

Consistent near-surface ocean warming since 1900 in two largely independent observing networks

Viktor Gouretski,¹ John Kennedy,² Tim Boyer,³ and Armin Köhl⁴

Received 5 July 2012; revised 27 August 2012; accepted 29 August 2012; published 5 October 2012.

[1] We compare historical global temperature time series, based on bias-adjusted sea-surface temperatures with independent temperature time series, for the upper 20 meter layer of the ocean based on the latest update of an historical hydrographic profile data set. Despite the two underlying data sets being different in number of data points, instrumentation and applied adjustments, both of the time series are consistent in showing an overall warming since 1900. We also extend records of temperature change in the upper 400 m back to 1900. Noting that the geographic coverage is limited prior to 1950, the temperature change in the 0–400 m layer is characterized by two periods of temperature increase between 1900 and 1940–45 and between 1970 and 2003, separated by a period of little change. **Citation:** Gouretski, V., J. Kennedy, T. Boyer, and A. Köhl (2012), Consistent near-surface ocean warming since 1900 in two largely independent observing networks, *Geophys. Res. Lett.*, 39, L19606, doi:10.1029/2012GL052975.

1. Introduction

[2] Numerous studies have identified an overall rise of the surface temperature of the Earth since the nineteenth century [Smith *et al.*, 2008; Hansen *et al.*, 2010; Morice *et al.*, 2012]. The global-average surface temperature is estimated from a composite dataset that includes both land- and sea-surface temperature (SST) observations. In addition to studies analyzing surface temperature data, collections of historical hydrographic temperature profiles have been used to estimate the change in heat content of the global oceans [Levitus *et al.*, 2005, 2009, 2012; Gouretski and Koltermann, 2007].

[3] Two main sources of uncertainty affect both the surface and subsurface time series based on in situ data. The first is related to insufficient data coverage both in space and time, with extremely irregular sampling in the earlier parts of the records. The second arises from instrumental biases which can be comparable in magnitude to real variability in the climate. Jones and Wigley [2010] identified biases in SST measurements as the most important remaining uncertainty associated with estimating global average temperature change.

Prior to the 1980s, SST measurements were mostly made using buckets or in the engine rooms of ships. Folland and Parker [1995] described systematic errors in SST observations associated with the use of uninsulated buckets for water sampling and developed adjustments. Uncompensated biases associated with a shift in the database from engine room measurements (relatively warm biased) to bucket measurements (relatively cold biased) occurred at the end of World War II and led to an apparent drop in observed SSTs in late 1945 [Thompson *et al.*, 2008]. More recent studies [Kennedy *et al.*, 2011a, 2011b] attempt to quantify SST biases and their associated uncertainties in the post war period. However, Kennedy *et al.* [2011b] note that “Until multiple, independent estimates of SST biases exist, a significant contribution to the total uncertainty will remain unexplored. This remains a key weakness of historical SST analysis”.

[4] Gouretski and Koltermann [2007] revealed significant biases both in the expendable BathyThermograph (XBT) and in the Mechanical BathyThermograph (MBT) data used to measure subsurface ocean temperatures. The effect of this instrumentation problem appeared as an artificial pattern of ocean warming around 1975–1985 in the Levitus *et al.* [2005] time series of ocean heat content within the upper 700 meters. Further studies have confirmed the general characteristics of the biases described by Gouretski and Koltermann [2007] and correction schemes have been developed for both MBT and XBT data [Wijffels *et al.*, 2008; Ishii and Kimoto, 2009; Levitus *et al.*, 2009; Gouretski and Reseghetti, 2010]. However, Lyman *et al.* [2010] showed that even in the recent record (1994–2008) the uncertainties of the bias adjustments applied to subsurface data were a major component of the total uncertainty in estimates of ocean heat content. It is often difficult to assess the effectiveness of bias adjustments in reducing the imprint of systematic errors in climate data because independent test data are rarely available. In this analysis an initial approach to resolve this uncertainty is made by comparing two independently derived estimates of near-surface ocean temperature. In addition, a time series of the mean temperature within the upper 400 meters of the world ocean is calculated back to 1900.

2. Data

2.1. Subsurface Hydrographic Data

[5] The global hydrographic database WOD09 [Boyer *et al.*, 2009] (downloaded December 2011) provided the majority of the subsurface hydrographic data used in this analysis. Additional profiles not included in WOD09 were downloaded (December 2011) from the International Council for the Exploration of the Sea (ICES) database, the Japanese oceanographic data centre, the Mediterranean Oceanic Database

¹KlimaCampus, University of Hamburg, Hamburg, Germany.

²Hadley Centre for Climate Prediction and Research, Met Office, Exeter, UK.

³National Oceanographic Data Center, NOAA, Silver Spring, Maryland, USA.

⁴Institute of Oceanography, University of Hamburg, Hamburg, Germany.

Corresponding author: V. Gouretski, KlimaCampus, University of Hamburg, Grindelberg 5, D-20144 Hamburg, Germany. (viktor.gouretski@zmaw.de)

(MODB), the German Oceanographic Datacenter (DOD), and taken from cruises conducted by the Institute for Marine Research of Hamburg University. WOD09 contributed 97.3% of the 7,615,223 temperature profiles with measurements at least 20 meters deep.

