GTSPP Annual Report for 2002

Prepared by Bob Keeley (keeley@meds-sdmm.dfo-mpo.gc.ca) with contributions by Charles Sun (NODC) and Loic Petit de la Villeon (SISMER).

1.0 Introduction

The Global Temperature Salinity Profile Project (GTSPP) is a joint World Meteorological Organization (WMO), and Intergovernmental Oceanographic Commission (IOC) project. Functionally, GTSPP reports to the Joint Commission on Oceanography and Marine Meteorology (JCOMM), a body sponsored by WMO and IOC and to the IOC Intergovernmental Oceanographic Data and Information Exchange committee.

Development of the GTSPP (then called the Global Temperature-Salinity Pilot Project) began in 1989. The short-term goal was to respond to the needs of the Tropical Ocean and Global Atmosphere (TOGA) Experiment and the World Ocean Circulation Experiment (WOCE) for temperature and salinity data. The longer term goal was to develop and implement an end to end data management system for temperature and salinity data, which could serve as a model for future oceanographic data management systems.

GTSPP went into operation in November 1990. The first version of the GTSPP Project Plan was published in the same year. Since that time, there have been many developments and some changes in direction including a decision by IOC/WMO to end the pilot phase and implement GTSPP as a permanent project.

GTSPP played a key role in the WOCE Upper Ocean Thermal Data Assembly Centre and contributed to the final WOCE Data Resource DVD Version 3. GTSPP is also an accepted part of the GOOS and recently has been accepted for participation in CLIVAR.

Many nations contribute data to the GTSPP and without their contributions the project could not exist. Contributions to the data management portion of GTSPP are provided by Australia, Canada, France, Germany, Japan and the USA. Scientists and data managers in these countries contribute their time and resources to ensure the continuing functioning of the project.

2.0 Objectives

The objectives of the GTSPP are as follows.

1. To provide a timely and complete data and information base of ocean temperature and salinity profile data of known and documented quality in support of global and local research programmes, national and international operational oceanography, and of other national requirements.

2. To improve data capture, data analysis, and exchange systems for temperature and salinity profile data by encouraging more participation by member states, by locating new sources of data from existing and new instruments and implementing the systems to capture and deliver the data, by taking full advantage of new computer and communications technologies, and by developing new services and products to enhance the usefulness of the GTSPP to clients and member states.

3. To develop and implement data flow monitoring systems to improve the capture and timeliness of GTSPP real time and delayed mode data, and to distribute information on the timeliness and completeness of GTSPP data bases so that bottlenecks in the data flow can

be identified and addressed.

4. To improve the state of databases of oceanographic temperature and salinity profile data by developing and applying improved quality control systems, by implementing new data centre tests for QC as appropriate for new instrumentation; by working with the scientific partners of GTSPP to train data centre staff and transfer scientific QC methods to the centre, and by feeding information on recurring errors to data collectors and submitters so that problems can be corrected at the source.

5. To facilitate the development and provision of a wide variety of useful data analyses, data and information products, and data sets to the GTSPP community of research, engineering, and operational clients.

3.0 GTSPP Operations

Figure 1 presents the data flows of national and international programmes within which GTSPP is placed. The boxes in the Figure represent generic centres. A given international JCOMM or IODE centre may fit within several boxes in carrying out its national and international responsibilities. The following sections discuss this figure in terms of essential elements of the GTSPP.

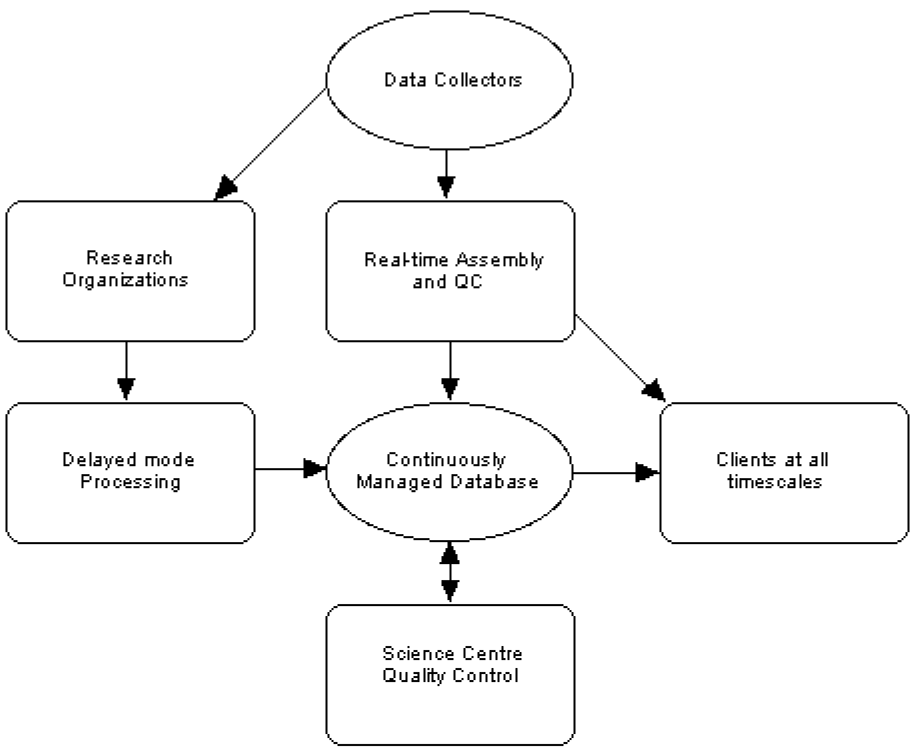


Figure 1: GTSPP data flow

3.1 Near Real Time and Operational Time Frame Data Acquisition

Near real time data acquisition within GTSPP depends on the GTS of the World Weather Watch of WMO and the telecommunications arrangements for BATHY and TESAC data established by JCOMM. Copies of other real time or operational time frame data sets are acquired from any other available sources via the Internet of other high-speed networks. The goal is to ensure that the most complete operational time frame data set is captured.

Figure 2 is a graphic representation of the GTSPP operational time frame data flow. The "data collectors" in the top boxes follow one of two procedures. In the first case the data are provided to GTS centres that place them on the GTS within minutes to days of its collection. In the second case the data are supplied to a national organization that forwards it to the real time centre in MEDS within a few days to a month of its collection.

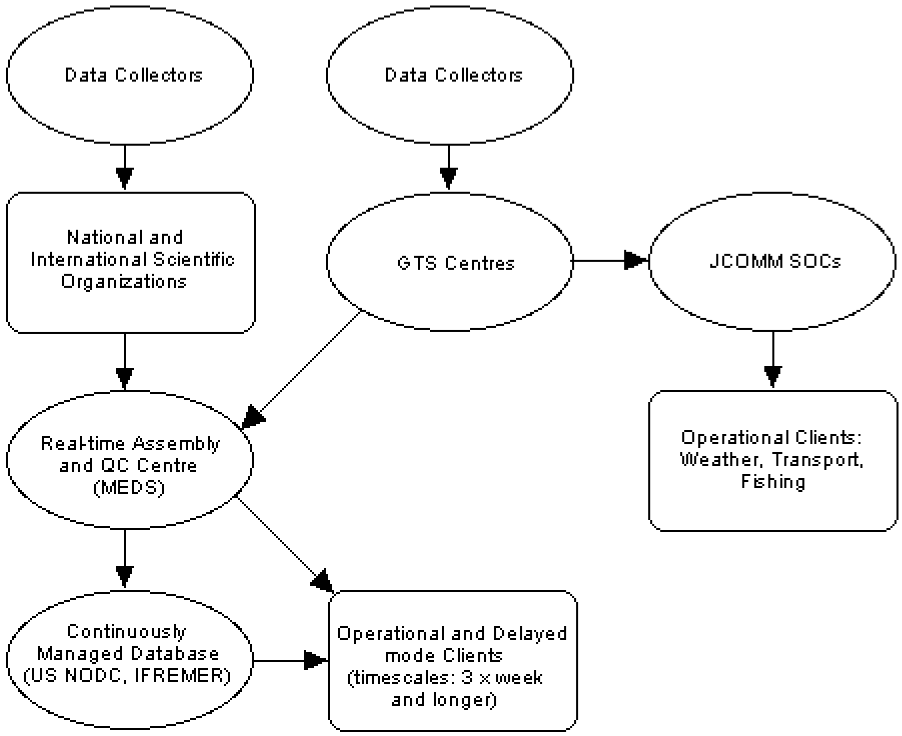


Figure 2: Real-time data flow

The real time data that are circulated on the GTS are acquired by MEDS and the Specialized Oceanographic Centres (SOCs) of JCOMM and by users of real time data who have access to the GTS. These users include meteorological and oceanographic centres that issue forecasts and warnings, centres that provide ship routing services, and centres that prepare real time products for the fishing industry.

MEDS compiles the global data set from the various sources, applies the documented GTSPP QC and duplicates removal procedures, and forwards the data to the US NODC three times per week. At NODC the data are added to the continuously managed database (CMD), on the same schedule. There are also several clients that receive copies of the data sent from MEDS three times per week. These are clients who do not need the data within hours but rather within a few days. By getting the data from the GTSPP Centre in MEDS they save having to operate computer systems to do quality control and duplicate removal.

The regular route for real-time data to the box marked "Operational Clients" in Figures 1 and 2 is not affected by GTSPP. This route provides for uninterrupted flow of data for weather and operational forecasting through the national weather services of member states. These centres need the data in hours rather than days.

3.2 Delayed Mode Data Acquisition

GTSPP utilizes, to the extent possible, the existing IODE data network and processing system to acquire and process delayed mode data. The box entitled "Delayed Mode & Historical Data" in figure 3 shows the delayed mode data flow in graphic form. The data flow into the continuously managed database is through a "Delayed Mode QC" process. This process is analogous to the QC carried out on the real-time data and conforms to the specifications of the GTSPP QC Manual. In some cases, where appropriate arrangements can be made this QC process exists and is performed in another national oceanographic data centre on behalf of NODC.

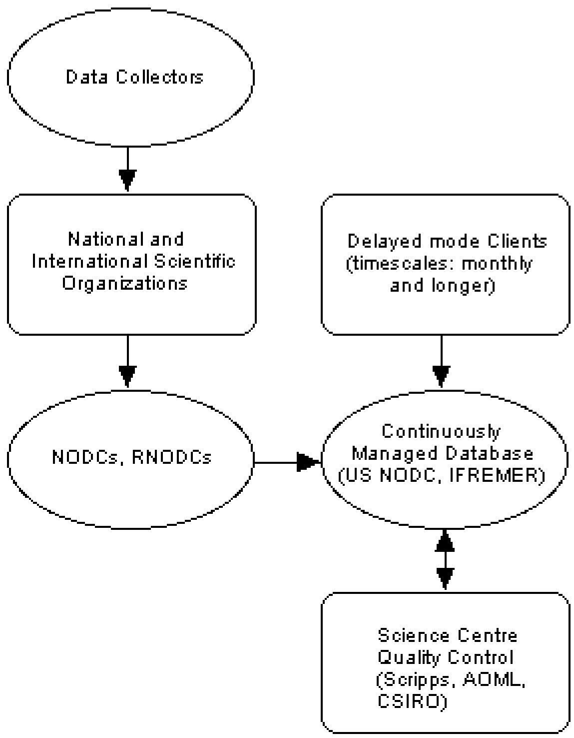


Figure 3: Delayed mode data flow

Having proceeded through the delayed mode QC process, the data then follow the same route as the real time data through the rest of the CMD process, however, on a different time schedule because of the more irregular times of arrival. During the merging of the data into the CMD, any duplicates occurring between near real-time and delayed mode data sources are identified with the highest resolution copy being retained as the active CMD version.

Acquisition of delayed mode data from the Principal Investigators is a priority for the GTSPP. The goal is to get the delayed mode data into the CMD within one year or less of its collection. An excellent way for any national oceanographic data centre to support GTSPP actively is to obtain national data sets of temperature and salinity data, apply GTSPP QC procedures, and submit them to the CMD.

