ARGO DATA MANAGEMENT

S. Pouliquen[1], C. Schmid[2], A. Wong[3], S. Guinehut[5], M. Belbeoch[4], & Argo Data Management Team

[1]*Ifremer, Centre de Brest, Plouzané, France*
[2] *NOAA/AOML , Miami, FL, USA*
[3]*UW, Seattle,,USA*
[4]*IOC, Toulouse, France*
[5]*CLS , Toulouse, France*

Abstract

During the past decade Argo has revolutionized the distribution of ocean data within the research communty [5]. People used to go to sea, acquire data, process them, submit one or more publications using these data and finally submit them to their national center that periodically transfers the new data to a World Data Center. WOCE had managed to reduce this exclusivity delay to about 2 years. With Argo, it was stated since the beginning that the data will be freely accessible in real-time both on GTS and Internet to serve the meteorological and oceanographic communities in operational and research capacities. Thanks to a great collaboration between the contributing teams, Argo managed to set up efficient and homogeneous data processing in real-time and in delayed-mode, as well as easy access through two Global Data Centers located in USA and France. Similar data system organization has then been endorsed by other components of the GOOS observing system.

1. THE CONTEXT

Over the past 10 years, the great technological advances in data storage, telecommunications and IT infrastructure at a global level have provided the platform on which real time and near-real time ocean data centers have been considerably enhanced in order to meet the needs of operational oceanography and research applications. Consolidated and more robust user requirements from ocean and atmospheric forecasting systems have also been derived based in part on the activities of GODAE and data assimilation systems have matured to the point where the use of large observational data sets in an operational context is now feasible. As a consequence, new or improved input data sets, e.g., of better quality and timeliness as well as with better characterization of data errors, and data products are now available. The Argo dataset has therefore become an important dataset for a lot of applications.

2. ARGO: A BREAKTHROUGH IN DATA MANAGEMENT AND DATA PROCESSING

The issue for Argo, apart from the obvious need to populate the ocean with suitable profiling floats, was to set up an information system that is able to provide a single entry point for data processed in national centers applying commonly defined quality control procedures at all steps of data processing. Two data streams have been identified (Fig. 1): a real-time data stream and a delayed-mode data stream. The real-time data stream delivers data that have been checked for gross errors and corrected in real-time if a correction is known, for example, a salinity drift determined in the delayed-mode process. The delayed-mode data stream delivers data that have been subjected to detailed scrutiny by oceanographic experts and adjusted based on comparison with high quality ship-based CTD data and climatologies.

The main actors identified in Argo data management are:

- DACs: The Data Assembly Centers receive the data via satellite transmission, decode and quality control the data according to a set of 17 real-time automatic tests agreed within Argo. Erroneous data are corrected if possible, flagged according to standardized rules, and then passed to the two global data centers and the GTS (Global Telecommunication System of WMO). Each float is under the responsibility of a unique DAC at all stages of its processing. Schmid et al. (2007) describe the data processing system developed by the US DAC.
- **GDACs:** The Global Data Assembly Centers, located at Coriolis/France and NRL/USA,) are in charge of collecting the processed Argo data from the 10 DACs and providing unique access to the best version of Argo profiles to the users. Data are available in a common NetCDF format both on FTP and WWW sites : (http://www.coriolis.eu.org) (http://www.usgodae.org/argo/argo.html). The two GDACs synchronize their database every day to ensure they provide access to the same dataset.
- **ARCs:** The Argo Regional Centers look at data from ocean basins to verify float data consistency and generate products. They provide basin-wide synthesis of all float data, other available data, and feedback to Argo scientists that are in charge of the delayed-mode quality control via the AIC.
- **AIC**: The Argo Information Centre, located in Toulouse/France, (http://argo.jcommops.org/), is in charge of information on the Argo program status. It monitors closely the Argo data distribution and acts also as a support centre to assist users, gather their

feedback on data quality and relay it to data producers.

- **Delayed-Mode Operators** are in charge of the delayed-mode processing of the float data in collaboration with the Argo scientists.