[6] The types of instrument used in this study are: 1) bottle casts, 2) Conductivity-Temperature-Depth (CTD) profilers, 3) Mechanical BathyThermographs (MBT), 4) eXpendable BathyThermographs (XBT), 5) Argo profiling floats, and 6) pinniped mounted sensors (sensors attached to marine mammals). Three instrument groups - the bottle cast, CTD and Argo float data - are characterized by a higher precision and together they constitute a reference dataset against which inhomogeneities associated with gradual changes in the mix of instrumentation can be assessed [Gouretski and Koltermann, 2007].

[7] The temperature profiles were interpolated onto 1-meter levels and integrated vertically to obtain point estimates of the mean temperature within the upper 20 meter and 400 meter layers. The data from the upper 20 meters were chosen because they mostly lie within the upper mixed layer [de Boyer Montégut et al., 2004] and are therefore most suitable for comparison with SST data. The 0–400 meter layer was chosen because the maximum sample depth is 460 meters for the numerous XBT probes of T4 and T6 types. Temperature profiles with a deepest measurement shallower than 400 meters were also used. The depth-averaged temperature for such profiles were converted into 0–400 meter averaged values using correlations between the mean temperatures in 0–400 meter layer and the mean temperature in shallower layers. These correlations were defined locally using high-resolution CTD temperature profiles. To reduce the possible geographical bias due to the generally higher observational density on the shelf ocean regions, areas shallower than 400 meters were excluded from the analysis of the 0–400 meter layer. In order to minimize the effects of biases in MBT and XBT data, the bias model developed by Gouretski and Reseghetti [2010] was used. This model takes into account both the thermal bias and the depth bias and allows for a better elimination of the original (predominantly positive) temperature bias in the XBT and MBT data. Before 1963, MBT data were routinely adjusted to SST measurements taken at the same time as the MBT deployment. Hazelworth [1966] showed that unadjusted temperatures from well-calibrated MBTs were closer to bottle (reversing thermometer) temperatures than those adjusted to SST, and the practice of adjustment was discontinued operationally in 1963. The subsurface MBT data used in this study were bias adjusted based on comparison with reversing thermometer (and CTD for later years) temperatures and depths [Gouretski and Reseghetti, 2010], in essence applying a broad-scale removal of the original SST adjustment. It should be noted, however, that no “standard bias model” exists, so that other bias-adjustment schemes might result in somewhat different time series and these differences constitute an important component of uncertainty.

2.2. Sea-Surface Temperature Data

[8] The SST data are an update of the Met Office Hadley Centre SST data set, HadSST3 [Kennedy et al., 2011a, 2011b]. SST observations were taken from version 2.5 of the International Comprehensive Ocean-atmosphere Data Set (ICOADS) [Woodruff et al., 2011] from 1900 to 2006.

Drifting buoy observations from 2007 to 2010 were taken from the NCEP real time GTS updates and supplemented with ship observations stored in the UK Met Office data base. The data were quality controlled and processed in the same way as for HadSST3 [Kennedy et al., 2011a]. Between 1900 and 2010, 244,104,080 SST observations passed quality control. ICOADS contains near-surface observations from the World Ocean Database (WOD05). Between 1900 and 1940 the percentage of WOD05 observations in ICOADS is around 1%. From 1940 to 1955 WOD05 observations constitute between 8 and 10% of observations. From 1955 to 1990, fewer than 5% of observations come from WOD05. Between 1990 and 2006 the fraction drops below 5%. Not all WOD05 observations are sub-surface measurements; the ICOADS documentation states that in some cases the SST from WOD05 is a simultaneous reference SST measurement. Because the subsurface observations contribute only a few percent of the total SST measurements, the SST and the subsurface datasets are considered to be nearly independent. The applied bias adjustments are completely independent. Meta data were taken from WMO Pub 47 (<http://www.wmo.int/pages/prog/www/ois/pub47/pub47-home.htm>) and more recent updates from ESURFMAR <http://esurfmar.meteo.fr/doc/vosmetadata/index.php>.

[9] Bias adjustments, for the varying mix of bucket, engine-intake and buoy measurements, were applied to the data according to the scheme of Kennedy et al. [2011b]. Uncertainties in the bias adjustments were represented by generating 100 versions of the adjustments - and therefore 100 gridded data sets - which had the parameters of the scheme randomly assigned within their uncertainty ranges. These parameters included the estimated fraction of incorrect metadata, the timing of the switch from uninsulated to insulated buckets, the relative biases between ship and buoy measurements, the size of the bucket adjustments and estimates of the size and temporal evolution of the engine-intake biases.

3. Method

3.1. Calculation of Temperature Anomalies

[10] For the subsurface temperature the ten-year period 2001–2010 was selected as a base period for anomaly calculations, because the profiling float data available in this decade provide a more even and complete spatial coverage than was available in previous decades. As the Argo global distribution was not achieved before 2006, there is still a mismatch between the hemispherical median observation dates, but is obviously less important compared to longer averaging periods used by Levitus et al. [2005] and Gouretski and Koltermann [2007] since the mean temperature anomaly for the last decade was relatively flat. The monthly climatological temperature fields for the base period were calculated on a 0.25° geographical grid, using an inverse-distance weighting function $w = (R^2 - r^2)/(R^2 + r^2)$ when averaging point temperature values within the radius of influence, $R = 111$ km, around each grid-point. r is the distance from the location of the observation to the grid-point. The point temperature anomalies were then calculated simply as a difference between the layer-averaged temperature and the respective monthly climatological value. To reduce the effect of outliers, the lowest and highest 2.5% of point temperature anomalies were discarded for a running 15-year window around each month January, 1900 through December, 2010. The subsurface point anomalies

were finally averaged onto a 5° grid using the median as the estimate for the grid-box average.

