Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/3/e1601545/DC1

Supplementary Materials for

Improved estimates of ocean heat content from 1960 to 2015

Lijing Cheng, Kevin E. Trenberth, John Fasullo, Tim Boyer, John Abraham, Jiang Zhu

Published 10 March 2017, *Sci. Adv.* **3**, e1601545 (2017)
DOI: 10.1126/sciadv.1601545

This PDF file includes:

- fig. S15. Frequency distribution of temperature anomalies for the years 1986 and 2015.
- fig. S16. Sampling error as calculated by two subsample methods.
- fig. S17. S/N ratio analysis for two methods of subsample test.
- fig. S18. Comparison between OHC and sea level change since 1993.
- fig. S19. Distribution of the ensemble anomalies.
- table S1. OHC trends obtained in this study for the 1960–1991 and 1992–2015 periods.
- table S2. Net OHC and EEI changes obtained in the current study compared with some independent estimates.

Supplementary Materials and Methods

1 Uncertainties in OHC estimation

Ocean heat content is calculated by integrating the three dimensional product of temperature, density and heat capacity. An illustration on how OHC is calculated from *in situ* temperature profiles can be found in *Cheng et al.* (*7*). Here the uncertainties in OHC calculations are discussed.

(i). Quality-Control (QC) of the data. As *in-situ* ocean temperature observations were collected by various instruments in different ocean conditions and on different platforms, it is unavoidable that some obtained measurements are erroneous and that the quality of measurements differs across the platforms. QC processes are designed to remove the spurious measurements. It is widely suspected that different QC process adapted in different groups impact the OHC estimates (*21*). There is no generally accepted best practice for QC though investigation is underway in the IQuOD project (http://www.iquod.org). Here, we apply QC flags in WOD13 dataset, which is the most widely used ocean subsurface dataset.

(ii). Instrumental bias. Some ocean instruments have systematic errors in their measurements, such as Mechanical Bathythermographs – MBTs and eXpendable Bathythermographs - XBTs. These biases can significantly impact the OHC estimates (*8, 22, 23, 31*), especially for decadal and long-term changes. Hence correction of these biases is one of the key tasks in OHC estimation. Corrections have been comprehensively discussed in recent studies such as (*21, 25*). In the present study, we correct the MBT bias using the *Ishii et al.* (*31*) method, and the XBT bias using the *Cheng et al.* (*26*) method. The latter is the recommended method by the XBT community (*25*), which takes account of all known factors affecting XBT biases. After correction, the XBT bias on decadal scale can be substantially reduced, but there may be some uncertainties on inter-annual scales due to the lack of data.

(iii). Choice of baseline - climatology. Ideally the choice of a baseline does not change the representation of long-term OHC variation. However, the observations are unevenly distributed in the ocean in the past, particularly before the Argo period (*27, 30*). The impact of the climatology originates from the irregular sampling and it is then dependent on the mapping method (*21*). A climatology constructed by using data with consistent geographical coverage is preferred, i.e. the Argo period climatology and WOA13 decadal climatologies. Alternatively, there is another option of removing the long-term trend in each grid cell before construction of the climatology. In the present study, we construct the climatology by using the data for the 1997–2005 period for both observations and models. However, use of such a short time period for the climatology could result in larger bias in the OHC estimate compared with a long-period climatology (*28, 30*). This bias, which arises from infilling data gaps, should be ascribed to the mapping method rather than the climatology

(iv). Mapping method. The mapping method defines how the global OHC is calculated from observations with incomplete global coverage. It relates to how data gaps are filled and the reconstructed field is smoothed. The mapping method is the biggest error source in the OHC estimate (*21*). Bias due to mapping in the OHC estimate has been recognized in recent studies (*28–30*), suggesting that a careful

evaluation of the available mapping methods is required. Here we introduce the general principles of a mapping method to give an impression of the error due to mapping. Most of the current available mapping methods follow a similar framework

$$\mathbf{X^a}(I) = \mathbf{X^b}(i) + \mathbf{W}\,(\mathbf{O}(j))$$

where the final analysis value ($\mathbf{X^a}$) in a grid i is a combination of a prior guess ($\mathbf{X^b}$) and a correction term according to the nearby observations ($\mathbf{O}$) in grid j. $\mathbf{W}$ is a weight function that defines how the nearby observations impact the analysis grid and how the analysis field is smoothed. Current available mapping methods use different prior guesses and methods to represent the covariance characteristics. These factors are responsible for their different performance.

In addition to mapping, a localization strategy is always applied which assumes that only data within a spatial area can be used during the analysis of a grid cell. The size of the area is defined by the influencing radius. Different mapping methods adopt different influencing radii, leading to their differences in their ability to filter noise and smooth data. This is also responsible for the differences in resultant OHC.

In summary, it has been well established that there are various sources of uncertainty in OHC estimates, which limits its accurate determination. Major progress has been made to correct the instrumental bias and the choice of methodologies in the past years. Before the community fully addresses the impact of QC on OHC calculation, it is reasonable to assume that the error of the spurious measurements is white noise and thus plays a minor role in the OHC estimation. The major remaining issue is whether a reliable mapping method is possible with negligible sampling bias, as discussed in this study.

2 Mapping method: iterative EnOI-DE/CMIP5 Method

Cheng and Zhu, (CZ16) provided an objective method for reconstructing the ocean temperature field for the upper 700m by utilizing CMIP5 simulations, which was named the Ensemble Optimum Interpolation method with Dynamic Ensemble of CMIP5 simulations (EnOI-DE/CMIP5). This method mainly uses the spatial covariance of model outputs, and the error in temporal information from models does not significantly impact the reconstruction. A large influencing radius of 20 degrees is applied in CZ16 to reconstruct the temperature fields within 0–700m layers, and here 25 degrees is used in the reconstructions for 700–2000m layers. Three iterative scans were performed by successively setting three different influencing radii (fig. S1):

(i). Scan-1 uses the traditional EnOI-DE/CMIP5 mapping by setting a large influence radius (25 degrees for 700–2000m layers and 20 degrees for 0–700m layers).

(ii). Scan-2 uses the reconstructed fields obtained by scan-1 as the prior guess. In this case, the results of the first scan inform the second scan, and provide the improved prior information. In scan-2, the influence radius is set to 8 degrees, consistent with the NCEI method and some other traditional methods. Note that NCEI method is also an iterative method with influencing radius of 8, 6 and 3 degrees.

(iii). Scan-3 uses the reconstructed fields of the scan-2 as the prior guess, but the influence radius is set to 4 degrees.

A comparison of the reconstructed fields after scan-1, scan-2 and scan-3 is shown in fig. S2. Smaller scale signals are included after scan-2 and scan-3 compared with scan-1.

Figure S3 presents the reconstruction of the subsampled fields according to the observation locations in February 1962, 1972, 1982, 1992, 2002, 2012, and the corresponding reconstructed fields, where the gridded temperature anomalies in August of 2011 are used as truth. The temperature anomaly pattern from 30°S to 65°N can be well reconstructed in various sampling patterns since 1960s. There are some uncertainties regarding the reconstructed field from 60°S to 30°S especially in the South Pacific, where there was extremely poor sampling. For example, in February 1972 (1962), less than 50 (20) grids contained observations in this region. Moreover, there are fluid eddies in that region which further affects accuracy. The zonal mean temperature anomalies for the truth and reconstructed fields are also presented in fig. S3. In August 2011, it shows warming signals from 50°S to 25°N, and from 40°N to 70°N, and cooling signals around 30°N and within 75°S–55°S. These large-scale structures can all be well reconstructed, especially north of 30°S. Larger uncertainty is found in the reconstructions around 30°N and 40°S, due to the presence of boundary currents and eddies.

In the summer season of the Southern Hemisphere, as there is less sea ice coverage than in winter, there are more observations available. Therefore, it is also valuable to check the reconstruction in the winter season of Southern Hemisphere. Figure S4 presents the reconstruction of the subsampled fields according to the sampling of September 1959, 1969, 1979, 1989, 1999, 2009, and the corresponding reconstructed fields, where the temperature anomalies in August of 2011 are used as truth. Because of the lack of data in ice-covered regions in the Southern Ocean (September), the reconstruction within 70°S–40°S contains larger errors than fig. S3.

3 Evaluating the mapping method

3.1 Choice of influencing radius

CZ16 used 20 degrees as the influencing radius for the upper 700m. For deeper layers a different approach was selected. A subsample test was used first to determine the influencing radius for 700–2000m layers. A depth of 1200m was selected here as an example, however the results using other depths were similar. The subsample test was run multiple times with different influencing radii from 4 to 36 degree with 4-degree increments together with a separate run without using a localization strategy.

The global mean temperature errors between the mapped and truth fields are shown by dots in fig. S5 as a function of the influencing radii. It appears that the error is reduced to near zero when the influencing radius is set to both ~15 and ~25 degrees. Consequently, 25 degrees was selected as the influencing radius to allow a broader propagation of information from the sampled to un-sampled regions, and also to ensure a global reconstruction.

This test indicates that the model ensemble in our method is able to provide a proper covariance even though there are only weak zonal-mean correlations (~0.2) and meridional-mean correlations (~0.0) up to 25 degrees away (Fig. 2 in the main text).

More results based on the subsample test in the following discussions confirm that there is insignificant error by using the large value of the influencing radius.

3.2. Global/Basin mean sampling error at different layers

The global/0-2000m averaged sampling error is shown in Fig. 3 in the main text; however, it is worthwhile investigating the sampling error for particular ocean layers, because there are different spatial and temporal variabilities. For instance, the inter-annual variation in the upper ocean (i.e. 0–300m) is dominated by ENSO variations, but below 300m, the inter-annual variation becomes much weaker. Also it is important to examine the error for six major ocean basins (fig. S6).

Figures S7–S10 shows the global/basin mean sampling error at several depths: 20m, 300m, 800m, and 1200m respectively.

