Simulated and observed variability in ocean temperature and heat content

K. M. AchutaRao*†, M. Ishii‡, B. D. Santer*, P. J. Gleckler*, K. E. Taylor*, T. P. Barnett§, D. W. Pierce§, R. J. Stouffer¶, and T. M. L. Wigley‖

*Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, CA 94550; ‡Frontier Research Center for Global Change, Japan Agency for Marine–Earth Science and Technology, Yokohama 236-0001, Japan; §Climate Research Division, Scripps Institution of Oceanography, La Jolla, CA 92037; ¶National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08542; and ‖National Center for Atmospheric Research, Boulder, CO 80307

Edited by Carl Wunsch, Massachusetts Institute of Technology, Cambridge, MA, and approved May 16, 2007 (received for review December 20, 2006)

Observations show both a pronounced increase in ocean heat content (OHC) over the second half of the 20th century and substantial OHC variability on interannual-to-decadal time scales. Although climate models are able to simulate overall changes in OHC, they are generally thought to underestimate the amplitude of OHC variability. Using simulations of 20th century climate performed with 13 numerical models, we demonstrate that the apparent discrepancy between modeled and observed variability is largely explained by accounting for changes in observational coverage and instrumentation and by including the effects of volcanic eruptions. Our work does not support the recent claim that the 0- to 700-m layer of the global ocean experienced a substantial OHC decrease over the 2003 to 2005 time period. We show that the 2003–2005 cooling is largely an artifact of a systematic change in the observing system, with the deployment of Argo floats reducing a warm bias in the original observing system.

climate | models | observations | ocean heat content

Observations suggest that the world's oceans were responsible for most of the heat content increase in the earth's climate system between 1955 and 1998 (1). This increase is embedded in substantial variability on interannual-to-decadal time scales. State-of-the-art climate models have been able to replicate both the overall increase in ocean heat content (OHC) during this period and its horizontal and vertical structure (2–7). Such detection and attribution studies have identified a large anthropogenic component in the observed changes and find that the "noise" of natural climate variability is an inadequate explanation for these changes.

The credibility of these results is strongly dependent on the reliability of natural variability estimates, particularly on the multidecadal time scales against which a slowly evolving anthropogenic signal must be discerned. This low-frequency noise information cannot be obtained from the relatively short (45- to 50-year) observational record and is typically estimated from model "undisturbed earth" experiments ("control runs"), which assume no changes in greenhouse gases or other external forcings (8). Several studies have reported that models may significantly underestimate the observed OHC variability (3, 9, 10), raising concerns about the reliability of detection and attribution findings (11, 12).

Although observational estimates of OHC change given in the 2005 World Ocean Atlas (WOA-2005) (1) are based on millions of individual temperature measurements, these measurements are unevenly distributed in space and time. Until recently, many portions of the global ocean were poorly sampled. To reconstruct the true (but unknown) four-dimensional structure of global ocean temperature and OHC changes, it is necessary to "infill" missing data. This has been done using either statistical approaches (1, 3, 13, 14) or physically based ocean models (15).

Because there is no unique solution to the infilling problem, and in view of concerns that previously applied statistical infilling approaches may alter ocean temperature variability (7, 16), it is preferable to restrict comparisons of modeled and observed variability to the actually observed portions of the ocean and, hence, to volume-averaged ocean temperature rather than OHC. This type of "model subsampling" strategy has been used in recent detection and attribution work (5, 6).

A previous study (16) employed model results from control runs and an idealized climate change experiment (17) to investigate the impact of incomplete space- and time-varying observational data coverage on simulated estimates of ocean temperature variability. Results were reported from eight different atmosphere/ocean general circulation models. Subsampling spatially complete model data with the observational data coverage mask amplified the temporal variability of ocean temperatures. In the control runs, the variability estimated from subsampled data was below variability levels in the subsampled observations. In the idealized experiment with 1%/year atmospheric CO_2 increases, however, the simulated variability of subsampled data was consistently larger than observed, primarily because of the unrealistically large CO_2 forcing (compared with the estimated observed forcing).

To evaluate the ability of models to simulate the observed amplitude of ocean temperature variability, it is therefore important to analyze model experiments that employ realistic estimates of historical forcings and to account for observational coverage and instrumentation changes. We consider all three issues here and address uncertainties in both model results and in the observations themselves.

Model and Observational Data

We examine a suite of recently completed climate model simulations carried out in support of the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Unlike the idealized experiments used in ref. 16, the simulations of 20th-century climate change [designated "20c3m" runs in the World Climate Research Program's Coupled Model Intercomparison Project Phase 3 (WCRP CMIP3) data archive] include estimated

Author contributions: K.M.A., B.D.S., K.E.T., T.P.B., R.J.S., and T.M.L.W. designed research; K.M.A., B.D.S., P.J.G., K.E.T., and D.W.P. performed research; M.I. contributed new reagents/analytic tools; K.M.A., M.I., P.J.G., and D.W.P. analyzed data; and K.M.A., B.D.S., P.J.G., K.E.T., T.P.B., D.W.P., R.J.S., and T.M.L.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: CTD, conductivity–temperature–depth; MBT, mechanical bathythermograph; OHC, ocean heat content; XBT, expendable bathythermograph.