[11] In HadSST3 [Kennedy *et al.*, 2011b], anomalies for each SST observation were calculated relative to a 1961–1990 SST climatology [Rayner *et al.*, 2006]. In order to make a direct comparison with the subsurface analysis, anomalies relative to the period 2001–2010 were calculated and it is these that are used in our analysis. This was done separately for each month by calculating the average anomaly for each 5×5 grid box for that month for the period 2001–2010. The 12 monthly 2001–2010 mean anomaly fields were then used to calculate anomalies relative to 2001–2010 for all months from 1900 to 2010.

3.2. Calculation of Temperature Anomaly Time Series and Uncertainties

[12] Monthly global SST time series were calculated by taking an area-weighted average of the available 5 degree grid boxes in each month. A monthly time series was generated for each of the 100 realizations and the spread in the estimates represents the uncertainty associated with the bias adjustments. The uncertainties associated with measurement errors, under-sampling within a grid box and limited observational coverage were calculated following the method of Kennedy *et al.* [2011a] and added to the bias uncertainty as if they were independent errors. The uncertainty due to limited observational coverage was estimated by comparing complete and sub-sampled fields from the globally complete SST analysis of Rayner *et al.* [2003]. For the 5-year running average (see section 4), the conservative assumption was made that the measurement and sampling uncertainty was as large as that of an annual value.

[13] The subsurface monthly time series were also calculated by taking an area-weighted average of the sampled 5 degree grid boxes. Since the subsurface sampling of the global ocean in the earlier decades of the record is extremely sparse we do not attempt to fill data gaps using interpolation methods, because the global time series obtained in this way are sensitive to the choice of infilling method [Carson and Harrison, 2008]. We estimate the uncertainty of the global subsurface anomaly due to limited observational coverage using the global GECCO ocean synthesis (German contribution to Estimating the Circulation and Climate of the Ocean) [Köhl and Stammer, 2008]. GECCO provides an estimate of the time evolving ocean circulation between 1952 and 2001 that is consistent with the dynamics embedded in a 1^0 ocean general circulation model driven by NCEP surface fluxes. An adjoint model was used to reduce the model-data misfit by iteratively changing the surface fluxes and the initial temperature and salinity fields. The data used as constraints during this assimilation include several satellite data sets (e.g., sea-surface height, sea-surface temperature fields), surface drifter velocities, and in-situ hydrographic temperatures and salinity profiles from the WOA2001 database [Conkright *et al.*, 2002]. Although the temperature profiles that were assimilated were not bias corrected, the dynamical constraints of the analysis meant that the artificial pattern of ocean warming around 1975–1985 was not present in GECCO [Stammer *et al.*, 2010]. Subsurface time series were calculated from the GECCO output using: 1) only those boxes sampled in the historical record during each particular year and month and 2) using the full model data coverage. The standard deviation of the difference between the two time series was used as the measure

of uncertainty due to the irregular and incomplete sampling for that month. Similar to the SST measurements, both the irregular sampling and residual biases are important components of the overall error budget for the subsurface time series.

4. Results

4.1. SST Versus 0–20 m Near-Surface Analysis

[14] Figure 1a shows a time series of monthly global average near-surface ocean temperature anomalies. Series calculated from SST and from subsurface profiles within the upper 20 meters of the ocean show very similar interannual variability back to around 1945, although the 0–20 m series is generally cooler than the SST series between 1960 and 1975. Prior to 1960, the coverage of hydrographic observations (shown as insets in the main diagram) becomes much sparser. Despite this, the global-average SST and near-surface estimates continue to track each other, although the agreement becomes progressively worse earlier in the record as the sampling in the subsurface analysis becomes more regionally confined. As illustrated by Figure 1b the mean monthly absolute temperature time series for the sea surface and for the 0–20 m level diverge before 1940 with near-surface sampling being biased to higher latitude resulting in a colder mean temperature. Before 1940, both time series show a warming trend, but the rate of warming is higher in the near-surface data than it is in the SST. Figure 1c shows the estimated 2-sigma uncertainty range and the difference between the estimates. The differences mostly fall within the uncertainty range back to around 1930. Both time series show a temperature increase from 1900 to about 1945, a slight decrease to the mid-1970s, and a temperature rise to the end of the record. While it is possible that the agreement is fortuitous, the fact that two independently derived series should agree so closely, is an indication that there is a common signal and that it is being faithfully represented, albeit with some uncertainty, by the data. An alternative view of global temperatures is provided by creating histograms of the point anomalies (Figure 1d). The point anomalies are characterized by a uni-modal distribution with the majority of the values concentrated within a narrow band around the mode. As in Figure 1a, the histograms show an overall warming of the global ocean since 1900.