- At 20m (fig. S7), the global averaged temperature change reveals significant long-term warming and apparent ENSO variations on inter-annual scales. The sampling errors are around zero. Both the inter-annual and decadal variability are larger than the sampling error since 1960 (S/N ratio larger than 10 for decadal variability, larger than 8 for the inter-annual variability, fig. S7H, I). This indicates that both inter-annual and decadal variability can be accurately reconstructed near the sea surface for the global ocean. Similar conclusions can be drawn for the six major basins. In particular, the southern oceans show the smallest S/N ratio. The southern oceans near the surface experienced weaker long-term warming rate than most regions in the ocean. The lack of warming is a hot topic in climate community.
- At 300m (fig. S8), the temperature signals are also detectable from the sampling errors on both global and regional scales on decadal scale (with S/N ratio from 2:1 to 30:1, fig. S8H). However, the S/N ratio for the inter-annual variability (larger than 2 after 1970) is smaller than that in the upper 20m. This indicates that the inter-annual signals become more uncertain at 300m depth than near the sea surface.
- At 800m (fig. S9), the decadal variation can be more robustly detected since the 1990s; there is a dramatic increase of S/N ratio (larger than 2 after 1990s for global/regional signals). This change is mainly due to the WOCE (1990s) and Argo project (2000s). The S/N ratio for the inter-annual variations is larger than 2 after 2010. This again suggests the inter-annual signals in the reconstructed field are uncertain in the deep ocean (below 700m), which explains the elevated noise in the time series in the North Pacific and Indian Ocean (fig. S9E and F) compared to other basins.
- At 1200m (fig. S10), similar conclusions can be drawn compared with those at 800m. The southern oceans and tropical/ subtropical Atlantic Ocean show the largest S/N ratio (fig. S10H, I), because these two ocean basins experienced much quicker long-term warming after the 1960s than the other basins in the deep ocean. In addition, the North Atlantic also shows a clear decadal variation.

3.3. Local sampling error by subsample test

The geographical distributions of the mean sampling error and two standard deviations in each 1^o by 1^o grid are presented in fig. S11 for 0–2000m averages, and fig. S12 for 20m and 1600m depths. It is seen that

- As the mean sampling errors are around zero over the global ocean for 20m,

1600m depths and for 0–2000m averages, there is no significant regional bias.

- At 20m, there are larger 2σ sampling errors in the boundary currents regions, ACC regions, and the Eastern Pacific Ocean.
- At 1600m, larger 2σ sampling errors occur in the North Atlantic, Western Indian Ocean and ACC regions (south of 30ºS). There are much smaller 2σ sampling errors in the North Pacific than the other regions in the deep ocean, in contrast to the upper ocean (i.e. 20m).

3.4. Importance of the iterative strategy

The iterative strategy is expected to significantly improve the reconstruction by adding the ocean variability on different spatial scales. However, using an increasingly smaller influencing radius also affects the noise, including both unphysical noise due to instrumental errors, and small-scale ocean variability such as meso/sub-meso scale eddies. As we are interested in the large-scale (>1 degree in spatial distance) temperature changes (because we divide the ocean into a 1^o by 1^o grid), it is questionable whether this strategy will reduce the analysis error and positively affect the reconstruction.

- Figure S13 shows the 2σ sampling error for the subsample tests at 20m, 300m, 800m, 1200m and 1600m from scan-1 to scan-3. Significant reductions of 2σ sampling errors in scan-3 are found for each depth compared with the previous scan, suggesting a better reconstruction after 2000. It is interesting to note that in the deep ocean (i.e. 1200m and 1600m), the 2σ error increases in scan-2 and scan-3 compared with scan-1 before 2000. This is likely because the grid-by-grid variations in the deep ocean are mostly dominated by the instrumental error rather than the meso-scale variability. Hence the smoothing by the current method using a larger influencing radius could better reconstruct the large-scale ocean changes.
- Figures S7–10 also attach the global and basin mean/2σ sampling error for all scans. Reduction of mean and 2σ sampling error for scan-3 (green error bars) on both global/basin scales is compared with the results of scan-1 and scan-2 (blue and orange lines and error bars).

It is further questioned whether different scans will significantly modify the historical global OHC estimation. Figure S14 presents the OHC 0–700m, OHC 700–2000m based the reconstruction of scan1, scan2 and scan3 respectively. The OHC of scan-1 provides a slightly smoother OHC time series than scan-2 and scan-3, but the decadal and long-term changes are nearly identical.

4 Examining the underestimation of the current subsample strategy

In the subsample test, we used 1-degree gridded averaged observations as truth (named 1degree grid method). By averaging the available observations in each 1^o by 1^o grid, there should be a reduction in noise due to meso-scale ocean variance and from spurious errors. If there were more data in a grid box during the Argo period than the pre-Argo period, then the historical sampling error would be under-estimated, and vice versa. Therefore, this can have an impact because the observation system has changed. It is questioned whether this effect could significantly bias the estimate of sampling error.

As a first-order global average, there are more data in the Argo period than the pre-Argo period. But there will be instances in each and every month where there are

no data (~53% in fig. S15, at 20m) or less than 3 individual measurements (~72% in fig. S15, at 20m) in a 1degree grid in the Argo period. By comparison, in Jan1986, there are 71% grids without data and 84% grids with less than 3 data. Hence in most of the cases, it is difficult to average out the meso-scale signals in each grid box even in the Argo period. Furthermore, a good portion of pre-Argo year/month data are on or near continental shelves, in boundary currents, etc. which are areas where the Argo fleet does not have a significant presence.

To make sure that this effect does not impact the sampling error estimate (by our 1degree grid method). We applied a different subsample strategy (named raw-profile method). With this method, the truth values are the raw temperatures observed for each selected month during the Argo period (January and August from 2007 to 2014, similar to the 1degree grid method). Each truth field is subsampled by finding the nearest profiles related to the historical observations, and then these subsampled profiles are gridded and mapped by the current method. This test is applied for 20m and 1200m depths for the whole global domain, and it shows results consistent with the 1degree grid method (fig. S16). The 2σ sampling error, when using the raw-profile method, is 15% larger than the 1degree grid method at 20m. Below the sea surface, since there are more grids with less than 3 observations during the Argo period, the difference of the two methods should be less than 15%. 15% is the mean since 1950s, but there are some fluctuations for different sampling years (fig. S16). At 1200m, the difference between the two methods is negligible, ~3% (in fact the 1degree grid method results in a 3% larger error than the raw-profile method) (fig. S16).

Even though there is about 15% underestimation of sampling error within the 0–2000m (an upper bound), it does not impact the key conclusion. If the sampling error over 0–2000m were increased by 15%, we could still make the conclusion that the ocean temporal variation on decadal/multidecadal scale can be robustly reconstructed (S/N ratio ranges from 2:1 to 20:1), but the ocean inter-annual variability is comparable with sampling error (S/N ratio ranges from 0.2:1 to 5:1) (fig. S17).

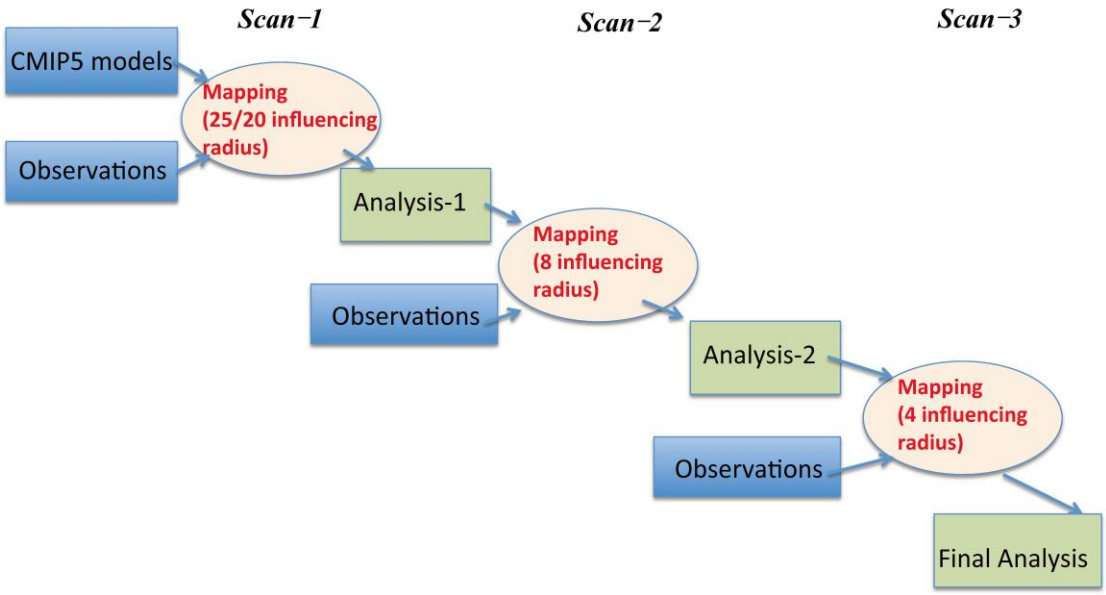

fig. S1. Illustration of the iterative EnOI-DE/CMIP5 method used in the current study.

