Building a global ocean profile database to estimate heat and freshwater variability for the World Ocean

T. P. Boyer, S. Levitus, J. I. Antonov, R. A. Locarnini, H. E. Garcia, D. Johnson, I . Smolyar, O. Baranova, R. Tatusko, M. Pitcher, R. Gelfeld, C. Forgy, A. Grodsky

NOAA/National Oceanographic Data Center/Ocean Climate Laboratory
1315 East-West Highway, Silver Spring, MD USA

corresponding author: T. P. Boyer, boyer@nodc.noaa.gov

Abstract: The *World Ocean Database 2001* (WOD01) was released in 2001 as part of the Global Ocean Data Archeology and Rescue (GODAR) project. It contains 7 million temperature profiles and 2 million salinity profiles from all parts of the world ocean from 1773 to July, 2001. The data is available on CD-ROM as well as a constantly updated version online. All temperature and salinity data on WOD01 have been quality controlled through rigorous procedures which include using the data to calculate climatologies over various time periods (annual, seasonal, monthly). For periods with sufficient geographic coverage, 1955 to 2003 for temperature, 1955 to 1998 for salinity, five year mean differences from the climatologcal monthly means where calculated, removing the annual cycle. The results were used to estimate decadal changes in heat content and freshwater budgets globally. The use of the WOD01 data in scientific research provided quality control of data and the discovery of problems which were not detected during prior quality control steps. This additional quality control was passed on in quality control flags included with the data. Results are detailed for the north Atlantic. Our results show warming over much of the Atlantic, reaching to 1500 m depth in the subtropical north Atlantic. The subtropical and tropical north Atlantic have been becoming more saline, while the subpolar gyre of the north Atlantic has been getting fresher.

1. Introduction

It has long been suspected that the world ocean may play a large role in the Earth's heat budget (Rossby 1959). Recently it has become apparent that changes in the oceans freshwater budget can also play a large role in global climate variability, to the extent of possibly shutting down the overturning circulation in the North Atlantic (Rahmstorf, 1995). But the lack of available subsurface temperature and salinity data have hindered attempts to quantify the oceans' heat and freshwater budgets. There have been many oceanographic cruises, merchant ships, navigation buoys, drifting meteorological buoys, etc. taking temperature and to a lesser extent, salinity measurements. However, most of these measurements were taken to investigate specific aspects of the physical or biological ocean and were never gathered together in sufficient numbers to create a global database sufficient to calculate global budgets for heat and freshwater. The Global Ocean Data Archeology and Rescue (GODAR) project is an attempt to gather together data from all possible sources into a uniform global dataset. However, simply gathering together all this data is just one aspect of the project. All the different instrumentation, recording methods, and uses of the data make for a very heterogenous collection of data. The second important aspect of the GODAR project is a careful quality control of the oceanographic data by oceanographers and data specialists. Scientific research using all available data during the compilation of the global dataset is one of the key aspects of the quality control procedure. The results of the GODAR project are the *World Ocean Database 2001* (WOD01) a global oceanographic profile dataset with quality control flags and all data in one format ; *the World Ocean Atlas 2001* (WOA01), a series of climatological mean felds of oceanographic variables at surface and subsurface depths based on the WOD01 dataset ; and peer reviewed scientific papers based on the WOD01 dataset. These later include papers on the decadal variability of heat and freshwater based on subsurface temperature and salinity measurements.

2. Overview of Data in WOD01

The WOD01 is a quality controlled collection of ocean profile, plankton, and ship-based surface measurements from 1773-present (Fig. 1).

