OceanObs'09
Draft Community White Paper

Title
The Data Management System for the Global Temperature and Salinity Profile Programme

Lead author
Charles Sun, Oceanographer, U.S. National Oceanographic Data Center
1315 East-West Highway, Silver Spring, MD 20910
E-mail: Charles.Sun@noaa.gov

Contributing authors
Ana Baldoni, Thierry Carval, Paul Chinn, Lisa Cowen, Gustova Goni, Vissa V. Gopalakrishna, Raul Guerrero, Norman Hall, Melanie Hamilton, Fengying Ji, Yoshiaki Kanno, Robert Keeley, Birgit Klein, Shaohua Lin, Giuseppe Manzella, Yoshiharu Nagaya, Franco Reseghetti, Lesley Rickards, Anh Tran, Ann Thresher, and Loic PETIT de la VILLEON

EXECUTIVE SUMMARY

The Global Temperature and Salinity Profile Programme (GTSPP) is a joint program of the International Oceanographic Data and Information Exchange committee (IODE) and the Joint Commission on Oceanography and Marine Meteorology (JCOMM) of the World Meteorological Organization (WMO) and the Intergovernmental Oceanographic Commission (IOC). Tasks in the GTSPP are shared amongst the participating countries including but not limited to Argentina, Australia, Canada, China, France, Germany, India, Italy, Japan, UK, and USA. Scientists and data managers in these countries contribute their time and resources to ensure the continued functioning of the program.

The GTSPP has four primary objectives: (1) to continue 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 programs, national and international operational oceanography, (2) to continue to implement data flow monitoring and reporting systems for improving the capture and timeliness of the GTSPP real-time and delayed mode data to prevent data losses, (3) to continue to

improve and implement agreed and uniform quality control and duplicates management systems by coordinating activities among all participating countries and (4) to facilitate the development and provision of a wide variety of useful data analyses, data and information products, and data sets.

The success of the GTSPP allows users access to the most up-to-date, highest quality and resolution data at the time of request by managing both real-time and delayed mode data. It delivers data to users through complete quality control (QC) and duplicates elimination procedures within 24 hours of data collection and has standardized quality control procedures for ocean temperature and salinity profiles that have been adopted in many other projects and countries.

The two most important activities that the GTSPP is undertaking are:

1. Management of expendable bathythermographs (XBTs) data: The GTSPP has preserved original XBT data in the GTSPP Continuously Managed Database (CMD). This is critical to examination and correction of the recently identified XBT depth bias problem, since the oceanographic community does not have consensus on what causes the bias or how to correct it;
2. Implementation of unique data tag identification: GTSPP has developed and implemented the algorithm of Cyclical Redundancy Check (CRC) as a way to uniquely tag data circulating on the Global Telecommunication System (GTS) and the original profiles used to create the real-time messages.

The paper describes a framework for developing and implementing operationally a state-of-the-art data and information portal with capabilities of exploring in-situ data from near real-time data streams and integrating the data streams with historical data. The paper also provides recommendations in order to ensure the interoperability of data and information systems, which have been developing in different countries.

BACKGROUND

The international oceanographic community's interest in creating a timely global ocean temperature and salinity dataset of known quality in support of the World Climate Research Programme (WCRP) dates back to the 1981 IODE meeting in Hamburg, Federal Republic of Germany. The community's interest led to preliminary discussions by the Australian Oceanographic Data Center (AODC), the Marine Environmental Data Service (MEDS), now the Integrated Science Data Management (ISDM), of Canada and the U.S. National Oceanographic Data Center (NODC) during the second Joint IOC–WMO Meeting of Experts on Integrated Global Ocean Services System (IGOSS)-IODE Data Flow in Ottawa, Canada in January 1988. The development of the GTSPP began in 1989, and went into operation in November 1990.

GTSPP INFRASTRUCTURE

The GTSPP is a collection of volunteer organizations, and therefore role adjustments are always needed to accommodate changes in levels of participation. Over the life of the program, there are consistently on the order of 5 to 7 countries, and up to 10 to 12 organizations participating in operations of the GTSPP. Figure 1 illustrates the GTSPP data flow and data management procedure. A given international JCOMM or IODE centre may fit within several boxes in carrying out its national and international responsibilities. Currently, the GTSPP consists of three components:

1. Global Telecommunication System (GTS): The WMO provides the use of the GTS for the transmission of oceanographic messages collected through various Panels in the JCOMM program. The GTSPP uses this service to acquire real-

time data exchanged this way. Real-time data processing services are provided by the ISDM.