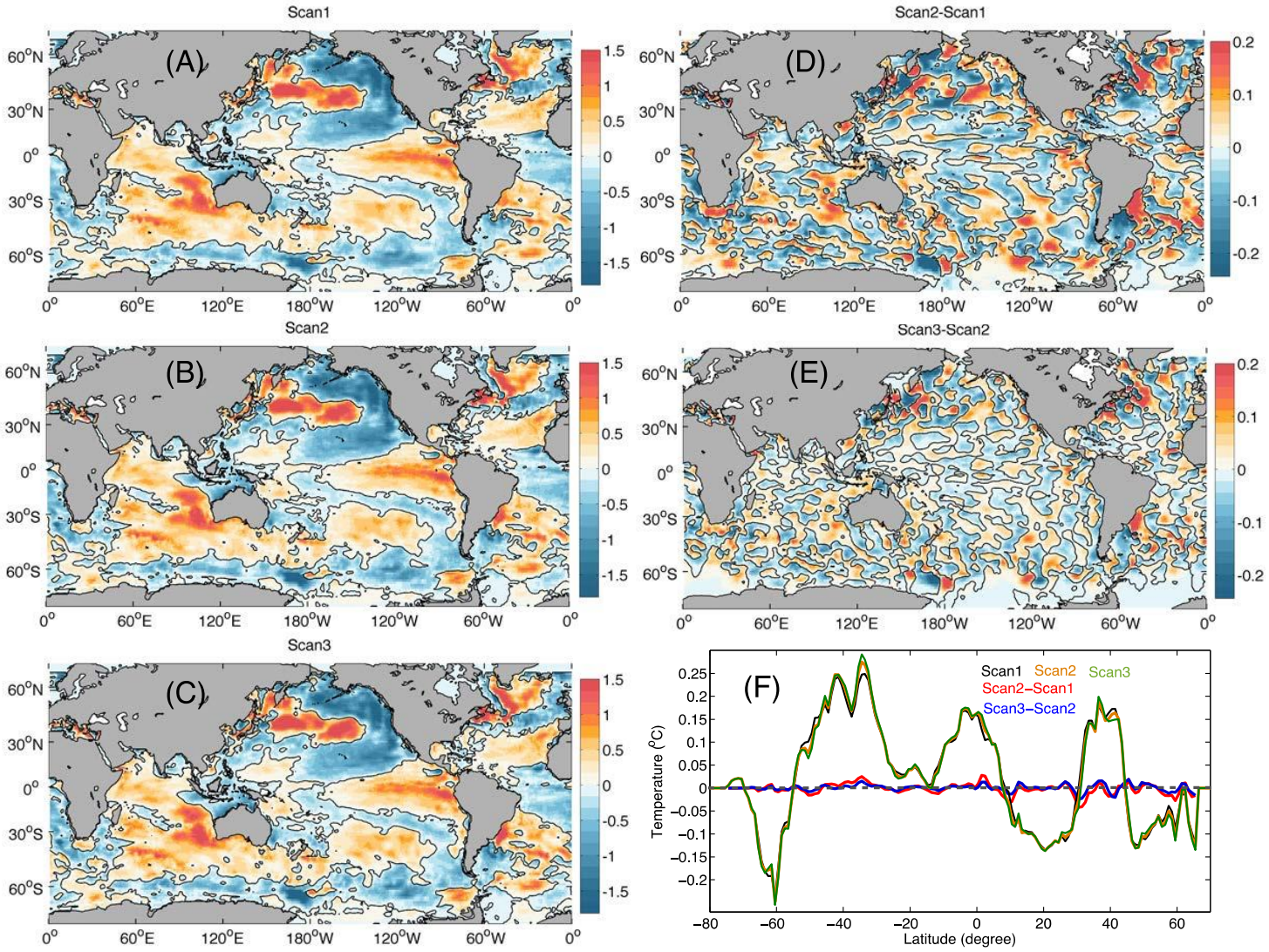

fig. S2. An example of the reconstructed fields after three iterative scans. The reconstructed field after (**A**) scan-1; (**B**) scan-2; and (**C**) scan-3 for 20m temperature anomaly in August 2012. (**D**) Temperature difference between scan-2 and scan-1; (**E**) temperature difference between scan-3 and scan-2; (**F**) zonal mean temperature anomalies for the three scans, and the differences between scan-2 and scan-1 (red) and scan-3 and scan-2 (blue).

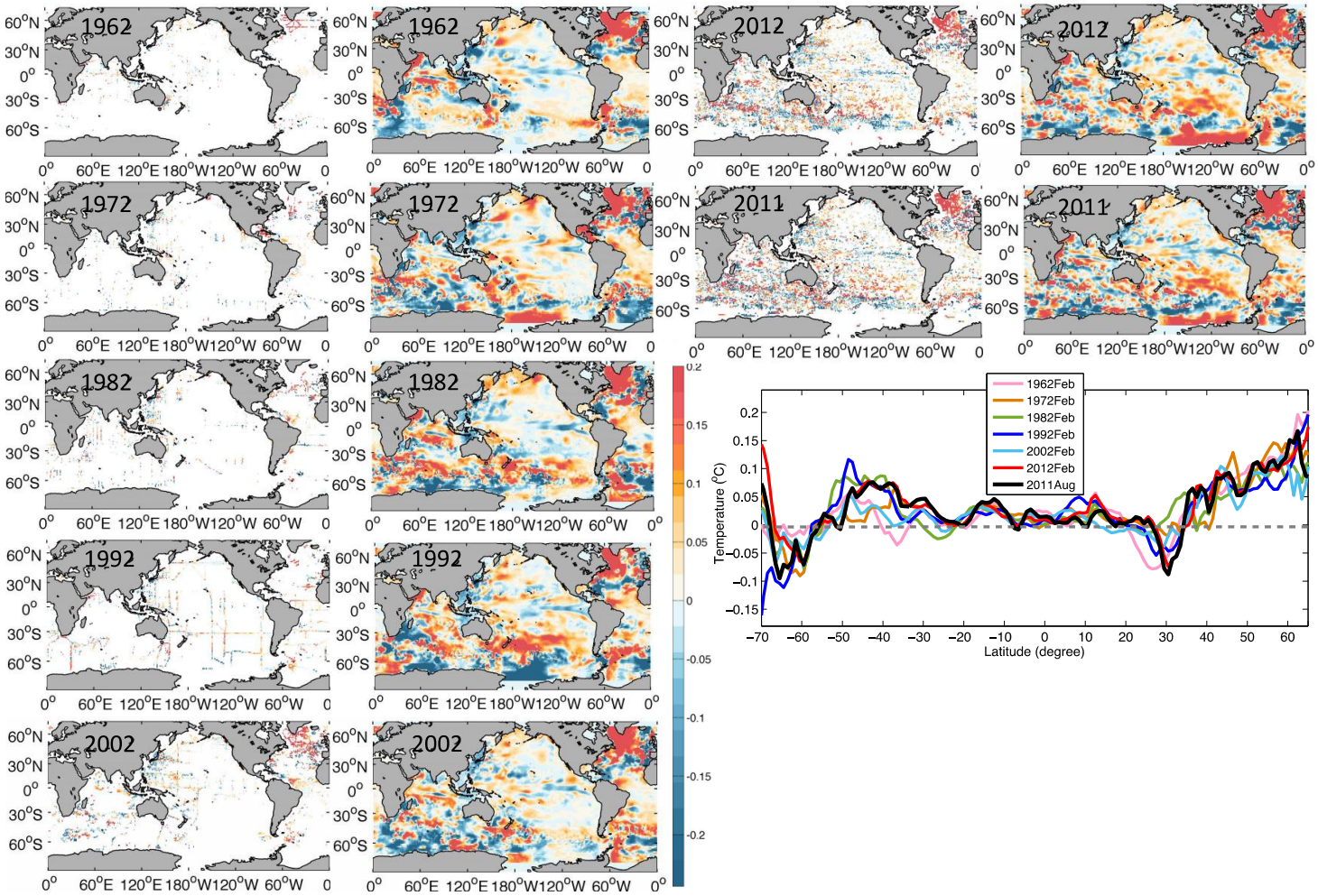

fig. S3. Reconstruction of temperature field at 1200 m in August 2011 for the historical sampling (in February). Subsampled temperature anomalies are shown on the left. The temperature anomaly field at 1200m in August 2011 was subsampled according to the location of observations in February of 1962, 1972, 1982, 1992, 2002, and 2012. The color shows the average temperature anomaly in each 1° by 1° grid. On the right, the fields mapped by using the proposed method are presented. The last panel shows the zonal mean temperature anomalies of the reconstructed fields compared with the truth (temperature anomaly field at 1200m in August 2011). The spatial correlations between the six reconstructions fields and the truth are 0.26 (1962), 0.47 (1972), 0.45 (1982), 0.48 (1992), 0.49 (2002), and 0.75 (2012). The color shows the average temperature anomaly in each 1^o by 1^o grid box, where there are data present, in the left hand plots.