[15] Figure 2a shows time series of 5-year running average near-surface ocean temperatures, which highlight the difference in low-frequency variability between the data sets. The higher trend in the subsurface analysis prior to 1945 is also seen in the SST analysis subsampled to have the same coverage as the subsurface analysis and is therefore likely to be due largely to poor geographical sampling. The subsurface data for the years before ca. 1920 have a strong geographical bias with the majority of the data coming from the (North) Atlantic Ocean (see Figure 1a). Both the near-surface and the subsampled surface time series indicate a warming of about 1.3 – 1.4°C since 1900, which is characteristic for the Atlantic Ocean. This is larger than for the whole global ocean, which amounts to around 0.7 – 0.8°C according to the more spatially complete SST time series. Levitus *et al.* [2005] also reported a higher warming rate for the Atlantic Ocean since the 1950s.

[16] The fact that the estimated uncertainties of the SST and near-surface analyses do not overlap implies that the coverage

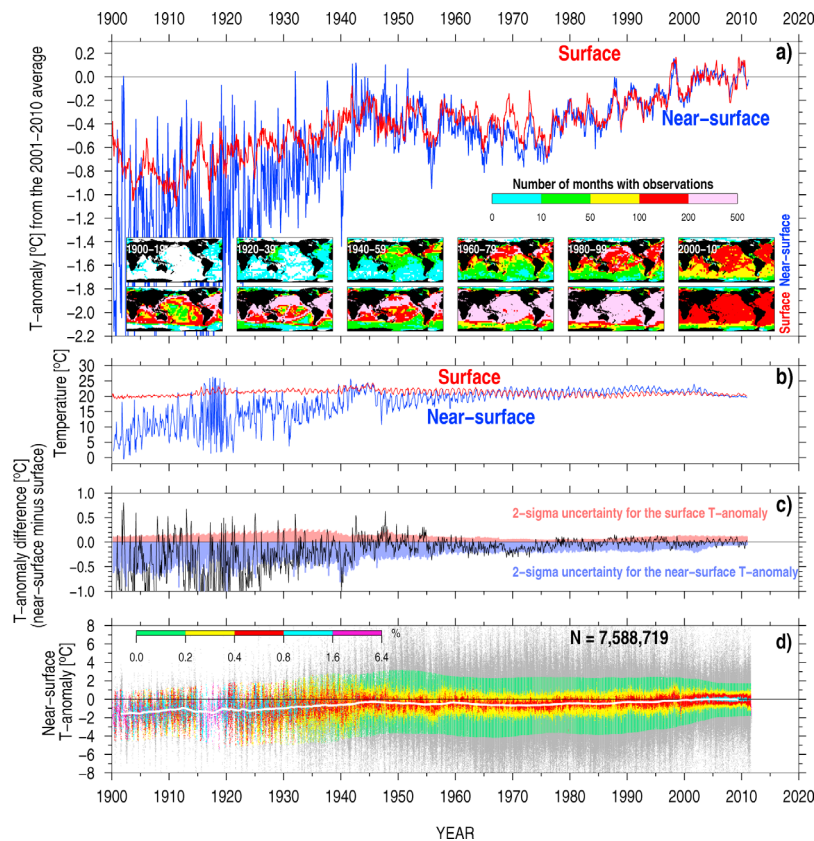


Figure 1. (a) Global temperature anomaly time series (monthly values relative to the average for 2001–2010) calculated for the sea surface (red) and for the near surface (0–20 m) layer (blue). Both time series are obtained by area-weighted averaging of all available monthly 5×5 -degree box anomalies. Inset maps show spatial distribution of the temperature observations/profiles and the number of observations/profiles within 5×5 -degree boxes for selected time periods. (b) Time series of the sea surface temperature (red) and of the 0–20 m layer temperature (blue). Both time series are obtained by weighted averaging of 5×5 -degree box absolute temperatures. (c) Difference between the sea surface and the subsurface temperature anomaly (black line) with respective two-sigma uncertainties shown in color. (d) Histograms of temperature anomalies for the 0–20 m layer based on temperature profiles from the global hydrographic database. Anomalies are calculated relative to the 10-year reference period 2001–2010. Color scale shows percent of the anomalies within $1 \text{ month} \times 0.01^\circ\text{C}$ bins from the total number of anomalies for each calendar month and year. Histogram bins colored in grey correspond to 5 percent of original anomalies rejected through the application of the median filter. White line denotes the position of the distribution mode (5-year running mean of monthly anomaly values). N gives the total number of the original temperature profiles.

uncertainties estimated by subsampling globally complete GECCO reanalyses do not capture the full uncertainty due to poor coverage in the early record. Figure 2a also shows the effect of using observed depths, or interpolated depths for estimating the temperature in upper 20 meters. The divergence between the two curves prior to 1945 arises from an abrupt change in the composition of the subsurface data set. Before 1940, all the data are from bottle cast temperature profiles, which have a relatively coarse vertical resolution. MBTs, which have a greater vertical resolution, were introduced during the Second World War greatly improving the accuracy with which the temperature structure of the upper layers could be determined. Consequently, the uncertainty of estimating the average temperature in the upper 20 meters, as represented by the difference between the two curves, is larger before 1940.

