Quality Control Processing in the UOT DAC System

Introduction

Quality control of the data in the UOT DAC system is handled at a number of centres. MEDS handles real-time, low resolution data and applies quality control processing on these data before forwarding them to the U.S. NODC. NODC handles the delayed mode, high resolution data. They also subject them to quality control procedures. The procedures they use are the same as those employed by MEDS except they do not undertake visual inspection of the data. The procedures of MEDS are described below.

The three science centres receive approximately one years worth of data at a time. They also have quality control procedures. Each has slightly different ways of handling the data. A brief description of their respective procedures follows as well.

Data Centre Quality Control Procedures for Ocean Profile Data

Data quality assurance (QC) at the Marine Environmental Data Service (MEDS) is a procedure of verification and validation. To validate the data, they are reformatted to MEDS internal processing format. In so doing, the data are checked to be readable and that they can be interpreted. Data which have format errors in the original source form, or which have invalid values (such as characters where numbers should be, or irrational contents of a parameters such as date/time, or profile depth order, are reviewed and corrected by programmers, by adjusting the reformat procedures to handle these inconsistencies.

Once the data are reformatted, the contents are QC'ed in order to verify that the numbers and codes actually represent physical quantities and that these are reasonable given the location and time of the observation. MEDS has adopted an approach which combines specialized computer code to organize, and test data values according to common rules with displays of the data as plots of ship tracks and data profiles with selection and editing capabilities to allow trained personnel to review and flag the data, or correct values where obvious. Typically, results of the QC procedure are the setting of flags or making corrections where data show instrument failures and human errors.

The subtler inaccuracies (such as those caused by instrument noise, or signal processing algorithms), and whether or not the observation is representative of the ambient conditions (by considering errors due to small scale variability, or inherent randomness in convective water flow) are apparent in the MEDS system. Automated tests have tolerances which allow for these inaccuracies. The QC technician quite often spots these problems and flags these where values go beyond reasonable bounds.

The Software System

Ocean observations at MEDS are reformatted, QC'ed, archived and retrieved from a system using the VMS operating system on a network of COMPAQ Alpha work stations. The programming languages are the DEC command language DCL, and the DEC implementation of an enhanced FORTRAN. The GKS library is used to generate graphics for the system.

The QC system is a fully automated pipeline of applications and executables capable of handling both real-time and delayed mode data. The real-time system is used to handle data received primarily through the GTS, and process through to the BATHY and TESAC archive files. The delayed mode system, which has differences to allow an operator to select an arbitrary file of formatted data, is used for the higher resolution data received well after data collection.

There are three main components to the quality control of ocean profile data. The first component examines the characteristics of the platform's track looking to identify errors in position and time. The second component, examines the various profiles of observations to identify values which appear to be in error. The third component, identifies duplications of profiles either by having received the data more than once, or because data of lower resolution (such as a BATHY message) will arrive, followed later by the XBT cast on which the BATHY message was based.

The Procedure for Checking a Platform's Track

The track QC procedure examines the position and date/time and, in the case of real-time messages, the call sign of the observations. To carry out these checks, the data are ordered by call sign (treated as the cruise number) and within each cruise, by date and time. Each cruise is passed through tests which check that the date is valid (including future and too far in the past), the latitude and longitude are valid, that the station location is not positioned over land (using a bathymetry file with values every 5 minutes of latitude and longitude and an algorithm that accounts for the resolution of the file), and that the inferred speed between stations is reasonable. Each cruise is plotted to show the cruise number, the track (with scales and land for reference) and, the platform speed from station to station (calculated from the time and space differences between stations).

At the same time as these displays are shown, the software tests for possible errors and presents the results directly to the screen. If a test fails, an appropriate error message and a scrollable and editable table of date/time, latitude/longitude, and their flags is shown. A QC technician then examines the plot (and error messages) and undertakes to assign flags or correct values. The interface allows the technician to select stations, edit values, and see the results of their changes, in order to experiment with solutions to find the logical reason (and fix) for the data. With real-time data, stations with different call signs may be merged into a single cruise if this is appropriate due to an erroneous call sign being reported.