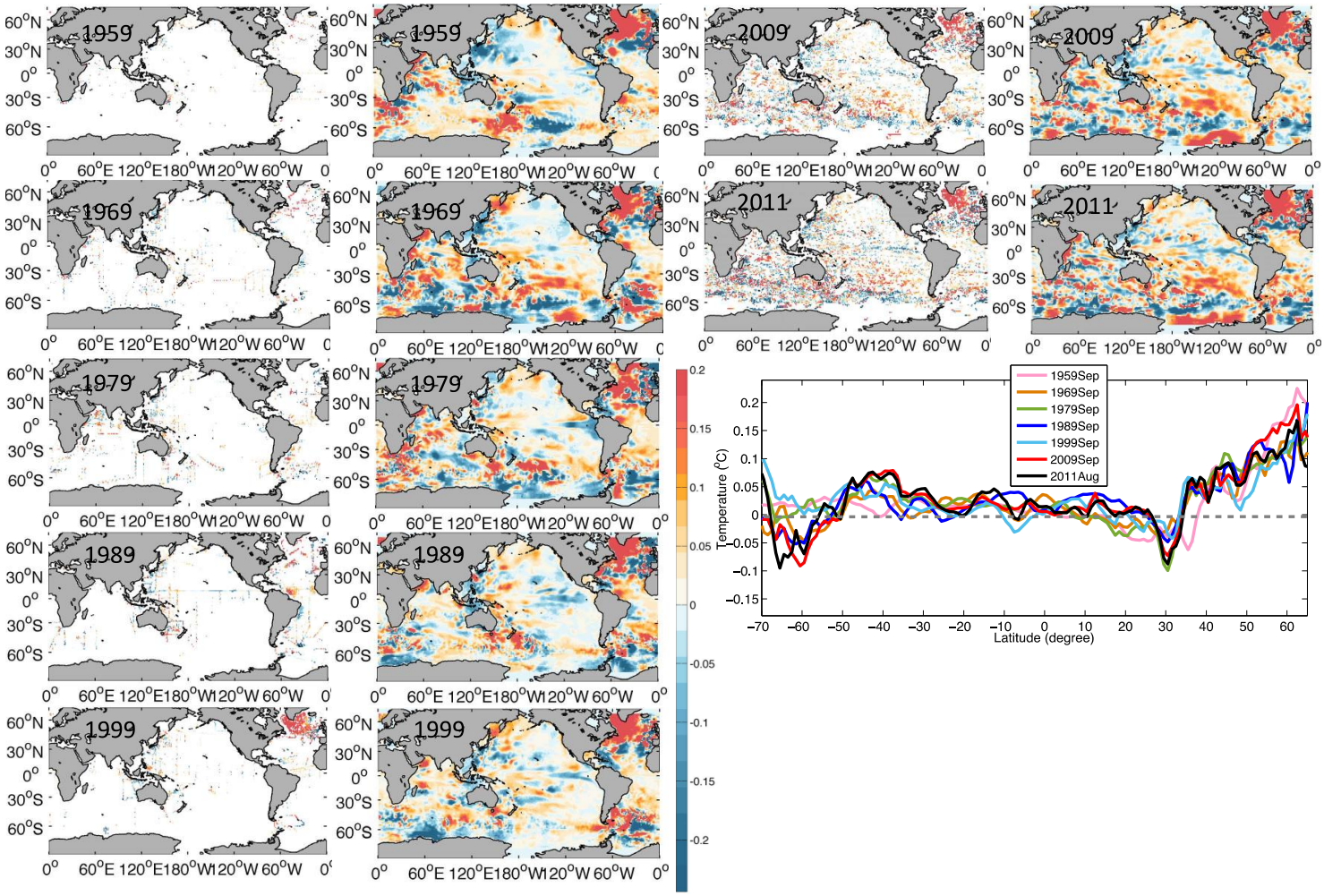

fig. S4. Reconstruction of temperature field at 1200 m in August 2011 for the historical sampling (in September). Subsampled temperature anomalies are shown on the left. The temperature anomalies at 1200m in August 2011 were subsampled according to the location of observations in September of 1959, 1969, 1979, 1989, 1999, and 2009. The color shows the average temperature anomaly in each 1° by 1° grid. On the right, the fields mapped by using the proposed method are presented. The final panel shows the zonal mean temperatures of the reconstructed fields associated with the truth (temperature anomaly field at 1200m in August 2011). The spatial correlations between the six reconstructions fields and the truth are 0.23 (1959), 0.42 (1969), 0.43 (1979), 0.44 (1989), 0.47 (1999), and 0.68 (2009). The color shows the average temperature anomaly in each 1^o by 1^o grid box, where there are data present, in the left hand plots.

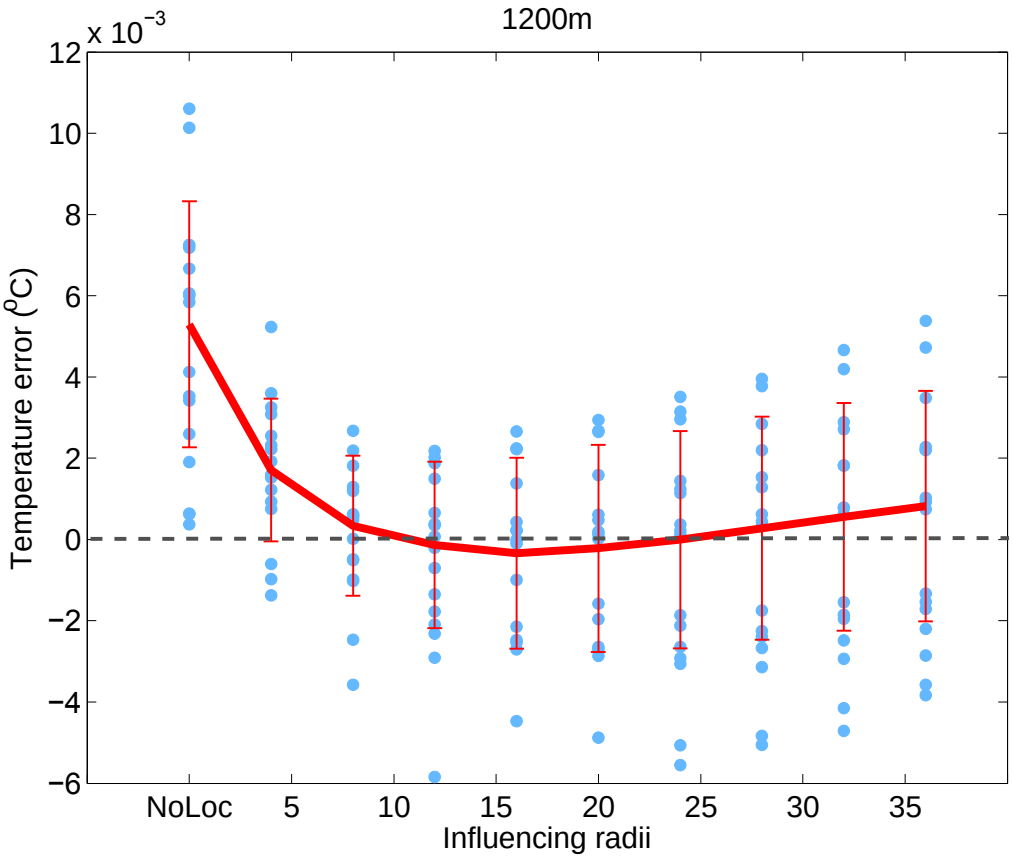

fig. S5. Mean temperature error as a function of different choices of the influencing radii between the reconstructed and truth fields. Each dot represents the averaged temperature error for each truth field and each influencing radius, where the errors at sampling years are averaged together. The red line and the error bar show the mean and standard deviation of the sampling error as a function of influencing radii, respectively.

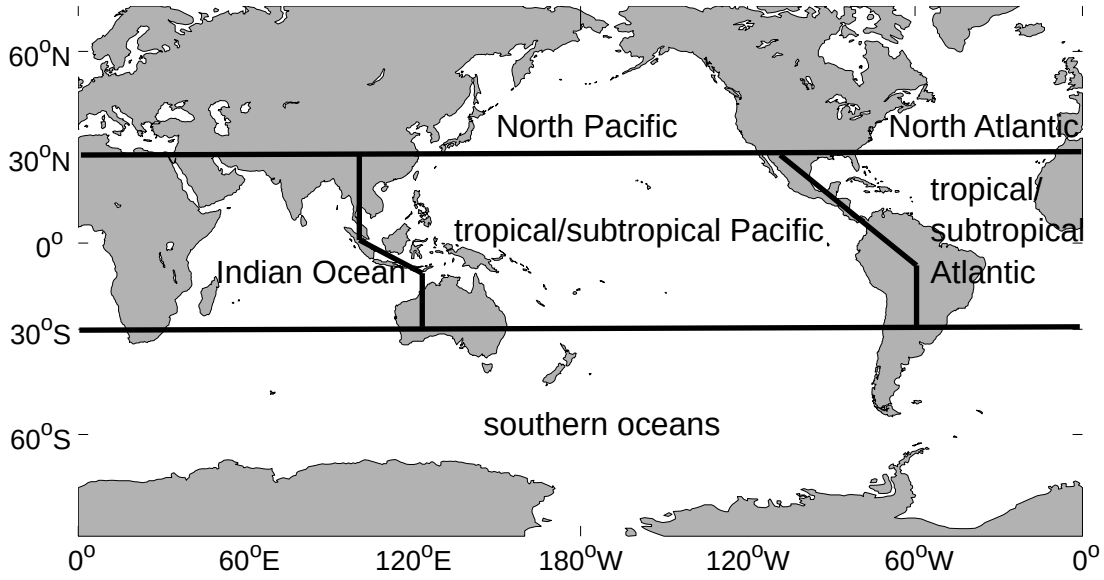

fig. S6. Six major ocean basins defined in this study.

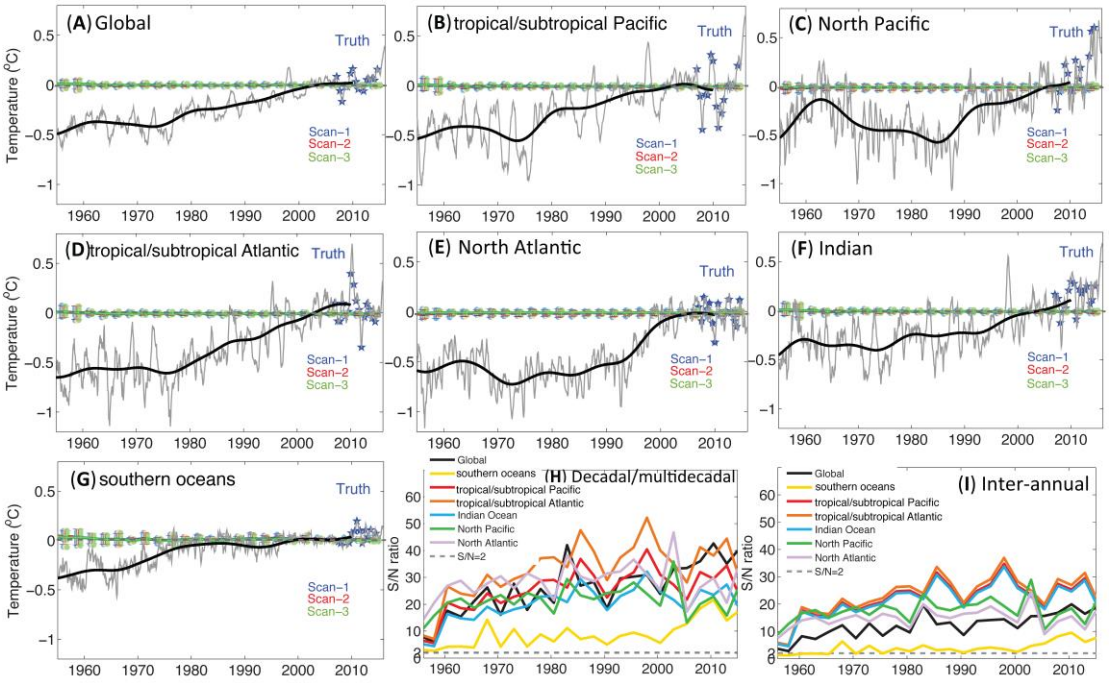

fig. S7. Global and basin-averaged sampling error compared with reconstructed temperature change at 20 m. Dots represent the sampling errors corresponding to 16 different truth fields (blue: scan-1; red: scan-2; green: scan-3), accompanied with the mean in solid lines and 2 standard deviation error bars (blue: scan-1; red: scan-2; green: scan-3). The grey line is the monthly temperature anomaly time series from 1955 to 2015 based on the current analysis, and the blue stars denote the 16 different truth fields. The dark black line is the time series after a 7-year low-pass filter. (**A**) Global; (**B**) tropical/subtropical Pacific; (**C**) North Pacific; (**D**) Indian Ocean; (**E**) tropical/subtropical Atlantic; (**F**) North Atlantic; (**G**) southern oceans. (**H** and **I**) The S/N ratio of the temporal variability of our reconstruction on (**H**) decadal scales and (**I**) inter-annual scales related to the sampling error.

