S E N S 2 0 0 7
Third Scientific Conference with International Participation
SPACE, ECOLOGY, NANOTECHNOLOGY, SAFETY
27–29 June 2007, Varna, Bulgaria

SOME ASPECTS OF QUALITY CONTROL OF TEMPERATURE DATA ACQUIRED WITH A CTD SYSTEM

Nadezhda Valcheva, Atanas Palazov

Institute of Oceanology - Bulgarian Academy of Sciences
e-mail: valcheva@io-bas.bg, palazov@io-bas.bg

Keywords: *Data quality control, sea temperature, CTD system*

Abstract: *Data quality control, or data validation, is an essential part of environmental data management, which primary objective is to ensure timely, efficient and open access to the best possible data, metadata and associated products, for use and re-use throughout their life-cycle and to prevent loss of data and associated information. There are several aspects of ocean data validation among which most important are primary automatic quality control of data and scientific/expert quality control.*
The discussion on the data quality measures in the paper is followed by examples of their application to temperature data obtained within the monitoring program of the Institute of Oceanology. Due to large variety of data sources only data acquired with the CTD Sea Bird 911plus system obtained in the period 2002-2006 are considered. The described quality control procedure validates the pressure (or depth) and temperature data as well as the principal metadata concerning the time-space coordinates of the CTD profile, as well as comprises checks on trends, expected correlations between variables, and comparisons with other data sources. Eventually, other variables, such as dissolved oxygen, chlorophyll, and nutrients are included as well. More specifically the oceanographic data editing and quality control examined here includes: spotting and identification of outliers or offsets; editing of data value and quality flagging; automatic range checks and manual or automatic editing; and climatologic expert assessment.

1. Introduction

Quality control is a vital prerequisite to scientific analysis of environmental data that ensures the information on how data was gathered, checked, processed, what errors were found, and how they have been corrected or flagged. It allows data from different sources to be integrated in long time series, and included in the oceanographic management system. Investigations of marine environment often require complex and large national and international research activities (Manual and Guides 26-Manual, 1993). Their outcome needs a data management plan which includes details about the data quality control in addition to a scientific and measurement plan. This quality control comprises all actions of the data originator in connection with data collection and validation. Only after these tests should the data be included in a database or distributed to users via international or national data exchange. The objective of data quality control is to ensure the data consistency within a single data set and within a collection of data sets, and to ensure that the quality and errors of the data are apparent to the user, who has sufficient information to assess its suitability for a task.

The four major aspects of oceanographic data validation are instrumentation checks and calibrations, complete documentation about the field measurements (location, duration of measurements; method of deployment, sampling scheme, etc.), automatic quality control and scientific or expert quality control which is the process of combining statistical analyses of the data with knowledge of historical means and of the relevant environment to make a scientifically based decision about the validity of each data point (UNESCO Guide, 1999).

Since the climatologic test constitutes an important step of quality assurance, there is need main hydrological features to be well-known.

The Black Sea has a specific hydrological structure that is due to the restricted water exchange with the other parts of the World Ocean more specifically with the Mediterranean through Bosphorus strait and the significant volume of fresh water inflow. River run-off is concentrated basically in the north-western part of the Black Sea while the warmer (13 - 16°C) and saltier (28 - 34‰) the Marmora Sea inflow – in the south-western part of the sea. The less dense waters circulate in the upper 50m layer through currents and turbulent processes. The main Marmora Sea water mass is denser and is distributed in deep water layers.

During the winter surface layer cools down to 6 - 7°C and the convective mixing spread to the depth of 50 - 60 m. Below the convective layer water temperature slightly increases up to 8.6 – 8.8°C at depth of 200m. At depth of 1000m it is 9°C and temperature remains stable to the bottom. In summer surface temperature is 23 - 25°C. Depth of surface mixed layer is 15 – 20m and rarely reaches 30 m depth after strong storms. It is followed by cold intermediate layer with absolute vertical minimum of temperature. Water temperature in this layer is near to winter temperature in convective layer. The seasonal thermocline of abrupt temperature diminution is formed between theses two layers. The cold intermediate layer is a unique feature of water stratification that in the Black Sea case is salinity dominated (Blatov et al., 1984).

2. Data and methods

2.1. Data description

The temperature data are derived through the CTD casts during the RV "Akademik" cruises in the Western Black Sea carried out in 2002-2006. The spatial distribution of observations conforms to the monitoring scheme of the Institute of Oceanology (Palazov *et al.,* 2005) shown in Fig. 1. The raw data are processed with SBE Data Processing tool and are consequently submitted to standard quality control procedures.

