Report from the First Meeting of the International Argo Data Management Group and Subsequent Actions

3 – 5 October 2000
IFREMER Center
Brest, France

Executive Summary

The primary objective of the first meeting of the International Argo Data Management (DM) Group was to begin standardization of DM procedures among the participating Argo DM Centers. Major recommendations from the meeting are given next by the primary topics considered during the session:

1) Data products

a) Statistics on network, float and sensor performance are needed to evaluate continuously the ability of Argo to meet user requirements
b) Argo profiles should be compared to neighboring profiles from other instruments, and analysis fields to ensure that a 'climate signal' is not being introduced into the observational record because of a new observing technique

2) Files and formats

a) All Argo data centers will use the same formats for meta data, GTS distribution and internet data exchange
b) Meta-data requirements must include a unique float identifier

3) Quality control and flags

a) All Centers should use the same automatic tests for real-time quality control
b)Delayed mode quality control should be standardized through a meeting of involved PI's.
c) Automatic real-time failure modes should be defined
d) Intercomparison of quality control procedures are needed
e) A standard flagging convention should be adopted

4) Data transmission and access

a) Immediate distribution of all Argo Float data in TESAC
b) Float operators, working with the Data Management Group, take steps to transition from)TESAC to BUFR format.
c) Data will also be available through internet (easy access to all users) and Ftp (more complicated access for expert users)
d) Position and some meta-data to be provided to the Argo Information Center to comply with IOC resolution XX-6.

5) Data Exchange

a) Argo should have two global data centers for redundancy (suggested sites, U.S. – GODAE Server and IFREMER, France)
b) Data Centers will coordinate activities including uniform access and filing conventions, periodic data – holding intercomparisons.

6) Profile location / trajectories

a) Speed – check tests should be a component of real-time quality control.
b) Method for estimating total surface trajectory is needed

7) Data archiving

a) Establish a long – term data repository at NODC

8) Meta data

a) Establish one meta-data file

9) Other Issues

a) The ad hoc Argo Data Management should become a formal sub-committee of the International Argo Steering Committee
b) Based on results from this meeting the Data Management Hand book should be updated.

Introduction

The first meeting of the ad hoc committee addressing data management (DM) issues for the international Argo experiment was held at the IFREMER Center in Brest, France from 3 to 5 October 2000. The list of participants is given in Appendix 1. and the agenda in Appendix 2.

Prior to the meeting a list of important issues was developed. For each issue, a discussion leader was chosen. The issues and leaders are :

1) Data products:	**Fabienne Gaillard**
2) Data files and formats:	**Gilbert Maudire**
3) Quality control and flags:	**Bob Keeley**
4) Data transmission and access:	**Etienne Charpentier, Sylvie Pouliquen**
5) Data exchange between Centers:	**Sylvie Pouliquen**
6) Trajectory data:	**Claudia Schmid**
7) Data archiving:	**Sylvie Pouliquen**
8) Meta data:	**Bob Molinari**
9) Other issues:	**Bob Molinari**

Each discussion leader prepared issue papers presented at the meeting. The issue papers are attached in Appendix 3. Discussion of the papers resulted in a series of recommendations and action items given next. Furthermore, since the meeting there has been some work done on the action items given below under status.

1) Data products: Fabienne Gaillard

Recommendations

Global information on the ARGO network

ARGO should provide information and statistics for feed back to P.I.s, manufacturers and deployement organisations in order to evaluate the performance of the sensors, vehicles and deployement methods. Among the usefull information we see:

- Maps of float positions (per month, seasons, or global)
- Maps of number of float days per deg. Square
- Characteristics of floats, performances: global data return percentages, transmission losses, stability of parking depth, lifetime, sensors stability

Comparison with reference datasets

Reference data sets must be defined, they can be:

- repeated CTD sections,
- repeated XBT section,
- long term stations,

Then, all data from floats found in the neigborhood of these reference dataset should be systematically compared, and results of the comparison saved.

Real time global comparisons with model forecast

The only global comparison of ARGO profiles that can be done in real time, is with the corresponding forecast issued by the operational models. It is not at all obvious that **A**RGO should do that because :

1. This work is systematically done by the modellers before assimilation anyway. This is not an extra work for them: they have the model forecast and they collect the datasets for asssimilation. ARGO would have to import the forecast.
2. Which model should ARGO choose and what is the meaning of the model-data discrepancy: data error or model deficiency? Here again, modellers are more qualified to answer.

Analyzed fields

In delayed mode, various methods of analysis can be proposed for providing three dimensional fields of T and S. These method can use different datasets, they also rely on statistics, or a priori information provided by climatology or models. The reference datasets defined above can also be used. The results of these analysis may have different goals:

1. Intercomparison of datasets and biases (or sensor drift) estimation. In that case, datasets may be used separately.

2. Produce the best estimate of the field. In that case all avalaible data, corrected from previously estimated biases are merged. All interesting derived quantities can be computed (heat content, mixed layer depth,….).

It seems difficult to perform those analysis on the global scale at once. They require expertise of the area and knowledge of other existing datasets (not always available in real time). But on the other hand a global analysis is needed. We must then:

1. Make sure that the whole world ocean is covered by regional or basin analysis
2. Define how to combine the results of the regional analysis

Interests have been expressed already by some participants for specific areas:

- Atlantic ocean: France
- Antarctic: UK and Germany
- Pacific and Indian

2) Files and Formats: Gilbert Maudire

Recommendations

All ARGO data centres will use the same formats for:

- meta data,
- GTS distribution,
- internet data exchange

Having a common format for gridded analyses appears to be less immediately essential since no mandatory QC test is based upon such a grid.

Meta data:

Initial list of meta data items proposed by Bob Molinari seems to be convenient. But, it has to be improved for:

- float unique identifier,
- calibration information according to the various float types. This includes data version and recalibration dates,
- mandatory/optional items,
- coded/textual items.

Providing a unique ARGO float identifier is crucial since the WMO numbers are reused (999) floats per area as a maximum in the present state). This could be done several ways:

- define a new identifier as data centre identifier + sequence number per centre. The uniqueness of the number has to be guaranteed by each data centre. (B. Keeley proposal),
- extented WMO buoy number to the maximum GTS/TESAC capabilities. In this case, Argo Information Centre must be in charge of maintaining the unicity of this number.

ASCII file seems to be appropriate for this metadata file. A structure has to be defined (XML, ...).

GTS distribution :

TESAC bulletins will be used in the near future to distribute profiles on the GTS as recommended by E. Charpentier.
Quality flags will not be distributed on GTS as TESAC messages do not include them.

E. Charpentier notes that TESAC will not be supported on GTS after 2005. But, transition to BUFR binary bulletins must be approved by ARGO data users and must take place in a concerted action of all data centres.

Internet data distribution:

Same format will be used for data exchange between data centres and data distribution on Internet.

NetCDF based format seems to be acceptable by all even if it may imply constraints to save disk space and network :

- homogeneous vertical sampling in the same file (this is more or less the case for profiler floats),
- separate files for data sets from T float and T-S floats.

The subset of meta data which have to be included in this file has to be defined.

Pressure, temperature and salinity will be the included parameters. Conductivity will be present if available.
Original values as well “best” values of each parameters will be distributed within the same file.

According to the QC issues, following items have to be included in the data file:

- quality flags for date and position (profile header),
- quality flag per parameter per immersion level (3 flags / immersion),
- history of applied quality checks,
- action undertaken on the data (position or date change, ...). The GTSPP way to manage this QC extra information is approved.
- A global QC status (i.e.: no QC performed, "real time" automatic checks performed, PI's checked, ...).

WOCE and PMEL/EPIC naming conventions for NetCDF dimensions and variables should be employed when they are defined and compatible.

Action Items

Action item 1: Sylvie Pouliquen will propose a NetCDF data format that addresses the meta data and QC issues.

Status: Appendix 4

Action item 2:
A survey will be undertaken by K. Schenebele to identify the level of original data to be archived: Service Argos hexa frames, raw probe data, and technical information.

Action item 3: Bob Molinari will undertake to improve the metadata list of items, identifying those, which are mandatory and coded. This list of items would mention the metadata subsets, which have to be distributed with each data sets (data header).

Action item 4: Etienne Charpentier will ask the DBCP and the WMO to extend the WMO buoy number.

3) Quality Control and Flags: Bob Keeley

Recommendations

- All Centers use same tests for automatic QC. Certain of the tests, to be determined, will automatically stop data going to GTS
- Additional test are allowed as long as they are documented
- QC. of delayed mode will be standardized through a meeting of involved PI's.
- All QC. must be well documented and available from a web site
- A test data set will be built to be used to intercompare the results of QC. procedures at different centers
- Since there will be some subjective testing, an intercomparison of QC. results should be carried out as soon as possible. (With delayed mode data)
- As much data as possible goes to the GTS in 24 hrs. Some can be delayed for manual scrutiny but should go to GTS as quickly as possible
- We do not feel QC flags need go on GTS at this time
- QC flags will be attached to position, time & observed values at each level.

A poll of users (ARGO Science team, GODAE, CLIVAR, GOOS….) will be done to see if they will make use of:

a) flags set at each level of a profile that describe both the specific test failed as well as the assessed data quality, or
b) a flag indicated data quality only.