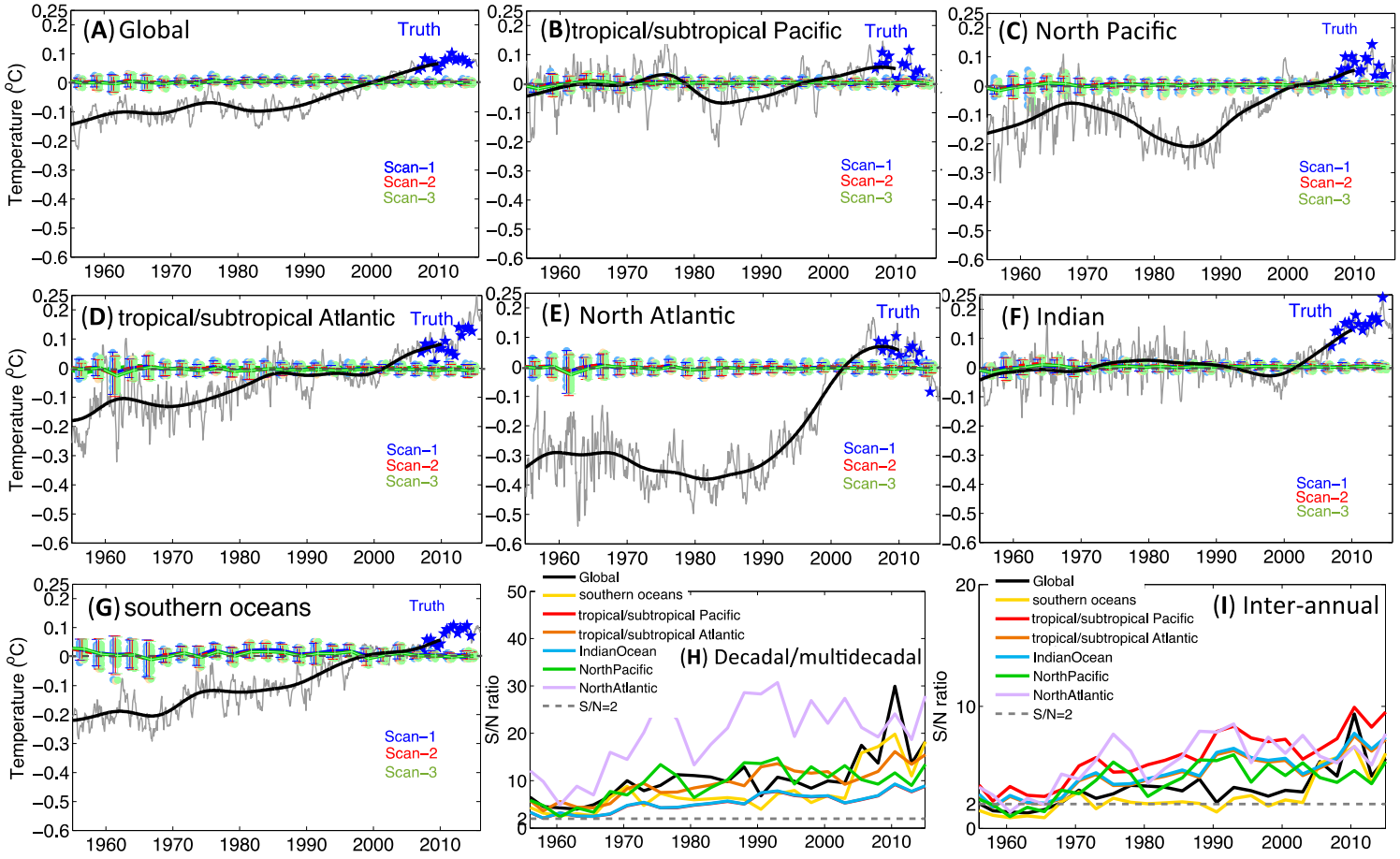

fig. S8. Global and basin-averaged sampling error compared with reconstructed temperature change at 300 m.

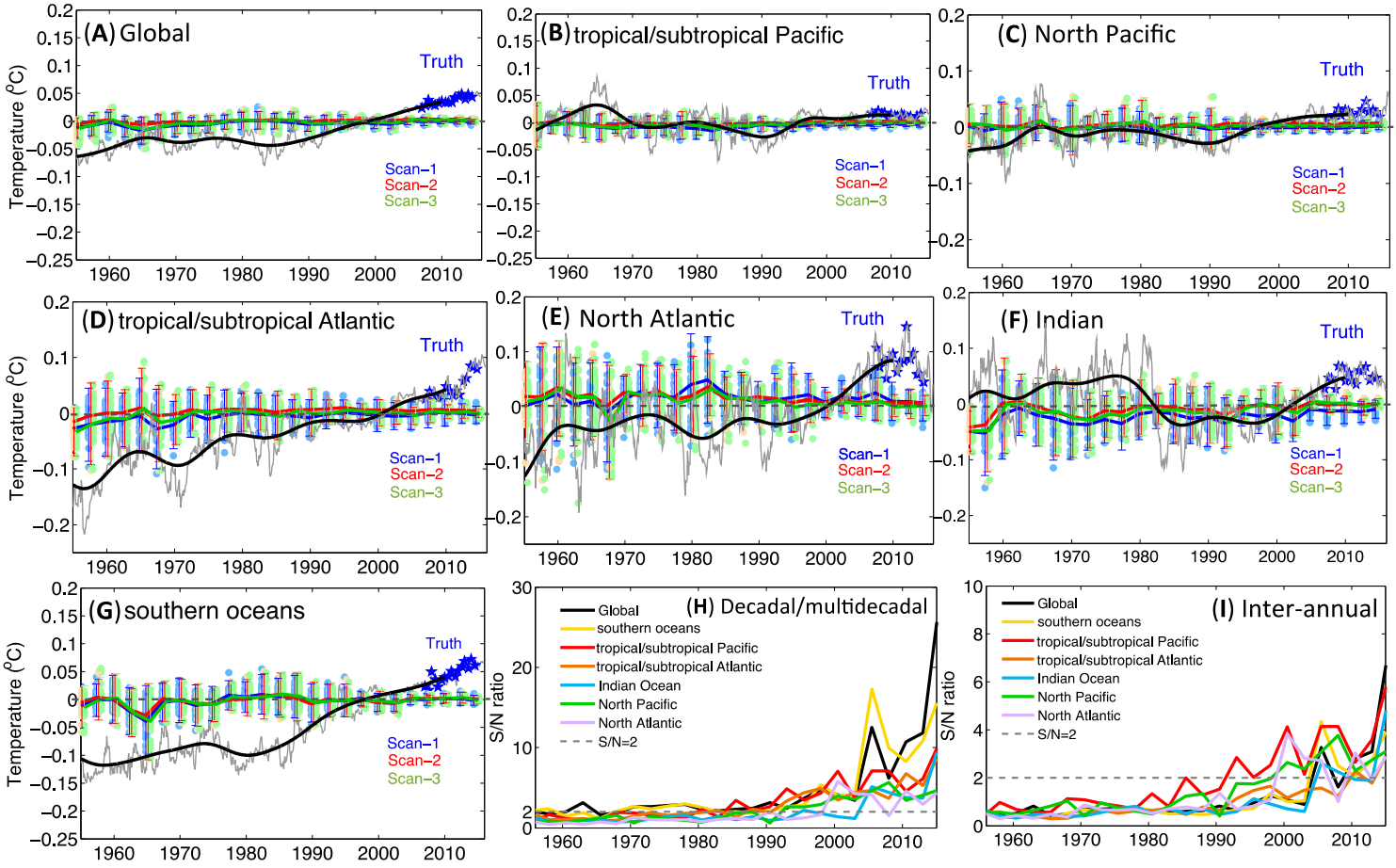

fig. S9. Global and basin-averaged sampling error compared with reconstructed temperature change at 800 m.

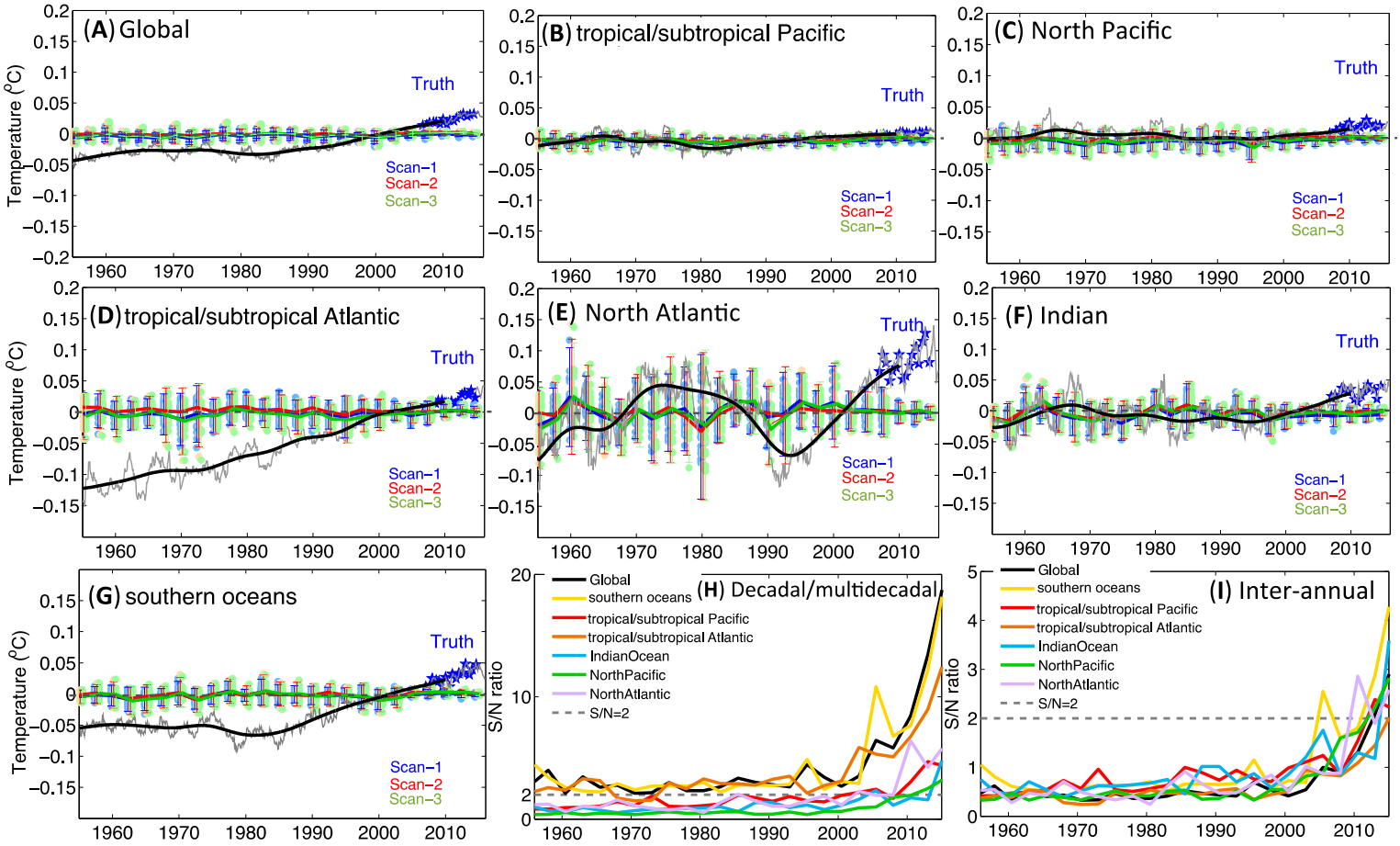

fig. S10. Global and basin-averaged sampling error compared with reconstructed temperature change at 1200 m.