†To whom correspondence should be addressed. E-mail: achutarao1@llnl.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0611375104/DC1.

historical changes in a variety of natural and anthropogenic forcings** [see ref. 8 and supporting information (SI) *Text*].

From WOA-2005, we computed the observed volume-averaged temperature changes over two different depth ranges: the top 700 m and top 3,000 m of the ocean. The observed data for these two depth ranges were available as annual and pentadal means, respectively. Volume-averaged temperature anomalies were calculated for both spatially complete ocean data (observed plus infilled) and for the portion of the ocean for which observations were available, yielding ΔT_o(Tot) and ΔT_o(Sub), respectively, where "o" is an observational result (see *SI Text* for further details). Estimates of ΔT_o(Tot) and ΔT_o(Sub) were also obtained from a second observational data set compiled by Ishii *et al.* (13) (henceforth, ISHII6.2), who used a different statistical procedure for infilling purposes. The ISHII6.2 data set is restricted to the top 700 m of the ocean.

After transforming the model ocean temperature data sets to the WOA-2005 grid, we first calculated simulated values of ΔT_m(Tot), where "m" is a model result, and then applied the WOA-2005 coverage mask (see *SI Text* for details of the masking procedure) to produce time series of ΔT_m(Sub). This was done for a total of 44 realizations of the 20c3m experiment, performed with 13 different climate models.

Model Performance in Simulating Variability

Fig. 1 shows time series of ΔT_o(Tot), ΔT_o(Sub), ΔT_m(Tot), and ΔT_m(Sub) for the upper 700 and 3,000 m of the global ocean for two selected atmosphere/ocean general circulation models. In the three realizations of the 20c3m experiment performed with the MIROC3.2(medres) model, the variability of ΔT_m(Tot) is noticeably smaller than the variability in ΔT_o(Tot) (Fig. 1 *A* and *C*). Subsampling the model data at the locations of observations substantially amplifies the simulated temporal variability (Fig. 1 *B* and *D*). In contrast, subsampling the five realizations of the CGCM3.1 (T47) 20c3m experiment yields only a small increase in decadal variability but enhances the ocean warming trend for the period 1955–1998 by a factor of ≈2. One important difference between the MIROC3.2(medres) and CGCM3.1(T47) experiments lies in the treatment of volcanic forcing, which is included in the former model but not in the latter.

Similar differences in variability and trends are apparent in the multimodel ensemble-mean ocean temperature changes estimated from models with and without volcanic forcing (V and No-V, respectively).†† For both the 0- to 700-m (Fig. 1 *A* and *B*) and 0- to 3,000-m (Fig. 1 *C* and *D*) layers, the average of the V model simulations has higher temporal variability than the No-V average but a smaller overall temperature trend. These results imply that (*i*) cooling caused by the inclusion of volcanic forcing offsets some of the greenhouse gas-induced ocean warming, thus reducing overall ocean temperature trends; and (*ii*) volcanic forcing is responsible for some of the decadal variability in volume-averaged ocean temperatures. Analyses of changes in sea surface temperature, subsurface temperatures, OHC, and sea level (8, 18–22) support these findings.

In Fig. 2, we illustrate the effect of subsampling on the temporal standard deviations (SDs) of global ocean temperature data. We