4.0 Progress to the end of 2002

The purpose of this section is to report on the performance of the GTSPP in meeting its objectives to the end of 2002.

4.1 Data Volumes

The GTSPP handles all real-time and delayed mode profile data with temperature and salinity measured. Real-time data in GTSPP are acquired from the Global Telecommunications System in the BATHY and TESAC codes forms supported by the WMO. Delayed mode data are contributed directly by member states of IOC.

The delivery of ocean data in real-time was initiated many years ago and administered by the IOC program called IGOSS. In 2001 operational oceanography programs of IOC and marine meteorological programs of the WMO were merged under the JCOMM. Under IGOSS, “real-time” was defined to allow data up to 30 days after collection to be included. This definition has persisted, even though, the trend is to shorten considerably the delays between observation and distribution.

In JCOMM, the BATHY and TESAC code forms are the ones used most often for distribution of ocean profile data on the GTS. Figure 4 shows the evolution of the use of these codes to make ocean data available. The dramatic change in mid 1999 shows the initiation of the Argo Project and the beginning of the use of TESAC to report profiles from robotic profiling floats. A review of the SOOP program in 1999 recommended a switch from broadcast sampling to line mode sampling. In principle, it was hoped that as many XBTs (exclusively reported using the BATHY code form) would be deployed along lines as formerly were deployed in broadcast mode. It is evident from the figure that the number of BATHY reports has declined since 1999.

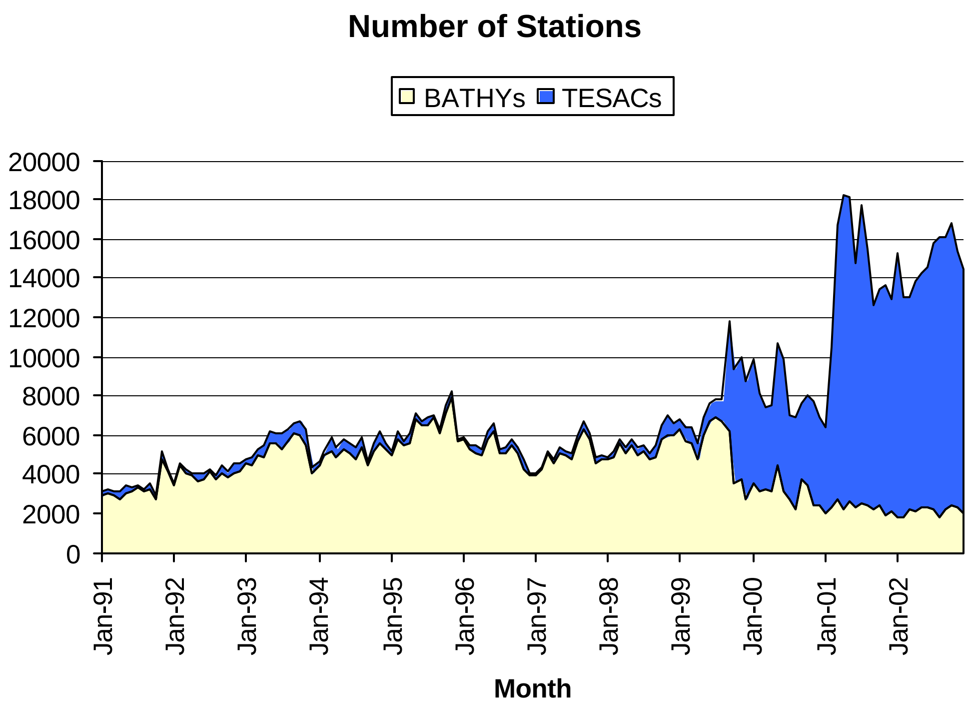


Figure 4: The number of stations reported as BATHYs and TESACs.

The next figure shows the kinds of instruments contributing data in delayed mode to the CMD. We can consider the delayed mode data in most cases to be of higher vertical resolution and higher precision in measurements. We have subdivided them into a few different types and looked at the number of stations of each type by year. Evidently, the majority of data are from XBTs and so have only temperature profiles. It is also evident that the volume of delayed mode data falls the closer we approach to the present. This reflects the time delays built in to higher resolution data arriving at archive centres. Note that later on, we show the balance of real-time to delayed mode data in the CMD. In some cases, real-time and delayed mode data have no difference in vertical resolution (such as for the presently operating profiling floats). We should also note that there are only a very few delayed mode profiles from profiling floats. The main reason for this is because that part of the Argo data system is still being developed.

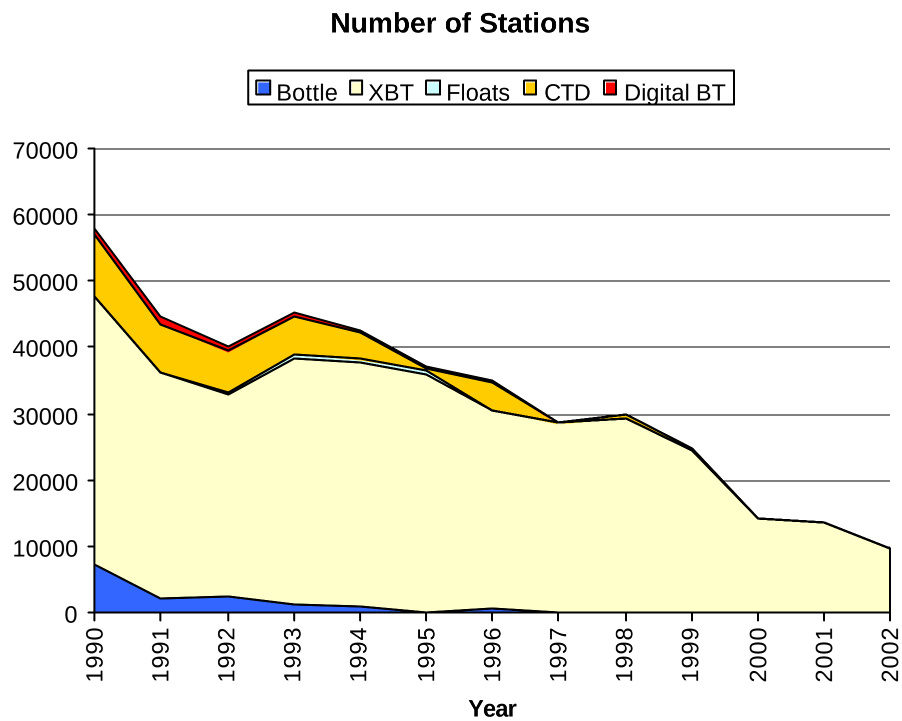


Figure 5: The number of delayed mode stations by instrument type in the CMD.

4.2 Completeness of delivery

When the GTSPP first began, it was suspected that data circulating on the GTS were lost at one or more points in the system. To test for this and to recover what might be lost, arrangements were made to have all BATHY and TESAC data gathered from the GTS at different sites and to send the data to MEDS separately from the GTS distribution. Three countries (four sites) volunteered in this effort.

In combining the data from these different sources, MEDS has to deal with the high level of duplication. It does so by examining data that are sampled within 5 minutes or 15 km of each other. An examination of the recorded values in the profiles is used to determine if a duplication exists or not. If a duplicate is found, only one of the profiles is retained. The selection of the data is based on a priority list of the sources. Figure 6 shows the numbers of stations by source that reside in the real-time archives.

If all was working well, each of the contributors would receive exactly the same data and the figure would be all light yellow (indicating MEDS received the same data as everyone else). It is clear that this is not the case. It is also clear that there have been improvements from the beginning of GTSPP, although we still find a small fraction of data appearing in data files provided from other sources that do not reach MEDS on the GTS. Some of the differences seen in more recent months stem from problems MEDS has had in its connections to the GTS. We have changed our connection and removed that problem.

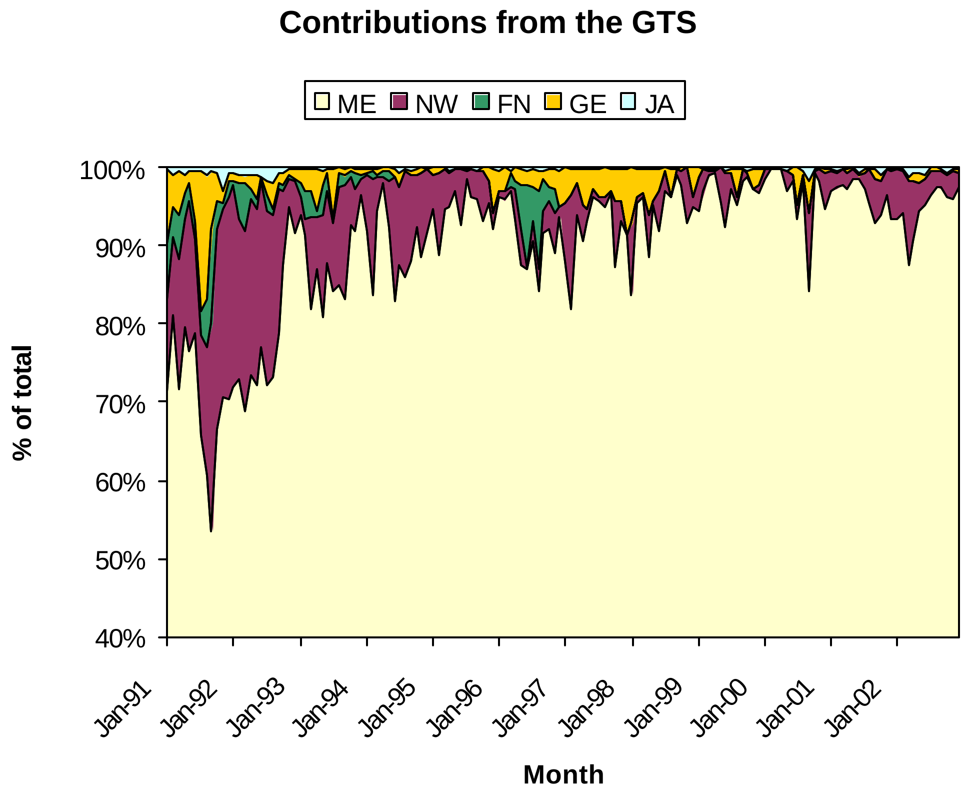


Figure 6: Contributions to the real-time archive by GTS source.

The figures below show the evolution of code forms used to report data on the GTS.

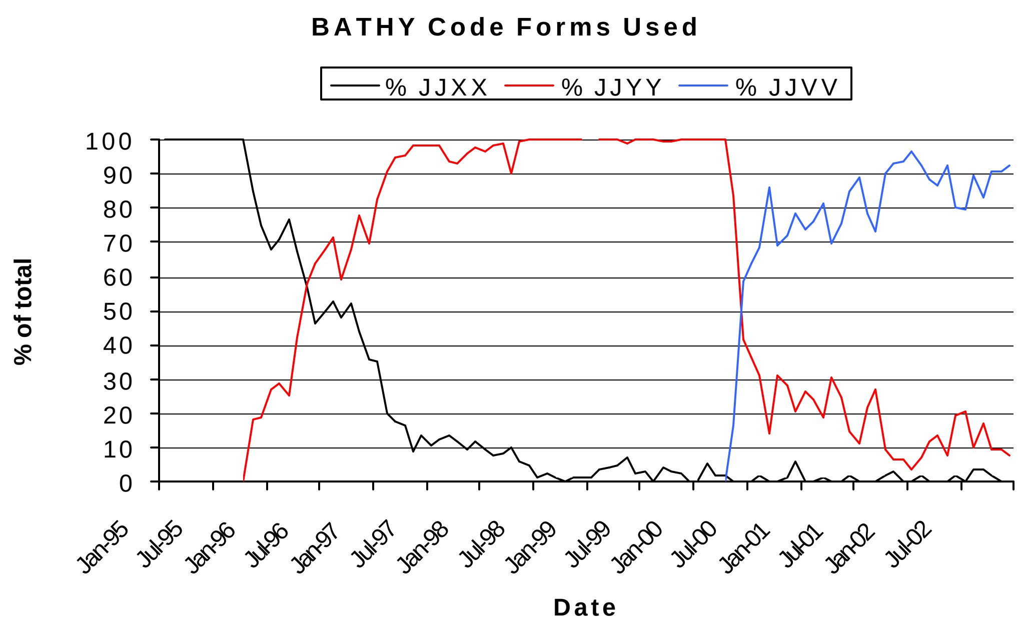


Figure 7: The percentage of total messages received using different code forms for BATHY data.