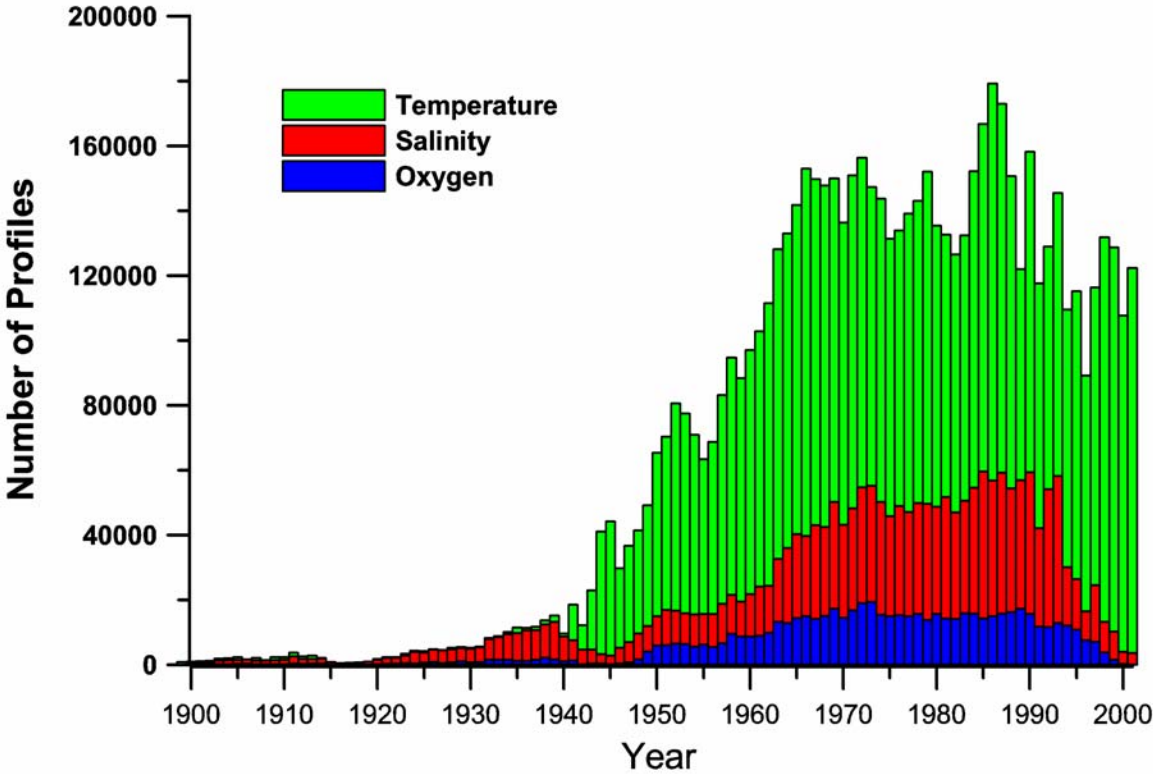

Figure 1. Number of Profiles per year in the World Ocean Database 2001

Before World War II there were relatively few oceanographic profiles. Starting in 1945, oceanographic measurements of temperature exceeded 40,000 profiles per year. There are many more temperature measurements than for any other oceanographic variable, with over 7 million in WOD01. There are also over 2 million salinity profiles, and more than 500,000 oxygen profiles. Other oceanographic variables, including nutrients (silicate, phosphorus, nitrogen), chlorophyll and plankton, and CO_2 variables (pH, alkalinity, CO_2 partial pressure) are available in WOD01 in smaller numbers. The drop in the number of profiles since 1990 reflects the sometimes time consuming process of receiving the data from researchers, who may have a proprietary period of between two and ten years for the data. The geographical distribution of measurements is governed by the sources of the data. Some of the data available in WOD01 comes from sources such as the World Ocean Circulation Experiment (WOCE, Fig 2a.) and GEOSECS (Fig 2b)

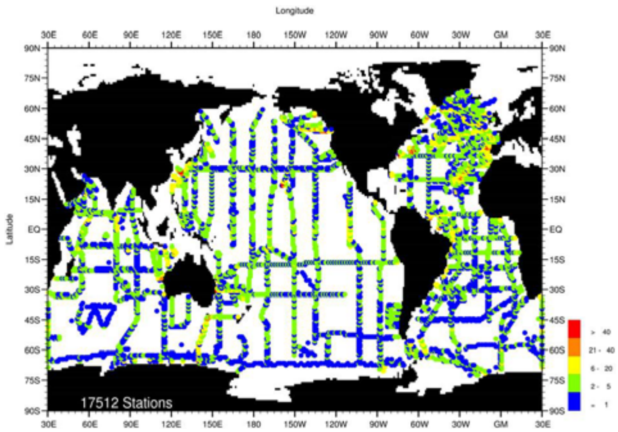

Figure 2a. Distribution of WOCE CTD data (1990-1998)