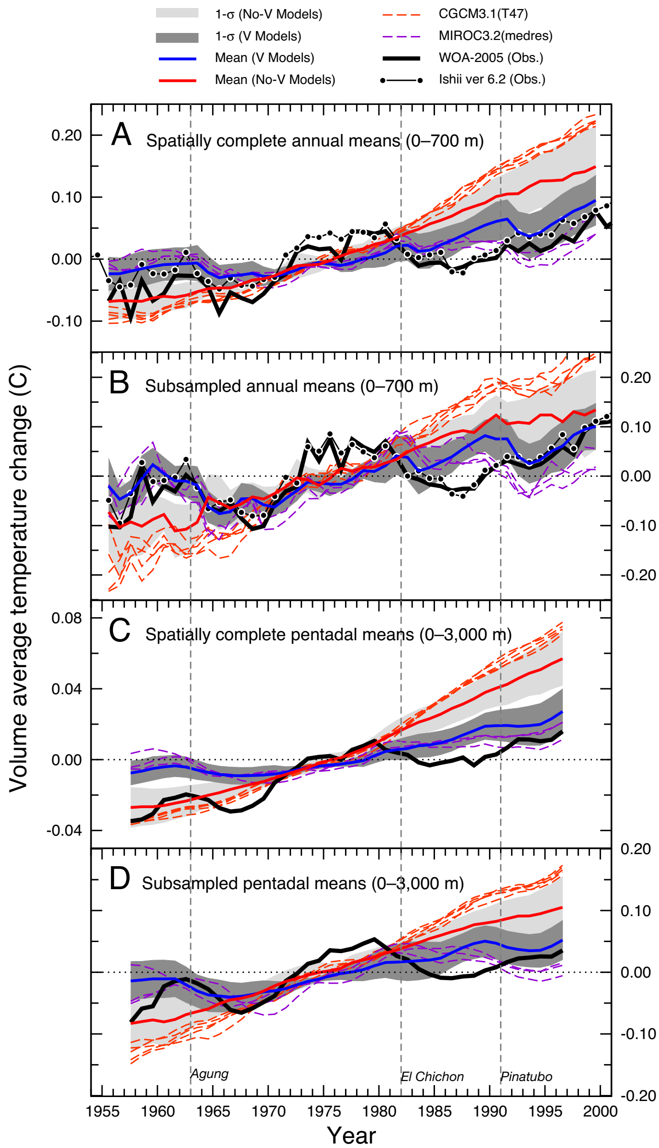

Fig. 1. Simulated and observed changes in volume-averaged temperature of the top 700 m (*A* and *B*) and 3,000 m (*C* and *D*) of the global ocean. Model results are from simulations of 20th century climate change performed with two atmosphere/ocean general circulation models: MIROC3.2 (medres) and CGCM3.1 (T47). Observations are from the WOA-2005 data set (1) and the ISHII6.2 data set (13). The ISHII6.2 data are available for 0–700 m only. Results are shown for both spatially complete temperatures (*A* and *C*) and temperatures subsampled with the WOA-2005 coverage mask (*B* and *D*). The multimodel V and No-V ensemble means†† are also plotted. These are based on 28 (16) realizations of the 20c3m experiment that included (excluded) volcanic forcing. Control run drift was removed from the model results (see *SI Text*). In both observations and models, the 0- to 700-m (0- to 3,000-m) temperature changes are annual (pentadal) means. The amplitude of the variability in the multimodel averages and observations should not be directly compared because averaging over realizations and models damps the unforced variability.†† Similarly, the SDs of the V and No-V results (represented by shaded envelopes) underestimate the true range of variability of individual model realizations. The observed variability is directly comparable to results from the individual MIROC3.2 (medres) and CGCM3.1(T47) realizations.

focus on the 0- to 700-m layer because (*i*) most of the observations are in the upper ocean, and the bulk of the observed increase in OHC since 1955 occurs in the 0- to 700-m layer; (*ii*) multiple observational estimates of changes in OHC and ocean temperature are available for the 0- to 700-m layer (12–14), whereas observa-

**Although all 13 modeling groups used very similar changes in well mixed greenhouse gases, the changes in other forcings were not prescribed as part of the experimental design. In practice, each group employed different combinations of 20th century forcings and often used different data sets for specifying individual forcings. End-dates for the experiments varied between groups and ranged from 1999 to 2003. Some modeling centers performed ensembles of the historical forcing simulation (see *SI Text* and SI Table 1). An ensemble contains multiple realizations of the same experiment, each starting from slightly different initial conditions but with identical changes in external forcings.

††We define $\bar{T}$ as the arithmetic mean of the ensemble means, i.e., $\bar{T} = 1/N \sum_{j=1}^{N} \bar{T}_j$, where N is the total number of models in the group (V or No-V) under consideration and T_j is the ensemble mean signal of the jth model. This weighting avoids placing undue emphasis on results from a single model with a large number of realizations. The intermodel SD is similarly defined based on the ensemble means (if available) from each model.

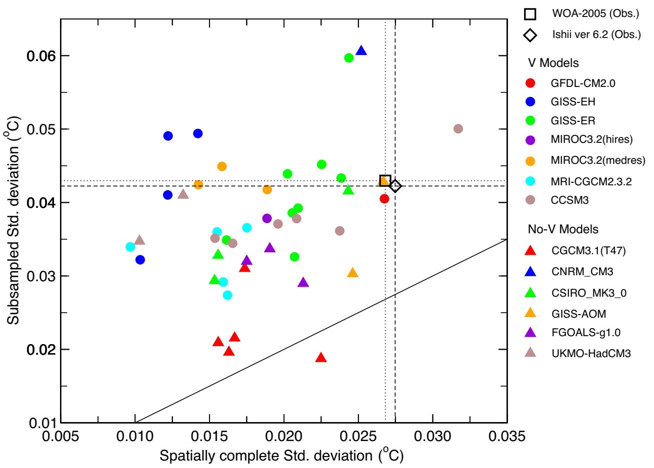

Fig. 2. Effect of subsampling on the temporal variability of 0- to 700-m volume-averaged ocean temperature changes. Results were calculated from temperature anomalies spatially averaged over the global ocean. The temporal SDs computed from spatially complete data (*x* axis) are plotted against the temporal SDs obtained from subsampled data (*y* axis). The WOA-2005 coverage mask was used to subsample both the model and ISHII6.2 temperature data. Colored symbols identify the 44 realizations of the Intergovernmental Panel on Climate Change 20c3m experiment. Colored circles (triangles) represent models that include (exclude) volcanic forcing. The diagonal line denotes equal SDs in the spatially complete and subsampled data. Horizontal and vertical lines mark the locations of observational results. Temporal SDs were calculated after removal of overall linear trends and are based on annual mean anomaly data for the period 1955–2000 (or 1999 in the case of model 20c3m experiments ending in that year). Note that the partitioning between V and No-V results is not as clear as in the multimodel averages in Fig. 1. This is because individual V and No-V models have differences in many other (nonvolcanic) forcings (see *SI Text*), as well as differences in climate sensitivity and the functional form of the forced OHC changes. All of these factors contribute to intermodel differences in the variability of the detrended ocean temperature data.

tional uncertainty is more difficult to assess for the 0- to 3,000-m layer; and (*iii*) recent detection and attribution studies (5, 6) have relied on temperature changes in the 0- to 700-m layer.