2. IODE Data Centers: Historical data are acquired either from other NODCs, or from cooperation with projects such as the Climate Variability and Predictability (CLIVAR), the World Ocean Database (WOD), and the Ship Of Opportunity Programme (SOOP).
3. Continuously Managed Database (CMD): The NODC provides data processing services for historical data and maintenance of the CMD (also known as the GTSPP archive). Historical data include both low resolution data from the GTS and the full resolution data from XBTs or CTDs from the ships that provided the real time low resolution data to the GTS, or fully processed and quality controlled data from other organizations.

DATA VOLUMES

The GTSPP continues to handle all real-time and delayed mode profile data with temperature and salinity measured. The real-time data in GTSPP are acquired from the GTS in the BATHY and TESAC codes forms supported by the WMO. In JCOMM, the BATHY and TESAC code forms are the ones used most often for distribution of ocean profile data on the GTS. Figure 2 shows the progression in 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. It is evident from the figure that the number of BATHY reports has declined since 1999 but appears to have stabilized or perhaps is slightly increasing. Over 2007–2008, the number of BATHYs reported steadily increased from 24,855 in 2007 to 27,775 in 2008, while the number of TESACs was 1,630,360 to the end of

2008, dramatically increased from 821,321 in 2007. A new data set of 6,869 CTD profiles (as of December 2008) derived from marine mammals was made available for the first time beginning in July 2008 [3]. The data are useful because they get high data return from areas between 60° S and 70°S, where data are very sparse.

The GTSPP data contain upper ocean temperature and salinity data and come mainly from profiling floats, XBTs, conductivity-temperature-depths (CTDs) and bottles. These have been subdivided into a few different types and presentations made of the number of stations of each type by year. Figure 3 shows the number of profiles contributed to the CMD, grouped by instrument type. It is clear that the majority of data were from XBTs (which primarily report profiles with temperature only) up until 1999, when the Argo Project began to report temperature and salinity profiles from profiling floats.

DATA QUALITY CONTROL

Data quality control is a procedure of verification and validation. Quality control of the data in the GTSPP system is handled at a number of centers. The ISDM handles near real-time, low-resolution data and applies quality control processing [4] to these data before forwarding them to the NODC. The NODC utilizes a "Delayed Mode Quality-Control (QC)" process analogous to the QC carried out on the real-time data and removes duplicates. The quality reviewing procedures generally include three steps: 1) cruise metadata checks, 2) profile data review and 3) quality control edit. The NODC developed and implemented data quality cruise editor (QCED) software, which allows an operator to view and edit temperature and salinity data from files in the GTSPP-ASCII format. This program allows checking/editing of position and time metadata, and other cruise identification metadata and also allows visual inspection and comparison of the profiles with climatology and neighboring stations.

At the end of the process, data will have passed the quality review process and will have been loaded into the GTSPP database with its data quality flags. The flags indicating data quality are those currently used in IGOSS processing with one extension. Table 1 describes the GTSPP data quality codes and their meaning.

Code	Meaning
	No quality control (QC) has been performed on this element.
0	No quality control (QC) has been performed on this element.
1	QC has been performed; element appears to be correct.
2	QC has been performed; element appears to be correct but is inconsistent with other elements.
3	QC has been performed; element appears to be doubtful.
4	QC has been performed; element appears to be erroneous.
5	The value has been changed as a result of QC.
6 ~ 8	Reserved for future use.
9	The value of the element is missing.

Table 1. GTSPP Data Quality Codes

However, because quality assessment is shared over processing centres, it is possible that data flagged as doubtful by one centre will be considered acceptable by another or vice versa. Flags can be changed by any processing centre as long as a record is kept of what the changes are. The use of the flagging scheme described here meets the stated requirements of the GTSPP. It is recognized that as new testing procedures are developed, it will be necessary to re-examine data. With version flags preserved with the data, it is possible to identify what has been done and therefore how best to approach the task of passing data through newer quality control procedures. Typically, results of the QC procedure are the setting of flags or making corrections where data illustrate instrument failures and human errors.

XBT DATA MANAGEMENT

In the mid 1990s it was found that some the fall rate (depth-time) equations provided by the manufacturers of expendable bathythermographs (XBTs) were not correct [2] [5].