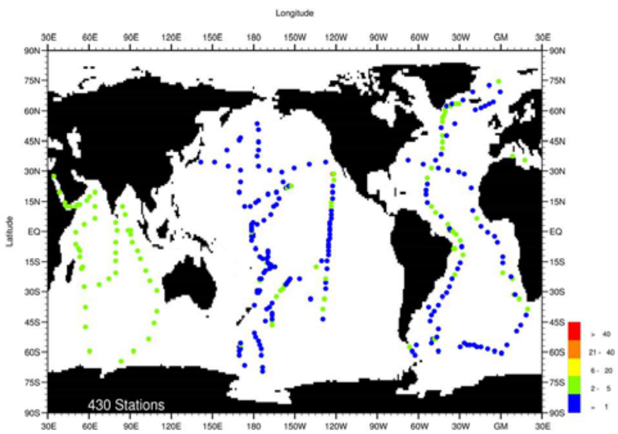

Figure 2b. Distribution of GEOSECS bottle and CTD data (1972-1979)

whose goal was to get a near-synoptic view of the ocean. Other sources include, the US National Marine Fisheries Service (2c) and the University of Hokkaido in Japan (Fig 2d), who have specific missions and areas of research..

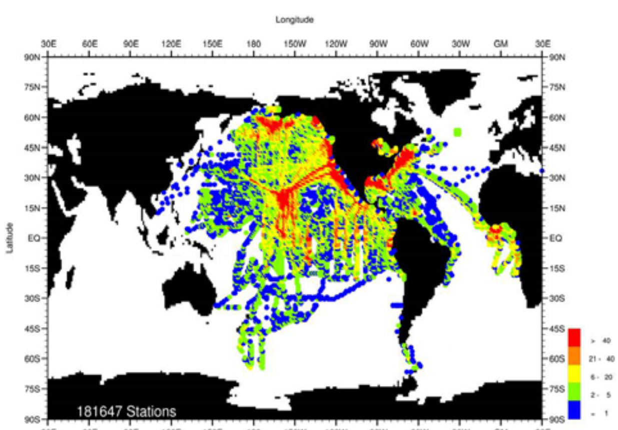

Figure 2c. Distribution of temperature data from NMFS (1965-1997)

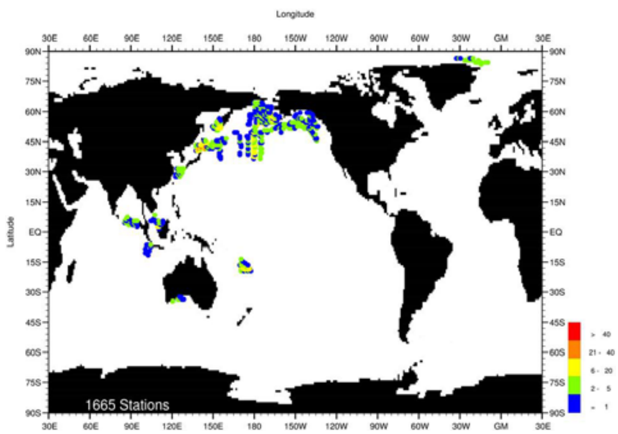

Figure 2d. Distribution of bottle data from Hokkaido University (1962-1994)

Merchant ships outfitted with measurement instruments are another source, which follow standard shipping routes. This leaves large areas of the ocean, especially the Southern Ocean relatively data sparse. In addition, historically, there are more oceanographic measurements near the surface than in the deep ocean (Fig 3).