[17] Also shown in Figure 2a are the series from the unadjusted data. The adjustments do not improve the agreement between the two data sets at all times, nor does the act of sub-sampling the SST data to have the same coverage as the

near-surface analysis. There are two possible reasons for this which might be acting together. The first possibility is that the 0–20 m layer and the sea-surface warmed and cooled at different rates though always in tandem. *Grotsky et al.* [2008] showed relative trends between SST and mixed-layer temperature for certain regions, but their analysis did not consider the possible effects of biases in the measurements. The second possibility is that the bias adjustments applied to one or the other data set are incorrect. If the differences are due to unresolved biases then they suggest an uncertainty of around 0.1°C at decadal time scales. It is interesting to consider on which side such a bias is most likely to lie. The bias adjustments applied to the SST data are generally larger than the adjustments applied to the hydrographic profiles. Furthermore, the adjustments applied to the bathythermographic profiles are calculated relative to reference datasets (based on CTD and bottle measurements) whereas the SST adjustments, based as they are on estimates of biases from the literature, are not. Independent subsets of the SST series based on adjusted bucket measurements and adjusted engine

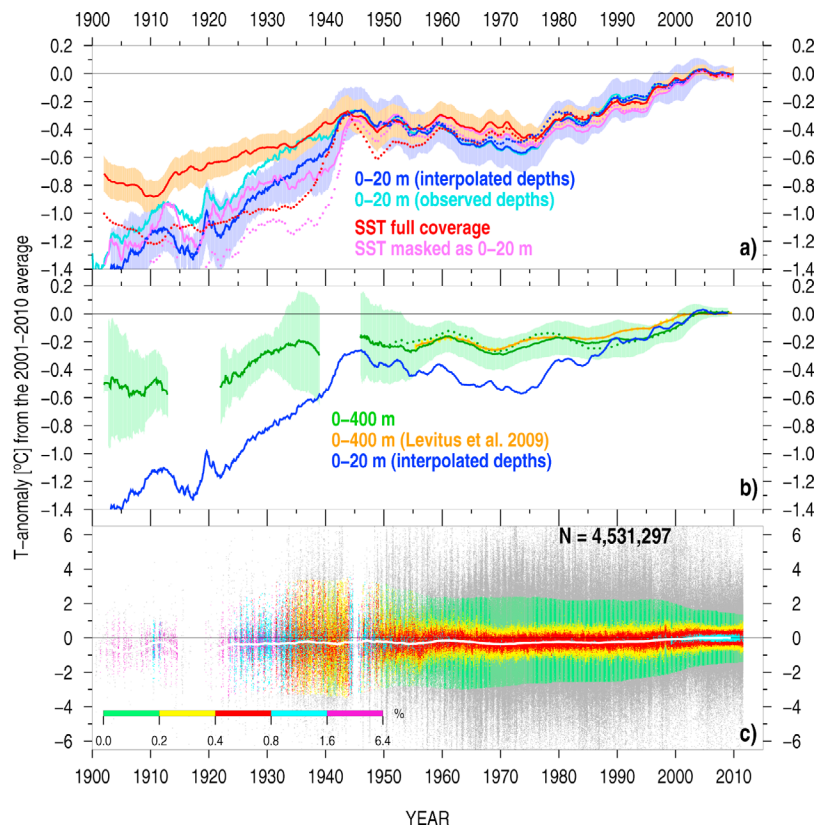


Figure 2. (a) Five-year running mean temperature anomalies estimated using the sea surface temperature data and the near-surface (0–20 m) temperature profiles. Red: anomaly for the full sea surface temperature dataset; magenta: anomaly for the sea surface temperature dataset sub-sampled to the coverage of the less abundant subsurface data; dark blue: temperature anomaly for the 0–20 m layer based on the vertically interpolated (1 meter) temperature profiles; cyan: the same time series based on the observed level data. Shading shows uncertainties due to the incomplete sampling for the 0–20 m time series (light blue) and to the sum of the uncertainties due to bias adjustment, measurement, sampling and coverage for the sea surface time series (light-orange). All solid-line curves correspond to the datasets adjusted for biases, whereas the dotted lines are based on the original unadjusted datasets. (b) Temperature anomaly time series for the 0–400 m layer estimated in this study (green) and similar to *Levitus et al.* [2009] (orange). Shading shows uncertainties due to the incomplete sampling for the 0–400 m (light-green) and the error bars show the formal standard error of the 5-year mean value for the 0–400 m by *Levitus et al.* [2009] (yellow). All solid-line curves correspond to the datasets adjusted for biases, whereas the dotted lines are based on the original unadjusted datasets. (c) The same as in Figure 1d but for the layer 0–400 m.

room measurements agree well on a global and hemispheric scale, but not perfectly. Collocated differences between the two suggest an uncertainty of around 0.1–0.2°C in the 1940s and 1950s, but a much smaller uncertainty from the 1960s onwards [*Kennedy et al.*, 2011b]. Although the subsurface measurements are adjusted relative to a reference data set, there is a possibility that existing adjustment schemes suffer from over-fitting, with spuriously good agreement occurring where simultaneous XBT and CTD measurements exist, but poorer performance where they do not. There are also systematic drifts in the average absolute temperature of the near-surface measurements (shown in Figure 1b) suggesting systematic changes in the water masses and geographical regions sampled. The issue of residual biases continues to be the focus of research both for the surface and subsurface data, and at present, there are no strong reasons to presume that one anomaly estimate is less biased than the other prior to World War II.

