How much is the ocean really warming?

Viktor Gouretski[1] and Klaus Peter Koltermann[2]

[1] We use a global hydrographic dataset to study the effect of instrument related biases on the estimates of long-term temperature changes in the global ocean since the 1950s. The largest discrepancies are found between the expendable bathythermographs (XBT) and bottle and CTD data, with XBT temperatures being positively biased by 0.2–0.4°C on average. Since the XBT data are the largest proportion of the dataset, this bias results in a significant World Ocean warming artefact when time periods before and after introduction of XBT are compared. Using bias-corrected XBT data we argue reduces the ocean heat content change since the 1950s by a factor of 0.62. Our estimate of the ocean heat content increase (0–3000 m) between 1957–66 and 1987–96 is $12.8 \cdot 10^{22}$ J. Because of imperfect sampling this estimate has an uncertainty of at least $8 \cdot 10^{22}$ J. **Citation:** Gouretski, V., and K. P. Koltermann (2007), How much is the ocean really warming?, *Geophys. Res. Lett.*, *34*, L01610, doi:10.1029/2006GL027834.

1. Introduction

[2] A number of studies have documented global climate change, providing evidence of the warming of both the atmosphere and the global ocean. Based on meteorological station data and sea surface temperature data an increase of the global surface temperature of about 0.61 ± 0.16°C between 1861 and 2000 is reported [*Folland et al.*, 2001]. The observational data set below the ocean surface is much sparser, with first deep water measurements dating from the end of 19th century and a basin-scale coverage first achieved during the International Geophysical Year 1957–59. In contrast to the sea-surface observations only a small fraction of the deep ocean is sampled each year, imposing severe limitation on the assessment of the long-term temperature variability. A global scale study by *Levitus et al.* [2005] (hereinafter referred to as LAB2005) revealed a progressive warming of the World Ocean with an increase of the heat content of $14.5 \cdot 10^{22}$J between 1957 and 1997 for the upper 3 km layer. However, the study does not take into account possible temperature biases associated with differing instrumentation.

[3] The problem of instrumental biases has been considered in meteorological literature in order to provide a more accurate estimation of climatic variations. For instance, *Folland and Parker* [1995] developed a physically based empirical technique for correcting historical surface temperature measurements for time-varying biases. Another temperature bias correction method for sea surface temperature was developed by *Smith and Reynolds* [2002] who estimate the bias in the global temperature anomaly to be about 0.2–0.4°C before the1940s. Here, we make a first estimate of the biases between oceanographic instrument types and use these biases in an attempt to provide improved ocean heat content estimates.

2. Data and Method

[4] The hydrographic data base for the current study consists of the World Ocean Database 2001 (WOD01) collection [*Conkright et al.*, 2002], added to by a complete set of the WOCE hydrographic data [*Gouretski and Koltermann*, 2004] and the North Atlantic Argo profiling float data [*Roemmich and Owens*, 2000] (an addition of 310,000 temperature profiles to the WOD01 collection used by LAB2005 were not available for this study). This large data base has observations from each of the five main instrumentation types: Mechanical (MBT) and Expandable (XBT) bathythermographs, hydrographic bottles (Nansen and Rosette sample bottles), Conductivity-Temperature-Depth (CTD) instruments, and profiling floats. In spite of the fact that a variety of techniques to measure water temperature were used since the beginning of oceanographic observations no global scale studies of instrument specific temperature biases exist. Inhomogeneous spatial and temporal sampling poses another problem in estimating long-term temperature variations in the global ocean.