Figure 7 shows that over the course of operations of the GTSPP, there have been three versions of the BATHY code form used. We can see that the transition to the latest form, JJVV, was dramatic at first but some 18 months later still only about 90% complete.

The situation for TESACs (shown in figure 8) is much better. The main reason for this is that much of the data reported in TESAC are generated from automated platforms. The software is usually operating at some central location on shore (rather than distributed on ships as is the typical case for BATHYs). So, if a change needs to be made to conform to a new code form, it is a relatively simple matter to do so at a few locations and to begin to use the new form quickly.

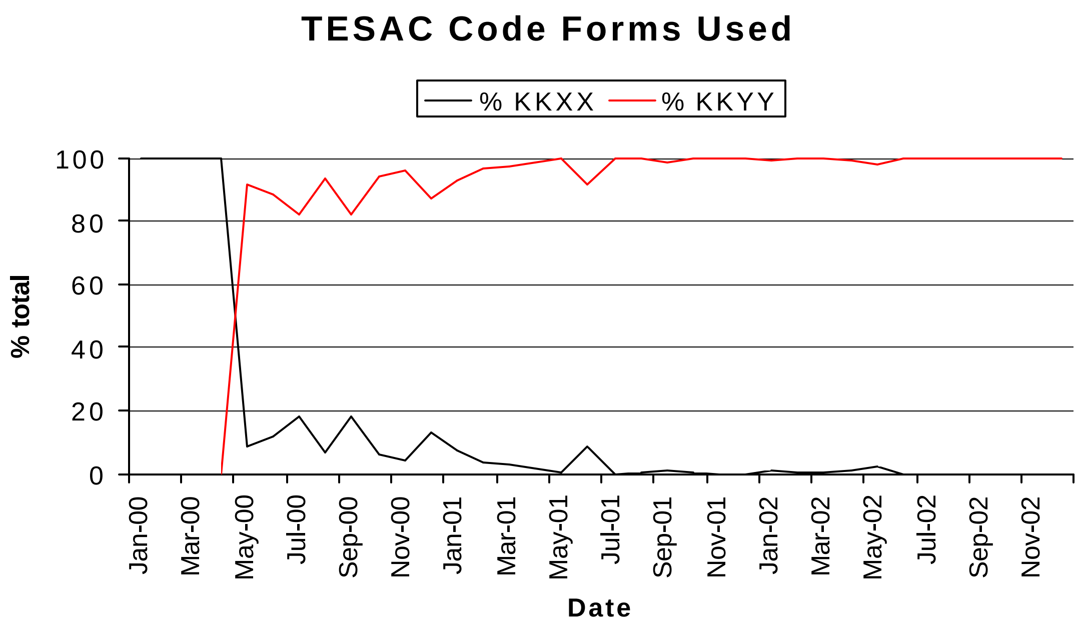


Figure 8: The percentage of total messages received using different code forms for TESAC data.

The next figure shows the relative proportion of real-time to delayed mode data present in the CMD. There are a number of things to take note of in this figure. First, it is evident that in only a few years have the delayed mode data arrived to replace the real-time even many years after the data collection. This tells us that even though we can look at time lags of delayed data coming to the CMD, figure 9 tells us that we still have a significant number of high resolution stations to recover. This assumes that we are able to match the real-time data to the delayed mode profiles as they arrive. This capability is not perfect and is something that is touched upon as part of ongoing work reported later.

Second, as expected, as we move towards the present, the number of stations of delayed mode data decreases and the number of real-time increases as a proportion of the total number of stations. This, too, is typical in that it can take years for delayed mode data to reach the archives. It is precisely because of these delays that GTSPP was started and to provides the combination of real-time and delayed mode data to any user when they request the data.

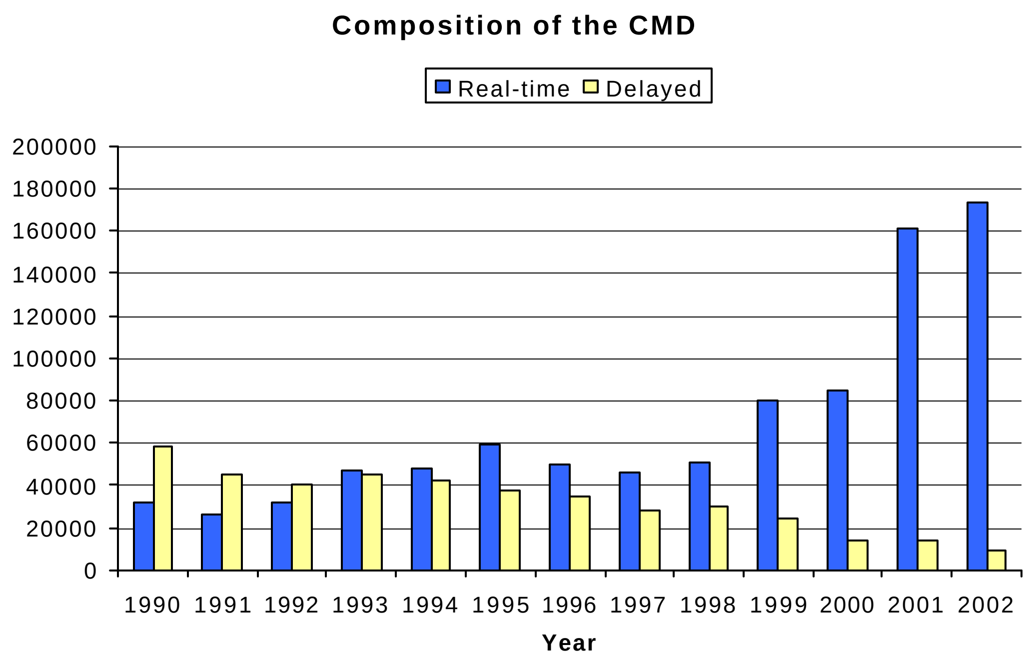


Figure 9: The volume of data in the CMD.

Finally, the graph shows spectacular growth in the number of real-time station from about 2000 to the present. Much of this is a direct result of the start of the Argo program. Note that in this case both temperature and salinity profiles are provided from 2000m to the surface. The vertical resolution varies with a typical profile having approximately 100 levels. This is all that will ever be returned from the floats and so the only difference between delayed mode and real-time profiles is in increased precision of the measurements and better quality control of the data. As of mid 2003, Argo is only about 25% of the way to its final array size. It is clear that it will be the dominant source of ocean profile data.

4.3 Timeliness of data

The management of data within the GTSPP is organized around the idea of a Continuously Managed Database. Clients of the CMD can receive data at any time and that are of the highest quality, and highest resolution available at the time of the request. Typically, the real-time data arrive first, and so become available first. As the delayed mode data arrive later, they replace the real-time data.

A variety of platforms report data. Each of these platforms has different systems by which data get ashore and to the GTS. While it is possible to look at the timeliness of reports as a function of the variety of platform types and instruments, it is more instructive to look at platform types that to some extent represent the extremes in timeliness. To this end, data arriving from ships can be considered the least automated (and so likely the slowest to arrive). At the other end are those data coming from automated platforms, of which we can take Argo as an example.

It is also possible to look at the time to get data to the GTS as well as the time to make data available from the CMD. The GTSPP goal is to make data available as rapidly as possible and so it is the time to make data available that is the more important. Consequently, we will show the difference between observation time and update time (equivalent to data being passed to GTSPP clients).

Another consideration is that the real-time collection and distribution of ocean profile data continues to operate on the principle that real-time is defined as any data up to 30 days after collection date. Thus, some contributors use ships to collect data, return back to their home port and then deliver data to the GTS to still make the 30 day cutoff. Although the trend these days is to move to more rapid data dissemination, those operating under these older principles still contribute to the data flow and this will impact the timeliness statistics.

The information shown in figure 10 shows that during the first years of GTSPP, roughly 10% of the data were available in the CMD 1 day after data collection. In the last year shown, 2002, this has jumped to about 40%. This is a very rapid change and much of it reflects changes in automation in data gathering and transmission. What is not evident here is that the data that are available from the CMD has undergone complete Data Centre quality control including visual inspection of every profile. More will be said about this later on in the report.

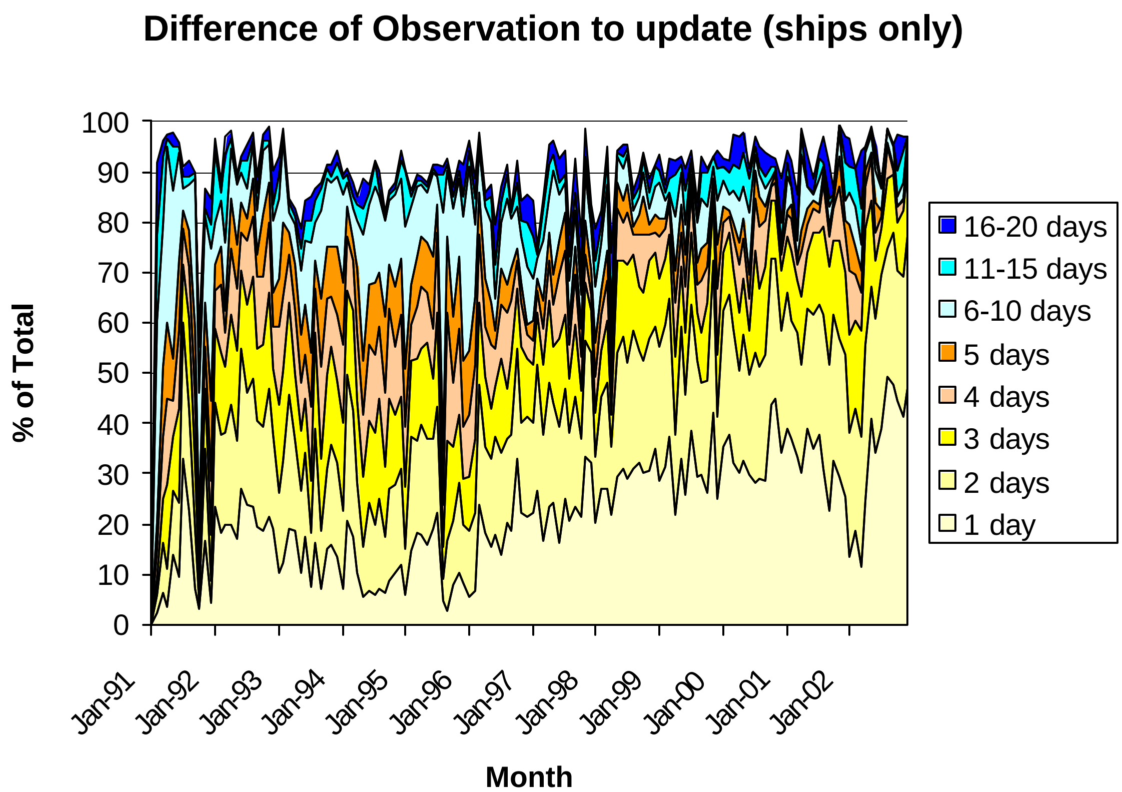


Figure 10: The time difference between observation and update to the CMD. This is generated from BATHYs only and only data reported from ships.

Figure 11 shows the same kind of display but now for profiling float data (the Argo project) reported as TESACs. Again, the time difference is measured between the observation time and the update time. Argo began earlier than 2000, but it was only at this time that a substantial number of floats began to be deployed. The Argo project has the stated goal to report all data to the GTS within 1 day of observation. As of the end of 2002 they were hovering about the 55% mark. The data management component of Argo is investigating improvements that can be made to substantially improve this figure.

