GODAE Ocean Data Quality Control Intercomparison Project

James Cummings[1], Gary Brassington[2], Robert Keeley[3], Matthew Martin[4], Thierry Carval[5]

[1]*Oceanography Division, Naval Research Laboratory, 7 Grace Hopper Ave Stop 2, Monterey, California, USA 93943, Email: james.cummings@nrlmry.navy.mil*
[2]*CAWCR, Bureau of Meteorology, Melbourne, Australia*
[3]*Department Fisheries and Oceans, Government of Canada, Ottawa, Ontario, Canada*
[4]*Met Office, Exeter, United Kingdom*
[5]*Ifremer, Technopole de Brest-Iroise, Plouzane, France*

1. Introduction

Ocean data quality control is a fundamental component of any ocean analysis/forecast system. Using or accepting erroneous data in the assimilation can cause an invalid conclusion to be made or an incorrect analysis. Alternatively, rejecting extreme, but valid, data can miss the detection of important events and anomalous features. Quality control, therefore, must correctly identify observations that are obviously in error, as well as the more difficult process of identifying measurements that fall within valid and reasonable ranges, but nevertheless are erroneous. It is likely that decisions made at the quality control step affect the success or failure of the entire analysis/forecast system.

Effective quality control requires a set of pre-established, standardized test procedures, with results of the procedures clearly associated with the data values. Effectiveness in turn depends on the reliability of the standard(s) and on the choices for measuring goodness of fit. Users of quality controlled data sets have a wide range of views on the most appropriate standards and on the appropriate "tightness of fit" demanded by the quality control procedures (too tight increases the chance of erroneously rejecting anomalous features; too loose increases the chance of accepting bad data). Indicators of data quality, therefore, must be useful for determining if the quality controlled observations are appropriate for a particular purpose.

2. Motivation

An integrated, end-to-end system must ensure that the results of the quality control procedures are recorded for independent analysis and later use. If the quality control is carried out well, then it can reduce the duplication of effort among the users of ocean data − value added is not lost or misinterpreted. At present, there are few agreed-upon standards for real-time ocean data quality control and very few cases where the procedures and results from the oceanographic centers have been compared. As the GODAE operational oceanographic community continues to develop a range of complex ocean analysis and prediction systems, it is important that procedures be developed for routinely assessing the effectiveness of ocean data quality control and for routinely exchanging statistics from the quality control processes at the operational centers.

Automated ocean data quality control involves specification of an observation error model that is used to determine "good" (consistent with the error model) and "bad" (violates the error model) observations. Evaluation of the effectiveness of an ocean data quality control system and the quality control procedures must include: (1) test of the hypothesis of the observation error model; and (2) tests of the quality of the decisions made using the observation error model. Typically, observation error models assume a normal probability distribution function. The validity of this assumption needs to be confirmed by formal statistical tests and by examining differences between automated and delayed-mode quality control outcomes on the same observation. The automated quality control system can be considered to have performed well if rejected observations are consistent with those modified or rejected by the delayed mode quality control. Delayed mode quality control outcomes of the Argo profiling float array are readily available and can be used in this evaluation.

3. Intercomparison Project

A workshop was organized prior to the Biarritz GODAE symposium in 2002 to discuss the potential and priorities for the exchange of information and collaboration on the quality control of ocean observations. The workshop was the initial step in a process that has evolved into a comprehensive ocean data quality control intercomparison project. Currently, outcomes of profile data quality control procedures from 5 oceanographic centers are available on the U.S. GODAE server: http://www.usgodae.org/ftp/outgoing/godae_qc. The contributing centers include: (1) U.S. Navy Fleet Numerical Meteorology and Oceanography Center (FNMOC); (2) U.K. Met Office (UKMO); (3) Canadian Integrated Science Data Management Branch (ISDM, formerly the Marine Environmental Data Service); (4) Australian Bureau of Meteorology (BoM); and (5) French Coriolis Data Center (Coriolis). Figure 1 illustrates the process of matching daily inputs of profile QC data from the centers and the creation of NetCDF formatted WMO call sign data files. A WMO call sign file contains the entire time history of the reporting platform and all of the quality control information that was used by the centers to determine

whether the observation was acceptable for use in an analysis/forecast system. The GODAE QC processing creates a new variable for each profile in a WMO call sign file that indicates if there were differences among the centers on the determination of profile quality. This so-called *conflict* variable allows a quick search capability to find cases where different QC decisions were made by the centers on the same observation. The WMO call sign files exist on the US GODAE server for the time period from 2004 to the present and are updated daily as new profile data quality control information is received from the centers.