[5] From about $49.4 \cdot 10^6$ temperature profiles a total of $142 \cdot 10^6$ point temperature anomalies have been computed relative to the reference climatology, which represents the mean state of the global ocean during a 10-year period between 1987 and 1996. Similar to the work of *Gouretski and Koltermann* [2004] reference fields of temperature and salinity were obtained on 45 levels using an optimal interpolation method. Since salinity is not available for the bathythermograph profiles, isopycnal averaging of the reference data was not possible, and optimal interpolation was applied on isobaric surfaces. In order to exclude seasonal signals a monthly climatology was used above 250 meter depth, while an annual climatology was used below that level. The reference period includes the World Ocean Circulation Experiment (WOCE) observational phase as well as pre-WOCE years and is characterised by the best global coverage ever achieved. Also the quality of the reference period data is superior to the earlier years according to the WOCE data quality assessment by *Gouretski and Jancke* [2001]. For bottle and CTD data a quality control procedure developed by *Gouretski and Jancke* [1999] was implemented, whereas a statistical check of XBT, MBT, and PFL data was performed for $2° \times 4°$ boxes. All analyses of large-scale temperatures recognize that the irregular distribution of point anomalies

---

[1]Klimawissenschaften, Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven, Germany.
[2]Intergovernmental Oceanographic Commission, UNESCO, Paris, France.
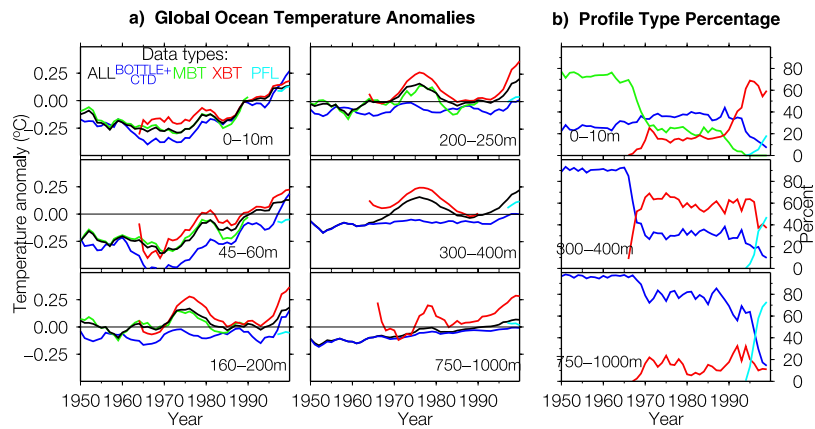
Figure 1. (a) Globally averaged temperature anomaly in selected layers based on different instrument types (ALL, all data types; MBT, mechanical bathythermograph; XBT, expandable bathythermograph; and PFL, profiling floats). (b) Percentage of profiles of different instrument types in selected layers.

requires some form of gridding, in order for the analysis not to be biased. In the vertical the ocean was subdivided into 24 layers and within each layer gridded (box-averaged) anomalies were produced on a $2° \times 4° \times 1$ year grid. The thickness of layers was chosen to increase from 10 meters near the surface to 500 m below 2000 m, because of a strong reduction of the number of observations and a decrease of natural variability with depth. A global monthly or even seasonal time binning is not possible due to insufficient sampling particularly in the deeper layers. Estimates of temperature anomalies for distinct seasons can be obtained for some data abundant regions of the North Atlantic and North Pacific Oceans, but this is beyond the scope of the present study. Point temperature anomalies were not calculated in the absence of reference data within a 275 km influence radius. Global temperature anomalies were calculated separately for each instrument type by averaging yearly box anomalies within each layer (Figure 1), indicating significant instrument related temperature offsets.

3. Instrumental Biases

[6] Introduced during the Second World War, MBTs consisted of a temperature recorder lowered to a depth and than winched up again, temperature values were read off the trace. Since the 1970s MBTs have gradually been phased out in favour of XBTs, which use copper wire to communicate the temperature measured by a thermistor during the drop back to the ship. Both types of the bathythermograph data dominate in the upper 250 (MBT) to 700 meters (XBT) (Figure 1b). The XBTs 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. The depth estimation based on a manufacturer fall-rate equation was shown to underestimate the sample depth, and corrections have been suggested in a number of studies [*Seaver and Kuleshov*, 1982; *Heinmiller et al.*, 1983; *Hanawa et al.*, 1994; *Kizu and Hanawa*, 2002]. Systematic temperature errors were found

to vary considerably, depending on the cruise, probe type and acquisition system (Table 1).