In the case of Argo, fully automated QC procedures are carried out on the data prior to submission to the GTS. Some delays are experienced when profiles fail the automated procedures and manual intervention is required. Other delays are introduced when data are corrupted during transmission and must be recovered manually.

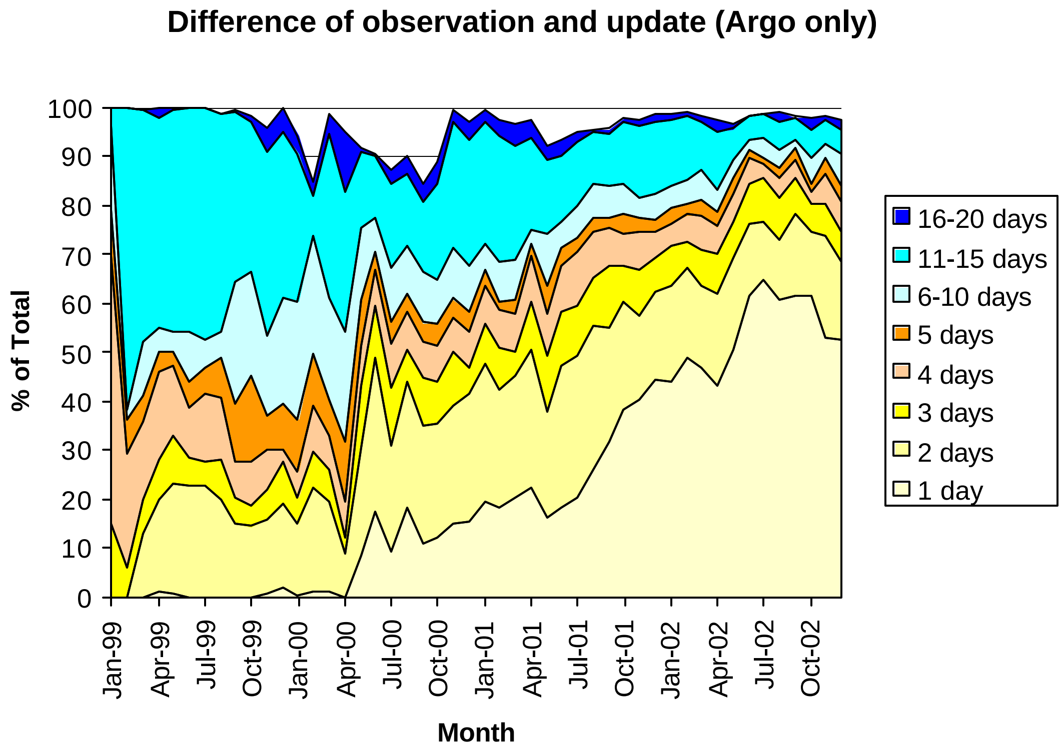


Figure 11: The time difference between observation and update to the CMD. This is generated from TESACs only and only data reported from profiling floats

Dealing with the same issue with delayed mode data is more difficult. We know that data can be at most 30 days old (or so) for real-time distribution. Any data older than this just does not get distributed. This makes for a clean cut off time and more importantly a clear upper limit to the volume of data expected.

For delayed mode data, the oldest date could be back to the time of the Challenger Expedition in 1873. As well, there is no known limit to the volume of data that may be received in delayed mode. Both of these make for imprecise limits against which we can measure the success of receiving delayed mode data.

Figure 12 shows statistics derived from the delayed mode data in the GTSPP archives. The number of delayed mode data decrease from past to present consistent with what is shown in figure 9. We also see in the early years of GTSPP, it was very common for data older than 5 years to be received by the project. In the mid to late 1990s, the major fraction of the data are received when they are 2 to 3 years old. In the more recent years, the delayed mode data that have arrived tend to do so within 1-2 years of their collection date.

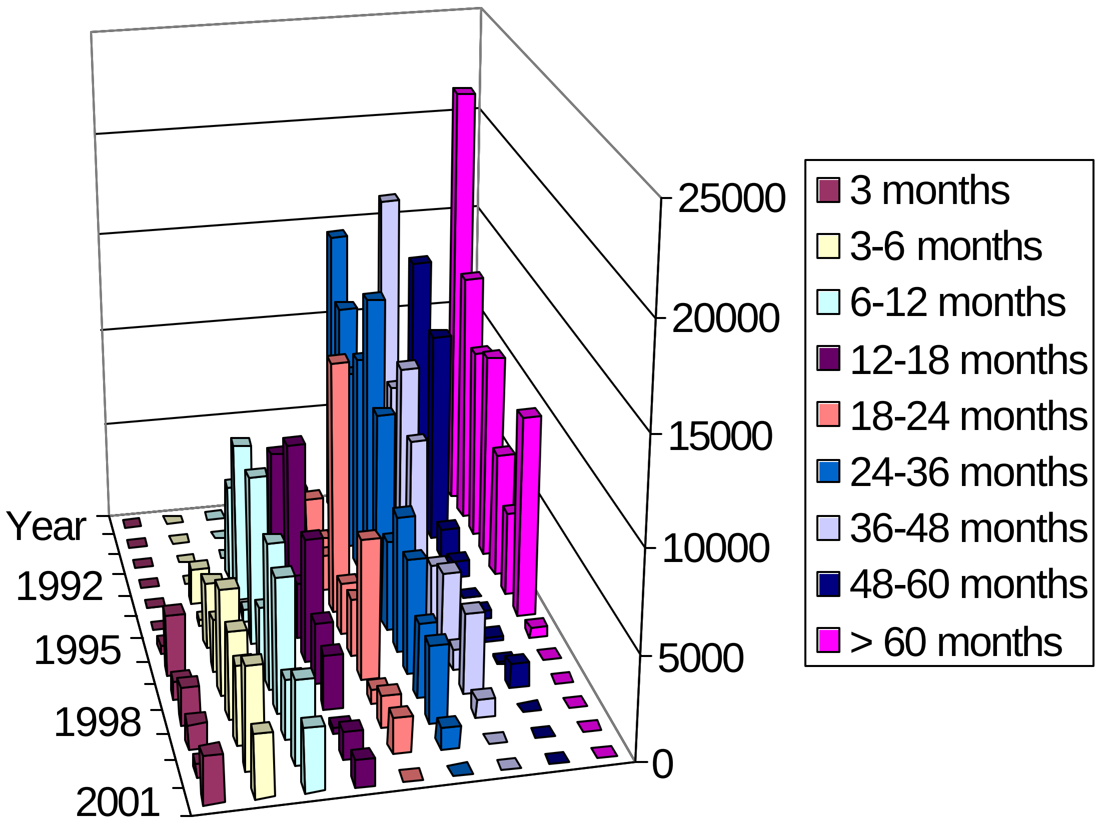


Figure 12: Timeliness of delayed mode data received at the CMD of GTSPP.

4.4 Data Quality

From the start, the GTSPP agreed to standardize the quality control procedures that were used and in ways that the quality information would be managed with the data. Within the GTSPP are both data centres and centres of oceanographic scientific expertise. Data centre QC is described in IOC Manuals and Guides #22 and is available on-line at http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/gtspp/qcmans/MG22/guide22_e.htm

Scientific quality control is provided by collaborating science centres. CSIRO, has produced a manual describing how to manage XBT data. It is available at http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/gtspp/qcmans/CSIRO/csiro_e.htm

In 1995, an intercomparison was done between the data center and science centre QC and a report may be found at http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/WOCE/WOCE_UOT/qcinterc_e.htm

All of the data resident in the CMD eventually passes through these two levels of scrutiny. The following figure shows the contents of the CMD where the relative volumes of data having gone through data centre QC and complete QC (science centre review) are shown.

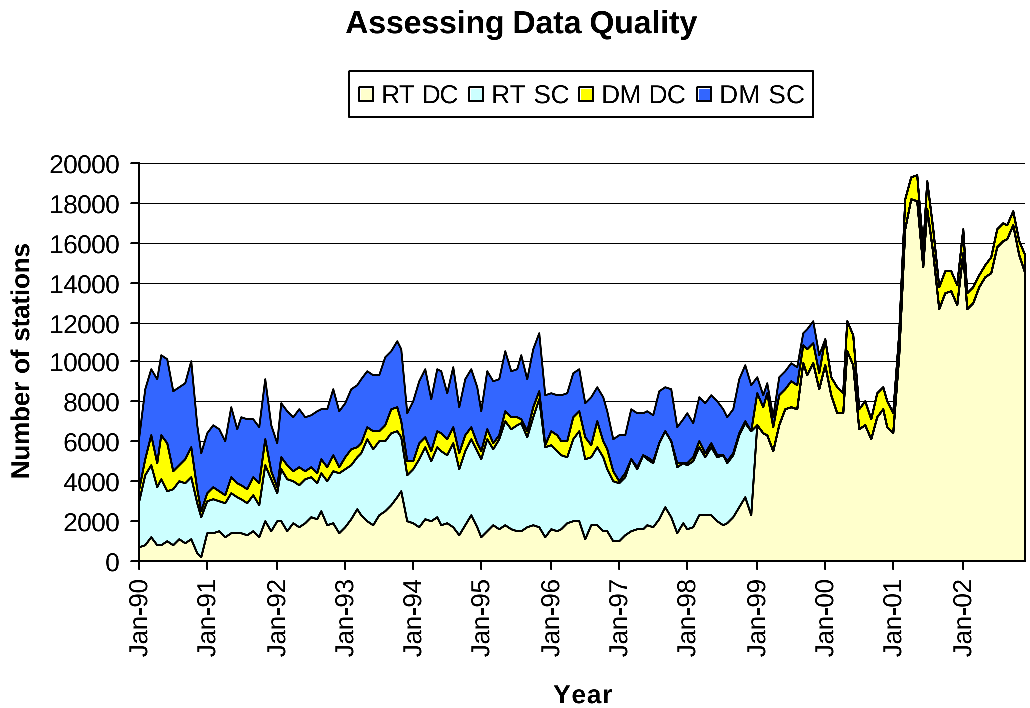


Figure 13: The numbers of real-time (RT) and delayed mode (DM) stations in the CMD having undergone quality control procedures at data centres (DC) and science centres (SC).

The review of data by science centres happens on a yearly basis, and there is always some fraction of data that escape this process. The large jump in the numbers of stations having passed just through data centre QC in January 1999 reflects the deadlines to meet requirements for publishing the WOCE Data Resource V3. We continue to pursue getting the data through science centre QC.

Because some users are interested in the relatively quick availability of real-time data, we can also show an analysis of the results of the data centre QC process (figures 14, 15, 16). Note that flag 3 means data are suspect, flag 4 means the data are considered wrong, and flag 5 means the original value received was changed to make it consistent with other data received from the same platform.

Figure 14 displays the percentage of the total number of stations (both BATHYs and TESACs) where some problems were found in the position. There has been some improvement over time with certain months having unusually large numbers of problems. Note that many of the position problems have been corrected. This is only done when it is possible to know the reason for the errors, or if by an examination of the problem station in the context of neighbouring stations from the same platform, we are confident of the change.

As can be seen, in most months, the number of stations affected are <1%. This reduction is largely the consequence of the rapid rise in use of TESACs resulting from the Argo program. Much of the Argo data (the goal is that all will) receives automatic quality control procedures before the data are inserted on the GTS. Because of this, the most serious errors are mostly eliminated from GTS distribution. This combines with the fact that in any month now there are upwards of 700 floats operating and returning about 2100 messages. This exceeds the number of BATHYs currently reporting.