In the ΔT_m(Tot) data, only one of the 20c3m realizations has higher variability than in ΔT_o(Tot). Without subsampling, we would therefore conclude that the models analyzed here systematically underestimate the observed variability of temperatures in the top 700 m of the global ocean. Limiting variability comparisons to observed portions of the global ocean produces a very different result: 6 of the 13 models have at least one 20c3m realization in which the SD of ΔT_m(Sub) $\geq$ ΔT_o(Sub), and there is no evidence of a fundamental mismatch in simulated and observed variability.

However, discrepancies still remain in the phasing of observed and simulated temperature anomalies, particularly in the 1970s (Fig. 1 *B* and *D*). The apparent increase in observed heat content in the 1970s has been interpreted as a manifestation of a phase change in the North Pacific Oscillation (23). However, a recent study by Gouretski and Koltermann (24) suggests that some component of the observed temperature changes during the 1970s is spurious and arises from a combination of two factors: (*i*) large warm biases in the expendable bathythermograph (XBT) data relative to the more reliable measurements obtained from both conductivity–temperature–depth (CTD) sensors and hydrographic bottles and (*ii*) a substantial increase in the spatial coverage of XBT data after the late 1960s.

Obviously, model simulations cannot capture spurious ocean temperature variability associated with instrumental biases. Furthermore, we are dealing with coupled model experiments in which sea surface temperature changes are predicted rather than prescribed, and we do not expect the simulations to reproduce the precise timing and amplitude of observed OHC fluctuations induced by the North Pacific Oscillation or other modes of natural internal variability (except by chance).

Variability on Different Time Scales

The SDs shown in Fig. 2 reflect modeled and observed variability behavior across a range of time scales. In the following, we provide more time scale-specific comparisons of modeled and observed variability for 0- to 700-m OHC changes over periods of 2, 5, and 10 years. Such partitioning is of considerable interest because concerns have been expressed regarding the ability of models to capture the observed amplitude of both the decadal and interannual time scale variability of 0- to 700-m OHC (11, 12).

For example, a recent study by Lyman *et al.* (12) reported that, over the period 2003–2005, the heat content of the upper 700 m of the ocean decreased by 3.2 ± 1.1 × 10^{22} J. This decrease occurred in the absence of any major volcanic eruption. Lyman *et al.* do not provide a specific physical explanation for the cooling and claim that interannual variability of this magnitude "is not adequately simulated in the current generation of coupled climate models used to study the impact of anthropogenic influences on climate." We assess this claim in the analysis given below.

Fig. 3 shows the simulated and observed sampling distributions of 2-year changes in 0- to 700-m global OHC. This enables a direct comparison with the observed OHC change for 2003–2005 estimated by Lyman *et al.* (12). In WOA-2005, ISHII6.2, and each of the 44 20c3m realizations, all possible (overlapping) 2-year changes in OHC were calculated for the period 1955–2000 (or for 1955–1999 in the case of 20c3m runs ending in 1999). For models with multiple 20c3m realizations, the ensemble-mean sampling distribution is plotted.

The observed distributions of 2-year changes in 0- to 700-m global OHC are very similar to those obtained with the V models (Fig. 3). Distributions generated from No-V models tend to be skewed positively relative to those estimated from observations and V models. The V models are able to capture the large OHC changes at the tails of the observed distributions. The claim that models are unable to replicate the observed interannual variability of 0- to 700-m global OHC (12) is, therefore, not supported by our analysis. Distributions of 0- to 700-m OHC changes on 5- and 10-year time scales are also highly similar in models and observations. This holds not only for the global ocean but also for individual ocean basins (see *SI Text* and SI Figs. 6–8).

Fig. 3 also shows observational estimates of 0- to 700-m global OHC changes for 2003–2005 from Lyman *et al.* (12) and ISHII6.2 (13). Each observational data set exists in two versions: with and without inclusion of information from the Argo temperature and salinity profiling floats. The deployment of Argo floats commenced in the Atlantic in 2000 and has rapidly ramped up to a total of 2,804 floats (as of April 3, 2007) and near-global coverage of the world's oceans (25).

Consider first the "With Argo" estimates of the 2003–2005 OHC changes. These range from −4.3 × 10^{22} J (ISHII6.2) to −3.2 × 10^{22} J [Lyman *et al.* (12)]. In the context of the sampling distributions of 2-year changes obtained from the full (1955–2000) ISHII6.2 and WOA-2005 data sets, these changes are unusual but not unprecedented. Three models (GFDL-CM2.0, GISS-ER, and CNRM-CM3) are capable of capturing 2-year OHC changes larger than the 2003–2005 OHC decrease in ISHII6.2, whereas five models (GFDL-CM2.0, GISS-ER, GISS-EH, CCSM3, and CNRM-CM3) simulate 2-year changes rivaling or exceeding the OHC decrease in Lyman *et al.*

Exclusion of Argo float data markedly reduces the estimated cooling. OHC changes in the "No Argo" data set versions range from −0.16 to −0.86 × 10^{22} J [in ISHII6.2 and Lyman *et al.* (12),