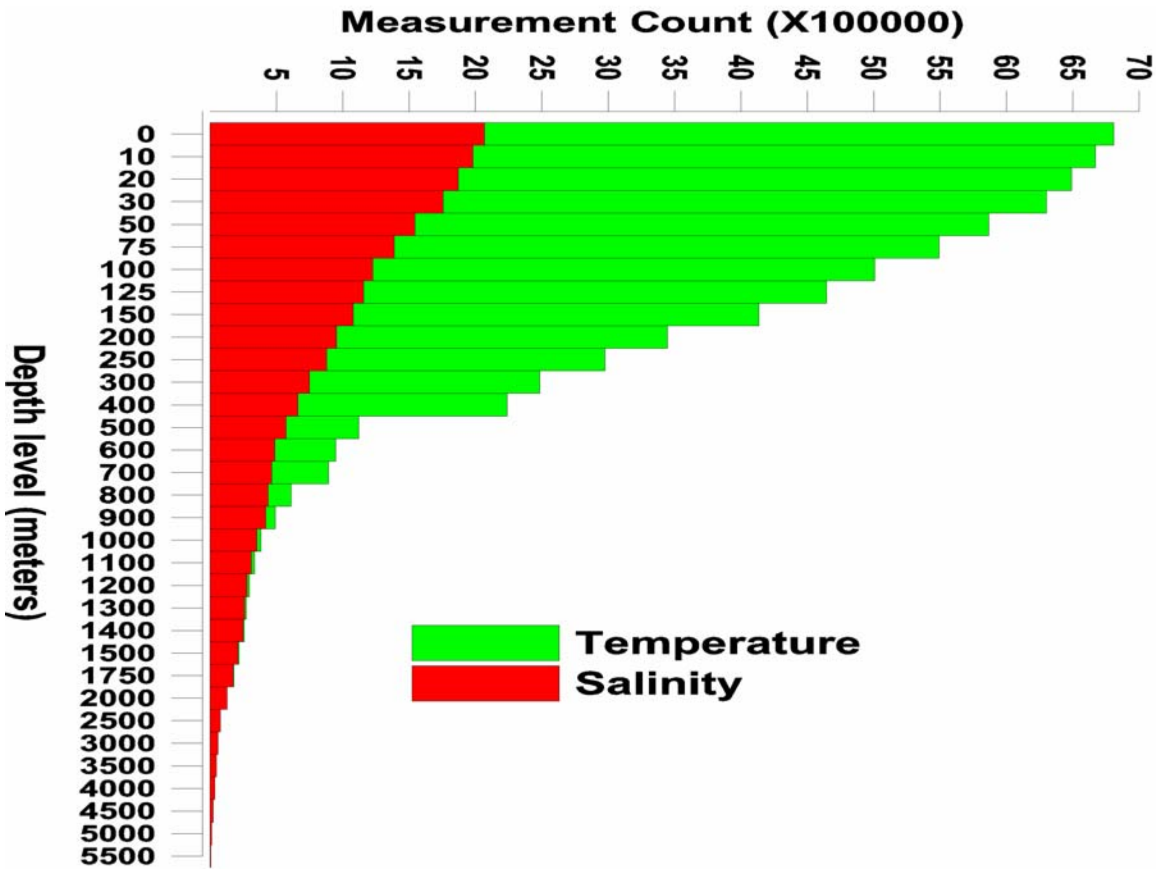

Figure 3. Distribution of temperature and salinity measurements with standard depths

Data is received regularly and in large quantities from National Oceanographic Data Centers from around the world. It is also found on the Internet, on index cards at research institutes, gathered from scientists and project managers, and from published cruise reports. Figure 4 shows Russian cruises in the Arctic from the 1920s and 1930s discovered in university and public libraries around the World, including the New York Public Library.

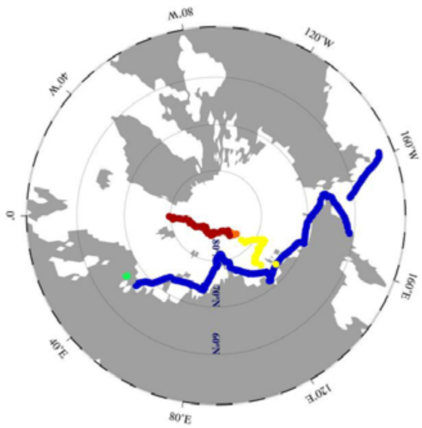

Figure 4. Arctic profile data (1920-1939) from cruise reports found in public/university Libraries.

3. Quality Control

Data from many different sources in many different formats presents special challenges to integration. Quality control of the data by experienced scientists is essential to assembling a dataset which can be used with confidence. The quality control of WOD01 consists of three stages each of which will be addressed briefly.

3.1 Initial quality control

Initial quality control consists of conversion of data from its original format to a homogenous format, running automated quality control tests and checking against the existing database for duplication. The automated quality control consists of very broad checks such as minimum and maximum ranges for each oceanographic variable, large gradient, inversion, and spike (large positive gradient followed by large negative gradient or vice-versa) checks, density inversion checks, and standard deviation checks. Checks are also conducted on ship speed between stations. The results of the automatic check can help in preserving the integrity of the data and metadata and identify possible systematic quality issues with incoming data. Preserving the integrity of the data and metadata means making sure no measured values are lost or altered and all available information necessary to understand the measurements is retained. Systematic problems can include incorrect units, shifted decimal, incorrect variable identification, incorrect ship identification. The problem may be in either the original data or the conversion process itself. Duplicate checks against the existing data are a very important and time-consuming aspect of the initial quality control. Receiving data from many different sources does result in receipt of data already found in the dataset. Scientific results can be biased by having multiple copies of the same data. Also, the same data from a variety of sources may be slightly or greatly different. For instance, data came to us recently which had been discovered in a box in a storage room. The data were then digitized. Some of the data we already had from a different source. The measured data values were nearly exact in the duplicate data, except the new values of silicate were much larger than the older version. When the submittor of the new data was informed, they found that they had converted the silicate values using the wrong conversion factor, resulting in erroneous values.