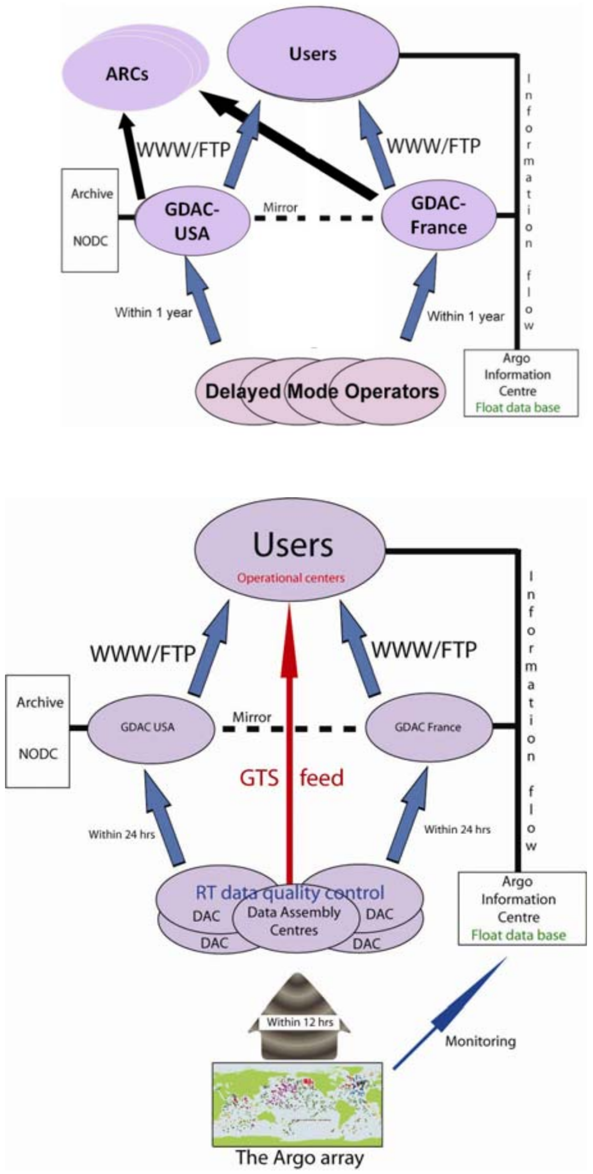

Figure 1: Argo data flows and actors for real-time (left) and delayed-mode (right) processing

This architecture has proven to be efficient, robust, able to serve both operational and research communities and sustainable in the long term. This model has been adopted by other international programs such as GOSUD (Global Oceanographic Surface Underway Data) and OceanSITES (Deep Ocean Eulerian observatories) which have both DACs and GDACs and have extended Argo NetCDF format to handle their data.

3. REAL TIME QUALITY CONTROL

To be able to serve operational users, Argo data have to be processed within 24h of collection in the best possible quality. This is achieved by using a set of automatic quality control tests that detects various errors, for example bad date, bad location, bad platform identification, stuck value, spike, gradient or density anomalies, gross salinity or temperature sensor drift (Fig. 2). The data that pass these tests are sent automatically on the GTS. The profiles are also sent to the GDACs but good and bad data are both provided and are accompanied by QC flags. A user should never use Argo data without looking at the QC flags. To be on the safe side, a user should only use data with QC flags equal to 1.

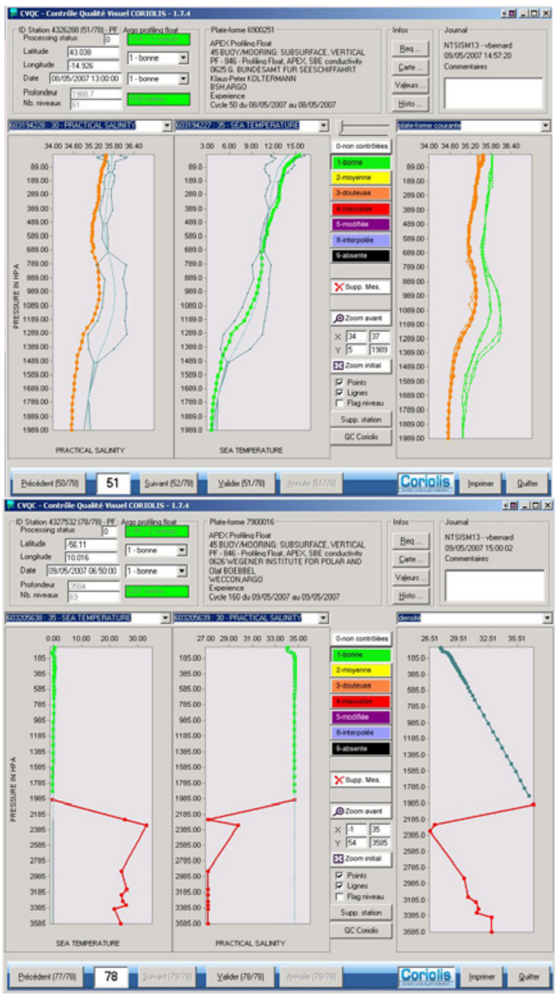

Figure 2: Examples of anomalies detected in real Time

However, real-time tests cannot detect all anomalies because:

- Real-time automatic test thresholds are a compromise in the sense that they are designed to let some bad data go through instead of stopping the distribution of good data.