The QC Procedure for Profile Checking

The profile checking software automatically tests each station profile, sets flags accordingly, and displays a plot of the profile and error messages for review and flagging by the QC technician. This is carried out in the following stages. First, a file of stations is opened by the software, and the technician uses a menu to move through this file, station by station. A station is read, and the profiles are identified and tested.

Tests include a group to examine global ranges, bathymetry, single valued profiles and monotonically increasing depths for all known parameter types (e.g. temperature, salinity, oxygen, nitrates, etc.). Next follow a set of statistical tests including regional range, global profile envelops, and a test against the Levitus Climatology. Other tests look for spikes, for gradients that are too pronounced, for density inversions (when temperature and salinity are present) and for temperature inversions (when only temperature is present).

Flags are set according to the severity of the test failure, based solely on the type of test. The profile is always plotted for examination by the technician. Where both temperature and salinity are present, both are plotted, and accompanied by a plot of calculated density. Flags are shown by graphical indicators. The QC technician examines the plot and sets flags by selecting points and menu items using a mouse. This interface provides a wide range of functionality, which allows the QC technician to list the station as a text file, to list flags and other specific information such as the climatology, to adjust scales and zoom, or to plot by arbitrary parameters (e.g. T/S Plot), and to show the cruise track, and location of the station in question.

The QC Procedure for Duplicates Checking

Duplicates checking is necessary to identify data which are versions of the same observation. Exact matches mean that one of the versions has no additional information, is redundant, and is usually deleted. Two or more data records are often found to be the same observation, but differ in their method of

analysis or reporting. In this case both records are kept, and all but the best one is flagged as a duplicate. For example, TESAC messages reported in real-time also arrive at MEDS in a much higher resolution form as the CTD profile itself in delayed mode, and bottle data used to calibrate a particular profile.

Duplicate handling at MEDS is a mix of software rules and algorithms, and a presentation for review and editing by the QC technician. The automatic step is carried out for a particular set of data by first determining the date/time range of the data. All data from the archive in the same range is retrieved and combined with the input data. The data are read through in date/time order, to select groups of stations which fall into a common time window and area (for MEDS this is 15 minutes of time and 5 km). Each of these groups is ordered by preference according to data type (CTD, XBT, TESAC, BATHY...), and their originator or institute from which they were delivered. Then, exact matches are removed, where the characteristics considered are the date/time, latitude/longitude, types and values of all profiles, and instrument type.

Where data are collected within the above defined space and time window but are not exact matches, the subsurface data are compared using an algorithm which selects from each station, a common profile type (e.g. temperature), without distinguishing the instrument type (CTD, XBT...), sets allowable tolerances for comparing the profiles, using the a table of accuracies for instrument type, compares the profiles, depth by depth, using linear interpolated values from the profile with the lower vertical resolution, to the exact depth value of the profile with the higher vertical resolution and returns a ratio of trials to failures for interpretation by the main program rules. Where duplication is proven, the duplicate checker uses program rules to select the best of the exact or inexact duplicate profiles, based on criteria evaluated in the previous steps.

Sometimes the automatic check can not determine if stations and profiles are duplicates, or if they are, which profile is the most desirable one. This often happens when data are very close in time and space, but actually different casts of the same instrument. It also may occur when a data originator delivers a "correction" or updated version of a station after their own revised analysis. These cases are isolated and reported (or displayed in the delayed mode system), for the QC technician to review and flag through an interactive session.

Quality Control Procedures Employed at Scripps.

Scientific quality control at the JEDA centers consists of examining each temperature profile in relation to its nearest neighbors and in relation to the Levitus (1982) climatology. For 1990, 2-4% of temperature profiles submitted to the NODC in both real-time and delayed-mode over the global ocean were reported to be in error by one of the three JEDA Centers. Quality control of historical temperature profiles for the ten years from 1980-1989 was conducted differently than with those procedures described above. Temperature anomalies at all observation locations were required to be less than four times the RMS of the anomalies. Also, preliminary monthly maps of anomalous temperature were constructed at the sea surface, 200 m and 400 m, and examined for outliers (i.e., anomalies that rise unreasonably above the background anomaly field). These were flagged as questionable and eliminated from further consideration. Details of these quality control procedures applied to historical observations is given in a series of annual reports for the period 1985-1989 by Pazan and White (1987, 1988, 1991a, 1991b, 1993).