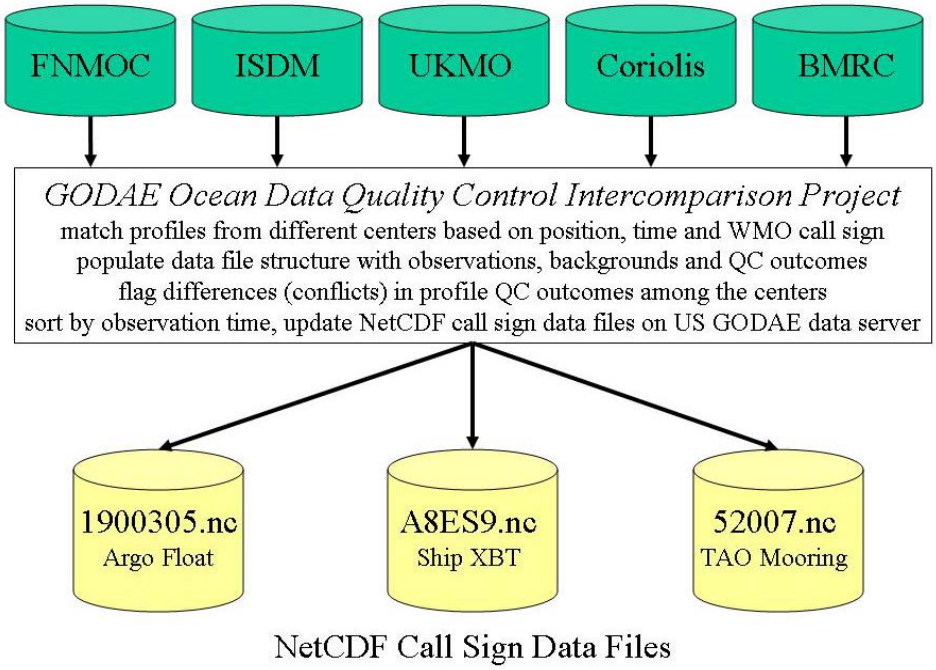

Figure 1. Schematic diagram of oceanographic center profile processing in the GODAE QC project showing input of center QC data files and production of NetCDF formatted WMO call sign data files. A WMO call sign file contains the entire time history of QC outcomes from the different centers for that platform.

4. Applications

The WMO call sign files have many applications. First, the call sign files allow GOOS data providers access to information about the fate of their data in GODAE analysis/forecast systems. Oceanographic centers are in the best position to operate ocean data quality control systems and the call sign data files provide a way to facilitate the relay of real-time QC information back to program managers and operators regarding utilization of the buoy, XBT, and profiling float observing networks in GODAE systems. Second, the WMO call sign files provide a way for the oceanographic centers to compare their ocean data quality control systems. The quality control procedures being used at the centers are expected to substantially vary depending upon the type of data being considered and whether extensive use is made of ocean model first guess fields or whether more specific tools (e.g., instrumentation error checks), cross validation checks, manual checks, and comparisons with climatology are used at the center. Finally, the time history aspect of the WMO call sign files provide a natural way to look at systematic problems (bias) in the reporting platform, such as sensor drift or calibration errors.

5. Future

All of the work to date has concentrated on building the facility to compare results from the different contributing centers. In a parallel development, there has been a push from international and national organizations to standardize a variety of procedures, including quality control procedures. A meeting held in Ostende, Belgium in 2007 developed a process for international cooperation to accept and recommend standards for community use (see http://oceandatastandards.org). For example, the procedures used by ISDM and Coriolis are very similar and are firming up for submission to this standards process.

A careful analysis of the quality control intercomparison results from the GODAE systems will be valuable in identifying advantages and weaknesses of particular ocean data quality control procedures. Once a thorough analysis of the intercomparison results are available it will be useful to organize a meeting of representatives from the GODAE centers to review the results, strike agreements on the successful procedures, document these procedures, and use the international process to encourage their adoption on a wider scale.

6. **Summary**

In this paper we review the development and future of the GODAE Ocean Data Quality Control Intercomparison Project. We describe the design and process of generating the WMO call sign data files using daily outputs of profile data quality control information from the oceanographic centers. The WMO call sign data files represent the starting point of all follow-on analysis and intercomparison of ocean data QC outcomes. The project is initially focusing on the ocean profile data, but the system can easily be expanded to include additional ocean data types, measurement variables, and analysis/application tools.