# BIO Argo Data Management Meeting, Hyderabad 12 and 13 November 2012

## General objectives of this meeting (Hervé Claustre)

The float platforms as well as biogeochemical sensors are mature enough (miniaturisation, low consumption) to consider the establishment of a Bio-Argo network. In parallel, the Bio-Argo community is organizing itself for the science aspects (several meetings, workshops, working groups during the last four years). "Bio-scientists" are participating in the Argo Science Team (AST) and for the first time last year, "Bio-data managers" participated in the Argo Data management Meeting.

The first four core variables have been selected through a community consensus. They are: dissolved oxygen concentration, nitrate concentration, chlorophyll-a concentration and backscattering coefficient. Other variables (e.g. pH, radiometry) could eventually also become core variables. However there is presently a need for a longer evaluation period for these variables.

There are several significant (in term of float numbers) projects funded (or near to be funded) to work on the biogeochemical Argo floats: remOcean (EU), Naos (France) INCOIS (India) and SOBOM (USA/NSF).

Historically, the "Bio"-community has not been used to managing huge datasets and to delivering publically data in real-time. Therefore a big challenge for the Bio-Argo community is to set up an efficient data management and distribution system. This is one of the main goals of the present Bio-Argo session as part of the ADMT meeting to pave the way for the progressive development and implementation of such a system.

## Argo Data management: history and objectives (Sylvie Pouliquen)

Sylvie Pouliquen, co-chair of the Argo Data Management committee, presented the Argo data system to the Bio-Argo community. The delivery of the Argo products is centralized at 2 GDACs (Global Data Acquisition Center) through two data streams: in near real time mode (24h), or in delayed mode (1,5 to 2 years). The data are stored in a common data format (NetCDF, CF compliant) and using a common vocabulary (Seadatanet) for naming parameters. The data are processed from telecommunication messages by 11 DACs spread all over the world applying common Real Time and Delayed mode Quality control procedures. Additional consistency checks using statistical methods are performed by Coriolis GDAC who emails the relevant DAC when a float fails the tests; it's the decision of the DAC to correct the profile or not and resubmit it to the GDAC. There are also 6 Argo Regional Centers (ARC) which perform delayed mode processing for orphan floats, organise CTD collection for the reference database used in DMQC and develop new products.

For each float the dataset is divided into 4 files: technical, profile, trajectory and metadata file. All data are provided to users with flags that indicate the measurement confidence (1 = Good, 4 = Bad, 2= probably good, 3 = probably bad, 0 = no QC). Moreover, when a correction is applied, it is provided in a PARAM_Adjusted field together with the PARAM field that provides the original data. This will allow reprocessing in the future without going back to the telecommunication messages.

The Argo community provides some recommendations for the Bio-Argo community: define clearly the procedure to perform Near Real Time and Delayed Mode processing, define some monitoring tools in order to know whether the datasets are reliable and put an effort in to data and metadata definitions.

There is a clear need to have Argo regional centres for Bio-Argo data, as these variables are more complicated than the T and S variables; it is in particular important to have expertise on some regions to check whether the data are good or not.

## What has been achieved for Argo-O2? (Virginie Thierry)

There are presently 200 floats equipped with O2 sensors (first one in 2002). Virginie Thierry highlighted the fact that the information provided by the Anderaa sensors and the Seabird sensors are different as well as the on board processing on the different platforms. At first, it is really important to agree on parameter names, units and on the different processing methods. A document was submitted in 2010. Every measurements transmitted by the sensor should be stored in the Netcdf file. The pre-deployment calibration is stored in the metadata file and the scientific calibration in the profile file for the adjusted procedure. Right now, there are still some decoding errors in the transmitted data and there is still a lot of work to define the delayed mode procedure.