[18] To further compare the two data collections, we produced two sets of decadal temperature anomaly maps based on the surface and near-surface data respectively (Figure 3).

These maps demonstrate that the first decade of the 21st century (2001–2010) was not uniformly warmer than previous decades. Before about 1920, the global ocean was almost everywhere colder than the reference decade of 2001–2010. After 1920, several regions of the global ocean were warmer than in the reference decade. The tropical regions of the East Pacific ocean were dominated by positive temperature anomalies during the 1980s and 1990s due to several strong El-Niño events (1982–83, 1986–87, 1991–92 and 1997–98). In contrast the period 2001–2010 was marked by relatively modest El Niño events and strong La Niña events (2000–01, 2007–08 and late 2010). Evidence of a similar anomaly pattern albeit of smaller magnitude can be identified in the same region during the 1930s and 1940s, possibly due to the protracted El Niño of the early 1940s combined with the positive phase of the PDO at the time. However the data coverage is much poorer in comparison to the later decades.

[19] Another large-scale pattern of positive anomalies (relative to 2001–2010) occurred in the Southern Ocean during 1970s to 1990s. Similar to the East Pacific, surface and near-surface water temperatures in this region were higher

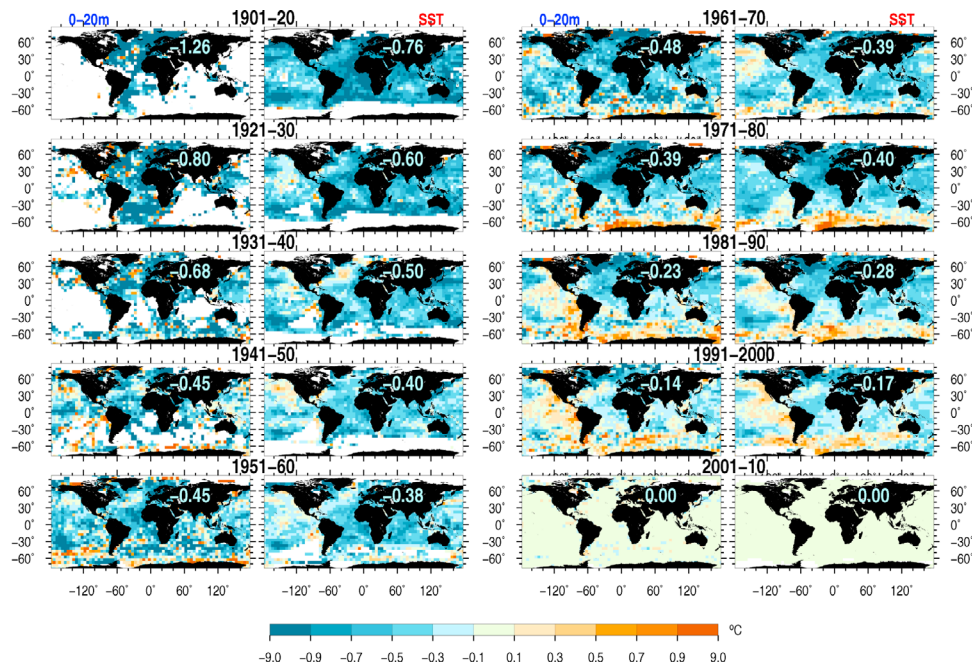


Figure 3. Decadal temperature anomaly (relative to 2001–2010) maps based on the subsurface (0–20 m) data (first and third columns from left) and on the sea surface temperature data (second and fourth columns). Maps are produced by averaging all temperature anomalies in each 5×5 -degree geographical square for the respective time period. Numbers shown on Asia correspond to the globally averaged 5×5 -degree box anomalies.

compared to the reference decade of 2001–2010. The belt of positive anomalies was most pronounced for the decades between 1970 and 2000 within the Atlantic and the Western Indian sectors of the Southern Ocean. It should be noted that the decadal maps for the layer 0–400 m (not shown) do not reveal any significant positive anomalies within the same regions of the Southern Ocean. The anomalies seem to be confined to the near-surface suggesting a different time evolution below the upper mixed layer. However, *Gille* [2002] reported a warming in the even deeper layers 700–1100 m between 1950s and 1980s. Thus, a rather abrupt cooling since the end of 1990s both in the East Pacific (connected to the weakening of El Niño and the shift to the negative phase of the Pacific Decadal Oscillation) and in the Southern Ocean [see also *Knight et al.*, 2009] may have contributed to a flattening of the global temperature anomaly series after about 2000. The flattening can be clearly seen in Figure 2a. It should be noted that the changes revealed in the Southern Ocean refer mostly to the period of the austral summer, since there have been only sporadic winter observations before the implementation of Argo floats.

4.2. Layer 0–400 m

[20] Sub-surface observations in the upper 400 meters are more limited in number. However, the good agreement between the independent temperature anomaly time series for the ocean surface and for the near-surface layer points to the possibility of monitoring secular changes in the deeper ocean. Extension of the analysis back to the beginning of the 20th century reveals evolution of the temperature average over the upper 400 meters that is similar to the near-surface time series, with two periods of warming separated by a period between about 1945 and 1970 when the oceans cooled slightly (Figure 2b).