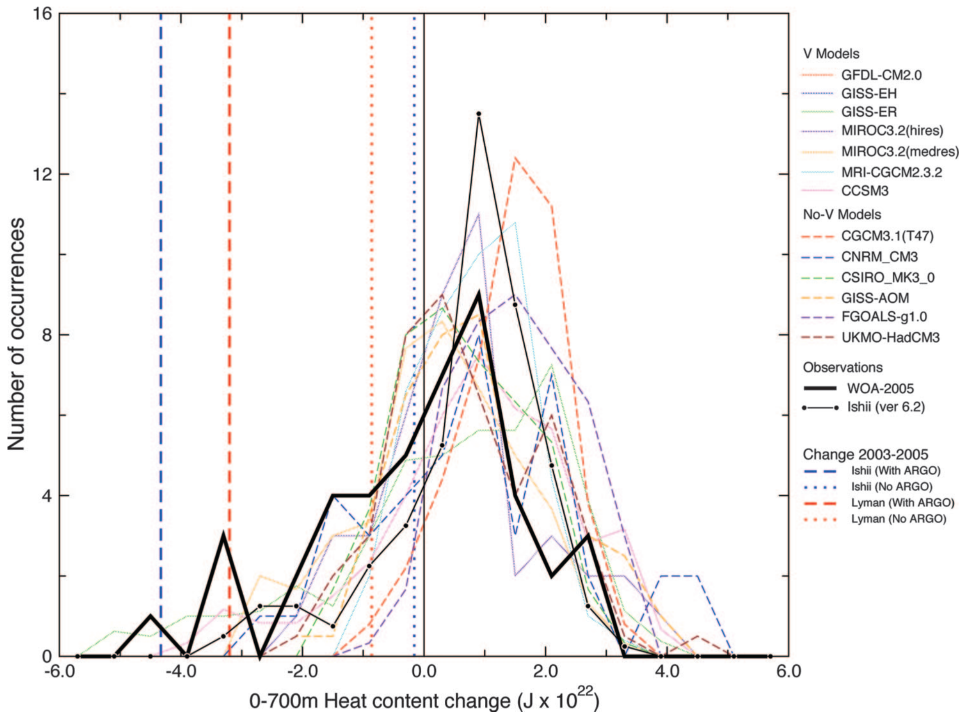

Fig. 3. Simulated and observed sampling distributions of 2-year changes in 0- to 700-m global OHC. Distributions were generated as described in the text. The observed sampling distributions are based on annual-mean OHC data from WOA-2005 (1) and ISHII6.2 (13). Vertical lines represent observed OHC changes over 2003–2005 in With Argo and No Argo versions of the ISHII6.2 and Lyman *et al.* (12) data sets. Model results are partitioned into V and No-V groups.

respectively]. As is visually obvious from Fig. 2, the No Argo OHC changes are not unusually large when compared with either observed or simulated sampling distributions of 2-year time scale OHC changes. The largest differences between the No Argo and With Argo changes estimated by ISHII6.2 are in the Southern Hemisphere (see SI Fig. 6).

One possible explanation for the pronounced differences between the With Argo and No Argo results is related to the