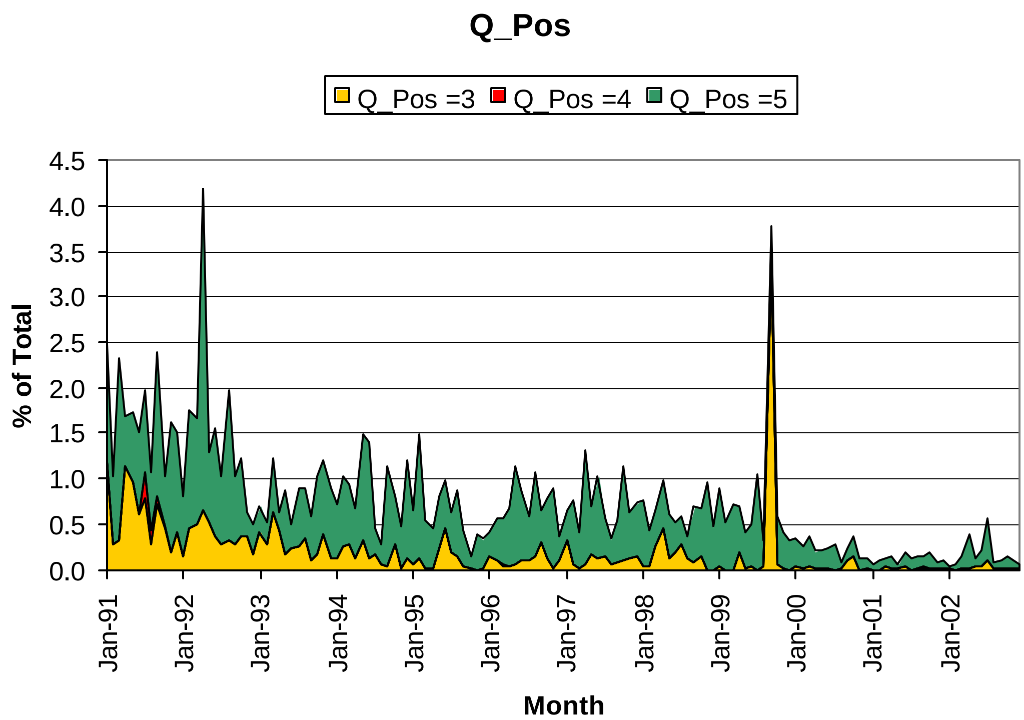


Figure 14: Percent of real-time stations with positions that had some identified problem.

The following figure (15) is a similar display to figure 14 but for the time of the station.

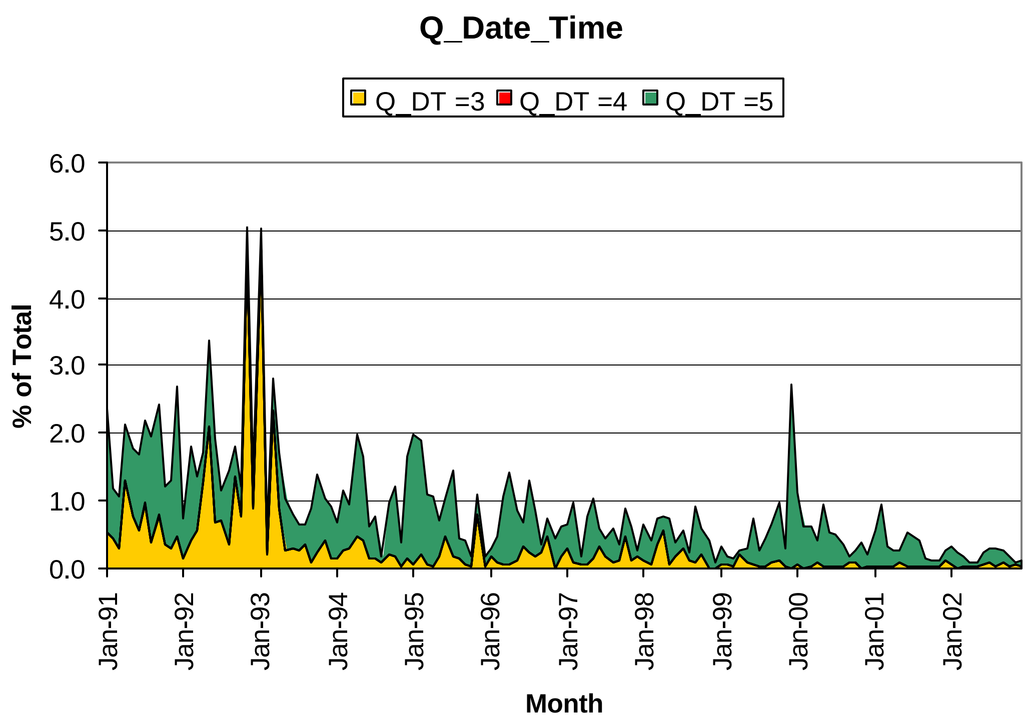


Figure 15: Percent of real-time stations with problems detected in the date or time.

In figure 15, we again see some improvements over the course of the GTSPP operations. Just as for positions, there are certain times when problems in recorded dates or times are more pronounced. Often these times are associated with the end of a year and in these cases are easily corrected. Again, the typical error rate is on the order of 1 or 2% of the total stations. In more recent months, the characteristics of the Argo data are starting to dominate the statistics. This is reflected in the steady reduction in time errors seen. In the most recent months all of the corrections noted have been for BATHY reports exclusively.

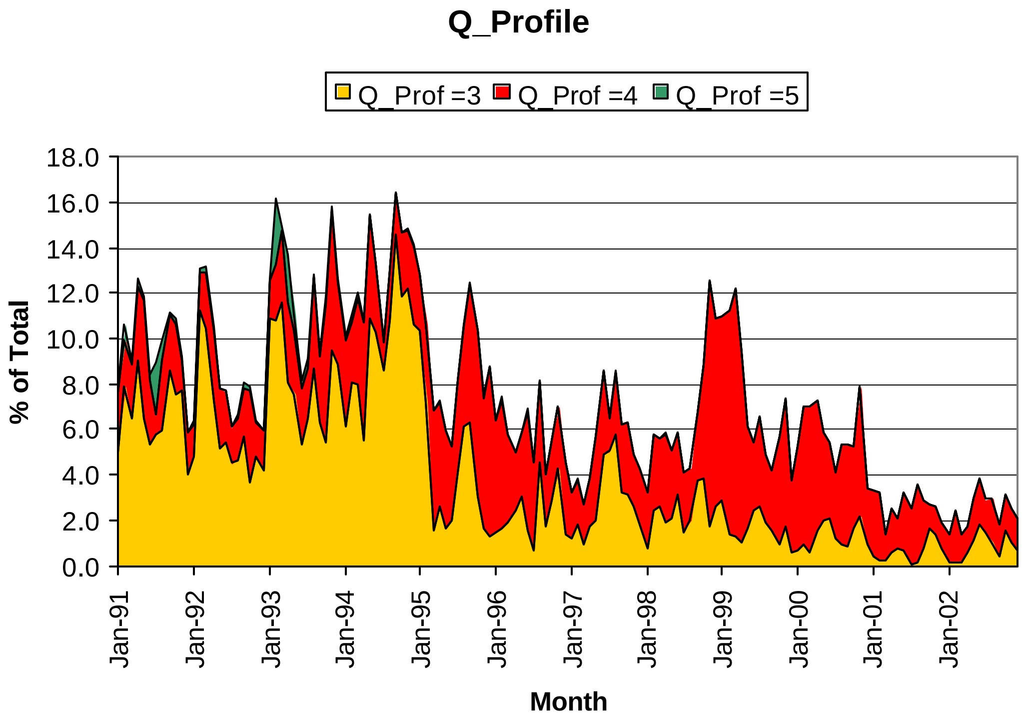


Figure 16: Percent of real-time profiles with a problem noted at one or more depths.

Figure 16 shows the rate of errors occurring in the BATHY and TESAC profiles themselves. A station is counted and shown if even one value in the profile appears to have a problem. There is an improvement over the course of the GTSPP with a significant change in 1995 when the incidence of flag 3 was substantially reduced. In late 1993, the GTSPP started to issue to operators a monthly report of problems seen in BATHYs and TESACs. At this time, BATHY reports dominated the statistics. It is tempting to interpret the reduction in errors as an impact of reporting errors back to operators. The delay between the introduction of the report and the fall in errors could be a result of the delays inherent in ship greeting activities and corrective steps being taken.

In more recent years, there has been a more or less steady decline in errors with another significant reduction noted about 2001. There is little doubt that this is a consequence of the number of profiles from the Argo program starting to dominate the statistics, and the automated quality control procedures reducing the number of erroneous values being reported to the GTS.

In looking at delayed mode data that have arrived at the CMD, we can generate similar charts as for real-time. Looking at the error rates on position (figure 17), we see they are typically about the 1.5 percent mark which is about the same as for real-time data. There are a few occasions where higher than normal rates of errors occur and these do seem to occur more often than for the real-time data. Just as for the real-time data, though, many of the errors in positions are readily correctable. Contrary to what is seen in the real-time data, there does not appear to be any systematic reduction in the rates of position errors.

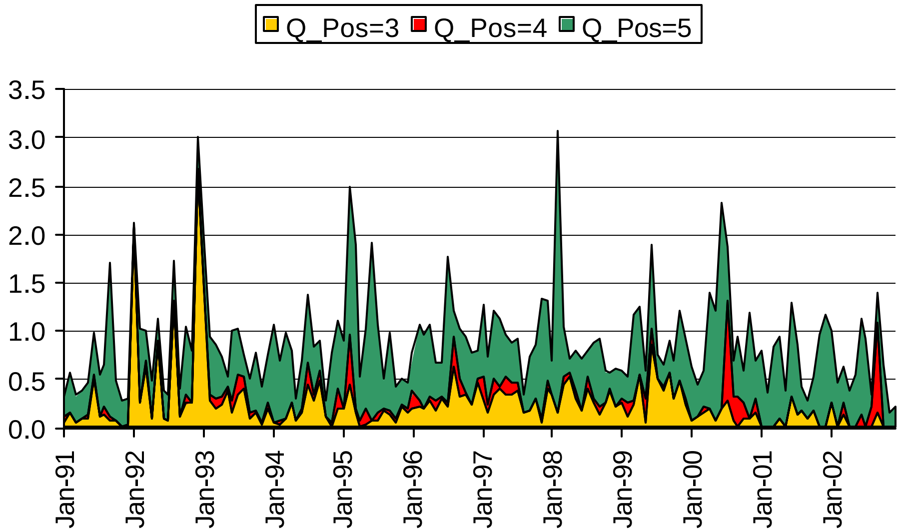


Figure 17: Percent of delayed mode stations with problems detected in the position.

Looking at the error rates in date and time (figure 18) for delayed mode data, we see typical error rates to be on the order of a couple of percent which is quite similar to the rates seen in real-time data. We see a peculiar spike around the middle of 1996, for which there is no explanation at present. Just as for the delayed mode position errors, and contrary to the error rates for real-time data, we see no diminishing in error rates on date and time.

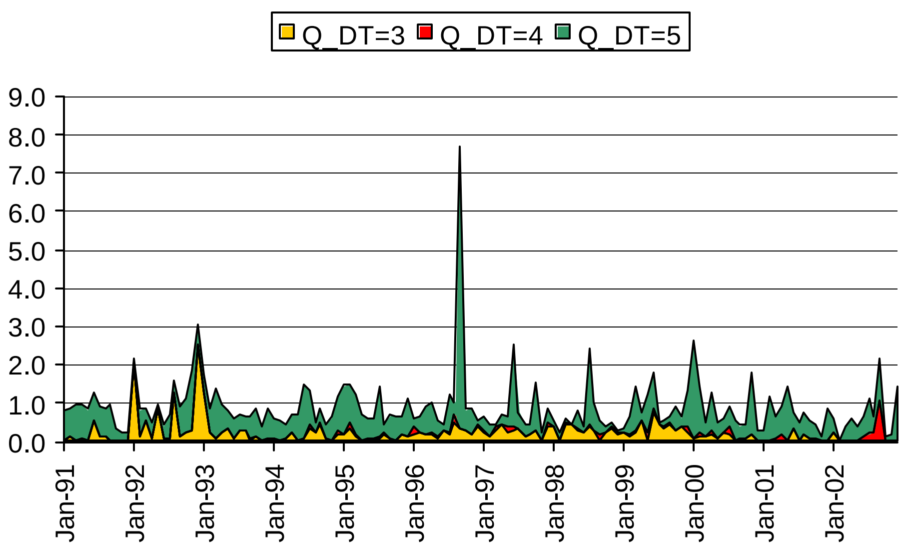


Figure 18: Percent of delayed mode stations with problems detected in the date or time.