For the Real Time Quality control the proposed tests are:

- Global range test
    - [0- 600] to Flag1
    - [-0,5 -0] flag 3 ( or 1?)
    - 4 for the rest
- Spike Test
    - Flag 4 for values over the threshold 50µm/kg for shallow water (<500m)
    - Flag 4 for values over 25µm/kg lower than 500m.
    - Suggested to normalize on pressure difference between 2 points (new test?)
- Gradient test
    - Flag 4 for values over the threshold 50µm/kg for shallow water (<500m)
    - Flag 4 for values over 25µm/kg lower 500m Suggested to normalize on pressure difference between 2 points (new test ?)
- Stuck value test at 0
- Regional range test: need to be studied
- Some T&S tests could be removed for O2 (rollover, bottom test, sensor drift )

## The nature of bio-geochemical measurements: the common and uncommon traits with T and S data management (Hervé Claustre)

Ocean biology and biogeochemistry is strongly dependent on physical forcing (while the reverse is not true). Therefore it is essential that Bio-data are tightly managed with physical ones. This is especially important as Biogeochemical data QC might sometimes require temperature and salinity to be computed to provide accurate result of some quantities (e.g. NO3).

One specificity of some biogeochemical measurements is that, although based on physical principles the quantity measured is modulated by the “biological” effect. This is for example the case for Chla fluorescence which quantifies the amount of red light emitted by phytoplankton Chla when excited by a certain quantity of blue light. However the intensity of this signal, although proportional to Chla concentration at the first order, is modulated by biological factors (species composition, nutrient and light, history of phytoplankton). This “biological” modulation is one of the challenging issues when dealing with the retrieval of accurate Chla concentration.

The range and the variability of biogeochemical variables are generally larger than for T and S. For example, in open ocean water Chla concentration ranges over three orders of magnitude. Getting the Chla value from fluorescence with an accuracy of 50 % is thus acceptable. At depth, some variables like chlorophyll or backscattering are generally expected to be very low (null value for Chla). Contrarily to T and S, background noise as well as spikes potentially have a biological information about the nature of the particles (average size, presence of large aggregates). It is therefore essential to keep the raw values in the netcdf file, in order to properly address this scientific aspect. Finally, the critical time scales of biology might sometimes require that the temporal resolution of profiles be shorter than 10 days to fit with some important events (e.g. bloom). Such requirements, however, might also take into consideration the requirement of the Argo mode (drift at 1000m, profiling from 2000 to the surface every 10 days).

## Sensor calibration and intercalibration (Antoine Poteau)

A system is proposed that allows the cross-comparison of 6 bio-optical sensors at the same time (SIBO: French acronym for System for Bio-Optical sensor Inter-calibration). Using such a system, factory calibration issues were identified for 10% of sensors (60 were tested; 3 sensors were sent back to manufacturer; 3 others had wrong calibration sheets). To improve the accuracy and the reliability of sensors, it is proposed, in particular, to systematically perform dark test profiles (using a black tape on the sensor) in a tank as well as intercalibration at sea with a gold/master reference (the calibration of this sensor would have to be extremely well documented).

## Chlorophyll a: Fabrizio d'Ortenzio real-time QC, delayed mode QC (conceptual)

This presentation deals with chlorophyll-a near real time and delayed mode procedures.  At depth (i.e. below the mixed layer in winter and below 400m in general) Chla concentration should be 0. This feature can be a way to correct for an offset of the sensor. But in some (rare) cases of very strong mixing or when float profiles are not sufficiently deep, correction of the Chla profile by “deep” values could not be adopted. Therefore an essential prerequisite for this correction is to get information about the density profile. Discussions arose about the possible correction for the depression of surface fluorescence around noon (a consequence of non-photo-chemical quenching, NPQ) and that does not correspond to a proportional decrease in the Chla concentration. Several ways to avoid this effect are discussed: (1) first, avoid profiling around noon; however such strategy would prevent match up with satellite data; (2) the second solution is to assume a homogenous property (including fluorescence) within the mixed layer as a way to correct the profile near the surface. A test would be based on the thickness of the mixed layer (calculated with T and S). In RT such an inhomogeneous profile in the mixed layer would be flagged as 3 (suspicious) while it could be corrected in delayed mode.

For the RT-QC the proposed tests are a global range test, a spike test (negative spike should be flagged 4 and positive spike should be flagged 3), a regional test, (eventually based on Longhurst bioregions).