[21] The uncertainty bounds due to imperfect sampling, estimated using the GECCO reanalysis, are very wide at the beginning of the analyzed time period and between 1930 and the mid-1950s. A reduced number of observations (especially during World War I and to a lesser degree during World War II) and a narrower geographical scope of observations during the wars make the global temperature anomaly estimates much less reliable. The observational gaps are most clearly seen in the point anomaly histogram (Figure 2c). For this reason, we omit the time periods 1913–1920 and 1939–1945 from the analysis. The introduction of profiling floats has led to a significant reduction in the sampling uncertainty after about 2003. Overall, our anomaly estimates based only on the geographical squares with observations suggest a warming of about 0.5°C for the upper 400 meter layer since the beginning of the 20th century. If these are used to estimate the temperature change of the 0–400 m layer across all the global oceans, the warming is between 0.1 and 0.9°C since 1900, or between 0.3 and 0.7°C since 1910. Our estimates of the temperature rise for the 0–400 m layer agree, within the rather broad uncertainty ranges, with the results by *Roemmich et al.* [2012] where the spatial mean warming of about 0.5°C since the Challenger expedition (1872–76) was found for the upper 366 meters. It is worth noting that, while the increase of the mean temperature anomaly for the 0–400 m is smaller than that of the 0–20 m layer, it represents a volume of water that is twenty times larger.

[22] Global analyses of MBT and XBT data [*Levitus et al.*, 2009; *Gouretski and Reseghetti*, 2010] demonstrate that the MBT and XBT bias correction schemes do bring them into a better agreement with the reference CTD and bottle data. The accuracies of the thermometers used on historical Nansen casts (which is the only subsurface data type before 1940) were about 0.05°C or better [*Wüst et al.*, 1932]. However,

the sample depth was often calculated from the length of the wire paid out and the angle of the wire to the vertical measured at the ship deck. This method could lead to large errors in sample depth, systematically underestimating the actually achieved depth and usually causing a respective warm bias in the data. However, introduction of the thermometric method of the depth estimation based on the simultaneous use of protected and unprotected reversing thermometers has essentially solved the depth bias problem [Wüst *et al.*, 1932]. Following the above considerations and taking into account significant differences between the full-coverage and masked SST time series we believe that the remaining uncertainties in our upper-ocean time series are largely due to imperfect sampling. However, the impact of residual or unknown biases can not be excluded.

[23] Levitus *et al.* [2009] estimate global temperature anomalies using extensive interpolation of the gridded anomaly fields. A modified version of the time series from Levitus *et al.* [2009], of mean temperature anomaly 0–400 m, averaged only over gridboxes with data, is shown in Figure 2b. Their time series suggests a more gradual warming since the 1950s than in our analysis. Differences between the two 0–400 m time series may be due to the XBT corrections applied, mapping techniques employed, and/or differences in data and quality control decisions for the data. For instance, modifying the mapping method described here to make it more similar to the method used by Levitus *et al.* [2009] by not applying the depth extension to shallow profiles results in a change in calculated mean temperature anomaly 0–400 m of $<0.05^{\circ}\text{C}$ for years prior to 1955, $<0.02^{\circ}\text{C}$ 1955–1970, and $<0.01^{\circ}\text{C}$ after 1970. Regarding bias corrections, Lyman *et al.* [2010] found that differences due to XBT bias corrections were the major factor in differences between ocean heat content estimates. A more detailed discussion of the differences between the two curves is beyond the scope of the present work. However, the differences between the time series are smaller than their respective uncertainty bounds and both time series show the same increase of about 0.2°C since the mid-1950s.

[24] As the global temperature time series available in the literature all start in the 1950's, no independent comparison could be made for this study before 1950. The comparison of the near-surface time series with the surface time series indicates that the GECCO-derived uncertainties might underestimate the sampling error on the global scale for years with extremely poor data coverage. However, our calculations do reveal a significant warming at least since the mid-1920s, when the German Atlantic Expedition (1925–27) [Wüst *et al.*, 1932] provided a good quality full-depth data set for the whole Atlantic Ocean between 20°N and 60°S . Consequently, the GECCO-derived uncertainties for the mean 0–400 m temperature suggest a much smaller sampling uncertainty during 1925–29.

5. Conclusions

[25] 1. The time series of the temperature anomalies within the upper 20-meter and 400-meter layers were extended to the beginning of the twentieth century, although there are gaps around the two world wars for the 0–400 m layer. Previous estimates started around 1950.

[26] 2. A good agreement is observed between the time series based on the sea surface and the near-surface data

respectively, but differences suggest either residual uncertainty of around 0.1C in the adjustments applied to minimize the effects of systematic errors, or actual differences between temperatures at the sea-surface and in the upper 20 meters.

[27] 3. The upper 400 meters of the ocean warmed by about $0.3\text{--}0.7^{\circ}\text{C}$ since 1910, with a central estimate around 0.5 to 0.6°C . The temperature change is characterized by two periods of stronger temperature increase between 1900 and 1940–45 and between 1970 and 2003, separated by a period of little change in the global average.

[28] 4. Decadal mean SST and 0–20 m layer anomalies calculated relative to the reference decade 2001–2010 give evidence of the general warming of the global ocean since 1900. However, large regions of the oceans have experienced cooling since the 1990s. Whereas cooling in the tropical Eastern Pacific ocean is associated with frequent La Nina events in the past decade, the cause of the cooling within the Southern Ocean remains unknown.