We have not shown the figures for error rates on profiles from delayed mode data. For these data, some data submitters choose to send all of the data collected and allow the error flagging procedures to indicate what data are useful. In some instances, profiles with data collected deeper than the design depth of the XBT, for example, show spikes that are retained in the data files. These are correctly flagged as wrong values. The consequence of this procedure, though, is that a large fraction of profiles receive at least one level with a flag indicating bad data. This tends to skew the comparison to the real-time data, where operators strive to send only reasonable data for real-time distribution.

4.5 Monitoring

The GTSPP has developed a number of tools that are used to monitor various aspects of the project. The displays already shown represent some of them. There are others that serve particular purposes.

Each month, MEDS produces a report that summarizes the BATHY and TESAC data received from Germany, Japan, the U.S. and MEDS own connection to the GTS. This is called the preliminary International Report and is distributed by email to interested parties. A shortened version of the report is shown in Annex 1 to illustrate its content. Each month's report can also be found at
http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/GTSPP/e_preint.shtml

Each month, MEDS carries out a review of all of the BATHY and TESAC data received with the goal of identifying platforms with consistent failures and notifying the operators so that corrective action can take place. Each report has the five components listed here.
1. A summary report of the data received with comments made about those platforms where more than 10% of the stations had problems.
2. A map showing the location of all of the data received during that month (see a sample in annex 2a)
3. A table that shows information and summaries of QC results for every platform reporting that month (see http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/WOCE/e_woce.html)
4. A map showing stations that reported on SOOP lines during the month. (see annex 2b for a sample)
5. A table identifying the platforms and SOOP lines sampled during that month (see annex 2c for a sample).

The report is sent by email to interested parties.

4.6 Products

The GTSPP was an important part of the WOCE Upper Ocean Thermal Data Assembly Centre. As such it contributed to the production of the various versions of the WOCE Data Resource. The final version was issued in November of 2002 and the UOT portion contributed over 1 million profiles. It is possible to order a copy of the DVD set or to see all of the data on-line at http://www.nodc.noaa.gov/woce_v3/

The GTSPP has updated its brochure that describes the program. Electronic versions are available from http://www.nodc.noaa.gov/GTSPP/document/gtspp/brochure/brochure.htm

The functions of the GTSPP are carried out by a number of centres as shown in figures 2 and 3. Web pages illustrating aspects of their contributions to the GTSPP include the following.
MEDS: http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/GTSPP/GTSPP_e.htm
US NODC: http://www.nodc.noaa.gov/GTSPP/gtspp-home.html

GTSPP has a number of client organizations that receive data from the project on a regular schedule. For some, this can be 3 times per week as the quality control and duplicates processing of the real-time data are completed. For others, it can be on a weekly or monthly basis when the latest data received since the last update is provided. Data are usually delivered through ftp and can be from the entire world or some smaller area. Clients interested in subscribing should contact the author.

The Science centres contribute scientific expertise to improve data quality and provide advice on how the GTSPP should evolve. They also use the data coming through GTSPP in the creation of ocean analyses. The following URLs provide a starting point to examine more of their work.

USA – Scripps: http://jedac.ucsd.edu/DATA_IMAGES/index.html
USA – AOML: http://www.aoml.noaa.gov/goos/
Australia: http://www.bom.gov.au/bmrc/ocean/JAFOOS/contents.html

4.7 Meeting JCOMM targets

Simple maps, such as shown in annex 2, show the locations of collected data. However, in order for the data to be useful in some applications, it is necessary to have a certain density of observations in space and time. JCOMM needs to measure how well its observation programmes are meeting sampling criteria for its clients.

In 1999 the Ocean Observations 99 meeting recommended that SOOP shift emphasis from broadcast to line-mode sampling. This report has already described the simple monitoring that is done by GTSPP to provide a month to month visual presentation of the success of sampling along lines.

A more comprehensive analysis has been designed and implemented at the JCOMMOPS site. (See http://www.brest.ird.fr/soopip/index.html).

In another development, the Ocean Observation Panel on Climate, OOPC, has set forth both time and space sampling criteria for different variables in order to meet the demands in monitoring climate. By itself, the GTSPP does not assemble the necessary suite of observations to define the measurement success for all of the variables treated by OOPC. However, GTSPP can deal in those areas that require profiles of temperature and salinity.

The OOPC requirements for measurements of upper ocean temperature and salinity require at least one observation every 30 days. For salinity, the spatial requirement is every 300 by 300 km while for temperature it is 200 by 500 km. In the Argo programme, the optimal sampling target has been set to be a T and S profile every 10 days and every 300 by 300 km. The GTSPP handles virtually all of the ocean profile measurements including those originating from the moored equatorial buoys. It is possible to examine the contributions from the different sources as well as derive a composite sampling density map for temperature and salinity. Before Argo began, the sampling was highly variable in both space and time. With the development of Argo, the sampling is becoming more uniform.

In figure 4 the number of BATHY and TESAC reports handled by the GTSPP as a function of time is shown. The figure in annex 2b shows the spatial distribution of the data received in a recent month. We wish to take both this spatial and temporal sampling and convert it to a figure that shows how well the present sampling program meets certain targets. The most well defined target for the broad scale sampling of the ocean is that defined by OOPC and more recently by Argo. For demonstration purposes, we have made an estimate of the density of T and S profiles by applying a Gaussian weighting function to the array of locations of data normalised by the same weights applied to a regular array of the size 300 x 300 km. So density = observed weight/reference weight. We have used both a single 10 day period and a period of 1 year to show the contrasts. A more detailed explanation of how the maps are generated is provided in annex 3.

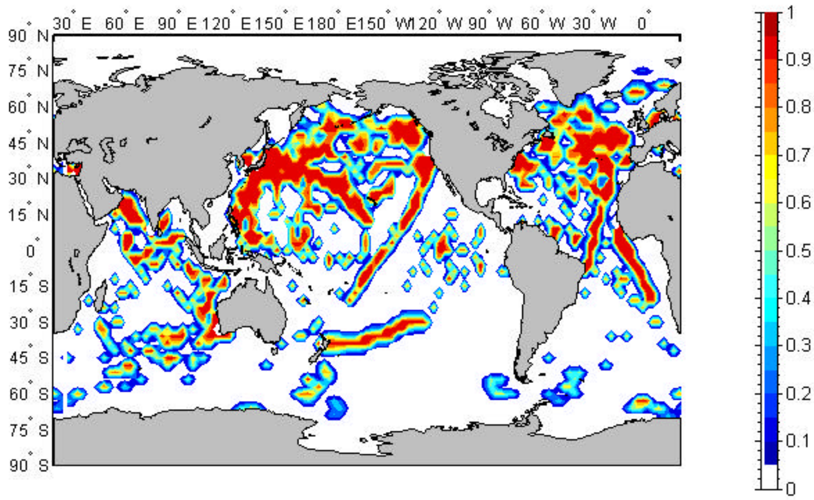


Figure 19: Density of temperature profiles sampled in a single 10 day period (May, 2003).

This figure (19) shows how well sampled the oceans are for just temperature profiles and in a single 10 day period in May, 2003. It is evident that along ship tracks the sampling goal is met. Also, in places where profiling floats are operating, and depending on their spacing, sampling is approaching the 100 percent desired. It is completely predictable, that for most of the ocean, the sampling goals are not being met. Because of variations in the number of data, there will be variations in these density maps from one 10 day period to the next. Within this limitation, a single map gives an approximate idea of how well the climate observing goals are being met at that time.

In figure 20 below, we have applied the same criterion for sampling but now applied it to a full year of data. In order for a particular area to be well sampled, it must have a profile in very 10 day period and every array cell over the course of the entire year. The most obvious result is a poorer success rate for meeting the observation goals. There are a few areas, such as the north eastern Pacific, where profiling floats have been operating for a long enough period of time that they are actually meeting the sampling targets. It is also true that along regularly sampled ship lines, such as off western Australia, the sampling is in the 60 to 80% range of the target. In other areas, such as off the south coast of Chile, even though there are profiling floats now operating, they have not been doing so long enough to have a measurable impact over a year.

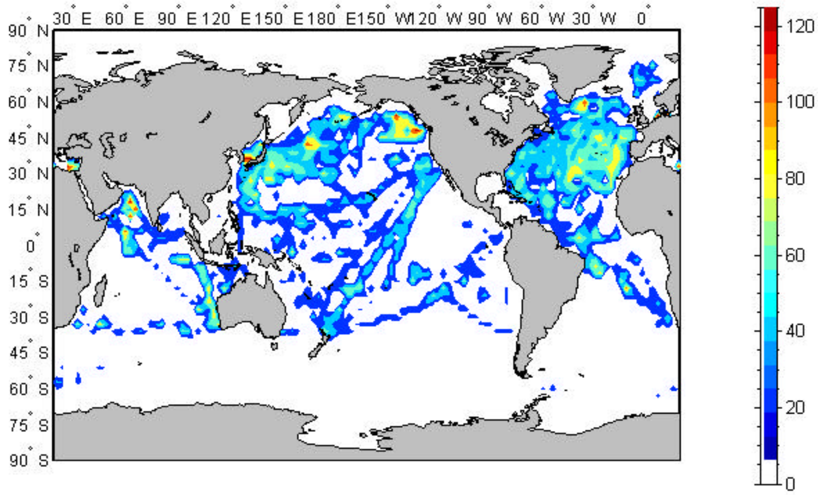


Figure 20: Density of temperature profiles sampled over the course of one year (May 2002 to April, 2003).

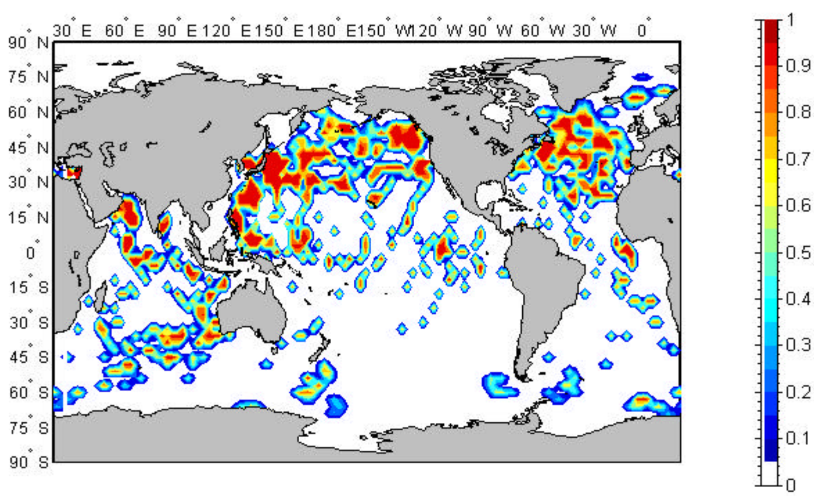


Figure 21: Density of temperature and salinity profiles sampled in a single 10 day period (May, 2003).

We have carried out the same analysis but this time requiring both temperature and salinity profiles to be present. Figures 21 and 22 are the result.