The possibility of correcting the chlorophyll data with ocean color was also presented and largely discussed. If the in situ float data set are corrected with satellite Chla concentration, the data sets are no longer independent. While this could have an advantage of producing a “Chla” for some dedicated operational applications, for scientific purposes alternative solutions have to be considered.

There was also a discussion on the stability of the sensors.

A recommendation of the Argo team to the the Bio-Argo community is to define the DMQC objective on the basis of clear scientific objectives.

## Backscattering and sidescattering: Emmanuel Boss

We measure backscatter in order to derive the POC/SPM (Particulate Organic Carbon, Suspended Particulate Material). The units are m-1 or NTU (for turbidity sensors). A (empirical) relationship has to be provided to transform backscatter into POC and SPM.

There is a clear need to be very careful with the definition and the storage of the metadata (calibration, model, year, wavelength range, angular range). Before deployment, the dark measurements and the slope should be checked (as Mass= (signal-dark) x slope)

.

For the Real Time Quality control the proposed tests are :

- Values should be positive (but be aware that there is a potential information in the noise and the spikes)
- Negative spikes are suspicious
- Climatology should provide the range at depth
- Others sensors (radiometers, beam transmissometer) could provide some constraints on the range
- From ~ 200 to depth, the value should decrease except if approaching bottom from below, or in the case of detached nepheloid layers near continental slope, which can result in the presence of isolated layers of increased scattering.
- Scattering should increase towards the sea surface (even some sub-surface maximum are observed)
- Scattering should increase towards the surface (even some sub-surface maximum are observed)
- Salinity might be needed to correct the signal in the open ocean (the backscattering by particles is derived from the total backscattering from which the water backscattering (dependent on salinity) is subtracted).
- If, at the surface (<3m),the signal is >1.2 times signal below (3-10m), it can be considered as suspicious (contamination by bubbles backscattering).

For the delayed mode QC, the value at great depth, away from the continents, should be (close to) 0. As for Chla concentration, it is also possible to correct in situ backscattering coefficient measured by float with the same quantity derived from remote sensing of Ocean color.

## Steve Piotrowicz Law of the seas

About 40% of the floats are in country EEZ (Exclusive Economic Zone). The coastal states could ask us to shut down the transmission (not possible for Argos2; possible for iridium) or at least not to distribute the data. An UNCLOS request clearance has to be submitted to operate in an EEZ like from research cruises. There is an automatic warning from AIC when a float is in a buffer zone and is likely to enter EEZ.

The agreement ICO XX-VI is about T, S and P. The tracking is done by AIC. In general countries trust Argo and the fact that the DAC will stop data transmission at their request.

For O2, nobody is presently complaining. Some significant concern might be for BIO (chla , $b_{bp}$ ) as it might be interpreted as living resources indicators. However it is also mentioned that synoptical information from remote sensing of Chla is much more efficient as an index of living resources (e.g. fisheries) than a float measurement. The safer approach for Argo would be to ask for clearance for floats with “Bio” packages.

### NO3: Ken Johnson sensor calibration and intercalibration, real-time-QC, delayed-time QC

There are presently 30 floats with nitrates, oxygen and some bio-optics instruments. The NO3 data are available at: www.mbari.org/chemsensor/floatviz.htm in ascii files.

The nitrate is a tracer of Net Community Production, it has a strong seasonal change which is more important than long-term change. By measuring nitrate, we can measure proxies of carbon cycling, not impacted by gas exchange. The sensor is a really complex instrument. Float transmits both measured observations and computes Nitrate on board. At shore everything can be recomputed from the values that are transmitted, i.e. an array of NO3 counts, the dark counts, and also the NO3 value. Precision of the instrument is +-0.2µM.

The different sensors, ISUS or SUNA (both working on the same optical principles), have the same data stream and mathematics and thus are expected to give the same results.

A calibration file for each sensor is needed for the wavelength array. To check the data quality, a hydro-Cast at the deployment should be performed.

In Near real time, from deep measurements, an offset could be detected and then the whole profile could be corrected. Note that slightly negative measurement (in the precision range) might potentially correspond to good data but would cause problems to models. For these applications, the nitrate value should be set at zero in the model when negative.