The computed depth errors are outside the manufacturer's specifications. The GTSPP has preserved XBT data and probe type and fall rate equation information in the CMD, if it was provided. Two new codes were created to retain depth correction information in the surface codes structure of the GTSPP data. The "DPC$" indicates the status of depth correction and the "FRA$" retains the conversion factor of 1.0336 if it was applied. Table 2 shows the "DPC$" code and its meaning. Having determined which

Code	Meaning
1	Known Probe Type, Needs Correction
2	Known Probe Type, No need to Correct
3	Unknown Probe Type, Not enough information to know what to do
4	Known XBT Probe Type, Correction was done
5	Unknown Probe Type, but a correction was done

Table 2 GTSPP DPC$ Code

profiles are from XBTs by querying the data type, the XBT probe type and the fall rate equation are stored in the XBT archives. One strategy for the fall rate correction is to simply multiply the existing depths by a factor of 1.0336. This was the technique employed with the multiplication factor stored in the file structure, when GTSPP played a key role in the WOCE and contributed to the final WOCE Data Resource DVD [6].

UNIQUE DATA TAG IDENTIFICATION

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 or slight errors in positions and times. 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. In addition, many ships choose to 'mask' their identify when reporting data to the GTS leaving a further level of uncertainty when trying to match real time to delayed mode data profiles. 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 in 2002. Because of rules imposed by WMO, no changes were permitted to BATHY or TESAC code forms, so it was not possible to add a unique identifier into real-time data transmissions. 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 Ship and Environmental (Data) Acquisition System (SEAS) program in the US have been collaborating to install the necessary software to implement the solution. The CRC is a 32 bit value based on the ASCII generated BATHY message of those values following the 8888 group and terminating at the equal (=) sign of the message.

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 transmitted to a land-based SEAS processing server and saved shipboard as the delayed mode record. The SEAS processing server builds two best real-time messages from the complete message. One is the usual BATHY message (GTS record) distributed on the GTS. The other real-time message, called a “real-time archive message (RTAM)”, has the same

GTS record but with the SEAS ID as well as the computed CRC of the GTS record attached. The RTAM record is sent to NODC to become part of the GTSPP operation.

The BATHY message reaches ISDM. ISDM computes a CRC from the BATHY message using the exact algorithm used by SEAS and attaches it to the record and sends this to NODC. Comparing the RTAM record to the ISDM GTSPP record completes the GTSPP real-time data flow. Figure 4 shows the statistics of the comparison. The results show that the CRC scheme caught up 94% of duplicates between 2005 and 2008. Both Australia and France has expressed interest in implementing the same CRC scheme for data originating from their platforms. GTSPP will continue to monitor these results to test how well the unique identification scheme performs.

CLIENTS AND SERVICES

The GTSPP provides tools and services that allow the public easy access to the GTSPP data both online and offline. The GTSPP sends additional operational data from non GTS sources to numerical weather forecasting services. Marine operations receive data in operational time frames for such operations as ship routing and fishing strategies from the GTSPP as well.   In addition, the GTSPP provides higher quality and more timely data sets of temperature and salinity observations which are used for seasonal to inter-annual forecasting. Finally, science and engineering users receive higher quality and more timely data sets for strategic studies and design.

The GTSPP data are available on-line through the GTSPP HTTP and FTP servers [1]. Users are allowed to view data file summaries, view station location plots, and download monthly real-time data or best copy data sets sorted by ocean and month of the year. All on-line data are distributed in the GTSPP ASCII format and the Network Common Data File (netCDF) format. Users, who only need a subset of the data of their interest, use the GTSPP Web Interface (GWI) [1] or submit a request to the GTSPP data manager via email at nodc.gtspp@noaa.gov .

The GTSPP continues to publish data on optical disc media. As the disc uses the ISO9660 standard with the RockRidge extension, the platform-independent files can, in principle, be read on all operating systems. The GTSPP data disc comes with a software package, known as the "Ocean Data Explorer" (ODE), for users to examine oceanographic data stored on the disc. The ODE (Figure 5) is a Java-based application that provides interactive graphical exploration and conversion of oceanographic vertical profile data stored on optical disc media.