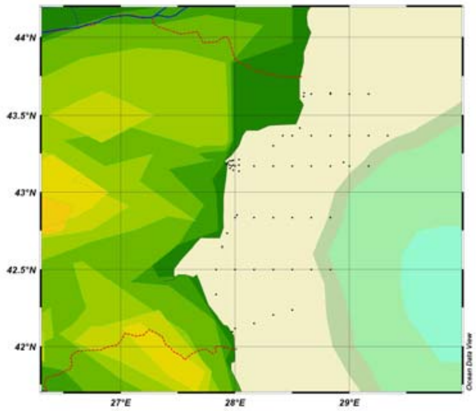


Figure 1. Station spatial distribution within the sea water monitoring scheme of the Institute of Oceanology.

The raw data are processed with *SBE Data Processing* tool and are consequently submitted to standard quality control procedures.

2.2. Quality control procedures

Current quality control procedure validates the pressure (or depth) and temperature as well as the principal metadata concerning the time and space coordinates of the CTD profile. A range of checks are carried out on the CTD data to ensure that they have been imported into the database correctly and without any loss of information (Manual and Guides 26-Manual, 1993). They include:

- Check header details (metadata) such as station numbers, date/time, latitude/longitude, instrument type, data type *etc*;
- Check ship speed between two consecutive stations (e.g., speed greater than 15 knots means wrong station date or wrong station location);
- Position on land test checks whether the profile position falls within the land polygons describing the area. Generally, for the Western part of the Black Sea the area covers from 27.25 to 29.5 in longitude and from 42 to 44 in latitude;
- Check pressure increasing finds record shallower than the preceding one;
- Check no data points below bottom depth;
- Broad range checks - check for extreme regional values compared with the minimum and maximum values for the region. The broad range check is performed before the narrow range check;
- Spikes detection;
- Narrow range check - comparison with pre-existing climatological statistics;
- Check of individual or groups of plot profiles for consistency;
- Check profiles versus climatology for the region.

As a result, to each numerical values of vertical profiles data sets, which are submitted to above described set of checks, is assigned an appropriate quality flag. No data are modified by the quality control procedure. If the data must be modified for some reason, these modifications are implemented outside the quality control procedure and the quality flags must be adjusted in consequence.

Several software packages performing data quality control have been developed in the framework of European and international projects. For the purposes of the present study quality checks are carried out using Ocean Data View (Schlitzer, 2006). This software offers wide range of data visualization and processing tools. The result of quality control is to add a quality flag to each numerical value. Quality flags are 0 for good data, 1 for unknown, 4 for questionable and 8 for bad data.

2.3. Climatology used

The best-known climatologic datasets are World Ocean Atlas editions 2005. However, this set does not include any CTD data for the Black Sea. Therefore, MEDAR/MEDATLAS regional climatology (Mediterranean and Black sea) is used. Although it does not offer sufficient data volume in its part for the Black Sea it is the only one available set for regional and sub-regional data quality validation.

MEDATLAS database of temperature salinity and bio-chemical parameters climatologic Atlas is created to rescue, safeguard and make available a comprehensive dataset of oceanographic parameters collected in the Mediterranean and Black Sea, through a wide co-operation of the Mediterranean and the Black Sea countries, where the need for marine database and appropriate data management is particularly crucial (Medatlas Group, 1994). The calculation scheme of climatic temperature and salinity fields included three steps: calculation of statistical characteristics (climatic norms) for 1°- squares; interpolation of the calculated data through objective analysis procedures; mapping and visualization.

3. Discussion

3.1. Preliminary tests

Calibration procedure of the CTD sensors is the first phase of the quality control. To obtain the highest quality data, necessary corrections need to be applied through pre- and post-cruise calibrations of pressure and temperature sensors in the laboratory. Further within this preliminary procedure, the header information is tested for correct date, time, position and other operator errors as described above. Special attention is paid to records with increase of pressure as they produce artificial spikes. They are a negligible part of the entire dataset; therefore, these minor recording errors are edited. The temperature profiles undergo visual inspection to identify and flag erroneous data within the profiles and erroneous profiles within the cruise.

3.2. Control checks against climatologic statistics

Control checks against climatologic statistics assumes primarily broad and narrow rang tests for outliers definition. The selected dataset is checked against broad and narrow range thresholds established within MEDAR/MEDATALAS Protocol (Maillard et al., 2001).