- Flag of O-5:   0 = NOT TESTED
  1 = GOOD
  2 = PROBABLY GOOD
  3 = PROBABLY BAD
  4 = BAD
  5 = CHANGE

Will be used consistent with flags definitions used in other programs

- Members of group should suggest additional tests, especially those tuned to failure modes of floats.
- Recommend centers establish a relationship with one or more users to get regular feedback on quality of data. A formal structure is better.
- Centers should /could assist PI's by preparing float "Performance" products (i.e. examine Buoyancy failures, etc.), but these do not set QC flags.
- Centers to prepare QC monitoring products (i.e. consistent failures,) to provide feedback to PI's.
- Recommend processing history be carried with data.
- Recommend position uncertainty be carried with data.

Action Item

- Action item: Keeley to recommend automatic failure test

Status: Appendix 5

4) Data transmission and access:
Etienne Charpentier and Sylvie Pouliquen

<u>Discussion</u>

WHO are the users?

Argo data interest at least 3 main categories of users who differ mainly on the timeliness requirements for the data:

- Meteorologists: need real-time data
- Operational Ocean Modelers/ Fisheries: need near real time qualified data: more or less within 24 hours (working day).
- Oceanographers: prefer highly qualified, high resolution and full depth data.

<u>Recommendations</u>

Data on GTS

Considering the fact that the meteorological community will abandon all character codes probably around 2005 the recommendations are:

- Immediate distributions of all Argo float data in TESAC.
- Float operators working within the DM group, take steps to transition to BUFR format instead of TESAC.

Before Argo data are actually distributed in BUFR, Argo operators must agree on exactly what information to distribute in BUFR reports (including resolutions). It might indeed be required to request new table entries or to modify existing ones. In addition, specific templates might be proposed in order to minimize the size of BUFR reports. Based on those requirements, the Argo Coordinator will then submit specific request to CBS. Meanwhile, float operators can work on the development of BUFR encoders or someone could be tasked to develop software that would automatically translate standard Argo data sets into BUFR (e.g. Net-CDF to BUFR).

Functionalities to be provided on Internet

The following set of functionalities needs to be provided to the users:

- distinguish different quality control procedures applied to the Argo data
- distinguish the recent from older data
- an easy recovery system for the users of all the Argo data
- possibility to subset these data according to some criteria

We have to provide both FTP and WWW access: WWW should provide an easy access to all users, FTP will be devoted to expert users or automatic access through software.

Data on the two Data Server FTP sites

- The best profiles and original data will be available on the FTP.
- The ftp server architecture will be refined between the two Argo Data Servers and proposed at the next meeting. It should be organized to fulfill user needs regarding data retrievials by batches using either geographical retrievials (by basin for instance) or time retrieving(by month/year…)

WWW functionalities

The following functionalities should be available on the two data server WWW sites:

- On line access to data
- On line Documentation (B. Molinari volunteered to coordinate Documentation aspects)
- Browsing tools
- QC information
- Maps
- Subsetting tools providing a multi-profile/multi-station files (Netcdf or ASCII)
  - time and space
  - periodicity (week, month, year)
  - mono or multi float files
  - sensor type (Provor, Alace…)
  - parameter (Temperature, Salinity…)

Data to submit to the Argo Information Center

To comply with the IOC XX-6 resolution, member states must be informed about the floats which enter their EEZs. This will be done through the AIC WWW site. Argo data centers must therefore routinely (real-time or daily) provide the AIC with the positions of all Argo floats. Minimal information to submit would include for all observations:

1. Float ID
2. Date and time of observation
3. Float position

This center also needs information on the float when it is launched. These data are a subset of the metadata file we discussed during the "Format discussion" The complete metadata files will be provided to AIC. MEDS offers to work with the AIC to provided this information on behalf of the DACS.

5) Data Exchange: Sylvie Pouliquen

Recommendations

Two Official Data Centers

Everybody agrees that we have to provide a unique access to all the Argo data. In the USA; a US-Godae Server already exists and should be able to fulfill this job. For security reason the group made the proposition to have a backup site. As Ifremer needs to have a rapid access to all the ARGO data to provide products to the community it is serving, Ifremer volunteered to be the second Argo data Server.

Relations with Argo Data Centers

These two servers will receive both Data (Near-real time and Delay mode) and Metadata from the different data centers. The data centers will post the data to these centers at the same rate as to the GTS. We agreed that the metadata will be exchanged in an "ASCII like format " and the data in netcdf. This format will be defined through the actions issued during the Format Discussion. It's up to each data center to retrieve all the Argo data from one of these data server if needed. Value added products generated by the different centers might also be available on these servers. Each data center will provide periodically (daily, weekly…) transmission logs to inform them on the data they should have received.

Relations with AIC

Metadata will be sent directly to AIC in the same format and at the same time it is sent to the two Argo data servers.

Relations between the two Argo Data servers

The two servers need to coordinate in order to:

- develop a uniform access to users
- to re-synchronize periodically if necessary
- Provide a file naming convention which will ease these exchanges with the data centers

Action item

Molinari will contact the operators of the US GODAE Server to establish their interest in this activity.

Status: Contact was made and the initial reaction was positive. Operators of the US GODAE Server will contact S. Pouliquen to establish activities of Argo Data Server.

6) Profile location / trajectory: Claudia Schmid

Recommendations

Real – time requirements:

- Check speed between first position and subsequent positions to find out if first position is good (up-profiling float). Similar procedure with last position of down profiling float.
- Use speed from all available positions and time log between profiling time and time of position fix to derive a measure for the position error.

Delayed mode requirements:

- Extrapolate surface trajectory to reduce position errors
- Give measure for position error based on the method used for the real-time version.
- Cokriging / tense spline (or other methods) can be used for extrapolation.
  Cokriging is very flexible.

Action Item

C. Schmid will conduct further tests and method comparisons to establish the most reliable method.

7) Data archiving: Sylvie Pouliquen

Recommendations

It is necessary to have a unique center that will archive long term the different data versions that have been distributed to users. NODC volunteers to perform this job retrieving this information from the two Global data centers.

Each regional data center will archive the original data from Argo messages, if funding is available. We have to decide what is the lowest level of data to be transferred to the Archive center.

Action item

The Argos messages but also all the software needed to decode them will have to be archived. Kurt Schnebele, who needs to perform this task for the V3 of Woce data will coordinate a survey with users and the regional data centers to determine which level of data needs to be archived centrally. He will produce an archive plan synthesizing these results.

8) META DATA: Bob Molinari

Action item

The exact structure of the meta-data file suggested by Molinari (Appendix 3.8) will be forwarded to participants for review.

9) Other Issues: Bob Molinari

Action item

1) The group agreed upon the need to formalize relations with the International Argo Steering Committee. (IASC)

Action Item

Molinari will forward a request to D. Roemmich, Chair of the IASC to recognize the DataManagement group as a formal subcommittee.

Status: Completed

2) Based on the results of this meeting, additions and changes to the DM Handbook will be made and distributed to the participants

Action Item

Working with other ‘volunteers’, Molinari will update the DM Handbook.

Appendix 1: AGENDA

Objective: To standardize methodologies of ARGO data centers and to document methods in an ARGO Data Management Handbook

Chair: Sylvie Pouliquen, Claudia Schmid
Co-chair: Bob Molinari

Format: Volunteer discussion leaders will review the particular issue, lead a discussion and summarize conclusions. Meeting attendees are expected to contribute ideas on issues before and during the discussion to the leaders.

Agenda:

3 October

0900: Local greetings
0915: Summary of meeting objectives, Chair and co-chair
0930: Data products: Fabienne Gaillard

Issues:
- examples from participants
- regional versus global

1045 Break
1115: Discussion of data formats: Leader, Gilbert Maudire

Issues:
meta-file i) standardize definitions
ii) develop look-up tables

header data i) meta-file items to be included
ii) other data (e.g., positions, time, etc.)
iii) unique profile id

data i) flags, errors
ii) two types of real-time data; (1) GTS (i.e., salinity only) and post-GTS, pre delayed mode qc (i.e., S and conductivity)
iii) differences between real and delayed mode

1215: Lunch
1400 : Discussion of data formats (continued)
1530 Break
1600: Quality control and flags: Leader, Bob Keeley

Issues:

- real time tests and flags
- delayed mode tests and flags
- degree of standardization
- compulsory versus optional tests

1700: End first day
=======================================

4 October

0900: Quality control and flags(continued): Leader, Bob Keeley
1030 Break
1100: Data transmission and access: Etienne Charpentier, Sylvie Pouliquen

Issues:
- GTS formats (including resolution)
- ftp/www/e-mail

1400: Data Exchange between Centers: Sylvie Pouliquen

Issues:
- formats (ASCII, netCDF)
- meta-files
- Real time versus delayed mode data
- Mirroring strategy
- how and where are original data stored?

1530: Break
1600: Trajectory data: Claudia Schmid

Issues:
- quality control
- accepted real-time position

1730: End of second day
====================================

5 October

0900: Data Archiving : Sylvie Pouliquen

Issues:
- Real time versus delayed mode (replace real-time with delayed mode profiles?)
- Multiple delayed mode profiles resulting from several qc iterations of the same profile (e.g., PI, basin-wide, re-analyses, etc.)

1030: Break
1100: Remaining issues
1200: Close workshop

Appendix 2.