PARTNERSHIPS

GTSPP collaborates with a number of international programs. In particular, it manages the XBT data collected by the operators of the SOOP, which is a subprogram of the Ship Observations Team (SOT) of JCOMM. GTSPP developed a strategy for linking XBT profiles to the SOOP XBT survey lines that were sampled and has been working closely with SOOP to assist in proper documentation of the XBT fall rate in the CMD. In addition, GTSPP also publishes a monthly ship report that highlights any errors found in the data so operations can monitor the performance of their ships' data collection systems. This is then sent to the operators for future action.

The GTSPP currently works with the WOD project and the CLIVAR-Carbon Hydrographic Office (CCHDO) to pull CCHDO data from the Internet quarterly to provide the fully quality controlled CTD data to the Argo CTD Reference Database used for delayed-mode quality control of Argo salinity data. It also collaborated with the Argo program to fix GTS reports from Argo floats that were reporting pressure instead of depth to the GTS.

RECOMMENDATIONS

The success of the GTSPP not only provides an end-to-end ocean temperature and salinity profile data management system to the public, but also demonstrates its capability of archiving and re-processing data. However, several recommendations are made with respect to data management. For example, the oceanographic community has just discovered that the well-known XBT fall rate problem is complicated by inadequate instrument metadata in the past. GTSPP records fall rate and instrument type in the GTSPP database and has the ability to reprocess them, but only if the metadata is provided by the data orginators. It is suggested that any correction to global archives be carried out in cooperation with other data centres around the world to ensure consistency and international standards. Other recommendations are:

Recommendation 1. Any extra metadata that is provided or present such as tests performed and failed, origins of the data stream, and data processing history, are critical aspects that should be retained by any future data management system.

Recommendation 2.A scheme of unique data "tagging" should be implemented immediately. Every original piece of data would be issued a unique identifier. This tag would never be separated from the data. Any subsequent processor would add further identifying information but would leave the original tag in place.

Recommendation 3. Documentation on data quality control and instrument calibration procedures should be available for public access.

Recommendation 4. It is a goal of the GTSPP to increase this participation by the international community. All member states can participate more actively in the GTSPP through their normal JCOMM/IODE activities. To formalize this participation, member states can write to the IOC Secretariat, the WMO Secretariat, or to the Chair of the GTSPP Steering Group outlining the areas in which they wish to contribute.

Recommendation 5. In order to assure interoperability and common services to a wider community of users, cohesive collaborations with on going projects on data management and information systems such as the Data Management and Communications Subsystem of the Integrated Ocean Observing System (IOOS) and SeaDataNet are highly recommended.

CONCLUSIONS

This paper describes the experience of the development and implementation of the GTSPP in the past two decades. Data are provided by networks in operational time frames and as global monthly data sets on the Internet. The unique features of the GTSPP include: (1) it unifies all temperature and salinity profile data into a common structure and therefore a common output, which is inter-operatable and extendable, (2) it sets standards for quality control of temperature and salinity profile data, (3) it documents data processing history, and (4) it carries complete metadata descriptions of every record.

The user comment feedback has been overwhelmingly positive, and shows that this program is a valuable resource for research and scientific work in support of climate assessment and ecosystem management.

ACKNOWLEDGEMENTS

Many agencies have played important roles in the development of the GTSPP system. Each participating agency carries out a number of functions in handling the data for the GTSPP. The most important contributors are the collectors of the original data. Without their efforts, this compilation of data and information would not have been possible. This work was partially supported by the NODC management and NOAA PRIDE (Pacific Region Integrated Data Enterprise) grant #PRIDE05-04 and #PRIDE05-07.

REFERENCES

1. Global Temperature-Salinity Profile Programme, June 2006. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Oceanographic Data Center, Silver Spring, Maryland, 20910. Date of Access, http://www.nodc.noaa.gov/GTSPP/