- Real-time tests are conducted on individual floats only and therefore cannot detect systematic errors that can only be revealed by comparison with other independent data.

So users must be cautious while using real-time data for applications that need a high level of accuracy.

4. DELAYED MODE QUALITY CONTROL

One of the primary objectives of Argo's data management system is to provide research-quality Argo data in a timely fashion. This is done in delayed-mode through the combined use of statistical tools and validation by scientific experts. The central task is the estimation and correction of multi-year calibration drift in salinity due to bio-fouling or other causes (Fig. 3). This task is conducted by comparing the time series from each Argo float with nearby reference data (Wong *et al.*, 2003 [7]; Böhme and Send, 2005 [2]; Owens and Wong, 2009 [4]). The reference dataset used for this comparison is made up primarily of high-quality shipboard CTD data from research cruises, and is supplemented by the more plentiful dataset of previously-verified Argo data. Scientific judgment and regional expertise come into play whenever the reference data provide ambiguous or possibly outdated information, and if nearby Argo data tell a different story.

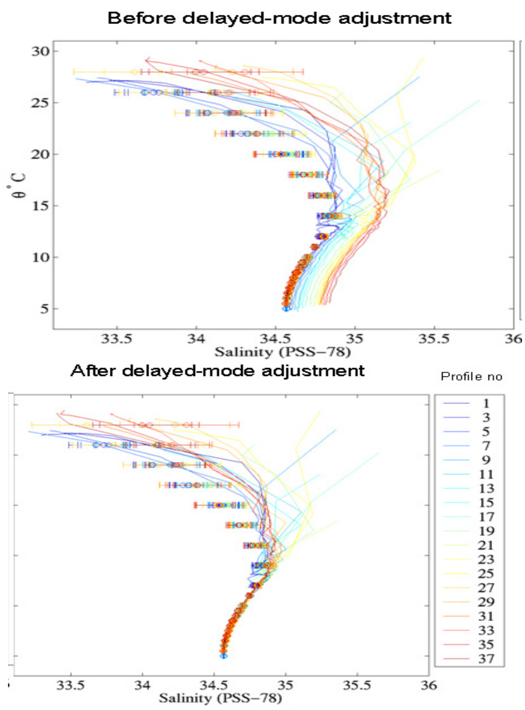

Figure 3 Example of an Argo float whose salinity measurements (solid lines) have drifted towards higher values over time. Sensor drift is removed in delayed-mode by weighted least squares fit to statistical salinity estimates from reference data (circles with error bars).

Fortunately, the accuracy and stability of Argo salinity sensors exceed original expectations, with most instruments showing no detectable calibration drift for the first several years of deployment. Successful development of stable low-power salinity sensors by Sea-Bird Electronics, in partnership with the Argo Program, has made high data quality possible.

5. TOWARDS AN ENHANCED ARGO DATASET

In addition to the real-time tests and delayed-mode adjustments, the Argo quality control process includes a third level of QC tests, some still under development, for identification of systematic and random errors. These include (i) comparison of Argo data with climatological mean and variability, (ii) comparison of satellite altimetric height with steric height from sequences of Argo profiles (Guinehut *et al.*, 2009) to flag suspect instruments for further examination, and (iii) comparison of nearby floats ("buddies") of differing type, origin, or age to reveal systematic differences. All of these tests become more useful and accurate as the Argo dataset grows and its statistics are better known.

The Guinehut et al. (2009) [3] method compares co-located sea level anomalies (SLA) from altimeter measurements and dynamic height anomalies (DHA) calculated from Argo temperature and salinity profiles for each Argo float time series. By exploiting the correlation that exists between the two data sets and a priori statistical information on their differences, altimeter measurements can be used to extract random or systematic errors in the Argo float time series. Different types of anomalies (drift, bias, spikes, etc) have been identified (Fig 4). As dynamic height includes integrated effects of pressure, temperature and salinity, the method allows a quick look at the general behaviour of the float time series but further examination of suspect instruments is needed.