[7] To estimate instrument temperature biases differences between different instrument anomalies were formed for the boxes with at least three point anomalies for each instrument type. These differences were than averaged around the globe within each layer and for each year (Figure 2). There is no requirement for the observations in any box to be close in time, other than that they be in the same year. Both types of bathythermographs exhibit positive offsets at all depths with respect both to the CTD/bottle casts and profiling floats. On average the MBT temperatures agree better with CTD/bottle data than do the XBT data. Before about 1957 a large positive bias (>0.5°C) in the MBT temperatures is observed in the upper 30–50 meter layer and below about 150 m. A substantial improvement occurred in the end of 1950s, with maximum offsets at about 50–100 m depth being less than 0.3°C. A good agreement between the two data types is achieved after 1980, with no significant temperature bias obvious. Unfortunately, we are not aware of any MBT-CTD/bottle data inter-comparison studies. A possible explanation of the positive bias could be a hysteresis of the diaphragm which is used in MBT instruments to sense the pressure, or a response lag in a xylene-filled copper tube which served as a temperature sensor.

[8] As shown in Figure 2, XBT data are positively biased, with two periods of especially high biases: before 1983 and after about 1995. The biases are largest between about 50 and 250 meters, and below 1000 m. Occurrence of positive temperature biases in the XBT data was reported in the literature for a number of inter-comparison experiments, during which XBT profiles were compared against the collocated CTD casts (Table 1). We note, that in this study the original XBT profiles as in the WOD01 were used and no further fall rate correction were applied.

4. Global Time Series and Comparison With *Levitus et al.* [2005]

[9] Yearly grid-point anomalies were averaged in each layer to derive global temperature and heat content anomalies. It is interesting to compare our results with those of LAB2005, who used an extended WOA01 data set and a

Table 1. XBT Temperature Biases From XBT/CTD Inter-Comparison Experiments

Author	Data Acquisition Date	Temperature Offset
W. Wood (Practical accuracy of Sippican T-7 XBT's, unpublished manuscript, 1976)	?	Positive offset
Heinmiller et al. [1983]	1973–79	0.13–0.19°C
Bailey et al. [1989]	Historical archive	Positive T-drift with depth
Wright and Szabados [1989]	?	0.11–0.24°C
Boyd and Linzell [1993]	1991	0.07°C
Hallock and Teague [1992]	1990	0.1°C
Schmeiser [2000]	2000	0.15°C
Roth [2001]	2001	0.08°C
Boedeker [2001]	2001	0.09°C
Fang [2002]	2002	0.025–0.107°C
Kizu and Hanawa [2002]	1985–2001	up to 1°C
Dixon [2003]	2003	0.13°C
Laird [2006]	2006	0.04°C
F. Reseghetti et al. (Improved quality check procedures of XBT profiles in MFS-VOS, submitted to Ocean Sciences, 2006)	2004	O(0.05°C) below 400 m 0.2–2.8°C in the thermocline
This Study	1965–2000	0.28°C (depth-time average)

different method for anomaly calculations. Also, before interpolating to standard levels LAB2005 corrected the depths of the originator's XBT profiles using a new drop-rate equation [Hanawa et al., 1994]. For the purpose of comparison we repeated LAB2005 calculations using their original 1° × 1° temperature anomaly fields and following their procedure as outlined by J. Antonov (personal communication, 2005). Though the temporal development of temperature anomalies is qualitatively similar in both studies (except for the deepest layers), the range of the temperature (heat content) change between the first (1957–1966) and the last (1987–1996) decade is a factor of 1.4 (0–400 m layer) to 2.2 (0–3000m layer) larger in our calculations when all data types are used (Figure 3). We explain this difference due to (1) a strong smoothing in the LAB2005 analysis scheme, and (2) by the use of a zero anomaly in their first-guess fields. As noted by Hurrell and Trenberth [1999], optimal interpolation methods can be biased towards zero anomaly, and so underestimate climatic signals. Because a number of boxes not sampled is considerably larger in the deeper layer (Figure 3), the influence of the zero first guess field is largest in the deep layers, leading to a decrease of the depth-integrated heat content amplitude. The agreement between our and LAB2005 time series improves considerably if only 1° × 1° boxes with observations are used for spatial averaging (for such boxes the optimally analysed anomaly field is less influenced by a zero-anomaly first guess field).