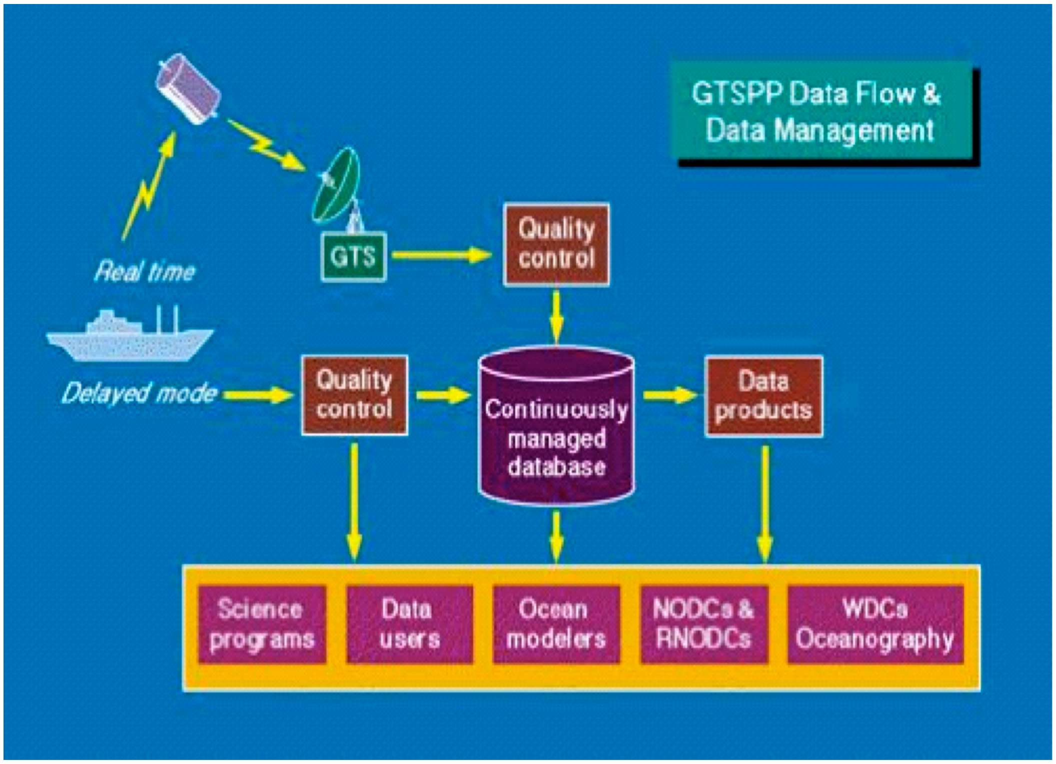
2. Hanawa, K., P. Rual, R. Bailey, A. Sy, and M. Szabados (1995). A new depth-time equation for Sippican or TSK T-7, T-6 and T-4 expendable bathythermographs (XBT). Deep Sea Res. I, 42, 1423–1451.

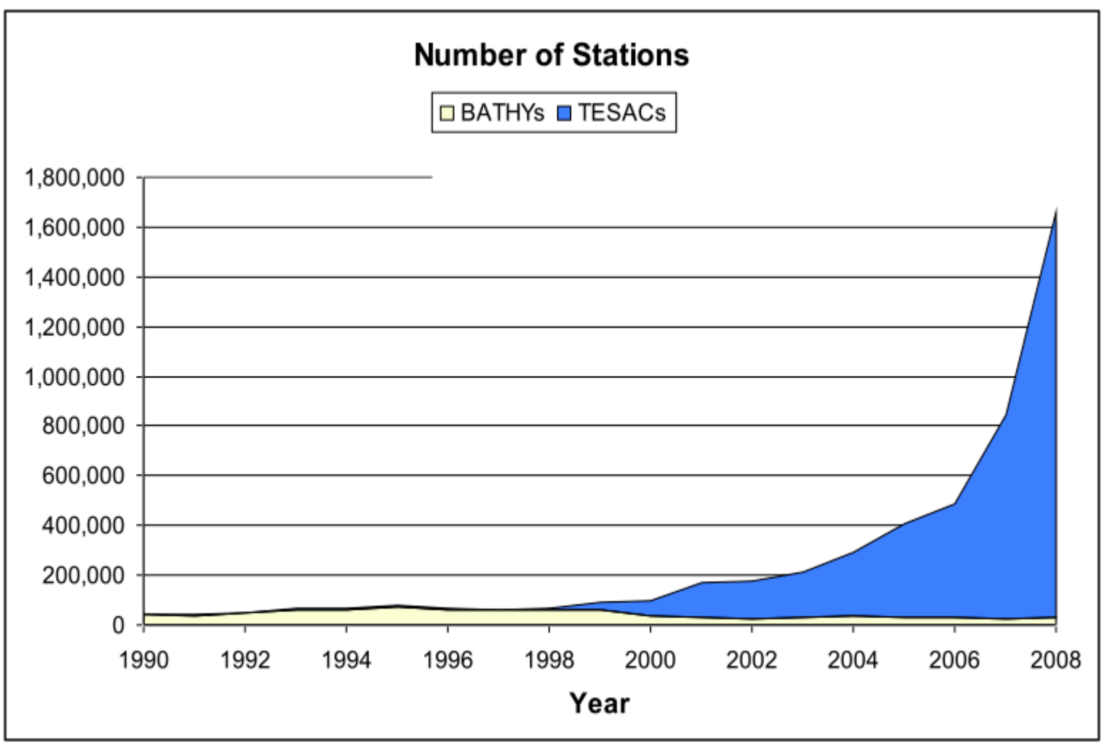
3. Hooker, S. K. And I. L. Boyd (2003). Salinity sensors on seals: use of marine predators to carry CTD data loggers. Deep-Sea Research, Part I, 50, 927–939.