LIST OF PARTICIPANTS (31)

Sylvie Pouliquen (chair)	IFREMER	France
Claudia Schmid (chair)	AOML	USA
Bob Molinari (co-chair)	AOML	USA
Bob Keeley	MEDS	Canada
Hee-Dong Jeong	Korea Oceanographic Data Centre	Korea
Pat Cousineau	MEDS	Canada
Dave Hartley	UK Hydrographic Center	UK
Lesley Rickards	BODC	UK
Kurt Schenebele	NODC	USA
Melanie Hamilton	NODC	USA
Charles Sun	NODC	USA
Etienne Charpentier	WMO-IOC	France
Mark Ignaszewski	Fleet Numerical Oceanographical Center	USA
K Takeuchi	Hokkaido University	Japan
Y Takatsuki	JAMSTEC	Japan
T Yoshida	JMA	Japan
Roger Goldsmith	WHOI	USA
Michèle Fichaut	IFREMER	France
Gilbert Maudire	IFREMER	France
Fabienne Gaillard	IFREMER	France
Thierry Carval	IFREMER	France
Loïc Petit de la Villéon	IFREMER	France
Juergen Fisher	Kiel University	Germany
Christopher Kyle Rushing	Naval Oceanographic Office	USA
Deborah Toca Bird	Naval Oceanographic Office	USA
Eric Duporte	EPSHOM	France
Catherine Maillard	IFREMER	France
Christophe Vrignaud	EPSHOM / CMO	France
Guillaume Du Réau	EPSHOM / CMO	France
David Corman	EPSHOM / CMO	France
Pierre Blouch	Météo-France	France

Appendix 3.1: Data centers contributing to ARGO Products by Fabienne Gaillard

- **Should the centers distribute products ?**
  1. Is it usefull ?
  2. Is it necessary ?
  3. Is it dangerous ?

  If answer to 1 or 2 is yes:

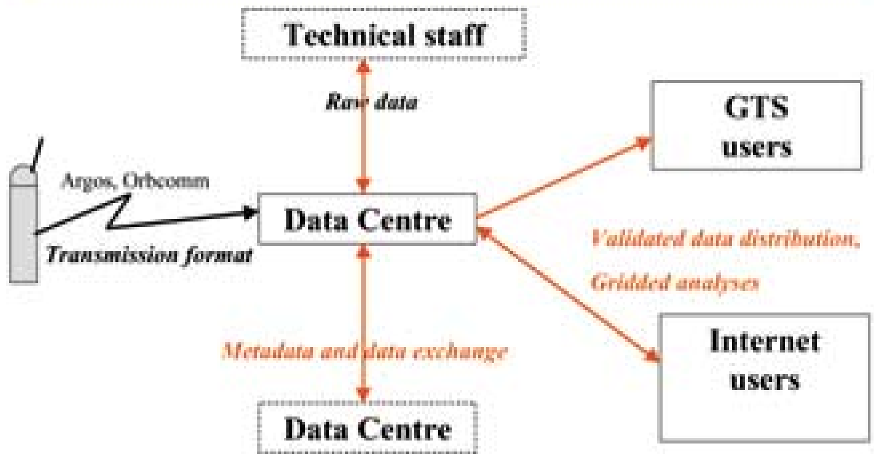
- **What ?**
- What type of products ?
- What time scales ?
- What are the users ?
- How?

Types of products

- Dataviewers:
  - Time series for a platform (T or S (z, t))
  - Trajectories
  - Statistics (number of floats days per deg. Square)
  - ...
- Analyzed fields
- 3D gridded data (Temperature, salinity)
- 2D gridded data: mixed layer depth
- ...
- Sensor drift estimates
- Comparisons
- Define reference points,
- Define reference sections
- Other type of reference quantity (heat fluxes at a given latitude, transports,
- New climatologies

Appendix 3.2: Data files and formats by Gilbert Maudire

Data flow



Argo data management meeting #1 Brest - 03-05/10/2000

ARGO files and formats

- Transmission format and raw data format
  - transmission dependant (Service Argos, OrbComm, ...)
  - float dependant
  - out of meeting concerns
- Focus on
  - metadata
  - data exchange
  - data distribution
- Objectives
  - identify the necessary formats and files
    - same format for data centre exchange and distribution to users
  - list the items to be included, item contents
    - contents often more constraining than format itself
  - define the formats

Argo data management meeting #1 Brest - 03-05/10/2000

Necessary formats

- Formats have to be defined for :
  - Metadata exchange
  - Data distribution on GTS
  - Data distribution and exchange on internet
  - Gridded analyse distribution

Metadata

- Items
  - float id (Argo id), ...
  - float provider name, institution, adress and country (provider id)
  - experiment name (EU project, ...)
  - float type : WMO code (probe type), manufacturer, serial number
  - probe descriptions : pressure, temperature, salinity manufacturer, type, calibration
  - float deployer name
  - deployment environment :
    platform (name or Call sign), platform type,
    launch position and date, CTD at deployment (y/n)
  - float cycle description :
    cycle period(s), drift depth, profile depth,
    profiles up/down, first down time
- Format

  ASCII file with key words (?) (cf Bob Molinari proposal)

Metadata / Data

- Complete metadata :
  - exchanged between data centres (at deployment time or on request)
  - disseminated to users on request
  - available on the Web or ftp
- A subset of metadata transmitted with data
  - subset to be defined according to national distribution policies :
    - float provider
    - experiment
    - ???

Argo data management meeting #1 Brest - 03-05/10/2000

Data

- Items
  - Header data : subset of metadata items
    - Float : Argo float id, float type (WMO), ...
    - Experiment, Provider identifier, ...
  - Profile identification
    - Cycle number / float, **A**scending/**D**escending
    - Position and date
  - Profile measurements
    - Pressure, Temperature, Salinity (Conductivity?)
  - Quality data
    - Flags : per immersion level, per profile
    - History : check log
    - Error estimation

Argo data management meeting #1 Brest - 03-05/10/2000

Data (cont.)

- Formats
  - GTS distribution
    - TESAC (now used by US, France, Japan for floats)
    - have to be discussed with GTS formats (tomorrow morning)
  - Data exchange / data distribution
    - NetCDF file (Coriolis multi-profile,…)
      - self documented files, direct access to data by "variable",
      - extensive use in oceanography (last Woce CD-Rom).
      - adapted if homogeneous number of immersions / profile
    - Flat ASCII file (Coriolis/EU-Medatlas, previous WOCE CD's,…)
    - Structured ASCII file (GTSPP, …)
    - ??? (J-Comm, …)
  - Proposal
    - Exchange : NetCDF multi-profile format (one file per day)
    - Distribution : two formats : NetCDF & flat ASCII files (cf Coriolis format)
    - NetCDF variable name = GF3 parameter name (TEMP, Error_TEMP, PSAL, …)

Argo data management meeting #1 Brest - 03-05/10/2000

Gridded analyses

- Used for quality control and analyses distribution
- Items
  - Metadata
    - Identification of the analyse
    - Identification of the provider
    - Analyse description
      - Spatial and temporal meshes
      - Immersion
  - Data
    - A table per interpolated parameter
    - A table per error estimation
- NetCDF format
  - well adapted for multi-dimension tables
  - Variable name = GF3 parameter name (TEMP, Error_TEMP, PSAL, …)

Argo data management meeting #1 Brest - 03-05/10/2000

Item contents

- Unique float identifier
  - Service ARGOS id number : not unique (two or four beacons)
  - WMO buoy ident : **A9xxx**

    A: area (1 digit)
    9: the buoy is a drifting float
    xxx: float number in the area

    → not large enough in the ARGO context
  - Proposal of **ARGO ident**

    → **A9yyyyy**
    where xxx=mod(yyyyy,999) (compatibility in first years)
- Quality flags and history
  - have to be discussed with quality control (this afternoon)
- Error estimation
  - have to be discussed with products (tomorrow evening)

Argo data management meeting #1 Brest - 03-05/10/2000

Item contents (cont...)

- Reference metadata (thesaurus)
  - Float

    Type, Recorder, profile reduction, ... (*WMO tables?*)
  - Platform (deployment)

    Call sign, Name, Type (aircraft, ship, ..) (*ITU tables?*)
  - Float provider

    Institution, country... (*NODC and IOC tables?*)
  - Probes

    Pressure, Temperature, Salinity (*to be defined*)
  - Parameters

    Name, ... (*GF3 table*)
  - Units, Latitude, Longitude, Date

    International rules, Degrees-North, Degrees-East,
    AAAAMMDDHH24MI (UT)

Argo data management meeting #1 Brest - 03-05/10/2000

Appendix 3.3: Quality control

Discussion paper from Bob Keeley

QC Discussion Points

QC of Real-time Profiles

Is it necessary that everyone apply the same suite of tests?

I suggest yes. If this does not happen, users of the data will get a mixed data set. This is the case with other data either those on the GTS now or received in delayed mode from one or other provider and we should strive to improve on this. The impact will be that certain procedures may require access to reference files, such as the same climatology or the same weekly NCEP files. Somehow all real-time QC centres need access.

Do we need to write some documentation that explains the tests? If this varies, does each centre need to do this?

I believe we need documentation and if there are variations, all must be well documented. The documentation should be detailed enough that a new centre can pick up the document and know exactly how to do the tests. The documentation need state what are the consequences of a test failure. For example, if the float seems to have a failing buoyancy, the observed values of the profile may be okay, it just doesn't get as deep or stays at the surface too long. In this case, failing the test does not cause observed values to be removed (in the case of TESAC coding) from the real-time data. For other tests, such as outside acceptable ranges, the documentation should state clearly when the point is removed from real-time transmission (if it is).

How do we keep documentation up to date?

If all use the same documentation, a single web site can be designated the master site for the documentation. This could be at the site of the Argo coordinator, but it could be elsewhere. It may be better at one of the QC sites since they will be closer to the workings of the software. Maintaining the web site also means maintaining the documentation.

How careful should we be about what goes out on the GTS? Do we err on the conservative side allowing only those points and profiles that we are sure are correct, or do we allow suspect data out?