The South Atlantic Argo Regional Centre has developed a system that allows comparisons of profiles from floats with various climatologies and nearby profiles from other instruments. The climatologies used are Levitus World Ocean Atlas 2005 (WOA 2005) and Navy GDEM3. The Argo climatology from Scripps Institution of Oceanography will be added in the future. Buddy comparisons are done with nearby CTDs, XBTs and floats. Information on floats can be obtained via tables [6], for example by Principal Investigator or DAC, and by summary plots (Fig. 5). The float specific pages display the locations and profile data. Profile specific pages for each float show the data used to derive the differences shown in the summary plots. A prototype version of the web page displaying the results is available here:
http://www.aoml.noaa.gov/phod/sardac/post_dmqc/delay_mode.html

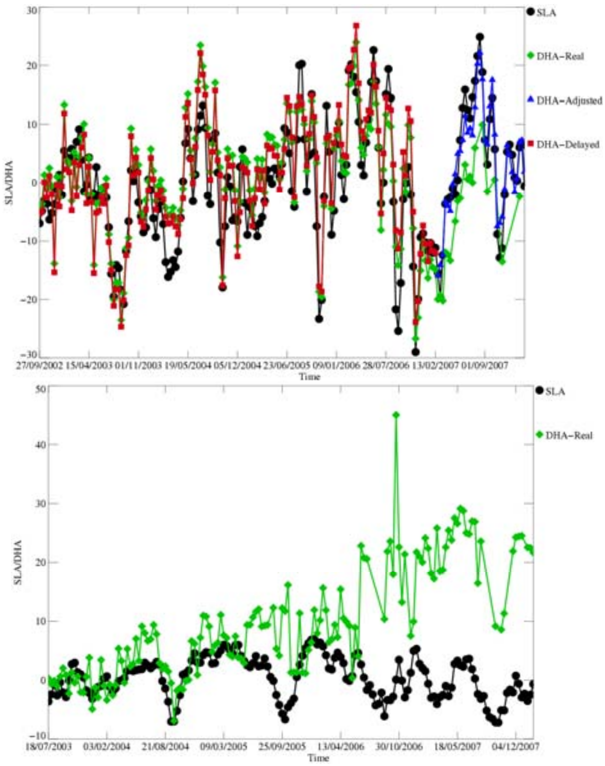

Figure 4: Examples of two Argo floats time series. The first one (left) shows perfect match between Argo steric height and altimeter height – the impact of the delayed-mode adjustment is also clearly visible. The second time series (right) shows that Argo steric height time series (green) has derived over time from the altimetric height (black).

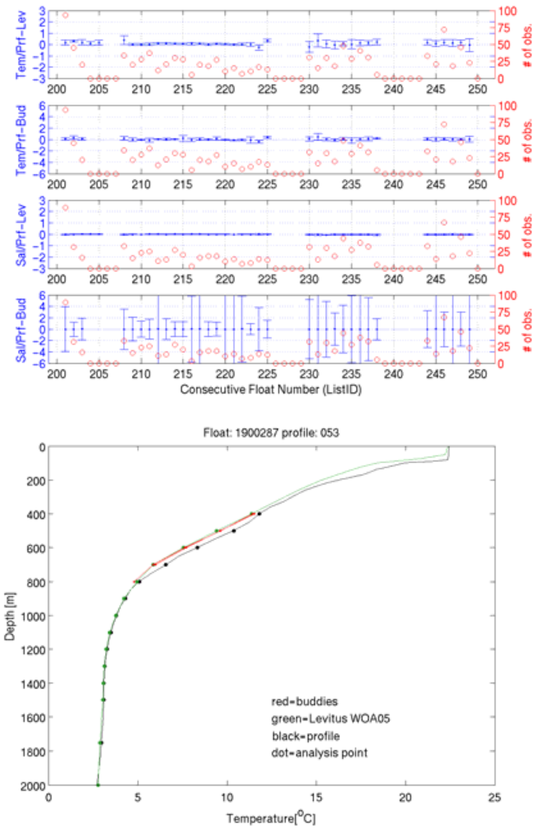

Figure 5: Left panels: Examples of a summary plot showing the temperature (upper two panels) and salinity (lower two panels) deviations with respect to the Levitus Climatology (WOA 2005, panels 1 & 3 from the top) and nearby profiles (panels 2 and 4 from the top). Right panel: Example of the comparison of a profile with climatology and its buddies. The statistics of the differences are provided online together with this image

6. AIC: ASSISTING DATA MANAGERS AND DATA USERS

Every float deployment is firstly and officially registered at the Argo Information Center (AIC) [1]. The registration process checks crucial metadata integrity (e.g. as float identifiers) to avoid future data distribution problems. Thanks to its direct connection with telecommunication providers, GTS centers, and the Argo GDACs, the AIC tracks in real-time every instrument that gives a pulse and verifies that its data are distributed as appropriate on both channels (GTS/GDACs). A set of routine checks permits correction of some errors in the data and metadata served by the GDACs. Data managers can check their data distribution status on-line, and are reminded regularly through the AIC monthly report to fix problems identified and to set up data distribution for new floats. As a consequence of this tracking, 99% of the Argo fleet distributes data as appropriate at a given time.