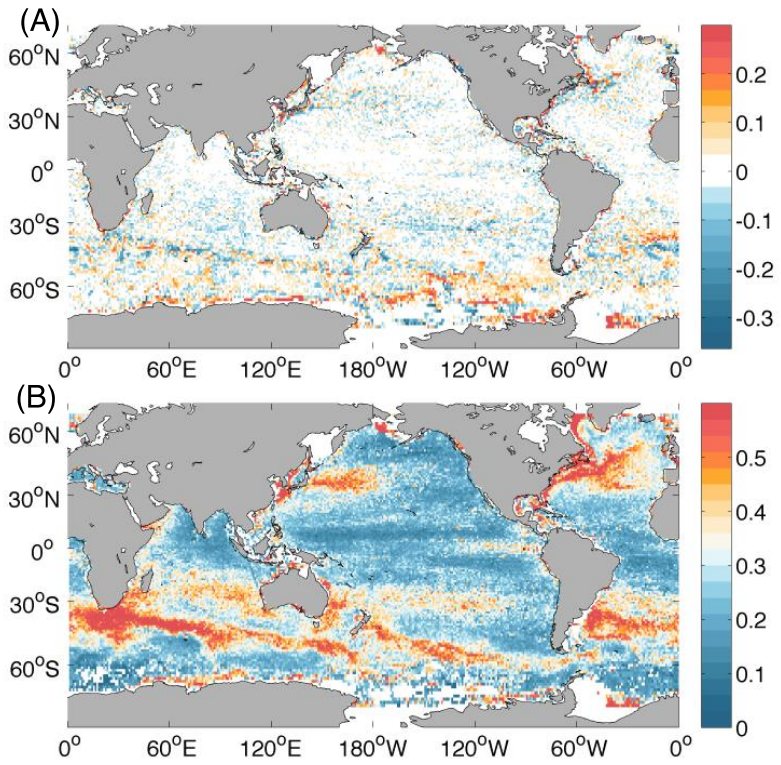

fig. S11. Geographical distribution of mean and 2σ sampling error for 0- to 2000-m average. (**A**) 1º by 1º gridded mean sampling error since 1960. (**B**) 2 standard deviations of the sampling error since 1960.

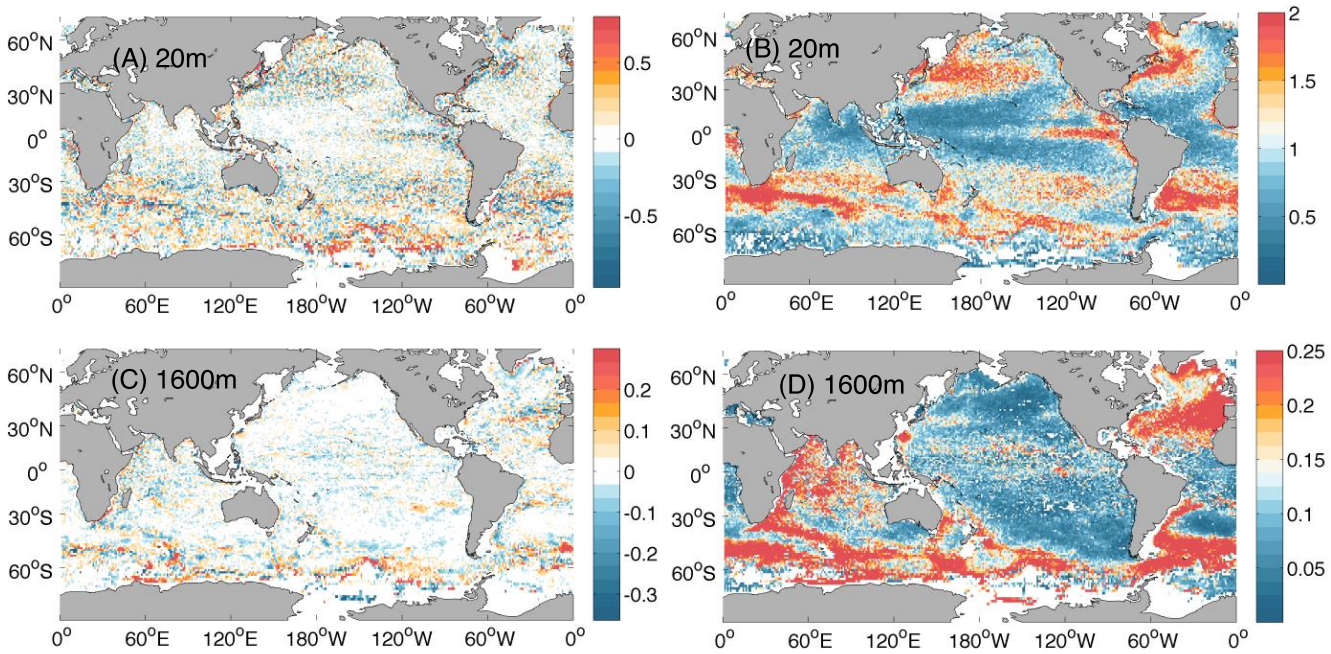

fig. S12. Geographical distribution of mean and 2σ sampling error in 1°-by-1° grid at 20 and 1600 m. (**A**) 1º by 1º gridded mean sampling error at 20m since 1960. (**B**) 2σ sampling error at 20m. (**C**) 1º by 1º gridded mean sampling error at 1600m since 1960. (**D**) 2σ sampling error at 1600m.

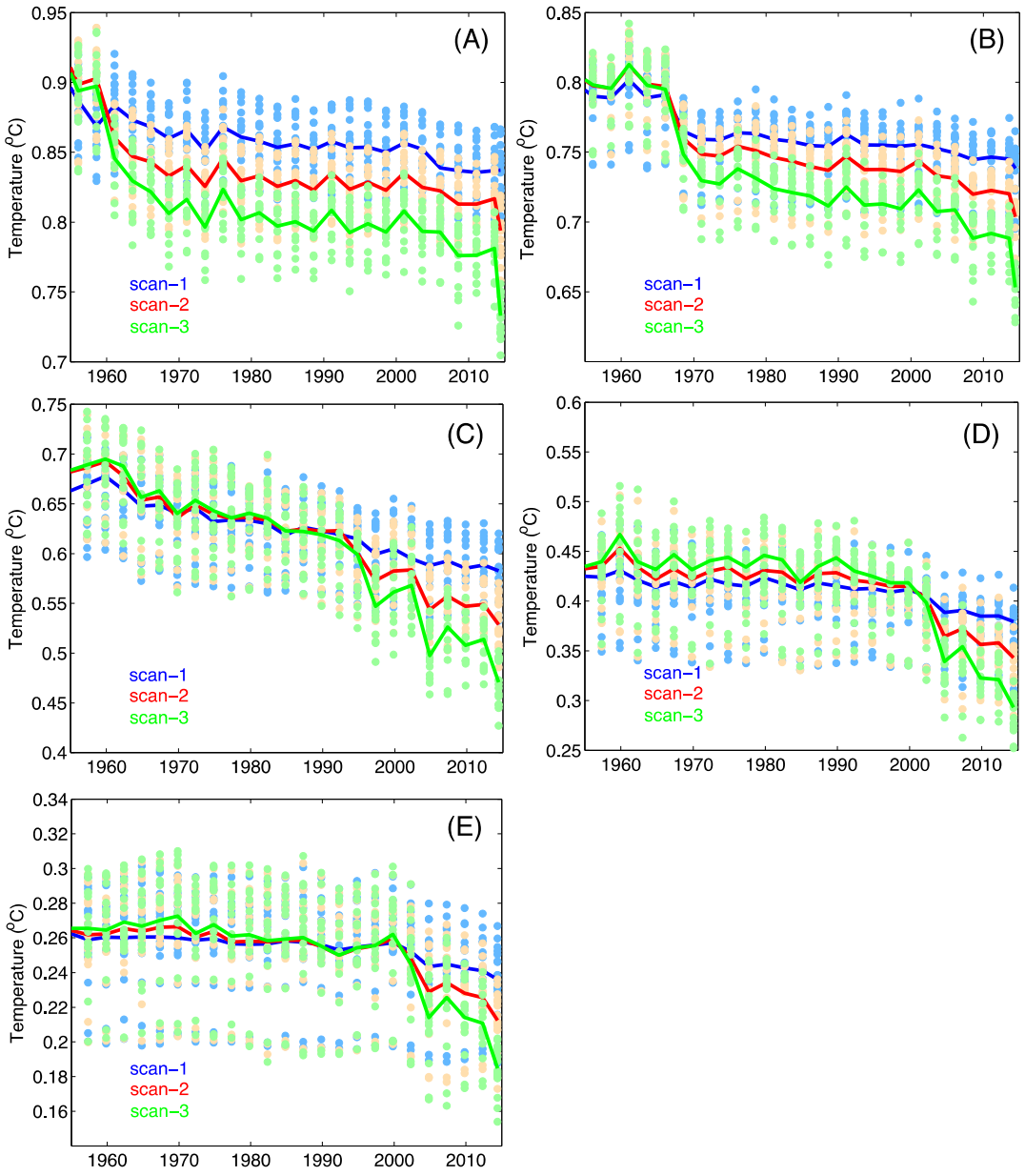

fig. S13. 2σ sampling error for different scans at different depths. (**A**) 20m, (**B**) 300m, (**C**) 800m, (**D**) 1200m and (**E**) 1600m. Each dot represents a 2 standard deviation of the sampling errors in 1° by 1° grid boxes for each truth field at each sampling year. Results for the 16 different truth fields are shown in different dots, with the mean shown in lines. Results of the scan-1 are shown in blue, scan-2 in red and scan-3 in green.