[10] Apart from systematic errors the important issue is an effect of varying sampling on large-scale temperature series. Random instrumental errors and meso-scale noise will cancel during the box averaging if the number of observations involved is sufficient. However, in the case of highly irregular and insufficient sampling the formal standard error will be no more representative for the total error. For the 2 × 4-degree binning only a few percent of the squares below 1500 meters are sampled each year (Figure 3). In order to assess the extent to which the more sparsely available data of earlier decades monitored the large-scale temperature averages, global anomalies were computed using sampling specific for each year between 1950 and 2000. This method was used by Jones et al. [1986] to investigate the effect of incomplete surface temperature data. The rms deviation from the full-sampling case is taken as a measure of the sampling error. As would be expected, the largest sampling errors of about $16 \cdot 10^{22}$ J occur for years before 1960s (Figure 3). Even after the 1960s a typical sampling error for the heat content in the 0–3000 m layer is about $8 \cdot 10^{22}$ J.

5. Discussion

[11] In most cases temperature offsets of O (0.1°C) are at the noise level, with respective data not being rejected
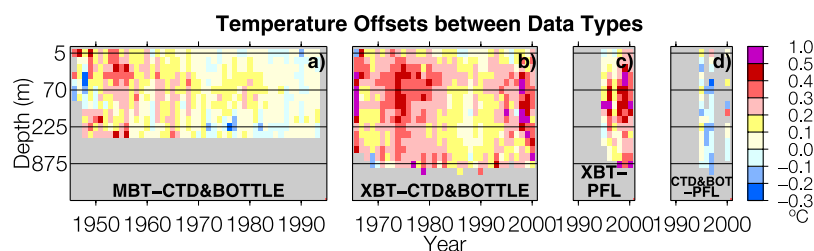

Temperature Offsets between Data Types

Figure 2. Yearly mean temperature offsets versus depth relative to CTD/bottle and profiling float data: (a) MBT vs. CTD/bottle; (b) XBT vs. CTD/bottle; (c) XBT vs. profiling floats; and (d) CTD/bottle vs. profiling floats. Calculations are based on at least 30 overlapping 2° × 4° boxes per year and level.
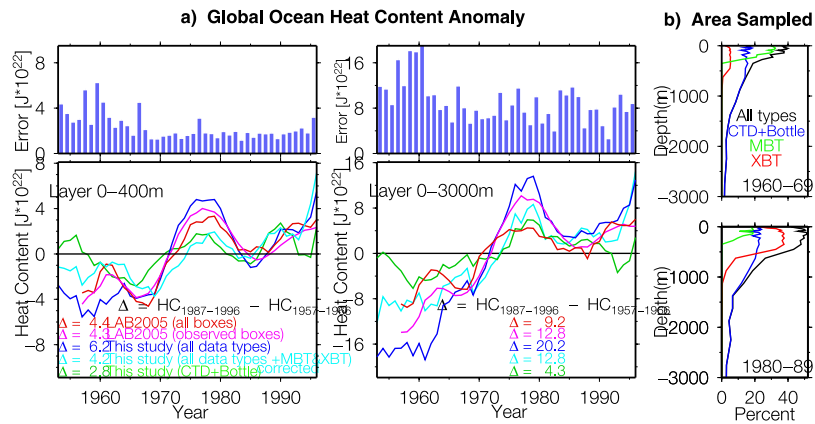
Figure 3. (a) Time series of the yearly ocean heat content anomaly (10^{22} J) for (left) 0–400 m and (right) 0–3000 m layers with respective sampling error estimates. Different estimates of the heat content increase between two decades (1957–1966 and 1987–1996) are given (triangles). (b) Mean percentage of the yearly sampled $2° \times 4°$ boxes for two decades (1960–69 and 1980–89).