Practically, I suspect we will allow suspect data out simply because the automated tests will not catch everything. We could use the monthly review that MEDS conducts to help trap those data that do get out to the GTS and therefore retrospectively catch problems. However, for some users this will be too late.

How do we communicate the results to PIs? Is this necessary?

Since the data that go through the hands of the automated QC will go both to the GTS and in full resolution to PIs the results of automated tests should accompany the data. This has implications for the exchange format. These results are both those that identify problems in the data and those that identify problems with failing floats.

Do insertion centres need to have a facility to block suspect data from going out on the GTS?

I can see this may be very useful but we would need to be careful. This is a facility that Service Argos offers for drifter data. We may want to provide a similar function. If so, we must be in constant touch with PIs to ensure good data are not blocked.

Can we specify now the QC tests to perform or how do we develop them?

AOML and MEDS have a precedent with the GTSPP test procedures. However, the experience is that automated tests can let obviously bad data through and we have not yet developed the foolproof suite of tests. One strategy would be to compile the tests that each of us recommends now, implement these then carry out a review in a year's time to see what improvements should be made. This will get us up and running, but promises a review. We should also make use of feedback from PIs doing the delayed QC as to what they found, how they found it, and automating the detection process.

What monitoring of data quality is required?

Since real-time QC is automated, and assuming all centres employ the same tests, an intercomparison should be done routinely where every center processes the same data through its automatic systems and results are compared. This will verify that test procedures are identical, or find variances due to changes made at one centre but not at others.

What feedbacks from users should be set up and how will this be used?

The real-time data will be employed in assimilation models and other endeavours. Through their use, problems in the data will be identified. QC centres need to get this information since it will impact QC procedures. This will be a positive and ongoing indicator of what improvements are needed. Being able to track which centre provided the data will be valuable since the "responsible" centre can take the lead in seeking solutions. I suggest each QC centre make a connection to a centre carrying out data assimilation. This pairing will work out how to exchange information. Argo meetings will provide a forum to discuss feedbacks from the modelers.

QC of Delayed Mode Profile Data

Is it necessary that everyone apply the same suite of tests?

Again, I suggest yes, but this may be less practical. The important thing is that the QC process be well documented. I would suggest a tracking system such as GTSPP which records who processed data and did what to the files. This will be useful later in sorting out versions of data and levels of QC applied.

Do we need to write some documentation that explains the tests? If this varies, does each insertion centre need to do this?

Again, documentation is crucial. Without clear and readily available documentation a user of the data will have to redo everything. Our experience is that while some users will do this anyway, others will accept some or all results, and add other tests specific to their needs.

Do we preserve quality flags with the data and if yes, how?

I advocate preserving the flags. Then, this becomes a format issue, but there are examples around of data sets with strategies for this. GTSPP is one I am familiar with and one I would recommend be considered.

What other information that affects the reliability and future use of the data do we need to keep?

This touches on the larger question of future use of the data and so goes beyond what might be considered quality control issues.

a. We need to keep a unique tag with each station to permit matching real-time to delayed mode data and the unequivocal substitution of delayed mode for real-time in the final archives.

b. We need to store information about tests performed and failed. As data are collected and archives built, QC systems will evolve and users (and archives) will want to know what tests were performed against data.

c. We need to store information about who processed the stations and what they did. This is helpful in knowing who did what to data, including QC.

d. We want to store sufficient information about the instrumentation so that should problems arise later about the performance of particular floats we have a chance to make corrections. We want to strive to not repeat the problems encountered when manufacturers fall rate equations for XBTs were found to be wrong. (At present, there is insufficient information about some XBTs that were used so that making corrections to them is difficult.

e. We want to record changes made in the data as a result of QC. However, we also want the original values to be kept, in case a user wishes to backtrack from changes.

f. We may wish to record not only assessment of data quality, but also precise information about a datum (such as which tests were failed) and information about the suspected cause of the failure or the reason why a suspicious datum (such as a temperature inversion) is considered okay.

What monitoring of data quality is required?

Since PIs will be involved in QC of delayed mode data, we can expect variations in results. Also, it may well be sensible for different centres/PIs to carry out different test procedures. I recommend that an intercomparison be carried out on the same delayed mode data set. The analysis of results should be carried out by one centre, and results distributed. A discussion of the results should take place at an appropriate Argo meeting. I would recommend this happens very soon after operations become "routine".

What feedbacks from users should be set up and how will this be used?

In delayed mode, this will be more difficult since the users will be more dispersed. However, if a regular user can be found, a QC centre should try to set up a liaison with them to provide feedback of data quality problems.

QC of Trajectory Data

The same questions for profile data apply here, however, there is no transmission of these data in real-time.

Flags

Do we need to send flags out with the real-time data?

Historically, this has not been done. The present GTS code forms (BATHY, TESAC) have no provision for this and are unlikely to change. There is the possibility of using BUFR, a binary format that is self describing. However, the real question is would the flags be useful to users who need the real-time rather than delayed mode data. I have no feel for this and so need help from the users.

What level of flagging do we need? Do we attach a flag to each value indicating the assessment of the quality of the value, do we attach a flag that indicates what tests were failed by that value, do we attach a flag that says what tests were failed and why to each value?

My own view is that a flag on a datum or grouping of data, such as a profile, should be used to indicate an assessment of the quality only. This will serve the majority of users. Of second importance is to identify which tests were failed at the detail of a grouping of data such as a profile. To say that a profile failed a freezing point test, for example, states that one or more points of the profile failed. The failed points will be marked with a flag indicating poor quality. Of course, a datum failing one test may also fail others, so a flag of poor quality does not uniquely indicate which test was failed. My view is that the number of users interested in this level of detail is low. In addition, those interested in this detail, likely will want to retest data that fail a particular test and are unlikely to use the fact that a single datum failed a test. Finally, recording a possible reason for a failure or a reason why a suspicious point is deemed good is again of lesser interest to people. It would be useful to record such information, to document systematic failures, provide clues to problems in test procedures, and to identify regions of the oceans where unusual features are common. However, this should be stored in such a way that it does not "clutter" access to the data and the simple quality assessment flags. In summary, I argue that all are desirable, but of decreasing utility to users. I would advocate a format that can preserve it all, but one that keeps the information beyond simple data quality, in a location of the format that does not impede access to the data and quality assessment flags.

Do we preserve flags with the delayed mode data?

Yes. Whatever we do to the data, the quality assessment should be kept with the data whether real-time or delayed. The format structures should be identical for real-time and delayed mode so all information can be recorded at whatever is the appropriate point in the processing.

What flag convention should we use?

I recommend the one used by GTSPP. It is pretty commonly used and simple. It is

0 = no QC done

1 = data judged good

2 = data judged probably good although there is some aspect that is troubling.

3 = probably bad with insufficient information to be sure it is bad

4 = bad data

5 = data value has been changed by QC procedures. The original value is found elsewhere in the data record.

GTSPP Tests

In GTSPP we view the data arriving from the GTS using the same software as we use for other profile data arriving in delayed mode. So, we use the routines as described in M&G#22. This provides a suite of tests looking at ranges, position checks, etc. Science centres of GTSPP employ other tests to further identify failures missed by these tests. You can see some documentation about these test procedures at http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Prog_Int/GTSPP/QC_e.htm

The GTSPP tests are not "tuned" to find failure modes particular to profiling floats. For example, I have seen profiles from floats with spikes at the surface. These are usually less than 1.0 degrees C and are not automatically found by the Spike Test. For profiling floats, we may want to tighten up such a test or devise instrument specific tests

Other Tests

I did a quick poll to see what other things are tested and how. I include ones here that are not already a part of GTSPP procedures.

Buoyancy Failure

If the buoyancy adjustment begins to fail then the float will spend increasing time at the sea surface and this will show up as larger drift segments. My own estimation is that such a failure is an indication of the coming demise of a float, but that the profile data returned from the float is still okay. This is information the PI may wish, but not something we need include in archived data.

Failure to reach Parking Depth

I view this as the same type as a buoyancy failure and would recommend the same handling.

T-S Failure

This uses a T/S plot procedure to look for a conductivity failure. I am not sure how this would be automated except through some sort of T/S climatology. Assuming there is such a test, it should result in quality flags assigned to data.

Anomaly Failure

This looks for unusual changes in anomalies of T and S. The float history is used to generate mean profiles and then anomalies computed for each profile based on the mean. Any abrupt or large changes would be viewed with suspicion. Again, I am not sure how we could automate such a test, but this might be appropriate for the high resolution data processing.

NCEP Comparison Failure

This compares the latest data to the previous week's temperature analysis from NCEP. It assumes NCEP is correct and if data are > 3 standard deviations from NCEP, they are considered bad.

Other Issues

I invite everyone to offer suggestions of other points that need discussion.

Appendix 3.4) Data transmission and access

a) Etienne Charpentier

Two issues are being discussed here:

1. Formats for GTS distribution of float data
2. Data to submit to the AIC

1. **Formats for GTS distribution of Argo float data**

   Only the following code formats may be used for GTS distribution of Argo float data. Other formats would not comply with GTS regulations. All data distributed on GTS are directly and automatically assimilated by numerical meteorological or oceanographic models operated by major meteorological centres. Any other format would oblige each centre to develop specific data assimilation schemes and would encourage other applications to define their specific formats with the risk of eventually having a multitude of formats to deal with.

   1. **Possible GTS formats:**

      FM 64-XI Ext. TESAC (KKYY): The floats presently reporting on GTS are using TESAC format (temperature/salinity), which permits a resolution of 0.01 Celsius and 0.01 part per thousand.