The following table defines areas considered by SIO to be marginal seas and for which no QC is undertaken.

Marginal Sea	Latitude Extent	Longitude Extent
South China Sea	8 S to 25 N	90 to 122 E
East China Sea	25 to 41 N	115 to 128 E

Sea of Japan	35 to 50 N	128 to 140 E
Sea of Okhotsk	45 to 65 N	135 to 155 E
Bering Sea	52 to 66 N	163 E to 159 W

Errors in Temperature Profiles

Instrumental error of the XBT is nominally 0.1 degree Celsius and 1% of depth. This error is less than the RMS of annual and interannual signals in upper ocean temperature. Limitations to measuring annual and interannual variability derive not from instrumental error, but from subgrid ambient noise due to internal waves, internal tides, mesoscale eddy activity (e.g., White et al., 1982). Often, ambient noise has a variance equal to or larger than that of the climatic signal. The strategy for suppressing this source of noise in interpolation (e.g., Gandin, 1993) is to increase the sampling density beyond what would be required to conduct the interpolation in the absence of noise. Sampling at 2 observations per decorrelation scale in three dimensions is usually enough to suppress subgrid ambient noise.

Instrumental bias in this global observation set is due to the systematic differences between temperature profiling instruments. Of special concern is the underestimation in depths of XBTs due to application of an inaccurate fall-rate formula over the past 30 years (Hanawa et al., 1994). For example, Kessler (1990) found depths of the 20 degree C isotherm in the western tropical Pacific underestimated by 2-4 m compared with Japanese digital bathythermograph depths. On the other hand, since XBTs constitute over 90% of the global population of temperature profiles from 1980-1992, this bias does not effect the anomalies nor their statistics. However, it does affect the mean statistics below approximately 200 meters.

Quality Control Procedures Employed at AOML.

Quality control (QC) procedures developed and implemented at AOML for the examination of real time XBT data are presented. The steps required to QC XBT data is outlined in a "cook book" format. The methods employed are primarily subjective as many of the stages involve interactive input from the user.

AOML is one of several Research Science Centers participating in a global effort to quality control oceanographic data collected from either research vessels or volunteer observing ships (VOS). AOML efforts are currently focused on XBT data collected in the Atlantic Ocean. The data being examined consists of "real-time", delayed, historical and Navy declassified modes.

The "real-time" data represents data obtained by the MEDS offices in Canada from various ships. This data has already been quality controlled by the MEDS office; nevertheless, AOML examines the data, without prejudice to the Canadian flags, and then compares its quality control flags against the Canadian flags. The delayed mode data represent both updated profiles submitted as part of the "real-time" data set and data that was not transmitted in "real-time" for 1990 . The historical data represent all of the XBT data that has been archived by NODC between 1966 and 1989. The Navy declassified data is the data recently released by the Navy covering the period 1985 through 1990.

The discussion that follows gives an overview of the procedures that are used to quality control any of the XBT data sets. For exemplary purposes, the "real-time" data set is discussed. NODC first obtains a data set from the MEDS office. AOML then receives an electronic mail message from NODC indicating that a particular month of data is available. AOML then copies the MEDS data over the SPAN network. The MEDS file is then rewritten into AOML's Indexed Sequential Access Method (ISAM) Database. This is a direct access "keyed" database for which XBTs are retrievable according to various keys, including: Latitude, Longitude, Time, NODC_Unique_ID_Number, NODC_Cruise_Consec_Number, MEDS_Station_Number, Ship_Radio_Call_sign.

Once the database has been updated, several programs are run to examine the XBT profiles. The quality control procedures presently employed consist of the following:

- a duplicate profile test
- a histogram test in which profiles are compared to climatological means and standard deviations to identify possible outliers
- a waterfall plot test in which profiles from a particular cruise are checked for profile to profile consistency
- a position and time test to see if profiles within a cruise are consistent
- a test in which vertical temperature sections of profiles within a cruise are plotted to insure internal consistency
- a mapping test in which monthly distributions of SST, temperature at 150m and the average temperature of the upper 400m are examined for "bulls-eyes"

In general, the AOML procedures can be characterized as more subjective than objective. There are no automatic flags established by any of these tests. Profiles that have been identified as suspicious are examined individually and flagged as appropriate. A flowchart and description of the programs used at AOML, to insure quality control of the XBT database in the Atlantic, follows.