through the quality control procedures. However, these offsets are not negligible for the estimation of the long-term temperature changes in the ocean, as the total temperature increase in the upper layers during the historical period is about a half-degree. For the dataset used in this study XBT temperature profiles represent about 20 to 60 and 40 to 70 percent of the total profiles in 0–10 and 300–400 meter layers respectively (Figure 1). As the XBT temperature observations dominate, they introduce a positive bias into the global temperature anomaly estimates, resulting in a larger apparent warming following their introduction in the late 1960s and early 1970s.

[12] To estimate the possible effects of the positive XBT-temperature bias on anomaly calculations year-mean depth-dependent corrections were introduced for each XBT temperature observation (the corrections were obtained by comparing XBT and CTD/Nansen bottle anomalies for overlapping $2° \times 4°$ boxes). Though a cruise-wise correction of the XBT is preferable, it remains beyond the scope of this study because of the absence of collocated high quality data and of the respective metadata (manufacturer, probe and acquisition system type). As expected, the adjustment of XBT data leads to a significant decrease of the global warming estimates. The heat content difference between the periods 1957–1966 (pre-XBT) and 1987–1996 is reduced by a factor of 0.68 and 0.60 for the layers 0–400 m and 0–3000 m respectively, when XBT temperature corrections are introduced (Figure 3). Such corrections if

applied would correspondingly reduce the estimate of the ocean warming in LAB2005 calculations.

[13] As noted by LAB2005, a heat content maximum is observed between about 1972 and 1983. This feature is also pronounced in our calculations (see Figures 1 and 4). *Gregory et al.* [2004] have cast doubt on the reality of this feature, but LAB2005 argue that the feature is real since the data coverage is excellent. Both LAB2005 and our calculations show a temperature maximum near 1978–80, more pronounced in the anomaly time series for the Pacific and Indian oceans. However, the feature can not be directly related neither to the Pacific Decadal Oscillation nor to the El-Niño/Southern Oscillation patterns. According to our calculations below about 100 m depth a strong positive anomaly is noticeable mostly for the XBT data and coincides with the time period of a particularly large XBT bias (Figure 2). The comparison of temperature anomalies for the "warm decade" 1973–1982 calculated separately for XBT and CTD/bottle data indicates a spatially uniform positive bias of the XBT data (Figure 4). Calculations of global anomalies using only CTD/bottle data also reproduce a maximum around 1975, but with a much smaller magnitude, suggesting that a positive XBT bias might be responsible for an exaggerated temperature signal during this decade.

[14] Unlike the MBT data, no gradual improvement in the offset between the XBT and CTD/bottle temperatures is observed. Indeed, a period after the main WOCE operational phase is characterized by higher offsets. According to
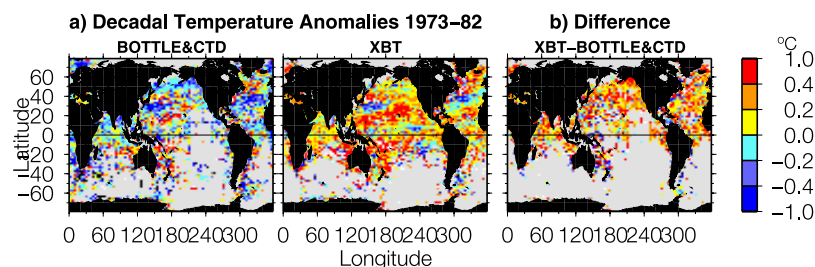

Figure 4. (a) Temperature anomalies in the layer 300–400 m averaged over the "warm decade" 1973–1982 for bottle/CTD and XBT data, and (b) temperature offset of XBT data relative to bottle/CTD data.