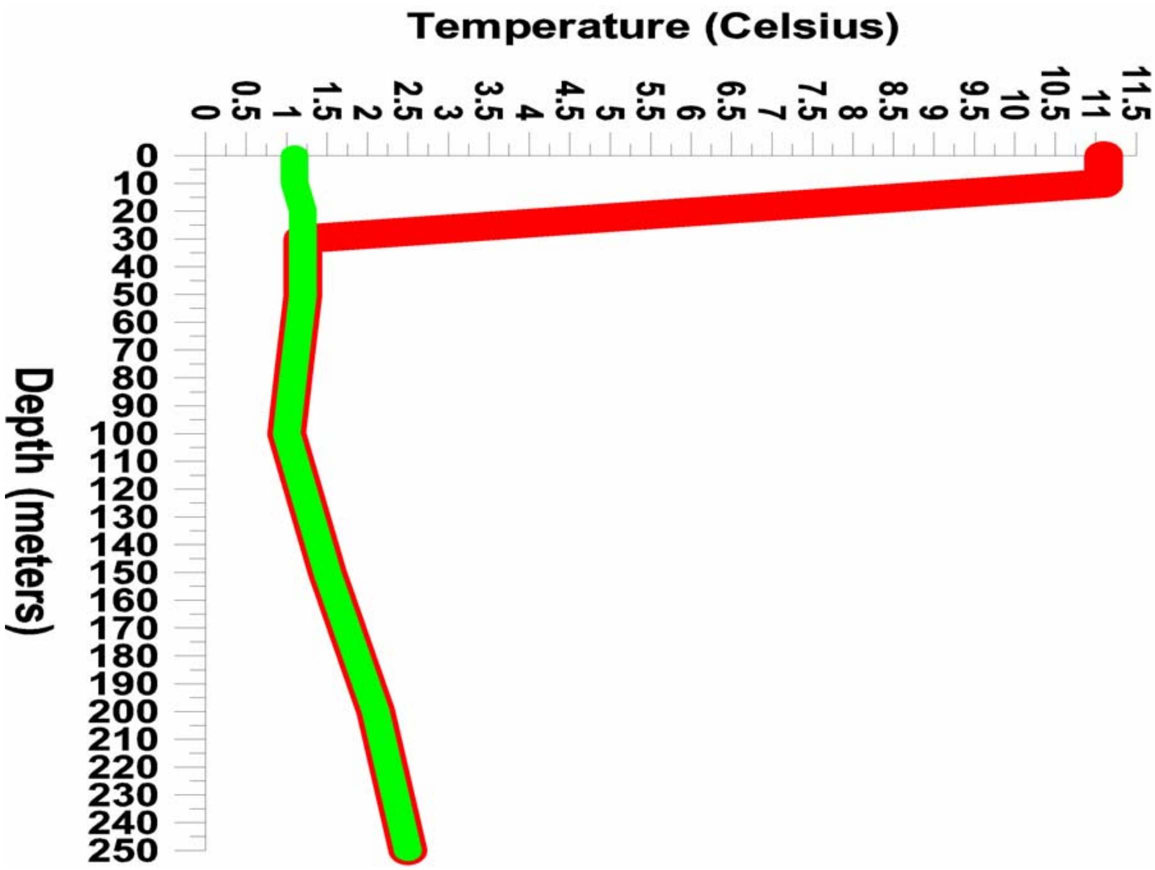

Figure 5. Duplicate temperature profiles with different values for temperature.

Another example is in Figure 5. A temperature profile taken in 1961 was received from the Russian Navy. The data from a Japanese source was already in the database. The latitudes and longitudes were different due to the Russian source actually reflecting degrees and minutes rather than decimal degrees. When the appropriate conversion is made, the latitudes were the same. In addition, the data are different by a factor of 10 for the first two measurements, due to a shift in the decimal. Both values were possible in this area of the ocean, but using the wrong (larger) value for temperature can lead to assumptions about a 10- to 20- meter mixed layer where none exists.