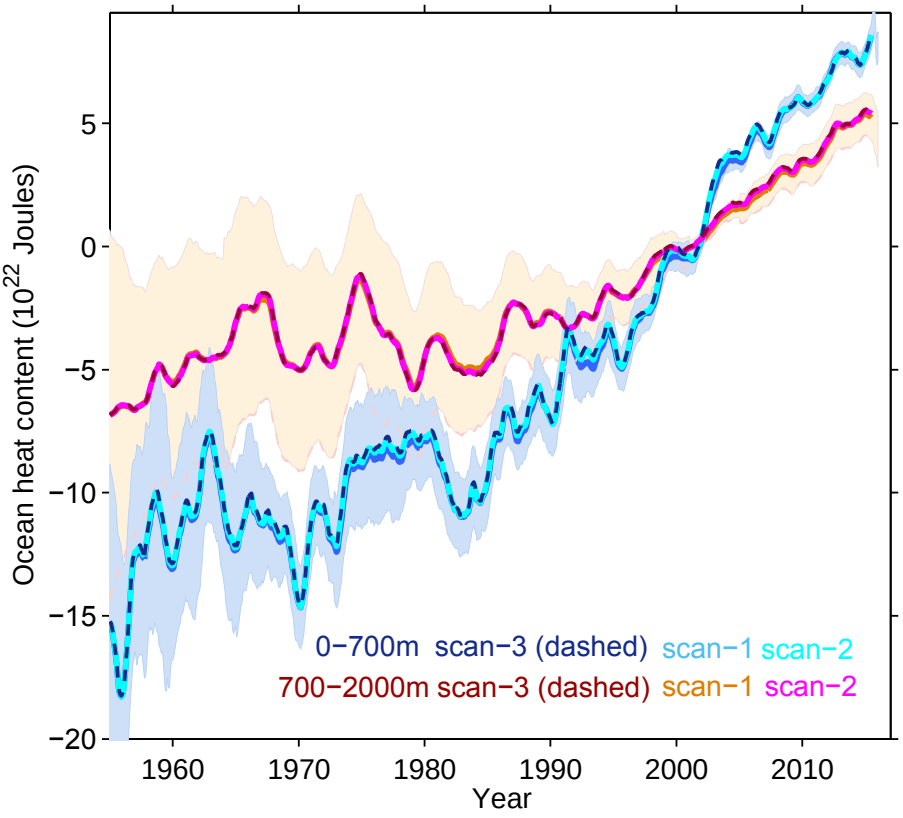

fig. S14. Global OHC time series for the reconstructions after scan 1, scan 2, and scan 3. Both OHC0-700m and OHC700-2000m are shown. The shading shows the 2σ error bar of the time series after scan-3.

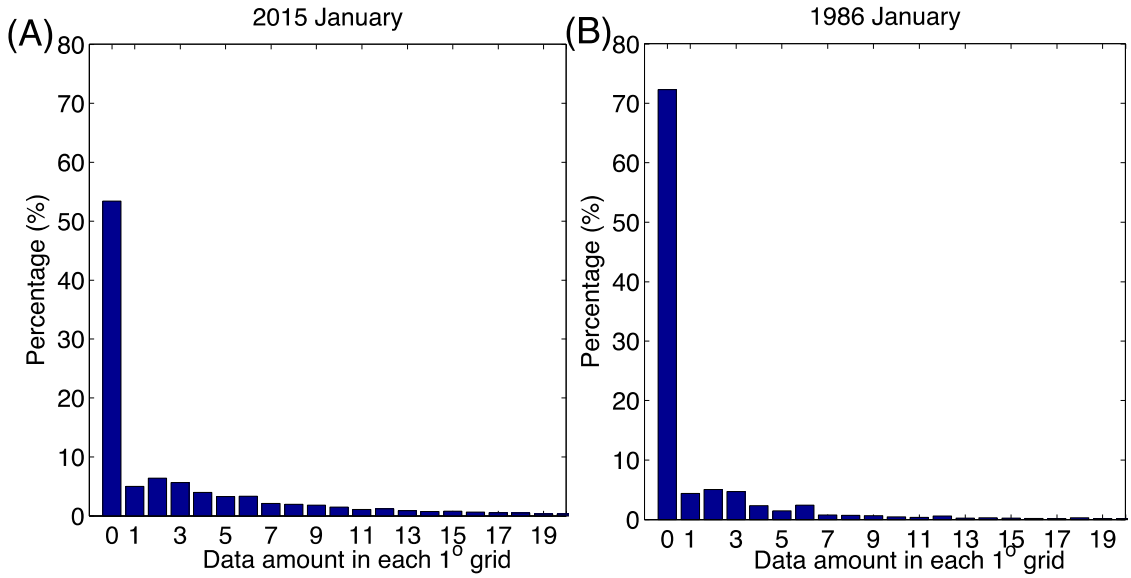

fig. S15. Frequency distribution of temperature anomalies for the years 1986 and 2015. Statistics of data amount in each 1° by 1° grid box at 20m depth in January 2015 (**A**) and January 1986 (**B**).

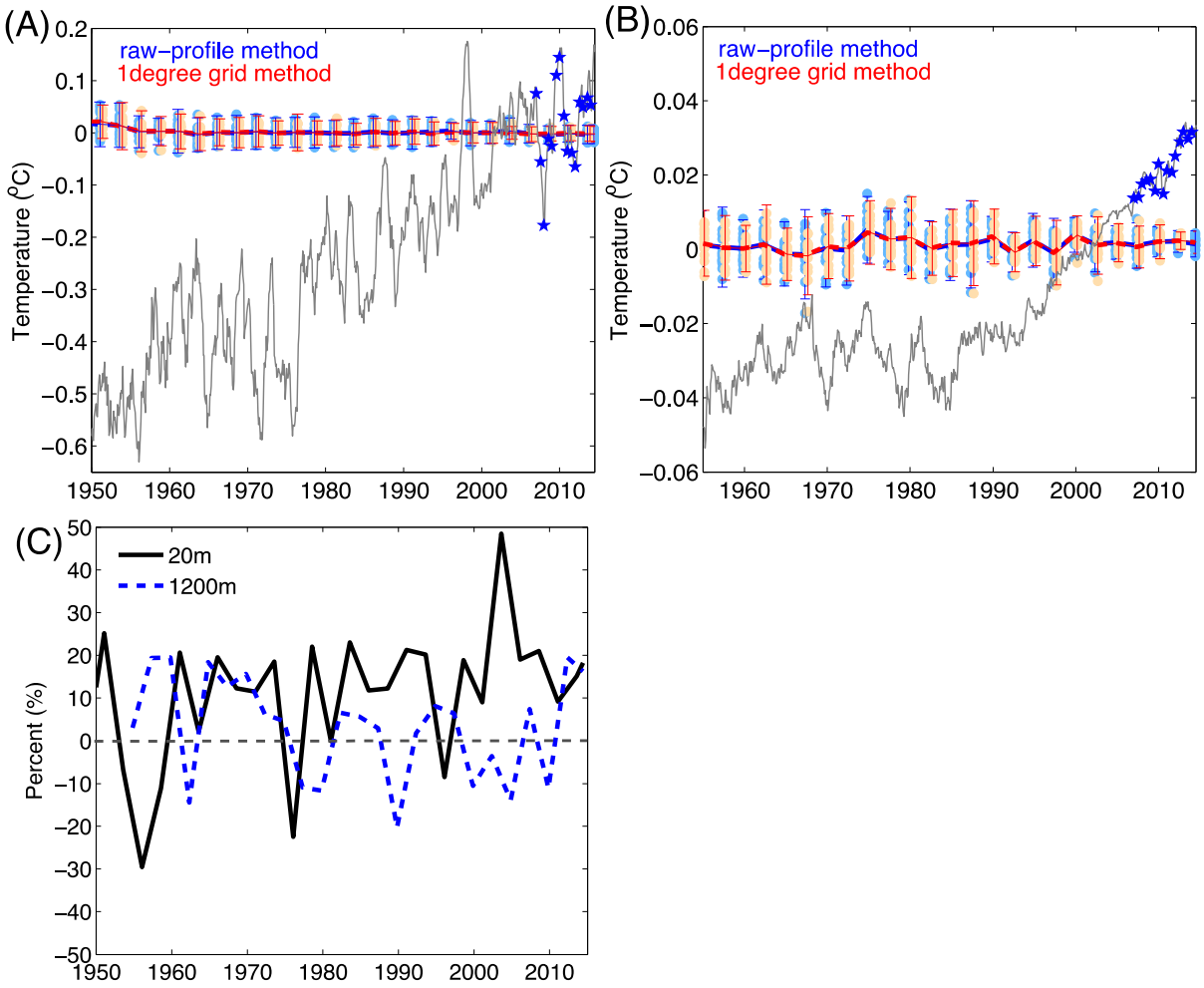

fig. S16. Sampling error as calculated by two subsample methods. Sampling error at 20m (**A**), and 1200m (**B**) by using two methods: raw-profile method (blue dots) and 1degree grid method (orange dots). The mean sampling error is shown in blue and red curves for the raw-profile method and 1degree grid method respectively. The 2σ sampling errors are also attached in error bars. The reconstructed temperature time series at 20m and 1200m are shown in the grey curves, and the truths are marked as stars. (**C**) shows the percentage difference between the two methods (the raw-profile method related to our 1degree grid method). Black solid line is for 20m, and blue dashed line is for 1200m.