      FM 63-XI Ext. BATHY (JJVV) only permits a resolution of 0.1 Celsius and 0.1 part per thousand. It is also limited to 20 points in the upper 500 meters. BATHY is used principally for XBTs.

      FM 18-X BUOY (ZZYY) code permits a resolution of 0.01 Celsius and 0.01 part per thousand. This code is used for drifting and moored buoys. It could also be used for floats.

      FM-94-X Ext. BUFR permits any resolution and provides for inclusion of associated data (e.g. QC flags), metadata as well as specific identification fields (e.g. Argos ID). Format is binary and permits compression, a feature which might be interesting for profile data. It is more complicated than character codes but encoders and decoders can be obtained from meteorological services

      FM-95-XI Ext. CREX permits any resolution and provides for inclusion of metadata as well as specific identification fields (e.g. Argos ID). This is a character code (i.e. a person can read it), which uses the same structure than BUFR. It also uses the same code tables than BUFR to describe the data. However, CREX does not permit compression nor associated fields (e.g. QC flags).

Summary table:

Format	Current version	Type	Temp. resolution	Sal. resolution	Depth resolution	Advantages	Drawbacks
TESAC	FM 64-XI Ext. (KKYY)	ASCII	0.01C	0.01 psu	1m. Significan points	Resolution	Rigid Poor depth resolution Eventually abandoned (2005)
BATHY	FM 63-XI Ext. (JJVV)	ASCII	0.1C	0.1 psu	1m. Significant points (limited to 20)	Widely used (XBTs)	Resolution (D,T,S) Rigid Limited to 20 points in the upper 500m Eventually abandoned (2005)
BUOY	FM 18-X (ZZYY)	ASCII	0.01C	0.01 psu	1m. Significant points (limited to 20)	Resolution Widely used (buoys)	Rigid Poor depth resolution Eventually abandoned (2005)
BUFR	FM 94-X Ext. (BUFR)	Binary Table driven	No limitation	No limitation	No limitation	Flexible Self describing Metadata QC flags Compression	Complex (encoders/decoders exist)
CREX	FM 95-XI Ext. (CREX)	ASCII Table driven	No limitation	No limitation	No limitation	Flexible Self describing Metadata Readable	Complex (encoders/decoders exist) Eventually abandoned (2009)

All code forms are formally documented in the WMO manual on codes, WMO No. 306, Part A - Alphanumeric codes, and part B & C - Binary codes & Common features to Binary and Alphanumeric Codes.

2. Choosing among character code:

Among character codes, BUOY and TESAC are better. TESAC might be preferable to BUOY because users looking for sub-surface temperature and salinity data would rather check TESAC reports first. TESAC is in fact dedicated to temperature and salinity profile data. So for float data, unless other instruments are installed on the floats, TESAC should be preferred to any other character code. Details regarding the BATHY and TESAC formats can be found via the SOOPIP web site at http://www.brest.ird.fr/soopip/gts.html. It is now extremely difficult to modify character codes in order to encode new variables for which there is presently no provision in those codes. This is because the Commission for Basic Systems (CBS) of WMO now encourages the use of table driven codes.

3. **Choosing among table driven codes:**

   Among table driven codes, BUFR is better than CREX for Argo needs because it permits encoding of QC flags and permits compression. CREX is a character code while BUFR is a binary code so CREX permits human readability but this is not a specific requirement for Argo data which are distributed on GTS primarily for direct data assimilation by numerical weather prediction models. CREX also will probably eventually be abandoned.

4. **Character codes versus table driven codes:**

   Character codes are indeed easier to implement because they are simple but all character codes will eventually be abandoned probably around 2005, i.e. when Argo will be fully implemented. On the other hand, encoders for BUFR can be obtained from NOAA (http://www.nws.noaa.gov/tdl/iwt/) or ECMWF. Considering above discussion we probably have to make a choice between TESAC and BUFR. According to requirements expressed by the Argo Science Team, it is clear that BUFR is preferable to TESAC because it permits encoding of Quality Control flags plus encoding of specific fields such as float id, Argos or Orbcomm ID, and profile ID. It is relatively easy to add new entries in BUFR or CREX tables. Those changes should be submitted to the CBS. For example, the Data Buoy Cooperation Panel succeeded in adding specific table entries for surface drifters. Same mechanisms can be used. Details regarding BUFR tables can be found via the WMO web site at http://www.wmo.ch/web/ddbs/Code-tables.html . Details regarding CREX tables can be found via the WMO web site at http://www.wmo.ch/web/ddbs/CREX-Tables.html

5. Proposed approach for GTS distribution of Argo data:

Considering above discussion, a proposed approach could be:

1. Immediate distribution of all Argo float data in TESAC.
2. Every float operator to take steps to eventually and gradually use BUFR instead of TESAC.

Before Argo data are actually distributed in BUFR, Argo operators must agree on what information exactly to distribute in BUFR reports and in what resolutions. It might indeed be required to request new table entries or to modify existing ones. In addition, specific templates might be proposed in order to minimize the size of BUFR reports. Based on those requirements, the Argo Coordinator will then submit s specific request to CBS. Meanwhile, float operators can work on the development of BUFR encoders or someone could be tasked to develop a software that would automatically translate standard Argo data sets into BUFR (e.g. Net-CDF to BUFR).

1. Data to submit to the Argo Information Center

1. Need for a list of operational floats

Meteorological or Oceanographic Centres assimilating Argo float data from the GTS data-flow do not necessarily know which of the floats are truly Argo floats, who operate the floats, and cannot easily access to crucial metadata (e.g. float type, stated sensor accuracies etc.). Typically, the only information they have to deduce the float operator is the WMO number. On the other hand, they can potentially provide float operators with valuable feed-back information regarding the quality of the float data (e.g. comparisons of the observations with the model first guess field or analysis). Maintaining a list of Argo floats, which would among other things include the WMO number, float ID, Argos or Orbcomm ID, the name of the float operator, and certain metadata is therefore important.

The list could also be used by the meteorological services to sort out truly Argo floats from the GTS dataflow and produce specific products, including statistics regarding the overall quality of Argo floats, delays, availability, etc.

Such a list must be centralized and easily accessible (e.g. via the web). It is proposed that such a list be maintained by the Argo Coordinator at the Argo Information Center (AIC). To realize this, float operators are invited to agree to submit information about the floats they deploy to the Argo Coordinator. Basically, information to provide would include:

1. Float ID
2. Satellite system used (Argos, Orbcomm..)
3. Formally Argo float (Y/N)
4. Telecomm ID (Argos or Orbcomm)
5. WMO number
6. GTS bulletin header
7. Deployment position
8. Parking depth
9. Cycle
10. T measured (Y/N) and resolution
11. S measured (Y/N) and resolution

After discussion on this topic at the Argo data management meeting in Brest, October 2000, the Argo coordinator will submit a proposal to the Argo Science Team regarding how to submit the data and in what format. This could be done in conjunction with the float deployment notification mechanism in order to avoid having float operators to submit the same information twice.

1. Float positions (IOC resolution XX-6, EEZ issue)

To comply with the IOC XX-6 resolution, member states must be informed about the floats, which enter their EEZs. This will be done through the AIC web site. Argo data centres must therefore routinely (real-time or daily) provide the AIC with the positions of all Argo floats. Minimal information to submit would include for all collected float observations:

1. Float ID
2. Date and time of observation
3. Float position

And in order to offer other products, additional details such as these listed below could be provided as well:

1. Profile depth
2. Availability of T (Y/N)
3. Availability of S (Y/N)

The meeting is invited (i) to agree on what data to submit, and (ii) suggest a mechanism for submitting the data routinely to the AIC.

b) **Sylvie Pouliquen**

NEAR Real Time ACCESS
Functionalities to be provided to the users

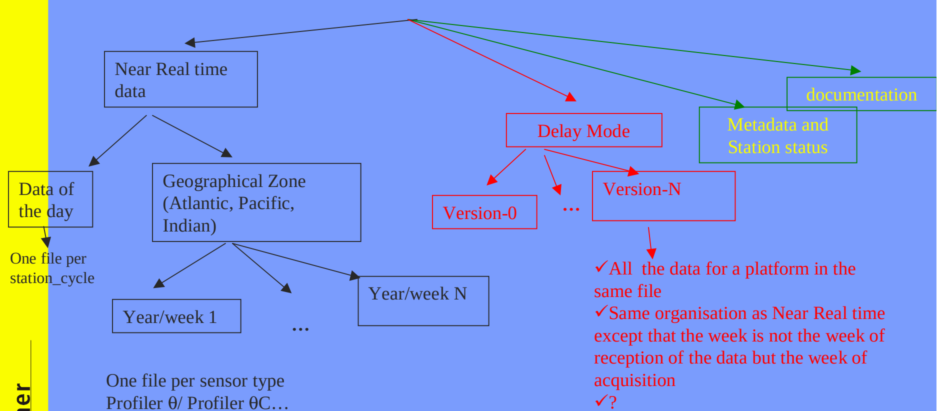
- WWW/FTP ➔ possibility to distinguish different qualification for ARGO data
- FTP ➔ within one category distinguishes the latest data put on the site from the oldest ones.
- Email ➔ Possibility to receive periodically the more recent data
- FTP/WWW ➔ Enable an easy recovery for the users of all the ARGO data
- WWW ➔ Possibility to subset these data according to predefined criteria:

WWW should provide an easy access to all users

FTP is devoted to expert users or automatic access through softwares

3

FTP Organisation
Data Format Netcdf and ASCII



4

WWW

- ✓On line access to data
- ✓On line Documentation
- ✓Browsing tools
- ✓QC information
- ✓Maps
- ✓Subsetting tools providing a multi-profile/multi-platform files (Netcdf or ASCII)
  - -Time and space
  - -Periodicity (week, month, year)
  - -Mono or multi float files
  - -Sensor type (Provor, Alace…)
  - -Parameter (Temperature, Salinity…)

Ifremer

5

Appendix 3.5 Data Exchange

Argo Data Management Workshop

Data Exchange

1

Data Exchange

GOAL: Provide in all data centres all the data available within the ARGO Program in order to ease user access

- 2 types of data to Exchange:

- Metadata:
  - ✓ exchanged once when a new experiment is settled
  - ✓ occasional additional transfers
- *The data centre in charge of the experiment/station issues an email to*
  - ✓ *warn the other centres that new metadata are available*
  - ✓ *or send the new metadata directly to the other centres*

2

Data Exchange
2 steps: Real Time / Delay Mode

- Real Time ⇒Mirroring strategy: within 24 hours the newest data have been exchanged between the centres.
- I have supposed that all the exchanges are made through the FTP servers

  ⇒ No interconnection between the databases

- At least 2 methodologies:
  - ✓ the blind one : every day each data centre update his "latest data " directory with the other centres' "latest data" ⇒ Need to eliminate the duplicates artificially created that way
  - ✓ Provider warning: the provider issues a formatted email containing the list of new files on his FTP server: Each centre uses this file to automatically retrieve these data.

3

Data Exchange

Delay Mode ⇒ New versions of datasets are exchanged when a complete reprocessing has been done

The process " Provider warning" can be used to exchange these types of data.

Perhaps only some versions of these datasets needs to be exchanged:

4

Appendix 3.6 Trajectory data

For PALACE floats using the ARGOS system the average time lag (based on the floats launched in the tropical Atlantic by AOML) between the first know position of a surface cycle and the start of the floats ascend is about 6.5 hours long (the standard deviation is about 2.5 hours). A similar time lag occurs between the last know position of a surface cycle and the start of the descend. These time lags cause position errors that are important for the profile locations and the float trajectories. The position errors can be reduced by extrapolating the surface drift of the float to cover the time between 'pump starts' and 'valve opens'. One method for such extrapolations is cokriging. The method is currently used in the tropical Atlantic, where it reduces the position errors considerably (the effectiveness was analyzed with drifter data, see Fig.3). In preparation for the data management workshop I applied the method to other ocean regions. The first tests with slightly changed structure functions (different time scales depending on the latitude) produced mixed results. The general tendency is that the extrapolation becomes less successful with increasing distance form the tropics (e.g. Fig 2., 4., and 1.). I think that this problem can be solved with a different structure function.

Fig 1.

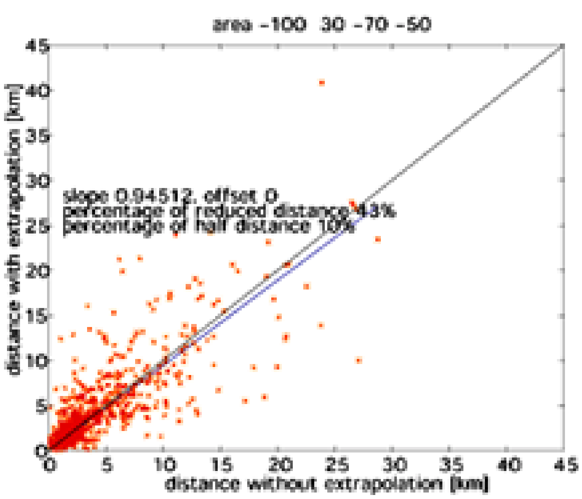
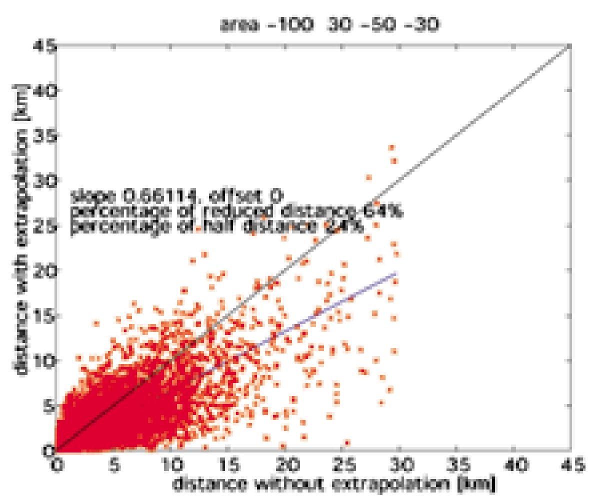


Fig 2.



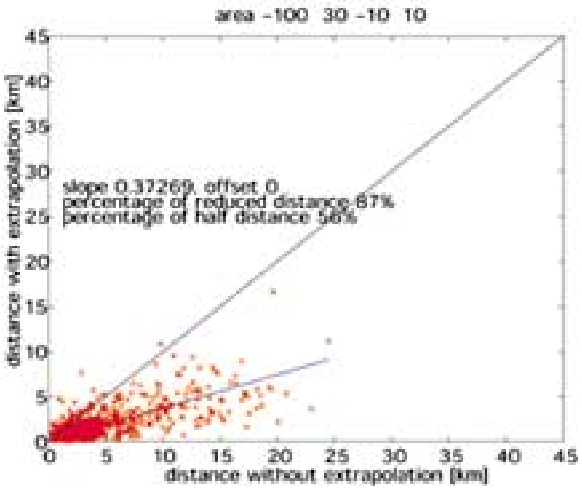


Fig 3.

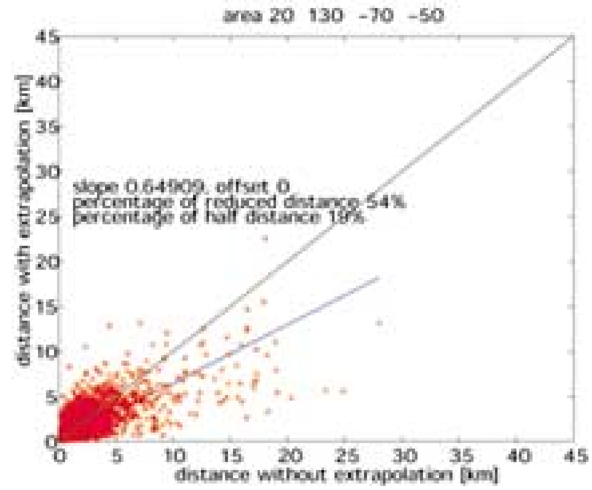


Fig 4.

Results from extrapolation of trajectories using cokriging. The results outside of the tropics are preliminary.

Appendix 3.7: Data Archiving (Sylvie Pouliquen)

Real Time

- These data are ephemera and will be replaced by the fully qualified data when it exists
- We have to provide the users with both datasets

Delay-Mode

- Do we agree that they will be multiplied delay-mode datasets?
- PI, Basin-wide, reanalysis…
- Each is centre has the responsibility to archive the datasets he generates.

Another issue: How to minimise the reprocessing made in the different centres?