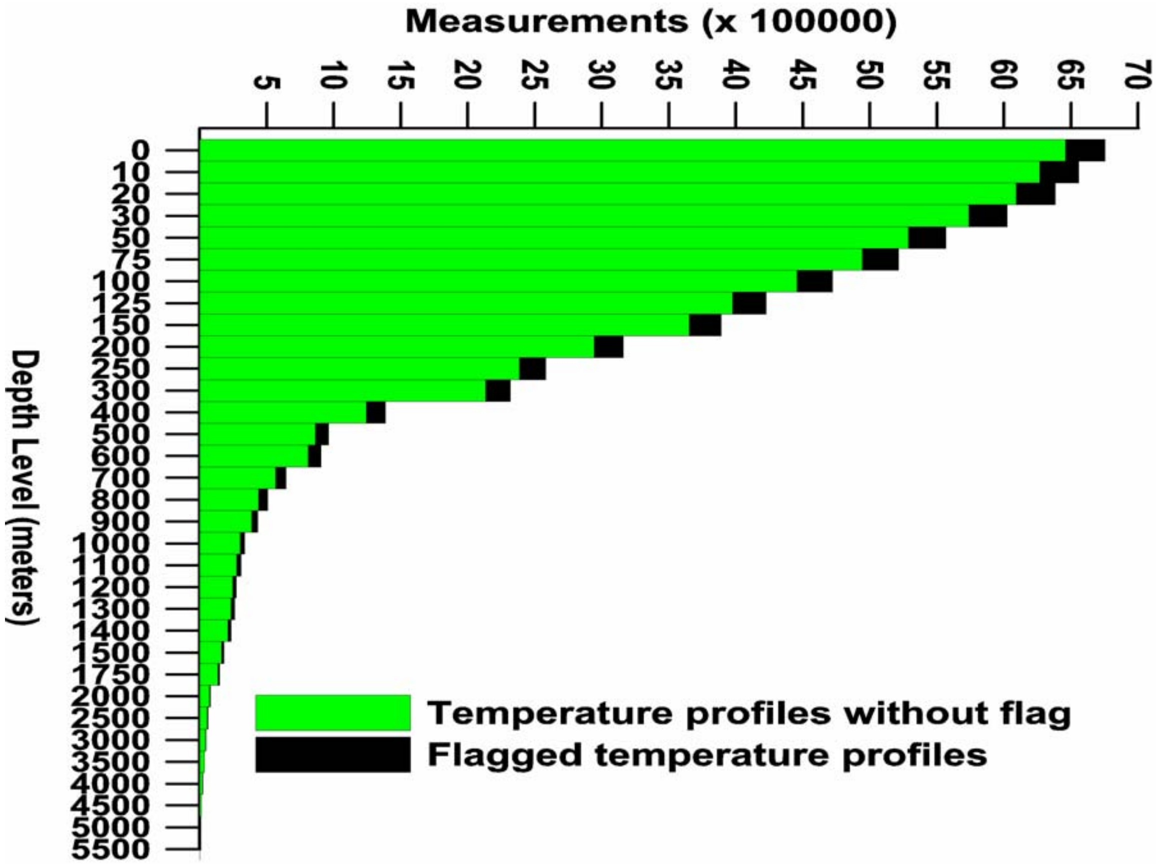

Figure 6. Temperature measurements with and without quality control flags.

Figure 6 shows the results of the automatic checks and initial quality control. Only ~3% of all temperature profiles are flagged completely. But, for the surface that means over 200,000 profiles have been flagged. Further quality control is necessary, but the automatic checks exclude a large subset of data from further consideration.

3.2 Calculation of climatologies

The second stage of quality control is the use of the entire dataset to calculate the mean conditions at 33 standard depths over the entire ocean. This leads to the discovery of problems in the data which are then investigated and corrected when possible, or flagged.

Figure 7a shows climatological mean temperature at 800 meters for the month of January before any quality control, except automatic quality control, is performed. Figure 7b shows the same field after additional quality control is performed.

