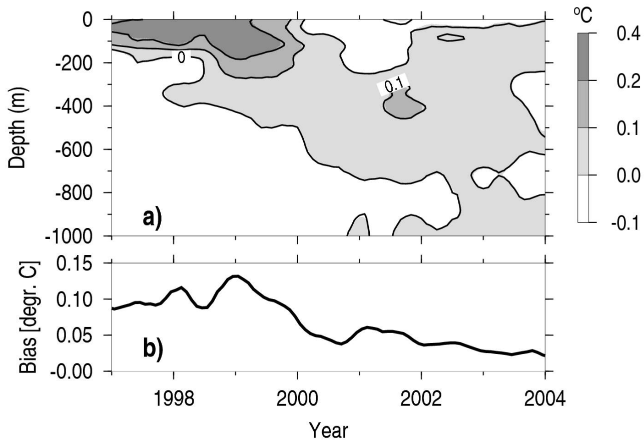

Fig. 12 Temperature bias of profiling float data: (a) at depths; (b) depth-averaged.

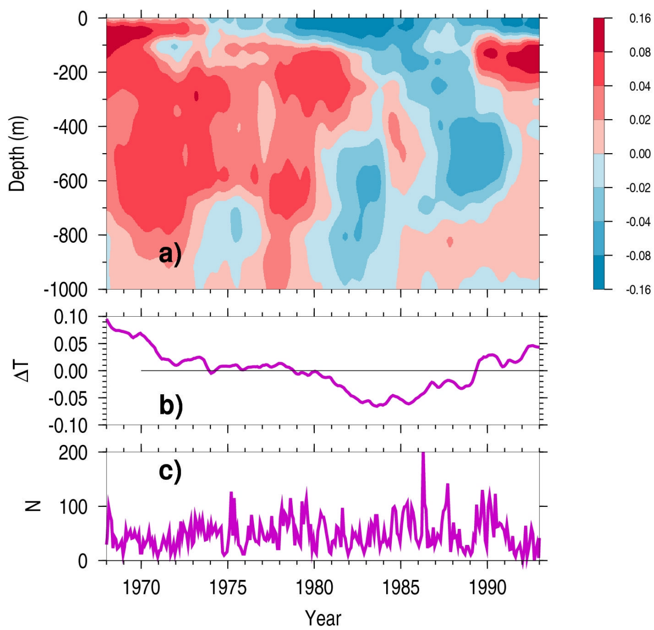

Fig.13. Temperature offset of Nansen bottle data relative to CTD data: (a) time evolution at depths; (b) depth-averaged offset; (c) number of overlapping 111x111 km boxes.

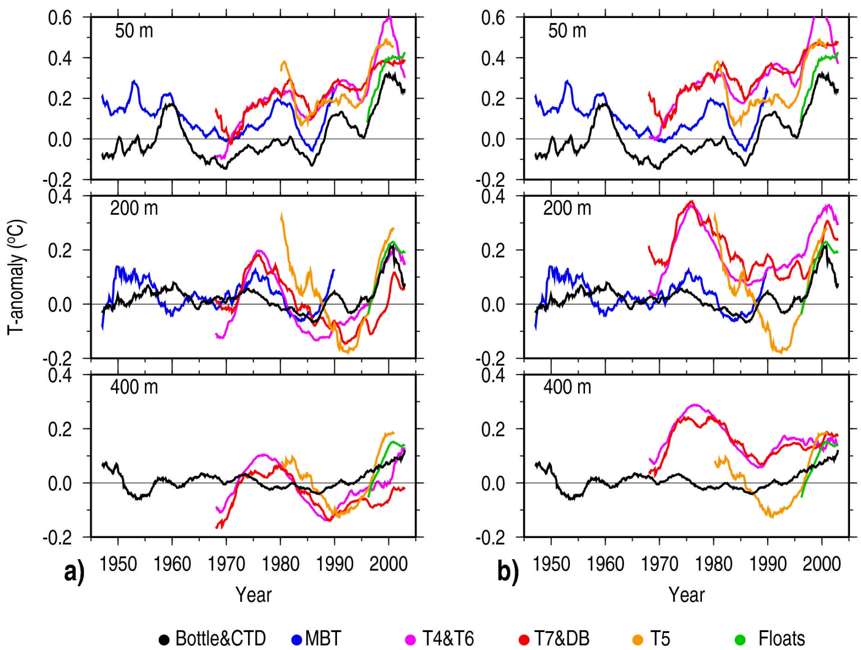

Fig. 14 Global averaged temperature anomalies at selected levels for different instrument types: (a) left panels, original data; (b) right panels, depth-corrected XBT data according to Hanawa et al., 1994.