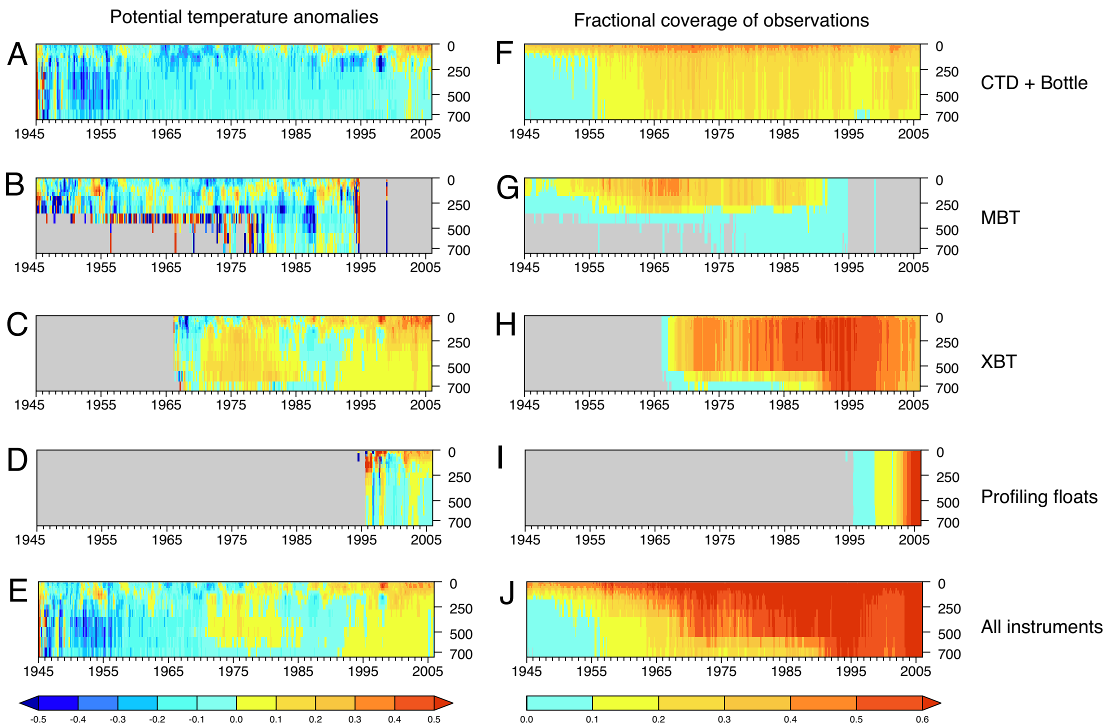

Fig. 4. Time series of layer-averaged monthly potential temperature anomalies and the area fraction of the global ocean observed at each level. (*A*–*E*) Potential temperature, separated by instrument type for CTD+Bottle (*A*), MBT (*B*), XBT (*C*), and profiling floats (*D*) and by combining all instruments (*E*). (*F*–*J*) Fractional areas of coverage, shown by instrument type for CTD+Bottle (*F*), MBT (*G*), XBT (*H*), and profiling floats (*I*) and by combining all instruments (*J*). Missing data are denoted by gray areas. The gridded anomalies were computed separately for each instrument type and by combining all instrument types, as in ref. 13, and were transformed to the same 2° latitude × 4° longitude grid used by Gouretski and Koltermann (24).

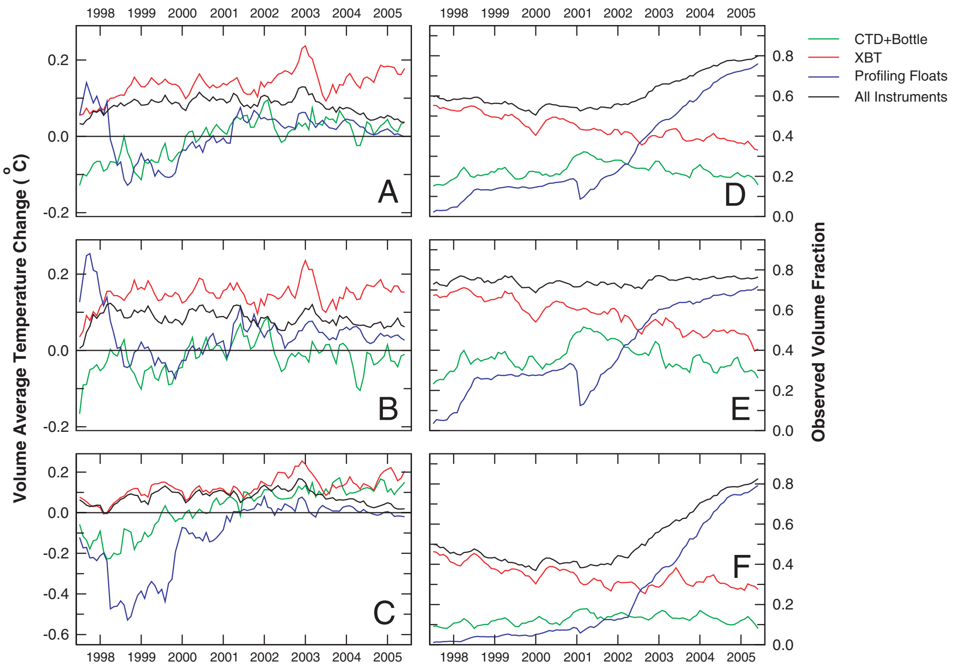

Fig. 5. Time series of volume-averaged potential temperature anomalies and coverage in the world's oceans, as measured by various instruments. (*A*–*C*) Volume-averaged temperature change in the global (*A*), Northern Hemisphere (*B*), and Southern Hemisphere (*C*) oceans. (*D*–*F*) Coverage (expressed as fractional volume of the basin) in the global (*D*), Northern Hemisphere (*E*), and Southern Hemisphere (*F*) oceans. The temperature anomalies and coverage fractions were computed from the data shown in Fig. 4. Note that because individual instruments may overlap in coverage with other instruments, coverage fractions from individual instruments do not add up to the ''all instruments'' coverage fraction.

large, systematic changes in the ocean observing system that occurred from 2003 to 2005 (see SI Fig. 9). Observational coverage increased dramatically over this 2-year period, primarily because of the large-scale deployment of Argo floats. The largest coverage increases are in the Southern Ocean, which was very poorly observed prior to 2004. There is considerable spatial correspondence between regions with large increases in the number of observations and regions with large decreases in OHC (SI Fig. 9 *C* and *F*).

The rapid change in observational coverage associated with the deployment of Argo floats is convolved with instrumental biases between the Argo profilers and the mix of XBT, CTD, and bottle data that formed the bulk of the observational coverage in the pre-Argo era. Gouretski and Koltermann (24) showed that XBT measurements (which constitute the majority of observations since the late 1960s) are biased warm relative to collocated, and more accurate, CTD+Bottle measurements and with respect to profiling floats.[‡‡]

To complement Gouretski and Koltermann's (24) analysis of biases in collocated data, we partitioned the observations used in producing ISHII6.2 according to instrument type [CTDs+Bottles, mechanical bathythermographs (MBTs), XBTs, and profiling floats] and then studied the relationship between the coverage changes of individual instrument types and the ocean temperature changes inferred from those instruments. Unlike in Gouretski and Koltermann, our XBT data include drop rate error corrections, which have the effect of increasing the warm bias of XBTs relative to other instrument types (24). The ocean temperature anomalies shown in Fig. 4 *A*–*E* were computed relative to a common WOA-2001 climatology, and the fractional observational coverage at each standard level is given in Fig. 4 *F*–*J*.