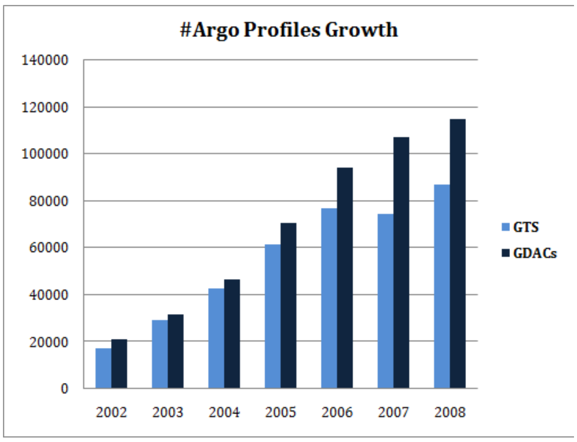

Figure 6 - 1: growth of float profiles on GTS and GDACs. Argo delivers more than 100 000 profiles per year. Less data are distributed on the GTS as existing format does not allow quality flags (grey-listed floats are excluded from GTS distribution and delayed mode profiles can't be distributed in a real-time system).

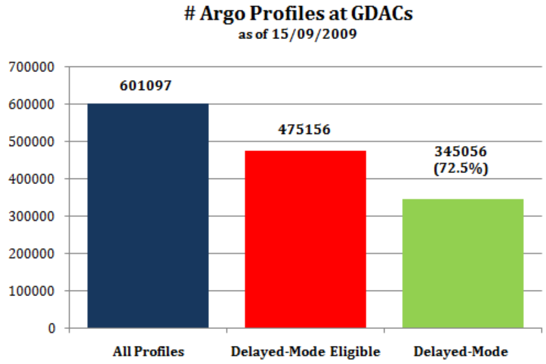

Figure 7 - 2: GDACs status as of September 2009: more than half-million profiles freely available. The Argo delayed-mode operators are putting a lot of efforts in optimizing the Argo dataset quality.

Argo data users are encouraged to use the AIC Support Centre (http://support.argo.net, support@argo.net) to obtain technical assistance or feedback on the data quality.

7. FUTURE IMPROVEMENTS

One of the main future challenges in Argo Data Management is detection of systematic errors in batches of floats with common origins. In the past, some of these "batch errors" have shown to be significant in affecting research results of global scales. The process for identifying these more subtle errors requires regional analyses of Argo data in relation to other independent data, and therefore is a multi-year research effort that is beyond the typical 12-mth timescale of delayed-mode qc. Improving the quality of Argo data on a global level is therefore an important task that lies ahead, and is central in realizing Argo's mission of providing accurate measurements of the physical state of the world's oceans.

8. REFERENCES

1. Belbeoch M. & Co-authors (2010). "The JCOMM in situ Observing Platform Support Centre: A decade of progress and remaining challenges", *in Proceedings of the "OceanObs'09: Sustained Ocean Observations and Information for Society" Conference (Vol. 2)*, Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306.

2. Böhme L. and U. Send (2005). "Objective analyses of hydrographic data for referencing profiling float salinities in highly variable environments", *Deep-Sea Research II*, **52/3-4**, 651-664.

3. Guinehut S., C. Coatanoan, A.-L. Dhomps, P.-Y. Le Traon and G. Larnicol (2009). "On the use of satellite altimeter data in Argo quality control". *J. Atmos. Oceanic Technol*. **26**: 395-402.

4. Owens W.B. and A.P.S. Wong (2009). "An improved calibration method for the drift of the conductivity sensor on autonomous CTD profiling floats by θ-S climatology". *Deep-Sea Research Part I: Oceanographic Research Papers*, **56(3)**, 450-457.

5. Roemmich & Co-authors (2010). "Argo: Observing the global ocean", *in Proceedings of the "OceanObs'09: Sustained Ocean Observations and Information for Society" Conference (Vol. 2)*, Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. and Stammer, D., Eds., ESA Publication WPP-306.

6. Schmid C., R.L. Molinari, R. Sabina, Y.-H. Daneshzadeh, X. Xia, E. Forteza and H. Yang (2007). "The Real-Time Data Management System for Argo Profiling Float Observations", *J. Atmos. Oceanic Technol.* ***24(9)***, 1608-1628.

7. Wong A.P.S., G.C. Johnson, and W.B. Owens (2003). "Delayed-mode calibration of autonomous CTD profiling float salinity data by θ-S climatology". *J. Atmos. Oceanic Technol*. **20**: 308-318.