A good example is the dataset obtained during the March 2003 cruise. Checking the data with the broad range tests results that all records falls into the prescribed interval. However, the narrow ranges test detects 292 outliers. All cases correspond to temperatures lower than 3ºC (Fig. 2a). Although such local cooling in shallow water coastal areas is possible, such temperature drop is unlikely to occur in particular considering the reported atmospheric conditions before and during the cruise: air temperature varying between 0 and 4ºC and wind speed 5-12 m/s blowing from the northern quarter. Moreover, such an extreme is neither reported in MEDAR/MEDATLAS nor recorded during another monitoring cruise of the Institute of Oceanology. Therefore, a questionable flag is assigned to corresponding data values.

a) b)

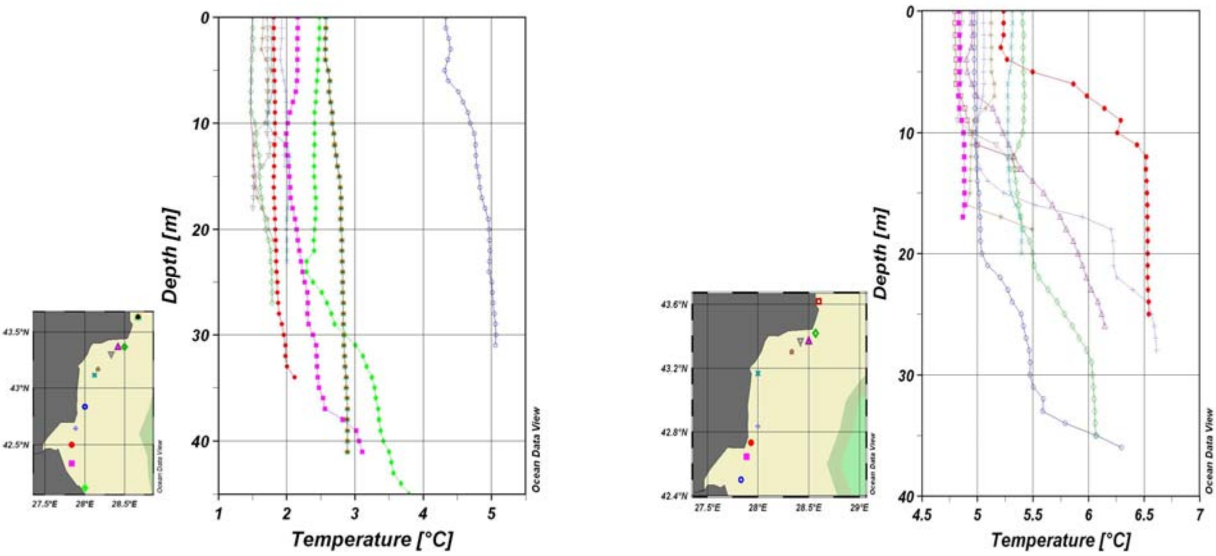


Figure 2. Comparison of doubtful profiles from March 2003 cruise (a) with those obtained during the March 2005 cruise (b) at similar atmospheric conditions.

The profile-to-profile comparison between doubtful and neighboring values shows that the supposed cooling process is isolated close to the shore. Further, evaluation of the same profiles against those obtained during the March 2005 cruise at similar atmospheric conditions (Fig. 2b) is another proof for the supposed inconsistency of the recorded profiles.