Further details of the procedures carried out at AOML are given in the section on quality control manuals.

Acknowledgements

The authors wish to thank Dr. Robert L. Molinari for his contribution to the completion of this report. This work was supported in part by a grant from NOAA's Long Term Ocean Observing Program.

References

Bailey. R.,A. Gronell, H. Phillips, G. Meyers, and E. Tanner (1994). CSIRO cookbook for Quality Control of Expendable Bathythermograph (XBT) Data. CSIRO Marine Laboratories Report No. 220, 75 pp. (In press).

Quality Control Procedures Employed at CSIRO.

by Rick Bailey, Director WOCE INDIAN OCEAN UOT/DAC, CSIRO Division of Marine Research, Hobart, Australia

The Indian Ocean Upper Ocean Thermal Data Assembly Centre (UOT/DAC) now called the Joint Australian Facility for Ocean Observation Systems, JAFOOS, is jointly operated by the CSIRO Division of Marine Research (CSIRO DMR, formerly Division of Oceanography), and the Australian Bureau of Meteorology Research Centre (BMRC). It is mainly operated by CSIRO DMR in Hobart, with assistance from BMRC in quality control (QC) and analysis software development. Apart from the production of a scientific quality controlled data set for use by WOCE participants by scientists with an intimate knowledge of the region, another objective of the Indian Ocean UOT/DAC is to pass on to operational data centres as much scientific and software expertise as possible.

Scientific quality control is the process of combining statistical analyses of the data (in this case, subsurface temperature profiles) with knowledge of historical means (climatology) and of the relevant environment (regional oceanography) to make a scientifically based decision about the validity of each data point. It is a vital step to be taken with any data set before scientific analysis can proceed. A knowledge of the recording instrument's performance and characteristics must also be used when evaluating the data.

CSIRO DMR and BMRC have developed a comprehensive software system for the scientific quality control of sub-surface ocean temperature data. This system, called QUEST (which stands for Quality

Evaluation of Subsurface Temperatures) combines the subsurface ocean temperature statistical analysis scheme developed by BMRC and CSIRO (Smith et al., 1991; Meyers et al., 1991; Phillips et al. 1990) with the quality control procedures of CSIRO as outlined in Bailey et al. (1994) (henceforth referred to as the CSIRO Cookbook). QUEST enables individual temperature profiles to be compared to climatology (Levitus), to an objective (statistical) analysis of all the available data, and to other profiles in the immediate area, in order to identify real features of a given region and to help distinguish between such features and erroneous data. All scientific decisions are recorded/flagged with the data according to the GTSPP, CSIRO Cookbook and WOCE flagging/coding schemes. Maps and sections of the analyzed data are used to complete the QC process.

The goals of scientific quality control at the Indian Ocean UOT DAC are:

● To identify and flag the profiles affected by malfunctioning of the instrumentation which have been missed by the originators and data centres;
● To recover data which has been erroneously flagged as bad by the originators and data centres, and which potentially represent climate signals;
● To identify the areas where vertical and horizontal temperature structure with small to medium scales occurs frequently (this in turn assists with the scientific QC process). The vertical and horizontal structure includes inversions, intrusions, steps, thermostads, eddies, low salinity layers at the surface due to rainfall and runoff, and unusual vertical structure associated with interactions between strong coastal currents and the continental slope.

Details of the test procedures used by CSIRO are given in the section on quality control manuals.

References

Bailey, R., A. Gronell, H. Phillips, G. Meyers and E. Tanner, 1994: CSIRO Cookbook for Quality Control of Expendable Bathythermograph (XBT) Data. CSIRO Marine Laboratories Report, 221, 75pp.

Meyers, G., H. Phillips, N. Smith and J. Sprintall, 1991: Space and time scales for optimal interpolation of temperature - Tropical Pacific Ocean. Prog. Oceanog., 28, 189-218.

Phillips, H., R.J.Bailey, and G.Meyers (1990): Design of an ocean temperature observing network in the seas north of Australia. Part II, Tropical Indian Ocean: Statistics. CSIRO Marine Laboratories Report, No.211.

Smith, N., J.E. Blomley and G. Meyers, 1991: A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. Prog. Oceanog., 28, 219-256.