4. Intergovernmental Oceanographic Commission, 1990: GTSPP Real-Time Quality Control Manul. IOC Manual and Guides 22, 128 pp.

5. Intergovernmental Oceanographic Commission, 1994: Calculation of new depth equations for expendable bathythermographs using a temperature-error-free method (application to Sippican/TSK T-7, T-6 and T-4 XBTs). IOC Tech. Series No. 42, 46 pp

6. WOCE Data Products Committee. 2002. WOCE Global Data: Upper Ocean Thermal Data, Version 3.0, WOCE International Project Office, WOCE Report No. 180/02, Southampton, UK. http://woce.nodc.noaa.gov


GTSPP Data Flow &
Data Management
Real time
Delayed mode
GTS
Quality control
Quality control
Continuously managed database
Data products
Science programs
Data users
Ocean modelers
NODCs & RNODCs
WDCs Oceanography


Number of Stations
BATHYs
TESACs
1,800,000
1,600,000
1,400,000
1,200,000
1,000,000
800,000
600,000
400,000
200,000
0
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
Year

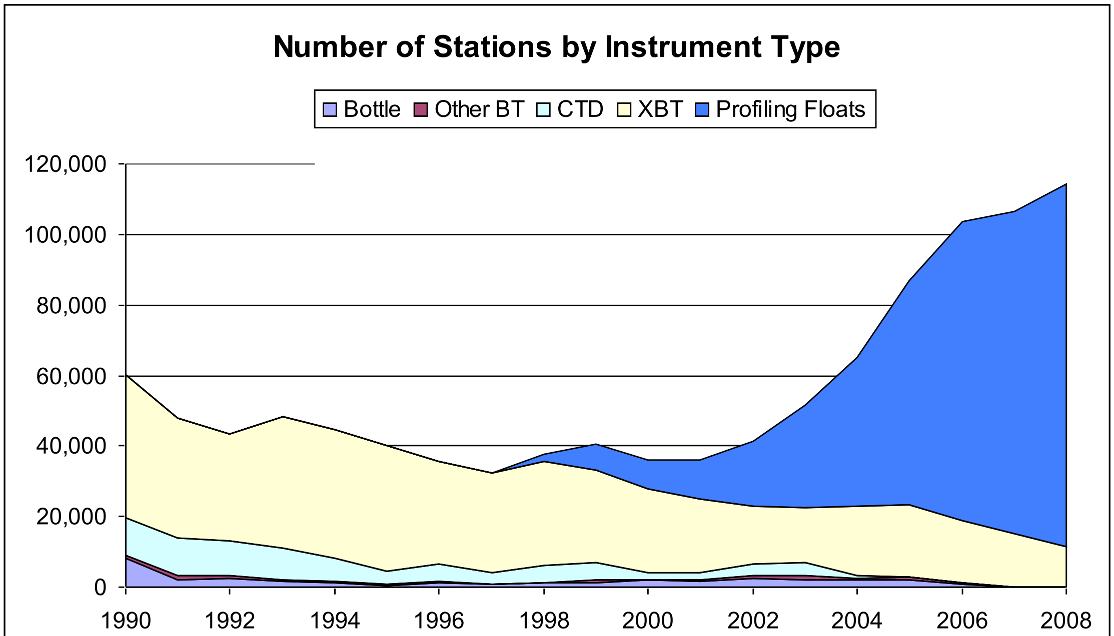


Figure 4 The statistics of the CRC scheme by comparing the real-time archive message and the GTS message.

Figure 3 The station number of historical data by instrument type in the CMD.