Even in the noncollocated data, a clear warm bias is seen in XBT measurements (Fig. 4*C*) relative to the CTD+Bottle and profiling float data (Fig. 4 *A* and *D*). The ocean warming in the 1970s is only manifest in XBT data and coincides with large changes in XBT coverage (Fig. 4 *C* and *H*; see also earlier discussion of Fig. 1). Because XBTs make up the largest fraction of ocean observations since the late 1960s, the XBT-inferred "warming" of the 1970s is also evident in the "all instruments" analysis (Fig. 4*E*).[§§]

The coverage fraction of profiling floats (Fig. 4 *I*) increased rapidly and exceeded that of XBTs after 2004 (25). The transition from XBT to Argo is most pronounced in the Southern Hemisphere (see Fig. 5 *C* and *F*). In the Northern Hemisphere, the shift from XBTs to profiling floats occurred more gradually, beginning in the late 1990s with the World Ocean Circulation Experiment. This transition from an XBT observing system that was biased warm to the Argo profilers results in an apparent cooling over the 2003–2005 period considered by Lyman *et al.* (12) (Figs. 4*E* and 5 *A*–*C*). It is remarkable that this cooling is only evident in the "all instruments" case; it does not occur in any of the individual instruments.

Summary and Conclusions

Our study has addressed the persistent concern that models may underestimate the true variability of ocean temperature and heat content. We find no evidence of such a systematic underestimate. We identify three major factors that need to be taken into account when comparing modeled and observed variability and show that the observed variability is difficult to estimate reliably.

[‡‡]The analysis of Gouretski and Koltermann (25) ended in 2001. In assessing profiler biases, therefore, it primarily focused on the pre-Argo generation of profilers used in the World Ocean Circulation Experiment.

[§§]We note, however, that there are also time-varying biases between the collocated XBT and CTD+Bottle data (25).

The first factor is incomplete and time-varying observational coverage. Until the advent of Argo data in the early 21st century, our view of the mean state and variability of ocean temperatures was based on incomplete observational coverage that varied geographically, with depth, and over time. Most analysts used statistical procedures to produce global-scale estimates of OHC changes. Different infilling choices yielded different estimates of OHC variability. Because of the uncertainties introduced by infilling, we compared modeled and observed ocean temperature variability at the actual locations of observations. Our analysis used results from 13 different climate models and focused on volume-averaged temperatures in the upper 700 m of the ocean. Subsampling model data at the locations of observations amplifies the ocean temperature variability in all models and ocean basins.

A second relevant factor is whether the model simulations incorporate volcanic forcings. Even in spatially complete model data, simulated variability is enhanced by the inclusion of volcanic forcing. When we subsample 20c3m runs with combined anthropogenic and volcanic forcing, we find no evidence of a fundamental discrepancy between simulated and observed ocean temperature variability.

For all major ocean basins, there was good agreement between the simulated and observed sampling distributions of OHC changes on 2-, 5-, and 10-year time scales (see SI Figs. 6–8). These results enhance our confidence in the reliability of previously published detection and attribution studies. Model-based variability estimates are an integral component of such work (2, 4, 5).

A third factor that influences variability in observational data sets is related to the complex interplay between the documented biases in different types of instruments (24) and systematic space–time changes in their relative contributions to the overall observing system. A recent study by Lyman *et al.* (12) claimed that the 0- to 700-m layer of the global ocean experienced a heat content decrease of $3.2 \pm 1.1 \times 10^{22}$ J over 2003–2005 and that models cannot replicate changes of this magnitude. Our analysis shows that the cooling found by Lyman *et al.* is spurious. At least five lines of evidence support this conclusion.

First, the main contribution to the large global OHC decrease over 2003–2005 is from the Southern Ocean, where Argo coverage increased dramatically after 2003 (Fig. 5*F*). The massive influx of Argo data reduces preexisting warm biases from XBT measurements. Second, there were no unusually large OHC decreases in the Northern Hemisphere oceans over the same period (see SI Fig. 6). Third, global OHC decreases are substantially smaller in No Argo versions of two observational data sets and are consistent with the magnitude of changes typically seen in model simulations (Fig. 3). Fourth, none of the individual instrument types show evidence of global- or hemispheric-scale cooling over the period analyzed by Lyman *et al.* (12). Finally, analyses of satellite data that are completely independent of *in situ* observations do not confirm such a decrease (26).

Our study does not directly address the accuracy of the Argo measurements. Within the next decade, Argo will vastly improve our knowledge of the oceans and their variability. However, some caution must be exercised in estimating global-scale OHC trends from an observing system that has undergone large and rapid increases in coverage and whose measurement biases have not been adequately quantified. As in the case of atmospheric reanalyses (27–29), there will be significant challenges in separating true ocean climate change from the effects of changes in the observing system itself. The large and time-varying inter-instrument biases discussed here, coupled with systematic changes in the spatial and temporal deployment of different instrument types, introduce significant uncertainty in estimates of the true variability of global ocean temperatures and heat content.

The authors of the original Lyman *et al*. paper (12) have now publicly acknowledged that their earlier finding of pronounced ocean cooling over 2003–2005 was spurious (30). Their unpublished analyses confirm that this "cooling" arose for reasons similar to those identified here.