For delayed mode, in order to establish a reference climatology, there are historical nitrate data in WOD09 (World Ocean Database 09, World Ocean Atlas).

## pH Ken Jonhson

The first float equipped with a pH sensors was deployed in September and seems to perform very well. In particular, the sensor accuracy (0.02) and precision (0.002) are good enough to allow ocean acidification trends to be detected after 5 years of time-series.

The weakness of these sensors is that each sensor has a different pressure response and must be calibrated at high Pressure.

### Additionnal measurements: CDOM, transmissiometry, radiometry: Hervé Claustre

Hervé Claustre presented additional variables that can eventually become core variables in the future:

The Attenuation coefficient (measured by a transmissometer) (660nm) is also a proxy of POC and SPM. It is very difficult to get an absolute measurements while there are issues with the collection angle of detection (apparently recently solved by manufacturers) and drift with time, as a result of accumulation of particles (“marine snow”) on the detection window. The value recorded at a sufficient depth (~1000 m) could be used as a reference to correct the drift in delayed mode. The attenuation coefficient measurement can reveal spikes in the signal which have the same significance as for Chla and backscattering (large aggregates).

The Radiometry (down welling irradiance) should be performed around noon (preferentially matching the satellite overpass), but this can induce some bias for other measurements (for example, if a float always measures at noon, the mean temperature computed from it will be biased)). Performing QC for these measurements has to take into consideration the presence of clouds in particular. A method combining the shape of the fluorescence profile and the radiometric profile at 490 nm has been proposed (related to the chlorophyll content, related to the diffuse attenuation coefficient at 490), to retrieve an accurate Chla as well as an irradiance profile free of cloud contamination.

The CDOM, measured by a CDOM fluorescence, is a by product of biological activity as well as a water mass tracer. At the moment and considering the available sensor used on floats, this measurement is associated with noisy profiles. The quality control should be analogous to that of Chla fluorescence, except that no information in spikes are expected. For the Delayed Mode, averaging should certainly be necessary because of the low signal/ratio.

## Review of actions to be undertaken:

| Action | what | who | ADMT (correspondant.) | Date |
|---|---|---|---|---|
| 1 | Write the minute of the meeting | Hervé, Catherine | Sylvie | 1 month |
| 2 | Contact seadata net for variable names (and close the action 45 of ADMT) | Catherine, Justin | | 2 weeks |
| 3 | Write the document "Processing Argo Chla data at the DAC level" (e.g. similar document produced by V. Thierry et al. for Argo-O2) | Catherine | Justin, Thierry | 6 months |
| 4 | Write the Chla QC document | Catherine | Christine | |
| 5 | Database harmonization: producing reference materiel => interaction with manufacturers | Tom & Antoine | Claudia | |
| 6 | Propose a Quenching test for Chla | Xiaogang & Clare | | |
| 7 | Propose a depth correction for Chla | Antoine, Haili, Xiaogang | | |
| 8 | Propose a spike test | Xiaogang, Hervé, Sandy | | |
| 9 | Write the document "Processing Argo backscattering data at the DAC level" (e.g. similar document produced by V. Thierry et al. for Argo-O2) | Emmanuel, Catherine | Justin, Thierry | 6 months |
| 10 | Writing the backscattering QC document | Emmanuel, Catherine | Christine | 6 months |

| | | | | |
|---|---|---|---|---|
| 11 | Write the document "Processing Argo NO3 data at the DAC level" (e.g. similar document produced by V. Thierry et al. for Argo-O2) | Ken, Catherine | Brian, Thierry | |
| 12 | NO3: review the RT test for Argo identify the useful, define new ones | Ken, Fabrizio | | 0-6 months |
| 13 | Writing the NO3 QC document | Ken, Catherine | | 6-12 months |
| 14 | Evaluation of the need for two Chlas: e.g. operational, blended synthetic Chla vs science Chla (AST action?) | Fabrizio, Hervé, Tom, Emmanuel, Sandy | to be discussed at AST | 3 months |
| 15 | Interact with IOCCG / maintain the link with spatial agency (e.g. yearly summary of Bio-Argo progress on their web site) (AST Action) | Hervé | Sylvie | 1 month |