Figure 2 XBT-temperature biases calculated relative to the CTD/bottle and profiling float data are in a very good agreement. Both comparisons reveal an XBT temperature bias of more than 0.5°C above 200 m after about 1995. Accordingly, a rapid increase of the temperature/heat content after 1994 (Figure 3) may be partly an artefact of this bias. A direct comparison between the CTD/bottle and float data (Figure 2) indicates a small negative CTD/bottle bias of about 0.03°C (the average value between 1994 and 2000). However, this value is indistinguishable from zero because of a smaller number of boxes with overlapping data due to a decreasing percentage of the CTD/bottle data since mid-1990s (Figure 1b).

[15] In our calculations no additional XBT fall-rate corrections were introduced. As noted by *Willis et al.* [2004], the application of this correction usually results in an overall warming of the profiles, since corrected depths become deeper and temperature generally decreases with depth. It means that the fall-rate corrected profiles would exhibit on average even higher biases with respect to the CTD/bottle data.

[16] Due to their wide introduction as a cost effective and easy measurement technique XBTs have now enormous impact upon international data bases. As our results show, a development of a proper correction procedure is necessary for an accurate estimation of the global warming. Comparison with LAB2005 results shows that the estimates of global warming are rather sensitive to the data base and analysis method chosen, especially for the deep ocean layers with inadequate sampling. Clearly instrumental biases are an important issue and further studies to refine estimates of these biases and their impact on ocean heat content are required. Finally, our best estimate of the increase of the global ocean heat content between 1957–66 and 1987–96 is $12.8 \pm 8.0 \cdot 10^{22}$ J with the XBT offsets corrected. However, using only the CTD and bottle data reduces this estimate to $4.3 \pm 8.0 \cdot 10^{22}$ J.