Appendix 3.8 METADATA FOR AN ARGO FLOAT by Claudia Schmid

```
----------------------------------------------------------------------------
internal ID number                      currently an integer
                                        might become alphanumeric
                                        (data center id + integer)
                                        or obsolete (if replaced by extended
                                        WMO number)
transmission ID number                  if ARGOS float: integer
                                        if ORBCOMM float: I don't know
transmission type                       character string
                                        ARGOS or ORBCOMM
WMO ID number                           integer
WMO instrument type (table 1770)        integer
WMO recorder type (table 4770)          integer
ARGOS program number                    integer
                                        Is this field obsolete for ORBCOMM
floats?
                                        If not we have to give it a new name
...
instrument type                         assigned by AOML for ARGO-USA
                                        purpose: tells software what to do
                                        Is there a requirement to coordinate
this
                                        between the data centers, or is the
                                        detailed information below
sufficient?
----------------------------------------------------------------------------
delay of first down time [hours]        real number
down time [days]                        real number
up time [hours]                         real number
transmission repetition rate [sec]      real number
clock drift (hours/hours)               real number
                                        obviously this allows only a linear
                              clock drift correction
nominal drift depth [dbar]              integer or real
max. departure from drift depth [dbar]  integer or real
                                        for floats that control it actively
nominal profile depth [dbar]            integer or real
start time [yyyy mm dd hh mm (Z)]       UTC date and time the float was
started
                                        (reset of float, not the setting of
the
                                        internal clock)
                                        1997 07 23 04 30
launch time [yyyy mm dd hh mm (Z)]      UTC date and time
                                        1997 07 23 10 41
launch position [lat lon]               real numbers (longitude is -180 to
180)
deployment platform type                RV, VOS, airplane (what else?)
deployment platform identification      for research vessels: name
                                        for VOS: name or radio code?
                                        for airplanes: radio code?
deployment cruise id                    for ships only?
CTD station(s)                          give station number or NaN
----------------------------------------------------------------------------
```

```
last cycle                             you will find the last cycle here if
                                       a float died
                                       otherwise the value is NaN
comment                                character string
                                       describe problems (reason for early
                                       death etc)
-----------------------------------------------------------------------------
temp. recalibration date [yyyy mm dd]  integers or empty
pres. recalibration date [yyyy mm dd]  integers or empty
-----------------------------------------------------------------------------
float manufacturer                     character string
float serial number                    alphanumeric
float provider                         first and last name
float provider institution             use common abbreviation
originating country                    USA
float deployer                         first and last name
float deployer institution             use common abbreviation
PI                                     first and last name
PI institution                         use common abbreviation
PI email                               character string
-----------------------------------------------------------------------------
conductivity sensor type               alphanumeric
conductivity sensor manufacturer       name
conductivity sensor serial number      alphanumeric
temperature sensor type                alphanumeric
temperature sensor manufacturer        name
temperature sensor serial number       alphanumeric
pressure sensor type                   alphanumeric
pressure sensor manufacturer           name
pressure sensor serial number          alphanumeric
battery type                           character string
initial battery voltage                real number
Board type                             alphanumeric
Board serial number                    alphanumeric
CPU serial number                      alphanumeric
CPU date [yyyy mm dd]                  integers
AUX serial number                      alphanumeric
AUX date [yyyy mm dd]                  integers
software version                       alphanumeric
GPS (Y/N)                              if Y give GPS type, serial number etc
-----------------------------------------------------------------------------
calib eq 1 for pressure                OPB2 = PB1+(PB2-PB1)/AV1
calib eq 2 for pressure                IF BIN<=PB1 P = (BIN-0.5)*PGAIN*BLOK
calib eq 3 for pressure                IF BIN>PB1 P = (BIN-PB1-
0.25)*AV1*PGAIN*BLOK+P(PB1)
calib eq 4 for pressure                IF BIN>OPB2 P = (BIN-OPB2-
0.25)*AV2*PGAIN*BLOK+P(OPB2)
calib coef for pressure                blok=10; av1=2; av2=5; pb1=26;
pb2=60; pgain=1.0446; poff=-104.42;
calib eq 1 for temperature             R=AT*(BT+TCNTS)/(CT+TCNTS)
calib coef for temperature             at=-27.39286; bt=-5863.026;
ct=4366.056;
calib eq 2 for temperature             T = (H1+H2*ln(R)+H3*ln(R)**3)**-1-
273.15
calib coef for temperature             h1=0.0011192; h2=0.000235665;
h3=0.82431e-7;
```

```
calib eq 1 for conductivity             C =
32+CCNTS*cdiv*4+asal+TCNTS*(bsal*csal*TCNTS)
calib coef for conductivity             cdiv=0.002; asal=-0.68479e+01;
bsal=0.41177e-02;csal=0.14414e-05;
calib eq 2 for conductivity             C = coff+cgain*C
calib coef for conductivity             coff=0.0; cgain=1.00;
calib eq 3 for conductivity             C = C*(1+(ctref-T)*7.5e-6+1.5E-8*P)
calib coef for conductivity             ctref=18.365;
calib eq 1 for cpu voltage              VCPU = CPUGAIN*CNTS+CPUOFF
calib coef for cpu voltage              cpugain=0.02703; cpuoff=0.18919;
calib eq 1 for pump voltage             VPMP = PMP_GAIN*CNTS+PMP_OFF
calib coef for pump voltage             pmp_gain=0.06061; pmp_off=1.0909;
-----------------------------------------------------------------------------

General remark:
The calibration coefficients are part of each profile and they vary
considerably between float types (the calibration information given
here is hopefully much more complicated than than any future
```
versions). Sylvie will work out a way to get them into netcdf ...

Appendix 4: CO-IF0013 : multi-profile Coriolis NetCdf format

Interface name	**multi-profile Coriolis NetCdf format**
Reference	CO-IF0013
Source	Function "NetCdf dissemination" CO-FO-05-05
Destination	Function "web dissemination" CO-FO-05-01
Operating mode	-
Frequence	-
Estimated volume	-
Last update	31/08/2000

Dimensions definitions

N_CODE = 4 ;

N_DATE_TIME = 19 ;

// Null terminated string

N_STRING64 = 64 ;

N_STRING16 = 16 ;

N_STRING8 = 8 ;

N_STRING4 = 4 ;

N_STRING2 = 2 ;

// These dimensions depend on data set

N_PARAM = 4 ;

mN_PROF = 80 ;

mN_ZLEV = 547 ;

mN_HISTORY = unlimited;

mN_SURFACE = unlimited;

Metadata

Metadata are coded as global attributes

```
// global attributes (given with their fillValue)
                // Date and comments on updates of this file (char*80)
                :Last_update = "";
                // File version (referencing the version of the format)
```

```
                    :Version = "V1.0";
                    // Data provider
                    :Data_provider = "IFREMER/CORIOLIS";
                    // References of experiment (char*80)
                    :Experiment_name = "";
                    :Project_name = "";
                    // Description of experiment (char*256)
                    :Experiment_description = "";
                    // Description of the geographical area (char*80)
                    :Geographical_area = "";
                    // Date of reference (convention : YYYYMMDDHH24MI)
                    // 195001010000 recommended
                    :Reference_date_time = "";
                    :Start_date = "";
                    :Stop_date = "";
                    // Limits of the experiment (decimal degrees ,positive =North,East)
                    :South_latitude = -999. ; // [-90,90] for latitudes
                    :North_latitude = -999. ;
                    :West_longitude = -999. ; // ]-180,180] for longitudes
                    :East_longitude = -999. ;
                    // Coordinate system
                    :Coord_system = "GEOGRAPHICAL";
                    // GF3 code of the immersion reference parameter for the file (DEPH, PRES, TIME)
                    :Reference_parameter = "";
                    // Data interpolated on fixed levels of reference parameterù ("YES" or "NO")
                    :Normalized_data = "";
                    // Data type = list of Roscop codes for the file (char*80)
                    :Data_type = "";
                    // Description of file constitution : data, user request... (char*1024)
                    :File_contents = "";
                    // Comments on file contents/modification (char*256)
                    :comments = "";
```