We thank the modeling groups for providing their data for analysis; the Program for Climate Model Diagnosis and Intercomparison for collecting and archiving the model output; the WCRP Working Group on Coupled Modeling for organizing the model data analysis activity; the editor, three anonymous reviewers, and Jonathan Gregory (University of Reading/Hadley Centre, Reading, U.K.) for valuable comments; Viktor Gouretski (Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany), Josh Willis (Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA), and John Lyman (National Oceanic and Atmospheric Administration/Pacific Marine Environmental Laboratory, Seattle, WA) for providing data and insights; Gary Strand (National Center for Atmospheric Research) for supplying ocean temperature data from the CCSM3 model; and Detelina Ivanova for additional technical help. The multimodel data archive is supported by the Office of Science, U.S. Department of Energy. This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under Contract W-7405-Eng-48. T.M.L.W. was supported by National Oceanic and Atmospheric Administration Office of Climate Programs ("Climate Change Data and Detection") Grant NA87GP0105.

1. Levitus S, Antonov JI, Boyer TP (2005) *Geophys Res Lett* 32:L02604.
2. Barnett TP, Pierce DW, Schnur R (2001) *Science* 292:270–274.
3. Levitus S, Antonov JI, Wang J, Delworth TL, Dixon KW, Broccoli AJ (2001) *Science* 292:267–270.
4. Reichert BK, Schnur R, Bengtsson L (2002) *Geophys Res Lett* 29:1525.
5. Barnett TP, Pierce DW, AchutaRao KM, Gleckler PJ, Santer BD, Gregory JM, Washington WM (2005) *Science* 309:284–287.
6. Pierce DW, Barnett TP, AchutaRao KM, Gleckler PJ, Gregory JM, Washington WM (2006) *J Clim* 19:1873–1900.
7. Gregory JM, Banks HT, Stott PA, Lowe JA, Palmer MD (2004) *Geophys Res Lett* 31:L15312.
8. Santer BD, Wigley TML, Gleckler PJ, Bonfils C, Wehner MF, AchutaRao KM, Barnett TP, Boyle JS, Brüggemann W, Fiorino M, *et al.* (2006) *Proc Natl Acad Sci USA* 103:13905–13910.
9. Sun S, Hansen JE (2003) *J Geophys Res* 16:2807–2826.
10. Hansen JE, Nazarenko L, Ruedy R, Sato M, Willis J, Del Genio A, Koch D, Lacis A, Lo K, Menon S, *et al.* (2005) *Science* 308:1431–1435.
11. Hegerl GC, Bindoff NL (2005) *Science* 309:254–255.
12. Lyman JM, Willis JK, Johnson GC (2006) *Geophys Res Lett* 33:L18604.
13. Ishii M, Kimoto M, Sakamoto K, Iwasaki S-I (2006) *J Oceanogr* 62:155–170.
14. Willis JK, Roemmich D, Cornuelle B (2004) *J Geophys Res* 109:C12036.
15. Carton JA, Giese BS, Grodsky SA (2005) *J Geophys Res* 110:C09006.
16. AchutaRao KM, Santer BD, Gleckler PJ, Taylor KE, Pierce DW, Barnett TP, Wigley TML (2006) *J Geophys Res* 111:C05019.
17. Meehl GA, Boer GJ, Covey C, Latif M, Stouffer RJ (2000) *Bull Am Meteorol Soc* 81:313–318.
18. Fyfe JC (2006) *Geophys Res Lett* 33:L19701.
19. Delworth TL, Ramaswamy V, Stenchikov GL (2005) *Geophys Res Lett* 32:L24709.
20. Gleckler PJ, AchutaRao KM, Gregory JM, Santer BD, Taylor KE, Wigley TML (2006) *Geophys Res Lett* 33:L17702.
21. Church JA, White NJ, Arblaster JM (2005) *Nature* 438:74–77.
22. Gregory JM, Lowe JA, Tett SFB (2006) *J Clim* 19:4576–4591.
23. Stephens C, Levitus S, Antonov JI, Boyer TP (2001) *Geophys Res Lett* 28:3721–3724.
24. Gouretski V, Koltermann KP (2007) *Geophys Res Lett* 34:L01610.
25. Gould JD, Roemmich S, Wijffels H, Freeland M, Ignaszewsky X, Jianping S, Pouliquen Y, Desaubies U, Send K, Radhakrishnan K, *et al.* (2004) *EOS, Trans Am Geophys Union*, 85:179, 190–191.
26. Lombard A, Garcia D, Ramillien G, Cazenave A, Biancale R, Lemoine JM, Flechtner F, Schmidt R, Ishii M (2007) *Earth Planet Sci Lett* 254:194–202.
27. Santer BD, Hnilo JJ, Wigley TML, Boyle JS, Doutriaux C, Fiorino M, Parker DE, Taylor KE (1999) *J Geophys Res* 104:6305–6333.
28. Pawson S, Fiorino M (1999) *Clim Dyn* 15:241–250.
29. Trenberth KE, Stepaniak DP, Hurrell JW, Fiorino M (2001) *J Clim* 14:1499–1510.
30. Schiermeier Q (2007) *Nature* 442:854–855.