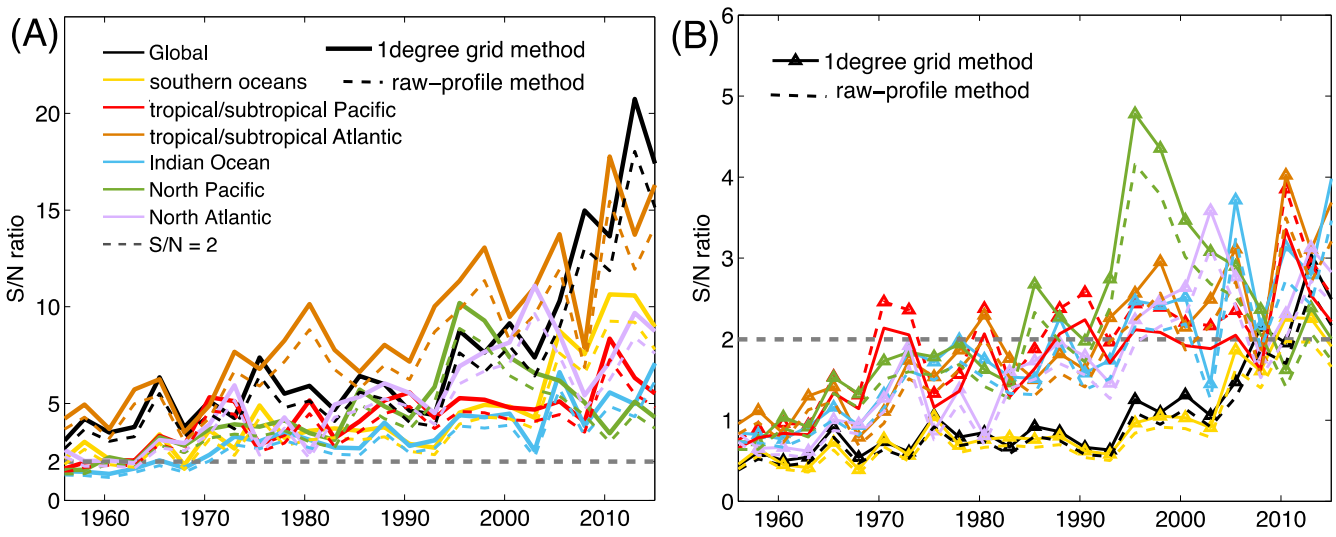

fig. S17. S/N ratio analysis for two methods of subsample test. S/N ratio for 0–2000m averaged temperature change is shown on global and regional scales. (**A**) S/N ratio for the decadal/multidecadal variation, and (**B**) for the inter-annual variation. The results using 1degree grid method (solid lines) and raw-profile method (assuming 15% larger sampling error) (dashed lines) are both shown for comparison.

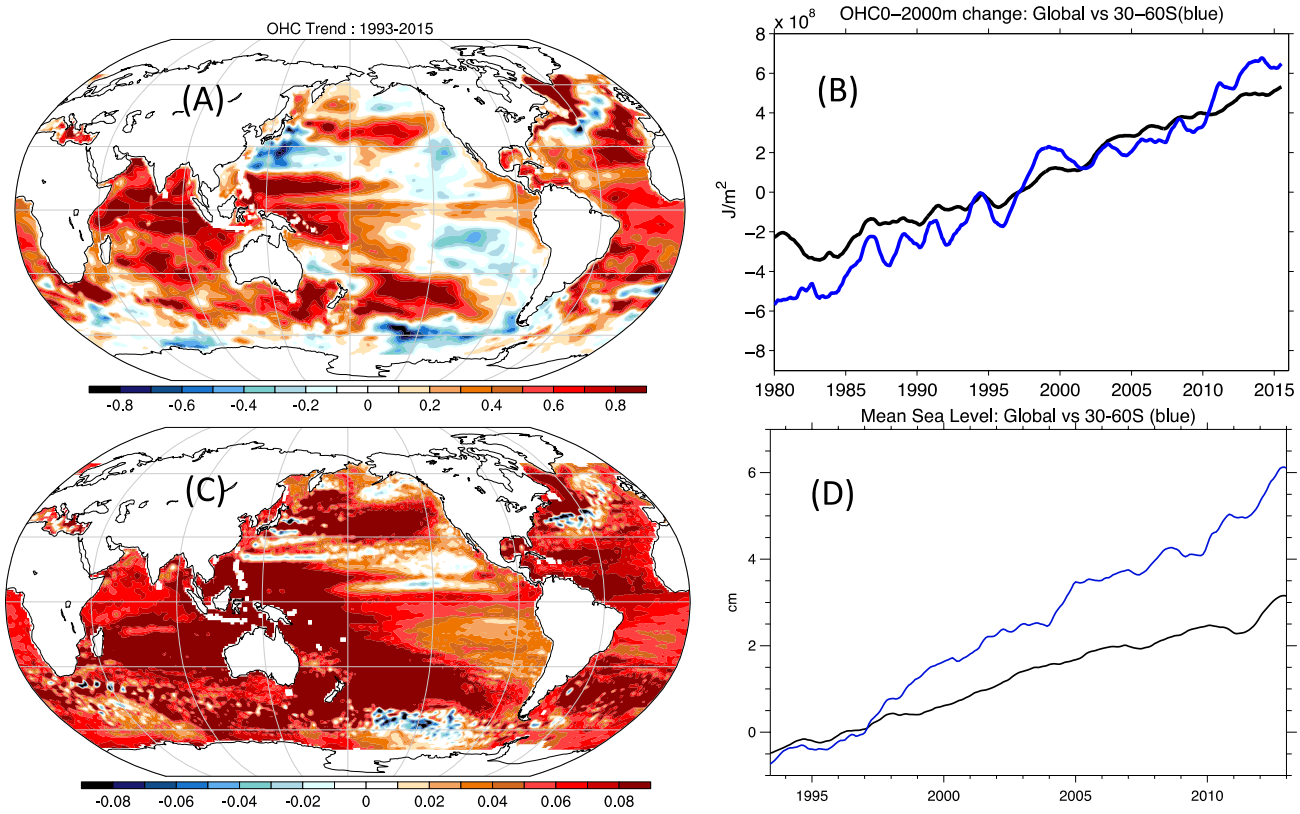

fig. S18. Comparison between OHC and sea level change since 1993. Linear trend of OHC0-2000m from 1993 to 2015 (in unit of 10^8 J/m^2) in (**A**) compared with sea level trend (in unit of m/decade) in (**C**). The time series of OHC0-2000m within 30°S-60°S and global ocean are shown in (**B**). Sea level time series are also presented in (**D**).

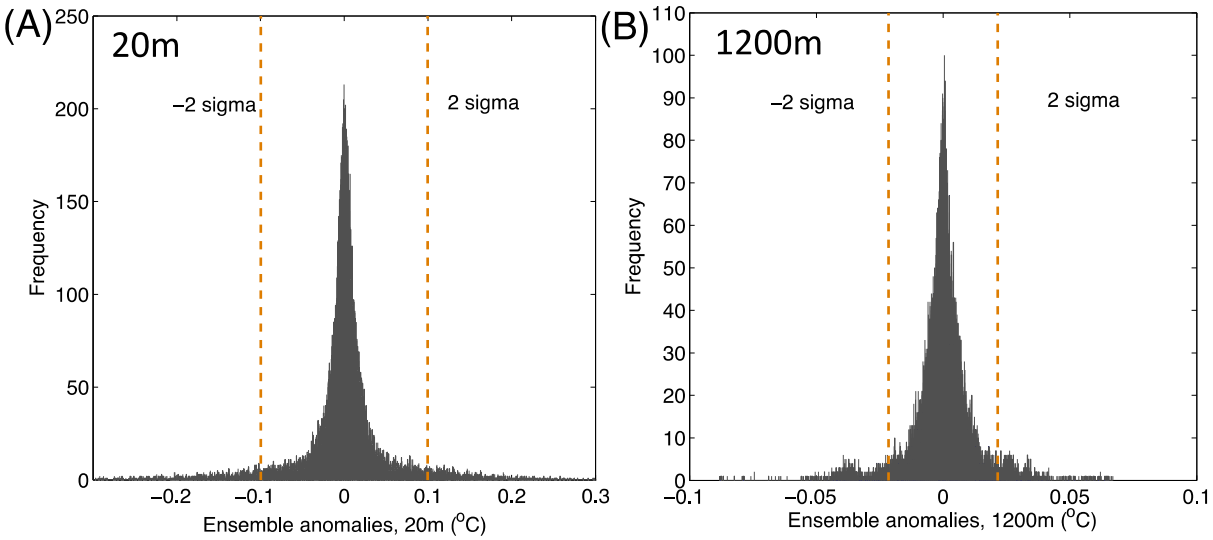

fig. S19. Distribution of the ensemble anomalies. Distribution of the ensemble anomalies of the current analyses at 20m (**A**) and 1200m (**B**). -2σ and 2σ are marked in orange dashed lines. The current mapping method is an ensemble method, so we use the ensemble anomalies to get the uncertainty estimate.

table S1. OHC trends obtained in this study for the 1960–1991 and 1992–2015 periods. The ocean heating rate is applied over the entire surface area of the Earth. Both OHC 0-700m and OHC 700-2000m are presented.

	OHC 0-700m ($\times 10^{22}$ J/yr)	**Ocean heating rate within 0-700m (W/m^2)**	**OHC 700m–2000m ($\times 10^{22}$ J/yr)**	**Ocean heating rate within 700–2000m (W/m^2)**
1960–1991 (linear trend)	0.15±0.08	0.09±0.05	0.04±0.08	0.02±0.05
1992–2015 (linear trend)	0.61±0.04	0.38±0.03	0.37±0.02	0.23±0.02

table S2. Net OHC and EEI changes obtained in the current study compared with some independent estimates. The independent estimates come from several published studies including *Cheng et al.* (*63*), *Allan et al.* (*65*), and IPCC-AR5 (*1*).

	Source	**Full-depth OHC ($\times 10^{22}$ J)**	**EEI ($\times 10^{22}$ J)**	
1960–2015	This study	33.5±7.0	36.0±7.5	
1970–2005	This study	26.5±4.8		
1970–2005	Adjusted observational OHCs and ORAS4 (*63*)	28.3±1.8		
1970–2005	CMIP5 (*63*)	26.6±4.4		
1993–2008	This study		18.7±1.1	
1993–2008	TOA (*65*)		16.7±17.2 (0.65±0.67 W/m^2)	
1971–2010	This study	28.8±4.4	31.0±4.7	
1971–2010	IPCC-AR5 (*1*)	25.5±6.1	27.4±7.8	