References

Bailey, R. J., H. E. Phillips, and G. Meyers (1989), Relevance to TOGA of systematic XBT errors, in *Proceedings of the Western Pacific International Meeting and Workshop on TOGA COARE*, edited by J. Picaut, R. Lukas, and T. Delcroix, pp. 775–784, Inst. Fr. de Rech. Sci. pour le Dev. en Coop., Noumea, New Caledonia.
Boedeker, S. (2001), Comparison of CTD and XBT temperature profiles, project report for OC3570, Nav. Postgrad. Sch., Monterey, Calif. (Available at http://www.weather.nps.navy.mil/~psguest/OC3570/CDROM/summer2001/Boedeker/)
Boyd, J. D., and R. S. Linzell (1993), The temperature and depth accuracy of Sippican T-5 XBTs, *J. Atmos. Oceanic Technol.*, 10, 128–136.
Conkright, M. E., et al. (2002), *World Ocean Database 2001*, vol. 1, *Introduction*, edited by S. Levitus, NOAA Atlas NESDIS 42, 159 pp., NOAA, Silver Spring, Md.
Dixon, J. S. (2003), A comparison of expendable bathythermograph and conductivity temperature depth profiles, project report for OC3570, Nav. Postgrad. Sch., Monterey, Calif. (Available at http://www.weather.nps.navy.mil/~psguest/OC3570/CDROM/winter2003/Dixon/)
Fang, C.-L. (2002), XBT/CTD comparisons, project report for OC3570, Nav. Postgrad. Sch., Monterey, Calif. (Available at http://www.weather.nps.navy.mil/~psguest/OC3570/CDROM/summer2002/Fang/)
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., et al. (2001), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.*, 28, 2621–2624.
Gouretski, V. V., and K. Jancke (1999), A description and quality assessment of the historical data for the South Pacific Ocean, *J. Atmos. Oceanic Technol.*, 16, 1791–1815.
Gouretski, V. V., and K. Jancke (2001), Systematic errors as the cause for an apparent deep water 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 [CD-ROM], *Ber. Bundesamt Seeschiffahrt Hydrogr. Rep. 35*, 52 pp., Bundesamt Seeschiffahrt Hydrogr., Hamburg, Germany.
Gregory, J. M., H. T. Banks, P. A. Scott, J. A. Lowe, and M. D. Palmer (2004), Simulated and observed decadal variability in ocean heat content, *Geophys. Res. Lett.*, 31, L15312, doi:10.1029/2004GL020258.
Hallock, Z. R., and W. J. Teague (1992), The fall rate of the T-7 XBT, *J. Atmos. Oceanic Technol.*, 9, 470–483.
Hanawa, K., P. Raul, 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), *Tech. Pap. Mar. Sci. 67*, 46 pp., UNESCO, Paris.
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.
Hurrell, J. W., and K. E. Trenberth (1999), Global sea surface temperature analyses: Multiple problems and their implications for climate analysis, modelling and reanalysis, *Bull. Am. Meteorol. Soc.*, 80, 2661–2678.
Jones, P. D., S. C. Raper, and T. M. L. Wigley (1986), Southern Hemisphere surface air temperature variations: 1951–1984, *J. Clim. Appl. Meteorol.*, 25, 1213–1230.
Kizu, S., and K. Hanawa (2002), Recorder-dependent temperature error of expendable bathythermograph, *J. Oceanogr.*, 58, 469–476.
Laird, A. (2006), XBT and CTD temperature measurement comparison and XBT and GDEM sound velocity profile comparison, project report for OC3570, Nav. Postgrad. Sch., Monterey, Calif. (Available at http://www.weather.nps.navy.mil/~psguest/OC3570/CDROM/winter2006/Laird/)
Levitus, S., J. Antonov, and T. Boyer (2005), Warming of the world ocean, 1955–2003, *Geophys. Res. Lett.*, 32, L02604, doi:10.1029/2004GL021592.
McPhaden, M. J., et al. (1998), The Tropical Ocean-Global Atmosphere observing system: A decade of progress, *J. Geophys. Res.*, 103(C7), 14,169–14,240.
Roemmich, D., and W. B. Owens (2000), The ARGO project: Global ocean observations for understanding and prediction of climate variability, *Oceanography*, 13(2), 45–50.
Roth, M. J. (2001), XBT and CTD temperature measurement comparison, quality of JJYY data and XBT data analysis of the mixed layer depth, project report for OC3570, Nav. Postgrad. Sch., Monterey, Calif. (Available at http://www.weather.nps.navy.mil/~psguest/OC3570/CDROM/winter2001/Roth/)
Schmeiser, G. (2000), XBT and CTD temperature measurement comparison, project report for OC3570, Nav. Postgrad. Sch., Monterey, Calif.
Seaver, G. A., and S. Kuleshov (1982), Experimental and analytical error of the expendable bathythermograph, *J. Phys. Oceanogr.*, 12, 592–600.
Smith, T. M., and R. W. Reynolds (2002), Bias corrections for historical sea surface temperatures based on marine air temperatures, *J. Clim.*, 15, 73–87.
Willis, J. K., D. Roemmich, and B. Cornuelle (2004), Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales, *J. Geophys. Res.*, 109, C12036, doi:10.1029/2003JC002260.
Wright, D., and M. Szabados (1989), Field evaluation of real-time XBT systems, *Oceans*, 89(5), 1621–1626.

────────────────
V. Gouretski, Klimawissenschaften, Alfred-Wegener-Institut für Polar- und Meeresforschung, Bussestrasse 24, Bremerhaven, D-27570 Germany. (vgouretski@awi-bremerhaven.de)
K. P. Koltermann, Intergovernmental Oceanographic Commission, UNESCO, 1 rue Miollis, Paris F-75732, France.