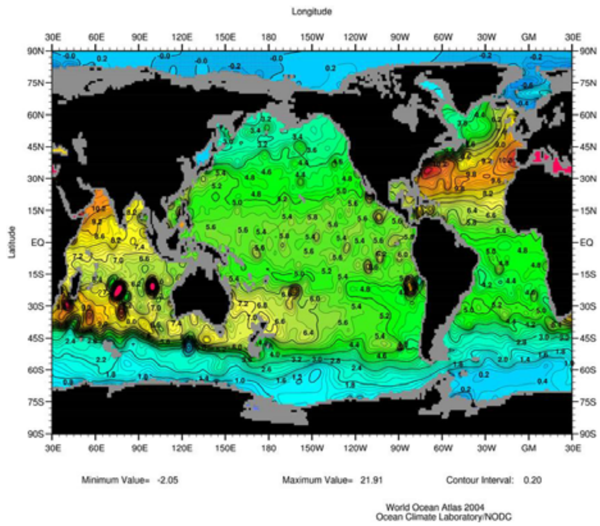

Figure 7a. January mean temperature at 800 m depth (only initial quality control)

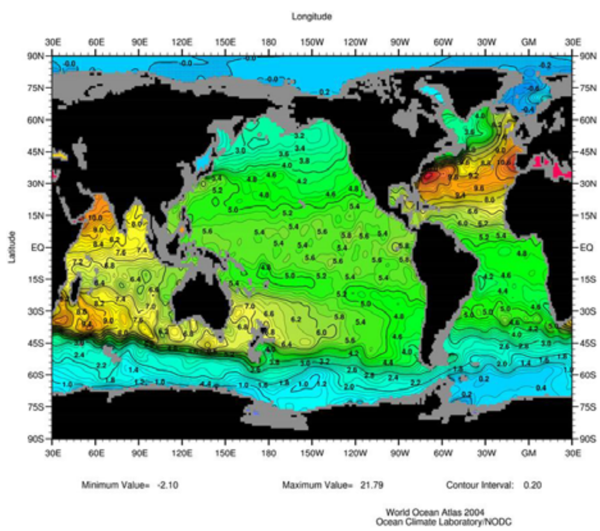

Figure 7b. January mean temperature at 800 m depth (with additional qc)

The second stage of quality control is crucial, because no amount of automatic checks or initial quality control is going to identify all anomalous data. The data responsible for causing the anomalous features in Figure 7a are identified and examined. The original data sets which contained the anomalous data are scrutinized for further error and the appropriate correction or flagging is applied. Numerous problems are identified at this stage. It is very important and not always easy to distinguish between problems in the data and anomalous oceanographic features, such as rings and eddies. After this stage of quality control, the dataset along with the quality control flags and the climatologies are released to the public. The climatologies are used as initial and boundary conditions for models and for checks against satellite and *in situ* data measurements.

3.3 Further Scientific Research

After the release of the WOD01 and WOA01, quality control continues. Comments from outside users are very important to identify problems not previously found. Further internal studies with the WOD01 continue to identify problems.

4. Keeping the User Community Informed

All the quality control for the WOD01 is useless if it is not communicated to the users of the data. The WOD01 was released online in 2001 sorted geographically (http://www.nodc.noaa.gov/OC5/WOD01/data2001.html) and temporally (http://www.nodc.noaa.gov/OC5/WOD01/data01y.html) . It was released in 2002 on compact disc (Conkright et al. 2002). The quality control flags are released with the data, so the user needs to exclude data based on the quality control flags they agree with. This was made alot easier in 2003 with the release of WODselect (http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html). This is an online tool for subsetting WOD01 data by geographic region, date, ship, project, measured variable, etc. It can also subset data by excluding all data with certain user-chosen flags or excluding data with any flag. This way the user can choose which quality control to accept WODselect and the geographically sorted data are updated monthly with the most current corrections and quality control flags. The updates are fully documented at http://www.nodc.noaa.gov/OC5/WOD01/pr_wod01.html, under

WOD01 Updates. If a user is interested only in corrections to position, date/time and measured values, all such updates are available separately under the same WOD01 Updates site.

5. Estimation of decadal change for heat and freshwater from WOD01

With a quality controlled database of historic and recent data, it is now possible to make estimates for heat and freshwater decadal changes from *in situ* measurements. These changes still can only be discussed in very broad terms and only back to around 1955 due to a lack of data for earlier time periods for some geographic areas. First, temperature and salinity anomaly data were binned by five-year periods between 1955 and 1998. Anomalies were calculated by subtracting each measurement from the climatological mean monthly value and averaging over a 1° latitude/longitude box. These anomalies were objectively analyzed. Heat content was calculated from the temperature anomalies. Then the linear trends were calculated for the zonal average for the Atlantic, Pacific, Indian, and World Ocean. The results for the North Atlantic are shown in Figures 8 for heat content and 9 for salinity. When salinity anomalies are converted to freshwater and integrated over the upper 3000 meters (not shown), a measure of freshwater change is achieved. The above results are found in Levitus *et al* (in revision) for heat and Boyer *et al* (submitted) for salinity. Levitus et al (in revision) is an update of Levitus et al. (2000) from an earlier dataset.