[29] **Acknowledgments.** V. Gouretski was supported through the Cluster of Excellence 'CLISAP' (EXC177), University of Hamburg, funded through the German Science Foundation. J. Kennedy was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). A. Köhl acknowledges support through the project Nordatlantik (03F0462E) funded by the German Science Foundation. We acknowledge Syd Levitus and John Antonov for their help in preparing the 0–400 m curves.

[30] The Editor thanks two anonymous reviewers for assisting in the evaluation of this paper.

References

- Boyer, T. P., *et al.* (2009), *World Ocean Database 2009*, vol. 1, *Introduction, NOAA Atlas NESDIS, 66*, edited by S. Levitus, 219 pp., NOAA, Silver Spring, Md.
- Carson, M., and D. E. Harrison (2008), Is the upper ocean warming? Comparison of 50-year trends from different analyses, *J. Clim.*, *21*, 2259–2268, doi:10.1175/2007JCLI2002.1.
- Conkright, M. E., *et al.* (2002), *World Ocean Database 2001*, vol. 1, *Introduction, NOAA Atlas NESDIS, vol. 42*, 159 pp., NOAA, Silver Spring, Md.
- de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *J. Geophys. Res.*, *109*, C12003, doi:10.1029/2004JC002378.
- 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, doi:10.1002/qj.49712152206.
- Gille, S. T. (2002), Warming of the Southern Ocean since the 1950s, *Science*, *295*, 1275–1277, doi:10.1126/science.1065863.
- Gouretski, V. V., and K. P. Koltermann (2007), How much is the Ocean really warming?, *Geophys. Res. Lett.*, *34*, L01610, doi:10.1029/2006GL027834.
- Gouretski, V., 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., Part I*, *57*, 812–833, doi:10.1016/j.dsr.2010.03.011.
- Grodsky, S. A., J. A. Carton, and H. Liu (2008), Comparison of bulk sea surface and mixed layer temperatures, *J. Geophys. Res.*, *113*, C10026, doi:10.1029/2008JC004871.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.
- Hazelworth, J. B. (1966), Quantitative analysis of some bathythermograph errors, *ASWEPS Rep. 11*, 21 pp., U.S. Nav. Oceanogr. Off., Washington, D. C.
- Ishii, M., and M. Kimoto (2009), Reevaluating 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.
- Jones, P. D., and T. M. L. Wigley (2010), Estimation of global temperature trends: What's important and what isn't, *Clim. Change*, *100*(1), 59–69, doi:10.1007/s10584-010-9836-3.
- Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011a), Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties, *J. Geophys. Res.*, *116*, D14103, doi:10.1029/2010JD015218.

- Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011b), Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, *J. Geophys. Res.*, *116*, D14104, doi:10.1029/2010JD015220.
- Knight, J., J. J. Kennedy, C. Folland, G. Harris, G. S. Jones, M. Palmer, D. Parker, A. Scaife, and P. Stott (2009), Do global temperature trends over the last decade falsify climate predictions?, *Bull. Am. Meteorol. Soc.*, *90*, S1–S196.
- Köhl, A., and D. Stammer (2008), Variability of the meridional overturning in the North Atlantic from the 50 years GECCO state estimation, *J. Phys. Oceanogr.*, *38*, 1913–1930, doi:10.1175/2008JPO3775.1.
- Levitus, S., J. Antonov, and T. Boyer (2005), Warming of the world ocean, 1955–2003, *Geophys. Res. Lett.*, *32*, L02604, doi:10.1029/2004GL021592.
- 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.
- Levitus, S., et al. (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.
- Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res.*, *117*, D08101, doi:10.1029/2011JD017187.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.
- Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett (2006), Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset, *J. Clim.*, *19*, 446–469, doi:10.1175/JCLI3637.1.
- Roemmich, D., W. J. Gould, and J. Gilson (2012), 135 years of global ocean warming between the *Challenger* expedition and the Argo Programme, *Nat. Clim. Change*, *2*, 425–428, doi:10.1038/nclimate1461.
- Smith, T. M., et al. (2008), Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880–2006), *J. Clim.*, *21*, 2283–2296, doi:10.1175/2007JCLI2100.1.
- Stammer, D., et al. (2010), Ocean variability evaluated from an ensemble of ocean syntheses, in *Proceedings of the OceanObs'09: Sustained Ocean Observations and Information for Society Conference (Vol. 1), Venice, Italy, 21–25 September 2009*, edited by J. Hall, D. E. Harrison, and D. Stammer, *ESA Publ.*, WPP-306, 11 pp.
- Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones (2008), A large discontinuity in the mid-twentieth century in observed global-mean surface temperature, *Nature*, *453*, 646–649, doi:10.1038/nature06982.
- 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 thermocline sea level rise, *J. Clim.*, *21*, 5657–5672, doi:10.1175/2008JCLI2290.1.
- Woodruff, S., et al. (2011), ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive, *Int. J. Climatol.*, *31*, 951–967, doi:10.1002/joc.2103.
- Wüst, G., G. Böhnecke, and H. Meyer (1932), *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition auf dem Forschungs- und Vermessungsschiff „Meteor“ 1925–1927*, vol. IV, 298 pp., Erste Teil. Ozeanogr. Methoden und Instrum., Berlin.