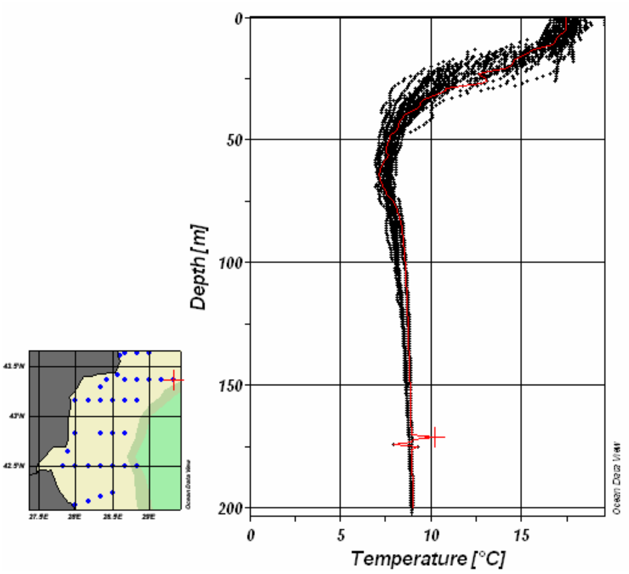


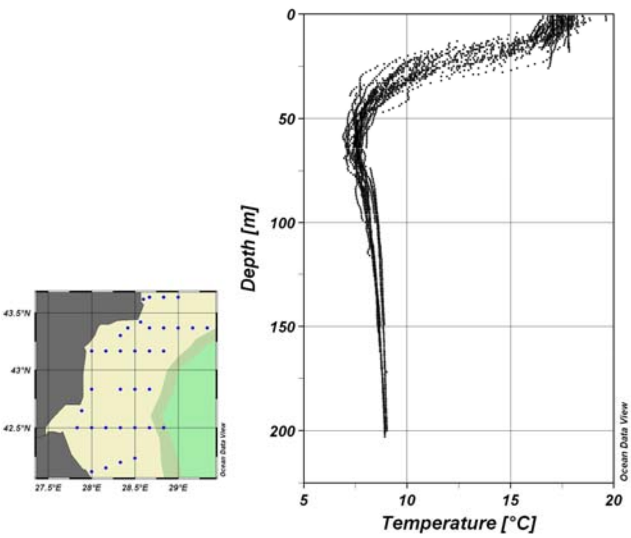
Figure 3. Example of spike detection for June 2004 cruise.

Another data test is associated with detection of spikes. Here, it is illustrated with vertical profiles derived during the cruise performed in June 2004 (Fig. 3). Two spikes are detected at depths between 150 and 200 m, which existence is ungrounded considering climatic vertical distribution of temperature. Estimated spike values are 1.18ºC и 1.4ºC, respectively. They exceed the threshold value set for the Black Sea. Therefore, flag for bad data is assigned. Both points that fail this test are bad and their quality control flags are set to 8. The corrected profile can be seen in Fig. 4a.

3.3. Comparison with regional climatology

The principle of the QC of any parameter is to compare the observations with the available statistics of the same parameter. These statistics vary from a region to another, and the checks are adjusted accordingly. The studied dataset is used for gross compared against MEDAR/MEDATLAS profiles for the Western part of the Black Sea.

a)



b)

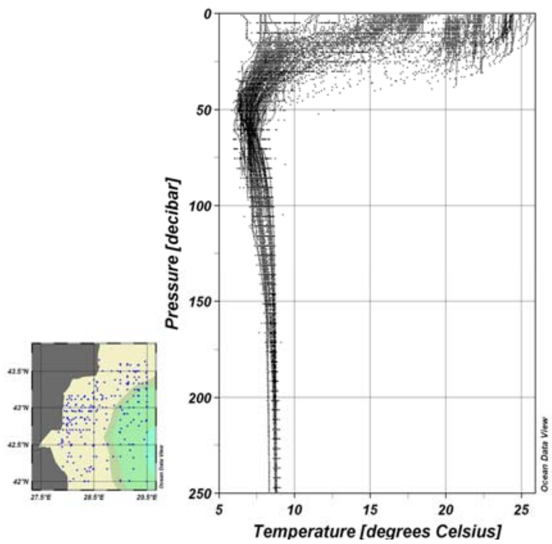


Figure 4. Comparison of CTD dataset with regional climatology: a) quality controlled cruise dataset; b) MEDAR/MEDATLAS excerpt for the Western Black Sea

Fig. 4 corroborates that all specific features of the Black Sea vertical structure are in conformance with the regional climatologic excerpt.

4. Conclusions

A methodic for primary checks and delayed mode scientific quality control is established encompassing various tests and checks that could be successfully applied to other types of sea data such as salinity and density derived by CTD system, bottles or XBTs. Results show that high quality data were collated during the studied period 2002-2006 that can comfortably be included in the database of Bulgarian ODC. Available regional climatology is in good agreement with the data obtained. However, there is need to be further expanded to provide tools for more reliable scientific assessment of the data quality in the Western Black Sea.

References:

1. B l a t o v A., N. B u l g a k o v, V. I v a n o v, 1984. Variability of the Black Sea Hydrophysical fields. Leningrad, Gidrometeoizdat, 238
2. Manual and Guides 26-Manual of quality control procedures for validation of oceanographic data, Prepared by: CEC: DG-XII, MAST and IOC: IODE, SC-93/WS-19 1993. UNESCO
3. M a i l l a r d C., M. F i c h a u t, MEDAR/MEDATLAS Group, MEDAR/MEDATLAS Protocol, Part I : Exchange format and quality checks for observed profiles, 2001.
4. Medatlas Group: Specifications for Mediterranean data banking and regional quality controls, IFREMER, Direction Scientifique, Sismer-Brest, SISMER/IS/94-014, pp. 29, 1994.
5. P a l a z o v A., H. S l a b a k o v H.D., H. S t a n c h e v - Monitoring programme of western Black Sea, Proceedings of the Workshop on “Black Sea Ecosystem in Support of Marine Conventions and Environmental Policies”, EUR 22176, European Communities, 2006, pp 40-46.
6. S c h l i t z e r R., Ocean Data View User’s Guide, 2006.
7. UNESCO: Guide to operational procedures for the collection and exchange of JCOMM oceanographic data - third revised version, Unesco Manuals and guides 3, pp. 35, 1999.