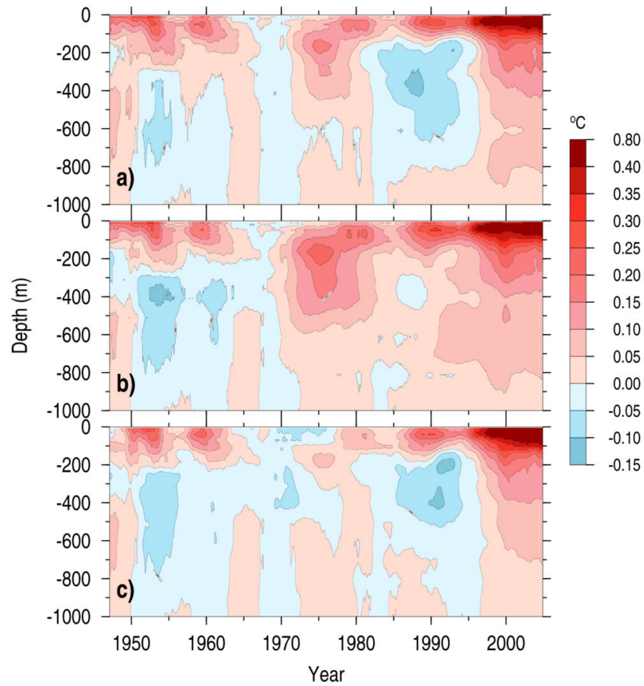

Fig.15. Global temperature anomaly based on blended data from all instrument types: a) original data; b)XBT depths corrected using Hanawa et al. (1994) fall rate equation for T-4, T-6, T-7, and Deep Blue probes, original data for other instrument types; c) MBT and XBT temperature- and depth-corrected data, original data for other instrument types.

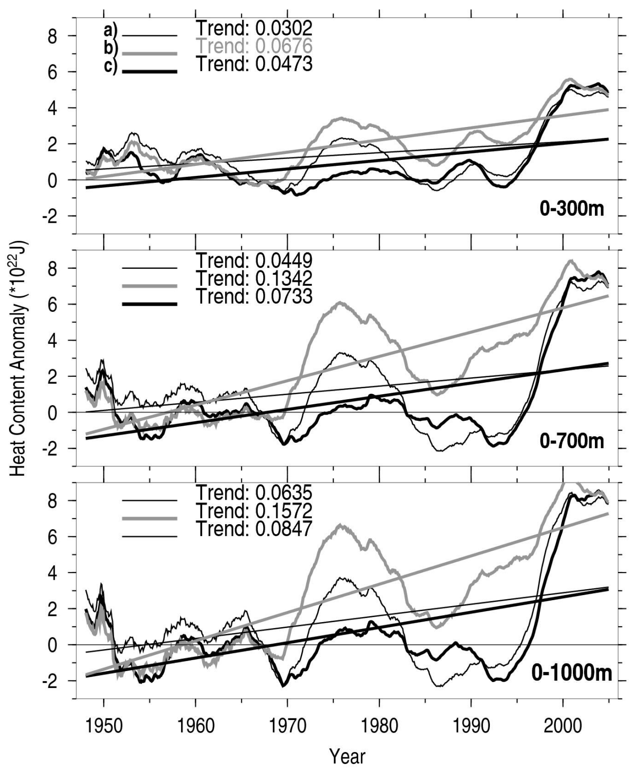

Fig.16. Global heat content anomaly in selected layers based on a composite data set of all instrument types: (a) – original data; (b) – XBT data with Hanawa et al. (1994) depth corrections for T-4, T-6, T-7 and Deep Blue probes, original data for other instrument types; (c) – MBT and XBT data corrected both for temperature and depth bias, original data for other data types. Linear trends are shown to illustrate the strong dependence of the trend slope on instrument biases.

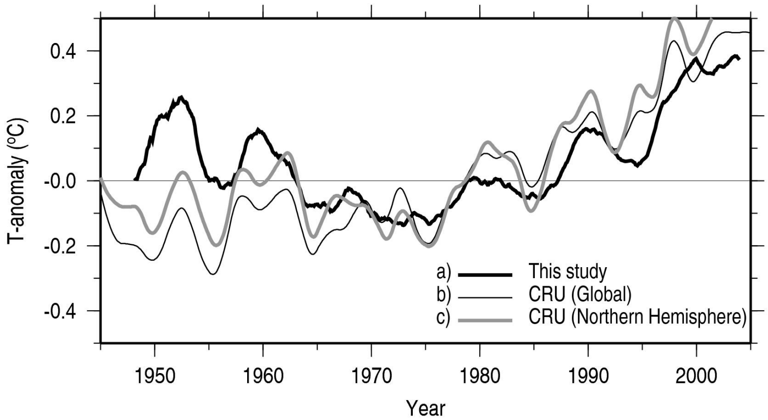

Fig.17. Comparison of surface temperature anomalies: (a) – sea surface, this study (all instrument types, bathythermograph data corrected both for temperature and depth bias); (b) – global mean temperature anomaly, (c) – Northern Hemisphere mean temperature anomaly (both time series are based on sea-surface temperatures and land surface air temperatures from the Climate Research Unit, University of East Anglia (http://www.cru.uea.ac.uk)).