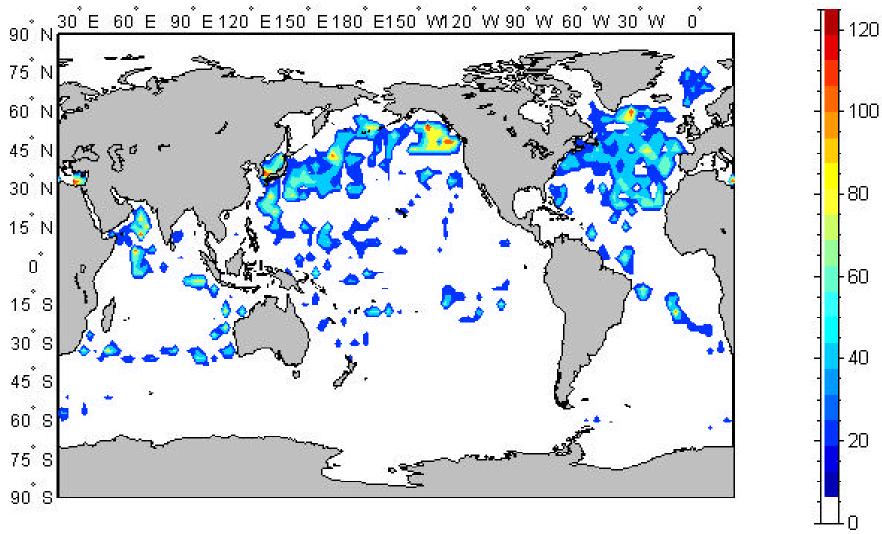


Figure 22: Density of temperature profiles sampled over the course of one year (May 2002 to April, 2003).

We see similar features as for temperature alone, except, of course, since there are fewer temperature and salinity measurements, the maps show even fewer areas where the sampling targets are being met. In this case, except for a few areas, the sampling is provided entirely from profiling floats.

Such figures are one way to show how well JCOMM programs are meeting the sampling requirements of clients. As long as clients can specify their needs in some quantifiable way, we should be able to create a display that indicates how well the goal is being met. It is important for JCOMM to work with clients to quantify their requirements, and then to translate these into metrics against which the observational programs of JCOMM are measured.

5.0 Partnerships

5.1 Argo

Argo data are presently being handled by the GTSPP system and so are entering the global archives in the same way as other data reported on the GTS and then in delayed mode. However, there is a closer association with Argo than this. The Argo data system relies on individual data assembly centers (DACs) to manage and contribute data both to the GTS and to the global data servers of Argo. Not all DACs begin operations with all capabilities in place. For some, the insertion of data to the GTS is handled by Service ARGOS while the contribution of the data to the global servers is delayed. GTSPP contributes the real-time data (having passed through GTSPP quality control procedures) to the global servers to provide at least a reduced form of the data at these servers until the originating DAC can start to send the data on their own.

The quality control procedures of the GTSPP were the starting point of the automated procedures employed in the Argo program. Although the GTSPP procedures had been developed for XBT data, with suitable modifications they are reasonably effective at catching errors in float data.

The main data centers operating in GTSPP all have a significant role in Argo. IFREMER operates one of the global data servers, NODC operates the Global Argo Data Repository (long term archive) and MEDS and IFREMER contribute resources in managing the programme through the co-chairs. The Japan Meteorological Agency and the CSIRO of Australia are contributors to both GTSPP and Argo.

5.2 JCOMM and GOOS

GTSPP started as a jointly sponsored program of WMO and IOC and so when JCOMM was formed it was adopted by the new commission. It reports through the Data Management Program Area but also contributes to the Ship Observation Team meetings. The experiences in data management gained from GTSPP operations over the last 13 years has been invaluable. We have implemented an operational program that put in place a large number of elements to ensure broad support. It continues to contribute this experience in the deliberations that JCOMM are undertaking to assemble a global observation system.

In the early days of GOOS, GTSPP was recognized as an important program that was delivering on some components needed. It was for this reason that it was accepted as an Initial Observing System component.

GTSPP provides the infrastructure support in data management that is required to move the data from collectors to users in the time frames and with the level of quality and consistency that is needed. It therefore supports both JCOMM and GOOS needs.

5.3 CLIVAR

GTSPP acted as the data system in support of the WOCE Upper Ocean Thermal Data Assembly Centre. This was a natural extension to the support provided for SOOPIP. Because of this participation, we have been invited to take part in CLIVAR. Initially, our contribution will be quite similar to that provided during WOCE. As the requirements for CLIVAR become clearer and different needs are expressed, operations of GTSPP will adjust.

6.0 Actions

6.1 SOT-1 Actions 4.4

(i) The Upper Ocean Thermal review noted that as capability was developed, the preference was for all XBT data to come ashore in real-time and full resolution. SOOPIP requested GTSPP to review the impact of this at its upcoming meeting in Australia and devise a scheme to handle this data stream.

This was discussed in a subsequent GTSPP meeting held in Australia. There is nothing within the GTSPP that limits the resolution of data being handled either in real-time or delayed mode. Certainly, in real-time, using the BATHY or TESAC code forms, there are restrictions on the resolution of the variables as well as a perceived restriction on the vertical resolution although some quite highly sampled profiles can be sent in real-time. In dealing with the SEAS data that will be coming ashore in full resolution, there are arrangements to have these data made available to GTSPP immediately. More details of the data flow are provided in section 6.2.

(ii) There presently exists a document that describes data QC procedures to be carried out on board ship before data are submitted to the GTS. In the case where full resolution data was coming ashore, the meeting asked GTSPP to review this to be sure it was still applicable. (Action: GTSPP with SOOPIP chair)

The chair of SOOPIP noted that within the SEAS software developments for sending full resolution data ashore, it was still the intention to use the existing, documented quality control procedures on board ship. Having the full resolution data available allows for the execution of other procedures, but would not negate the value of applying the existing procedures. Within GTSPP, the suite of quality control procedures is more extensive than those suggested for ship board application, and so these could be used prior to submitting data to the GTS. The experience is that procedures used against the real-time data are equally useful against higher resolution profiles as well.

A note of caution is in order. In the application of quality control procedures to real-time data, a certain fraction of the data will fail and these likely will be removed from real-time distribution. Some of these failures may be because measured values lie outside usual and expected ranges, but are still a true measure of oceanic conditions. It is important that all of these failed values are submitted in delayed mode to the appropriate archive centres so that these unusual values are not lost from the data record.

6.2 Implementing a Unique Data Identifier

One of the most difficult problems faced by the GTSPP has been in matching real-time and delayed mode data from the same original observation. The problems stem from reduced vertical and measurement resolution reported in real-time messages and from uncertainties in positions and times as demonstrated by the levels of position and time errors shown earlier. The delayed mode data may have these errors corrected and so matching real-time to delayed mode is not simply a matter of matching ship identifier, position and time. The GTSPP developed software that considers detailed comparisons of individual station data when real-time and delayed mode positions are within 5 km distance and 15 minutes of time to each other. It assumes that errors in these quantities are not large. In a number of cases, the assumption is borne out, but not in every case. So, although a degree of success has been attained in matching real-time and delayed mode data, there is still room for improvement.

A new strategy was discussed at a GTSPP meeting a few years ago. It was inspired by the Ocean Information Technology Pilot Project being undertaken by JCOMM and IODE. The solution was suggested by colleagues in Australia and hinges upon the use of a cyclic redundancy check (CRC) calculation. Since then, the GTSPP and the SEAS program in the US have been cooperating to install the necessary software to implement the solution.

The CRC will be incorporated into the US SEAS system. The CRC is a 32 bit value based on the ASCII generated BATHY message of those values following the 888 group and terminating at the equal (=) sign of the message. Development is concurrent with the development of the AOML automatic quality control software. It is expected to be implemented in time for the August 2003 high-density cruise of the Sea-Land Florida (KRHX). A meeting with NODC will determine the desired format and content of the real-time archive message along with details regarding the destination server and schedule. Paul Chinn is responsible for development, test, and implementation and can be contacted at Paul.Chinn@noaa.gov or 301-713-2790 x 289.

When an XBT is taken, SEAS shipboard software creates a binary record of the entire data stream, metadata, and computed unique SEAS ID for archive aboard ship. This is referred to as the “complete message”. The complete message is the delayed mode record sent to AOML and forwarded to NODC. SEAS shipboard software also creates a “best message” and SEAS ID for transmission to a land-based SEAS processing server.

The SEAS processing server builds two real-time messages from the best message. One is the usual BATHY record distributed on the GTS. The GTS record reaches MEDS and is incorporated into their GTSPP operation. MEDS computes a CRC from the BATHY message using the exact algorithm used by SEAS and attaches it to the record. The other real-time message is a real-time "archive message". The archive message is the same GTS record but has the SEAS ID and computed CRC of the GTS record attached. This archive record is sent to NODC and becomes part of their GTSPP data management operation.

NODC receives two SEAS records from NOAA. The real-time archive message (SEAS ID + CRC ID) and the delayed mode complete message (SEAS ID). Comparison of the SEAS ID completes the data flow from NOAA. NODC also receives a GTSPP record from MEDS which has the same CRC computed. Comparing the GTS CRC ID of the archive message to MEDS GTSPP record completes the GTSPP data flow.

This scheme should eliminate any of the uncertainty in matching real-time to delayed mode data, and so make an important improvement in resolving some of the duplication problems in the global ocean data archives. Assuming it works, it will be a model for handling real-time data.

6.2 Variable Coding

Within the GTSPP data format, names of the variables or metadata are represented by parameter codes (PCODES). The list is derived from the IODE GF3 code list. It forms the data dictionary of the project and is used by all participants.

MEDS has been the custodian of the list and it is now available on-line. The value of this is, of course, the ease with which participants and others may see existing content and request new codes to meet their needs. At the same time we have posted other code tables that are used by MEDS and are of interest to GTSPP. These are available from http://www.meds.sdmm.dfo-mpo.gc.ca/meds/About_MEDS/Code_Tables/list_e.htm

MEDS has also placed a detailed description of the data format used by GTSPP. It was necessary to develop a new format because of the extensive metadata and quality flags that we are carrying with the data. For more information see http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Databases/OCEAN/Format_e.htm

Annex 1: An abridged version of MEDS Monthly Preliminary International Report

GTSPP PRELIMINARY ANALYSIS OF INTERNATIONAL MONTHLY GTS DATA
(Data received at MEDS, US National Weather Service, BSH Germany and JMA Japan)

GTSPP Preliminary International GTS Data Flow Report, MAR 2003

STATISTICAL OVERVIEW REPORT

There were 1842 unique BATHYs and 3318 TESACs in the input file

STREAM_IDENT	GETE:	2395 TESACs	(72.2%)
STREAM_IDENT	GEBA:	1642 BATHYs	(89.1%)
STREAM_IDENT	MEBA:	1611 BATHYs	(87.5%)
STREAM_IDENT	JABA:	1598 BATHYs	(86.8%)
STREAM_IDENT	NWBA:	1730 BATHYs	(93.9%)
STREAM_IDENT	METE:	2735 TESACs	(82.4%)
STREAM_IDENT	NWTE:	2652 TESACs	(79.9%)
STREAM_IDENT	JATE:	2121 TESACs	(63.9%)

Receipt matrix by STREAM_IDENT for BATHY and TESAC messages

	Unique	GETE	GEBA	MEBA	JABA	NWBA	METE	NWTE	JATE
GETE	44	2395	0	0	0	0	1935	2075	1660
GEBA	17	0	1642	1579	1563	1547	0	0	0
MEBA	11	0	1579	1611	1537	1517	0	0	0
JABA	1	0	1563	1537	1598	1557	0	0	0
NWBA	153	0	1547	1517	1557	1730	0	0	0
METE	333	1935	0	0	0	0	2735	2113	1634
NWTE	40	2075	0	0	0	0	2113	2652	2048
JATE	0	1660	0	0	0	0	1634	2048	2121

Difference matrix by STREAM_IDENT for BATHY and TESAC messages

	Totals	GETE	GEBA	MEBA	JABA	NWBA	METE	NWTE	JATE
GETE	2395	0	0	0	0	0	460	320	735
GEBA	1642	0	0	63	79	95	0	0	0
MEBA	1611	0	32	0	74	94	0	0	0
JABA	1598	0	35	61	0	41	0	0	0
NWBA	1730	0	183	213	173	0	0	0	0
METE	2735	800	0	0	0	0	0	622	1101
NWTE	2652	577	0	0	0	0	539	0	604
JATE	2121	461	0	0	0	0	487	73	0

GTSPP Preliminary International GTS Data Flow Report, MAR 2003

GTS Header	No. BATHYs	No. TESACs
SOVD01 KWBC	0	2
SOVD01 RJTD	0	3
SOVD02 CWOW	0	194
SOVD83 KWBC	0	705
SOVE01 AMMC	7	0

Etc…………………………………………………………….