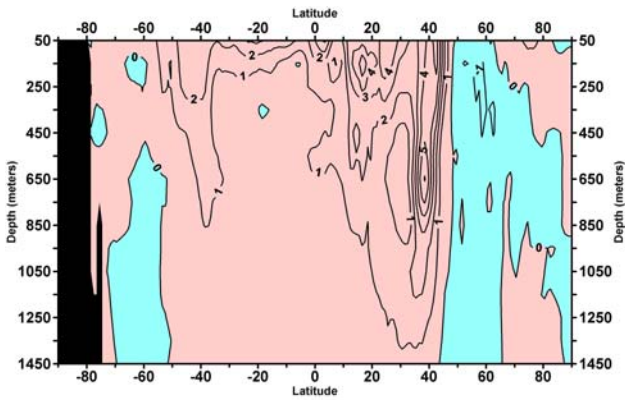

Figure 8. Linear trend in heat content (10^{18} Joules/year) 1955-2003, surface to 1500 m depth for the Atlantic Ocean (from one-year anomaly fields).

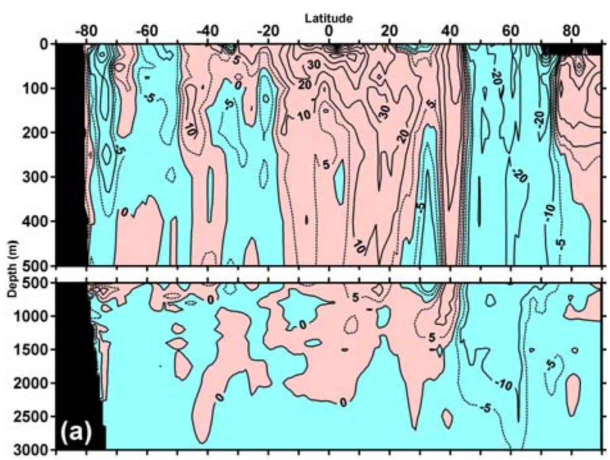

Figure 9. Linear trend in salt content (10^{-4}) 1955-1998, surface to 3000 m depth, for the Atlantic Ocean (from five-year anomaly fields).

6. Conclusions

There is now enough data to begin to understand the decadal changes in heat and freshwater from 1955 using the WOD01. A lot of work went into gathering and quality controlling the data in this dataset. A key component of the quality control was calculation of climatologies and other scientific research using the data in WOD01 as it became available. Quality control is an ongoing process involving the entire scientific community. The continued quality control of WOD01 is communicated to the user community through monthly updates on the internet. There still exists a lot of recent and historic oceanographic profile data which can be incorporated into a future edition of the WOD. The task for the future is to keep up with the increasing amount of data from new instruments such as profiling floats while integrating and improving the quality of historic data.

References

Boyer, T.P., S. Levitus, J.I. Antonov, R.A. Locarnini, and H.E. Garcia (2004), Linear trends in salinity for the World Ocean 1955-1998 (submitted to *Geo. Res. Letters*)

Conkright, M. E., J. I. Antonov, O. K. Baranova, T. P. Boyer, H. E. Garcia, R. Gelfeld, D. Johnson, R. A. Locarnini, P. P. Murphy, T. O'Brien, I. Smolyar, and C. Stephens (2002a), *World Ocean Database 2001, Volume 1: Introduction. NOAA Atlas NESDIS 42, U.S.*, Ed. S. Levitus, 159 pp., CD-ROMs, Government Printing Office, Washington, D.C.

Levitus, S., J. Antonov, T.P. Boyer, and C. Stephens (2000), Warming of the World Ocean, *Science*, *287*, 2225-2229.

Levitus, S., J.I. Antonov, T. P. Boyer (2004), Warming of the World Ocean, 1955-2003

(submitted to *Geo. Res. Letters*)

Rahmstorf, S. (1995), Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle, *Nature*, 378, 145-149

Rossby, C-G (1959), Current Problems in Meteorology, from Atmosphere and Sea in Motion, B. Bolin editor, The Rockefeller Institute Press, New York, NY, 9-50