CENTRES SUMMARY REPORT

For organization: GE

Headers Received			
SOVD02 CWOW	185	/	194
SOVE01 AMMC	7	/	7
etc…………………………………………			

Headers Not Received		
SOVD01 KWBC	0	/
SOVD01 RJTD	0	/
etc…………………………………………		

etc. for MEDS, US, Japan.

SHIP SUMMARY REPORT

Call Sign	BATHYs	TESACs	Headers
10004	108	0	SOVF01 EDZW
19019	0	4	SOVX10 KARS
3EZI6	197	0	SOVX01 KWBC
3FRY5	25	3	SOVX02 RJTD SOVX01 RJTD
3FRY9	19	0	SOVX01 KWBC

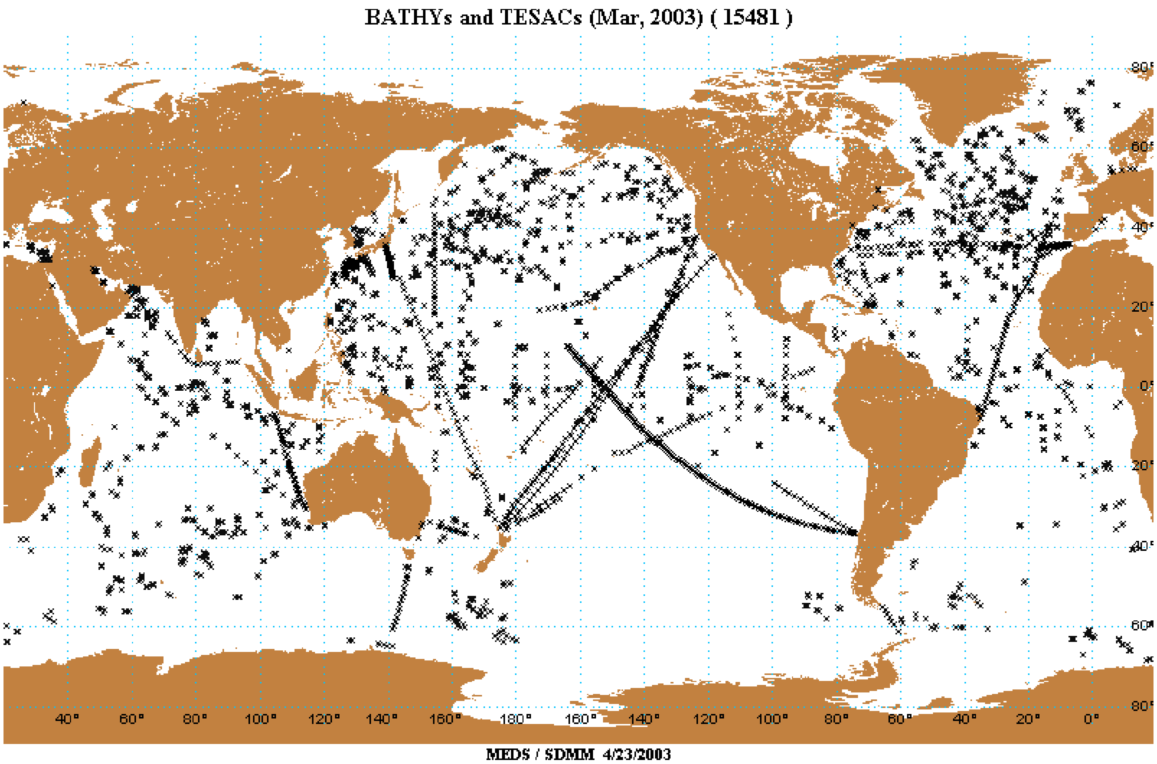
etc. for all other platforms

TIMELINESS REPORT
GTSPP Preliminary International GTS Data Flow Report, MAR 2003

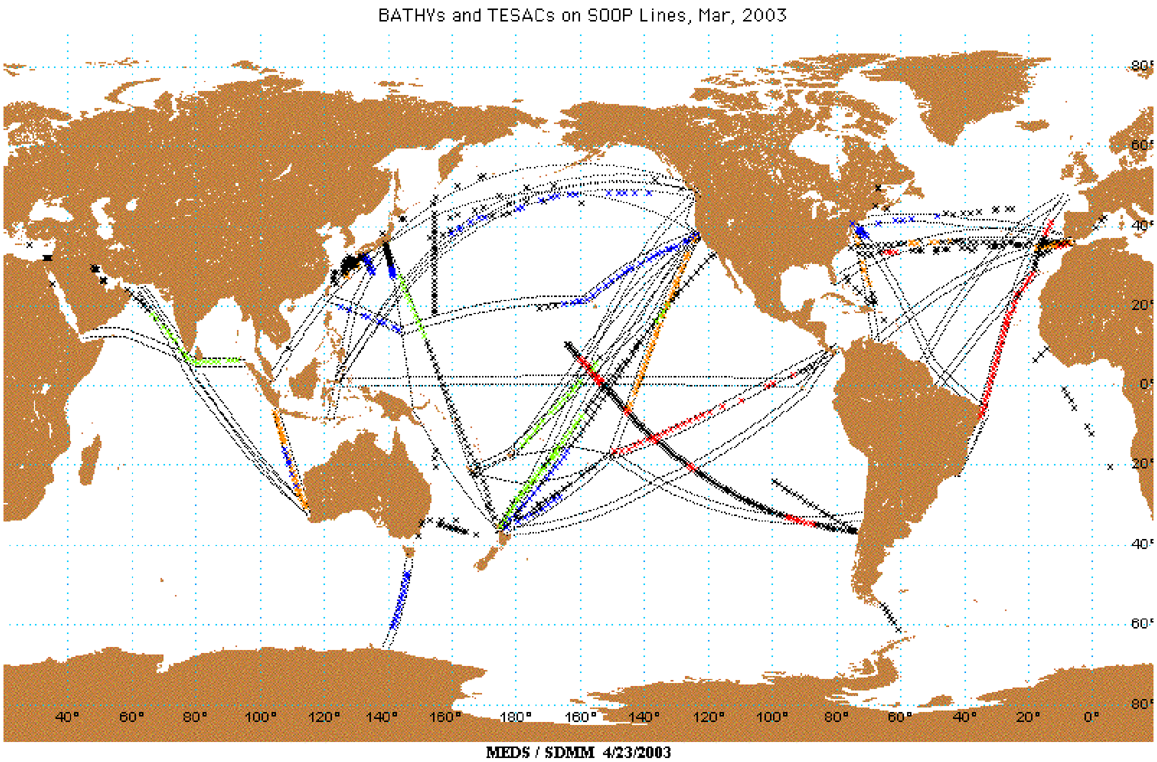
Days	Number	Percent of Total	Cumulative Number	Cumulative Percent
1	2498	54.1	2498	54.1
2	1039	22.5	3537	76.6
3	235	5.1	3772	81.7
4	164	3.6	3936	85.2
5	161	3.5	4097	88.7
6	110	2.4	4207	91.1
7	105	2.3	4312	93.4
8	90	1.9	4402	95.3
9	84	1.8	4486	97.1
10	30	0.6	4516	97.8
11	10	0.2	4526	98.0
12	17	0.4	4543	98.4
13	30	0.6	4573	99.0
14	8	0.2	4581	99.2
15	1	0.0	4582	99.2
16	3	0.1	4585	99.3
17	0	0.0	4585	99.3
18	3	0.1	4588	99.4
19	1	0.0	4589	99.4
20	5	0.1	4594	99.5
21	4	0.1	4598	99.6
22	6	0.1	4604	99.7
23	5	0.1	4609	99.8

24	4	0.1	4613	99.9
25	0	0.0	4613	99.9
26	3	0.1	4616	100.0
27	0	0.0	4616	100.0
28	0	0.0	4616	100.0
29	1	0.0	4617	100.0
30	1	0.0	4618	100.0

Annex 2a: A map showing locations of all BATHYs and TESACs collected in March, 2003.



Annex 2b: A sample map showing BATHYs and TESACs that collected data along SOOP lines in March, 2003.



Annex 2c: A sample table indicating which ships collected data along SOOP lines in March, 2003. This table accompanies the map shown in annex 2b.

Cruise #		Total # of Stations	Stations on SOOP Line(s)	Colour	SOOP Line(s)
3EZI6	03,	185 ,	43 ,	RED,	PX99 , PX07, PX31, PX24, PX18, PX17
3FRY5	03,	24 ,	18 ,	GREEN,	IX10 , IX09
3FRY9	03,	19 ,	19 ,	ORANGE,	AX09
9VRA	03,	52 ,	23 ,	BLUE,	PX12 , PX13, PX28
DACF	03,	51 ,	40 ,	RED,	AX11 , AX20
DDGY	03,	37 ,	26 ,	GREEN,	PX12 , PX07, PX31
ELES7	03,	32 ,	32 ,	ORANGE,	IX01
ELVX4	03,	15 ,	5 ,	BLUE,	AX04
ELVZ6	03,	43 ,	21 ,	RED,	PX17
ELZT3	03,	53 ,	25 ,	GREEN,	PX18 , PX13, PX07
FHZI	03,	14 ,	14 ,	BLUE,	IX28
FNCM	03,	11 ,	4 ,	RED,	AX09
H9TO	03,	37 ,	10 ,	GREEN,	PX05
JGKL	03,	123 ,	14 ,	ORANGE,	PX49
JHLO	03,	24 ,	10 ,	BLUE,	PX05
JPBN	03,	19 ,	14 ,	BLUE,	PX11 , PX49
KIRF	03,	20 ,	10 ,	ORANGE,	AX10
KRGB	03,	61 ,	53 ,	BLUE,	PX44 , PX85, PX01, PX26, PX37
PJJU	03,	10 ,	10 ,	BLUE,	AX29
V2FA2	03,	38 ,	38 ,	ORANGE,	PX18
VKLD	03,	6 ,	6 ,	BLUE,	IX01
WAUW	03,	29 ,	4 ,	RED,	AX07
WMLG	03,	35 ,	6 ,	ORANGE,	AX07

Annex 3: Creation of the weighted density array

The data for a single 10 day period starts with a count of the number of profiles in every 1 degree square in MEDS archives. For temperature only, we used data coming from both BATHY and TESAC code forms, while for temperature and salinity together, we only used the TESAC code form. These raw values are summed over each 3 x 3 degree square.

In each water area, the radial distance between two points is given by

Radial distance = $[111.2\sqrt{((\Delta\varphi)^2+(\Delta\lambda)^2\cos^2\lambda)}]$ km

where
The distance between two parallels = 111.2 km
The distance between any two meridians = $[111.2 \cos \lambda]$ km
λ is the average latitude between the two points
$\Delta\lambda$ is the absolute difference in latitude degrees
$\Delta\varphi$ is the absolute difference in longitude degrees

For every 3 x 3 degree square, these are weighted and summed for each element j, by adding the value of all other elements multiplied by a weight that decrease exponentially with the square of the distance

$$w_j = \sum_i C_{ij}\, e^{-(x_{ij}/d)^2}$$

where
d is the scale (set to 200 km)
$\mathbf{x_{ij}}=[111.2\sqrt{((\Delta\lambda)^2+(\Delta\varphi)^2\cos^2\lambda)}]$ km
$\Delta\lambda = \lambda_i - \lambda_j$
$\Delta\varphi = \varphi_i - \varphi_j$
This results in the actual weighted sampling array

We then do the same, assuming an ideal sampling of data; that is all 3 x 3 degrees are sampled according to the goals. The array thus obtained is the ideal weighted array. Its values range from 0 to 21, highest values are found at the highest northern latitude (87ºN) for geometrical reasons and since Antarctica occupies the highest southern latitudes.

We then divide every element of the actual array by its corresponding element of the ideal weighted array**.** We use a coastline map to mask the land values.