General information on each profile

```
// General information on each profiles
          char PARAMETERS(N_PARAM,N_CODE) ;
                    PARAMETERS:long_name = "List of parameters" ;
                    PARAMETERS:Conventions = "GF3 code of the indexed parameter among
(DEPH, PRES, TEMP, PSAL, CNDC)";
                    PARAMETERS:_FillValue = " " ;
          char EXPERIMENT_NAME(mN_PROF,N_STRING16) ;
                    VOYAGE_NAME:long_name = "Experiment or voyage name" ;
                    VOYAGE_NAME:Conventions = "Text" ;
                    VOYAGE_NAME:_FillValue = " " ;
          char PLATFORM_NUMBER(mN_PROF,N_STRING16) ;
                    PLATFORM_NUMBER:long_name = "Float unique number" ;
                    PLATFORM_NUMBER:Conventions = "To Be Defined (AIC or extended WMO?) " ;
                    PLATFORM_NUMBER:_FillValue = " " ;
          char PLATFORM_PROVIDER(mN_PROF,N_STRING16) ;
                    PLATFORM_PROVIDER:long_name = "Float provider" ;
                    PLATFORM_PROVIDER:Conventions = "Text (to be defined ?) " ;
                    PLATFORM_PROVIDER:_FillValue = " " ;
          int STATION_NUMBER(mN_PROF) ;
                    STATION_NUMBER:long_name = " Station or float cycle number" ;
                    STATION_NUMBER:Conventions = "Cycle number for floats" ;
                    STATION_NUMBER:_FillValue = 99999 ;
          char DIRECTION(mN_PROF) ;
                    DIRECTION:long_name = "Direction of the station" ;
                    DIRECTION:Conventions = "A:Ascending Profile, D: Descending profile " ;
                    DIRECTION:_FillValue = " " ;
          char DATA_CENTRE(mN_PROF,N_STRING2) ;
                    DATA_CENTRE:long_name = "Data centre in charge of float real-time processing" ;
                    DATA_CENTRE:Conventions = "GTSPP table" ;
                    DATA_CENTRE:_FillValue = " " ;
          char REFERENCE(mN_PROF,N_STRING8) ;
```

```
		REFERENCE:long_name = "Reference for each profile in data centre" ;
		REFERENCE:Conventions = "To Be Defined" ;
		REFERENCE:_FillValue = " " ;
	char INST_TYPE(mN_PROF,N_STRING4) ;
		INST_TYPE:long_name = "Instrument type for profile" ;
		INST_TYPE:Conventions = "OMM probe code" ;
		INST_TYPE:_FillValue = " " ;
	char REC_TYPE(mN_PROF,N_STRING4) ;
		REC_TYPE:long_name = "Recoder type for profiles" ;
		REC_TYPE:Conventions = "OMM recorder code" ;
		REC_TYPE:_FillValue = " " ;
	char DATE(mN_PROF,N_DATE_TIME) ;
		DATE:long_name = "Dates of each profile" ;
		DATE:Conventions = "YYYYMMDDHH24MISS" ;
		DATE:_FillValue = " " ;
		DATE:epic_code = ??? ;
	float BOTTOM_DEPTH(mN_PROF) ;
		BOTTOM_DEPTH:long_name = "Bottom depth of each profile" ;
		BOTTOM_DEPTH:unit= "m" ;
		BOTTOM_DEPTH:Conventions = "not mandatory, always positive" ;
		BOTTOM_DEPTH:_FillValue = -99999.f ;
		BOTTOM_DEPTH:valid_min = 0 ;
		BOTTOM_DEPTH:valid_max = 15000 ;
	double JULD(mN_PROF) ;
		JULD:long_name = "Julian day of each profile relative to REFERENCE_DATE" ;
		JULD:units = "julian days" ;
		JULD:Conventions = "Relative julian days with decimal part (as parts of day)" ;
		JULD:_FillValue = -99999. ;
		JULD:epic_code = ??? ;
	double LATITUDE(mN_PROF) ;
		LATITUDE:long_name = "Latitude of each profile" ;
		LATITUDE:units = "degrees_north" ;
		LATITUDE:_FillValue = -99999.f ;
		LATITUDE:valid_min = -90.f ;
		LATITUDE:valid_max = 90.f ;
		LATITUDE:epic_code = ???;
	double LONGITUDE(mN_PROF) ;
		LONGITUDE:long_name = "Longitude of each profile" ;
		LONGITUDE:units = "degrees_east" ;
		LONGITUDE:_FillValue = -99999.f ;
		LONGITUDE:valid_min = -180.f ;
		LONGITUDE:valid_max = 180.f ;
		LONGITUDE:epic_code = ???;
	char Q_DATE(mN_PROF) ;
		Q_DATE:long_name = "Quality on Date and Time" ;
		Q_DATE:Conventions = "Q where Q =[0,9]" ;
		Q_DATE:_FillValue = "9" ;
	char Q_POSITION(mN_PROF) ;
		Q_POSITION:long_name = "Quality on position (latitude and longitude)" ;
		Q_POSITION:Conventions = "Q where Q =[0,9]" ;
		Q_POSITION:_FillValue = "9" ;
	char POSITIONNING_SYSTEM(mN_PROF, mN_STRING16) ;
		POSITIONNING_SYSTEM:long_name = "Positionning system" ;
		POSITIONNING_SYSTEM:Conventions = "argos, gps, ..." ;
		POSITIONNING_SYSTEM:_FillValue = " " ;
	char Q_BOTTOM(mN_PROF) ;
		Q_BOTTOM:long_name = "Quality on bottom depth" ;
		Q_BOTTOM:Conventions = "Q where Q =[0,9], not mandatory" ;
		Q_BOTTOM:_FillValue = "9" ;
	char Q_DEPTH(mN_PROF) ;
		Q_DEPTH:long_name = "Quality on bottom depth" ;
		Q_DEPTH:Conventions = "Q where Q =[0,9], not mandatory" ;
		Q_DEPTH:_FillValue = "9" ;
	char Q_PROFILE_PRES(mN_PROF) ;
		Q_PROFILE_PRES:long_name = "Global quality flag of pressure profile" ;
		Q_PROFILE_PRES:Conventions = "Q where Q =[0,9]" ;
		Q_PROFILE_PRES:_FillValue = "9" ;
	char Q_PROFILE_DEPH(mN_PROF) ;
		Q_PROFILE_DEPH:long_name = "Global quality flag of depth profile" ;
		Q_PROFILE_DEPH:Conventions = "Q where Q =[0,9]" ;
```

```
Q_PROFILE_DEPH:_FillValue = "9" ;
char Q_PROFILE_TEMP(mN_PROF) ;
Q_PROFILE_TEMP:long_name = "Global quality flag of temperature profile" ;
Q_PROFILE_TEMP:Conventions = "Q where Q =[0,9]" ;
Q_PROFILE_TEMP:_FillValue = "9" ;
char Q_PROFILE_PSAL(mN_PROF) ;
Q_PROFILE_PSAL:long_name = "Global quality flag of practical salinity profile" ;
Q_PROFILE_PSAL:Conventions = "Q where Q =[0,9]" ;
Q_PROFILE_PSAL:_FillValue = "9" ;
char Q_PROFILE_CNDC(mN_PROF) ;
Q_PROFILE_CNDC:long_name = "Global quality flag of practical salinity profile" ;
Q_PROFILE_CNDC:Conventions = "Q where Q =[0,9]" ;
Q_PROFILE_CNDC:_FillValue = "9" ;
int Q_CHECK_SUMMARY(mN_PROF) ;
Q_CHECK_SUMMARY:long_name = "Bit string identifying passed quality checks" ;
Q_CHECK_SUMMARY:Conventions = "GTSPP check ident : 1 bit per check" ;
Q_CHECK_SUMMARY:_FillValue = 0x000000 ;
int HISTORY_COUNT(mN_PROF) ;
HISTORY_COUNT:long_name = "Number of history data rows associated with the profile" ;
HISTORY_COUNT:_FillValue = 0 ;
int HISTORY_INDEX(mN_PROF) ;
HISTORY_INDEX:long_name = "Index of first history data row associated with the profile" ;
HISTORY_INDEX:_FillValue = "9" ;
int SURFACE_COUNT(mN_PROF) ;
SURFACE_COUNT:long_name = "Number of surface data rows associated with the profile" ;
SURFACE_COUNT:_FillValue = 0 ;
int SURFACE_INDEX(mN_PROF) ;
SURFACE _INDEX:long_name = "Index of first surface data row associated with the profile" ;
SURFACE _INDEX:_FillValue = "9" ;
```

Profile data

/ / Data for profiles

/ / Variable naming conventions:

/ / Measured value : GF3 code (mandatory if measured and mentionned in PARAMETERS),

/ / Associated QC flag : "QC_" + GF3 (mandatory if measured and mentionned in PARAMETERS),

/ / Error estimation : "Error_" + GF3 (not mandatory)

```
float PRES(mN_ZLEV, mN_PROF) ;
PRES:long_name = "Pressure (sensor drift removed)" ;
PRES:_FillValue = 9999.f ;
PRES:units = "decibar = 1000 Pa" ;
PRES:valid_min = 0. ;
PRES:valid_max = 15000. ;
DATE:epic_code = ??? ;
char QC_PRES(mN_ZLEV, mN_PROF) ;
Q_PRES:long_name = "Global quality on pressure" ;
Q_PRES:Conventions = "Q where Q =[0,9]" ;
Q_PRES:_FillValue = "9" ;
float DEPH(mN_ZLEV, mN_PROF) ;
DEPH:long_name = "Depth below sea surface" ;
DEPH:_FillValue = 9999.f ;
DEPH:units = "metres" ;
DEPH:valid_min = 0. ;
DEPH:valid_max = 15000. ;
DEPH:comment = "Not mandatory, computed form PRES, TEMP, PSAL";
DEPH:epic_code = ??? ;
char QC_DEPH(mN_ZLEV, mN_PROF) ;
Q_DEPH:long_name = "Global quality on depth" ;
Q_DEPH:Conventions = "Q where Q =[0,9]" ;
Q_DEPH:_FillValue = "9" ;
float TEMP(mN_ZLEV, mN_PROF) ;
TEMP:long_name = "Temperature in situ T90 scale (corrected value)" ;
TEMP:_FillValue = 9999.f ;
```

```
TEMP:units = "degres Celsius" ;
TEMP:valid_min = -3. ;
TEMP:valid_max = 40. ;
TEMP:comment = "value corrected from sensor drift";
TEMP:epic_code = ??? ;
float OBSERVED_TEMP(mN_ZLEV, mN_PROF) ;
OBSERVED_TEMP:long_name = "Temperature in situ T90 scale (original value)" ;
OBSERVED_TEMP:_FillValue = 9999.f ;
OBSERVED_TEMP:units = "degres Celsius" ;
OBSERVED_TEMP:valid_min = -3. ;
OBSERVED_TEMP:valid_max = 40. ;
OBSERVED_TEMP:comment = "in situ measurement";
char QC_TEMP(mN_ZLEV, mN_PROF) ;
Q_TEMP:long_name = "Global quality on temperature" ;
Q_TEMP:Conventions = "Q where Q =[0,9]" ;
Q_TEMP:_FillValue = "9" ;
float PSAL(mN_ZLEV, mN_PROF) ;
PSAL:long_name = "Practical salinity, sal78" ;
PSAL:_FillValue = 9999.f ;
PSAL:units = "P.S.U." ;
PSAL:valid_min = 0. ;
PSAL:valid_max = 60. ;
PSAL:comment = "sensor drift has been removed";
PSAL:epic_code = ??? ;
float OBSERVED_PSAL(mN_ZLEV, mN_PROF) ;
OBSERVED_PSAL:long_name = "Practical salinity, sal78 (in situ measurement)" ;
OBSERVED_PSAL:_FillValue = 9999.f ;
OBSERVED_PSAL:units = "P.S.U." ;
OBSERVED_PSAL:valid_min = 0. ;
OBSERVED_PSAL:valid_max = 60. ;
OBSERVED_PSAL:comment = "in situ measurement, mandatory if salinity (conductivity) measured";
char QC_PSAL(mN_ZLEV, mN_PROF) ;
Q_PSAL:long_name = "practical salinity  quality flag" ;
Q_PSAL:Conventions = "Q where Q =[0,9]" ;
Q_PSAL:_FillValue = "9" ;
float CNDC(mN_ZLEV, mN_PROF) ;
CNDC:long_name = "Electrical conductivity" ;
CNDC:_FillValue = 9999.f ;
CNDC:units = "mmho/cm " ;
CNDC:valid_min = 0. ;
CNDC:valid_max = 60. ;
CNDC:comment = "not mandatory, avalaible only if transmitted";
CNDC:epic_code = ??? ;
char QC_CNDC(mN_ZLEV, mN_PROF) ;
Q_CNDC:long_name = "electrical conductivity quality flag" ;
Q_CNDC:Conventions = "Q where Q =[0,9]" ;
Q_CNDC:_FillValue = "9" ;
Q_CNDC:comment = " not mandatory, flag duplicated from PSAL quality checks";
```

History data

// History on profiles QC

// Based on GTSPP format

// One history row = same index for (INSTITUTION, SOFTWARE, SOFTWARE_RELEASE,
// ACTION, PARAMETER, IMMERSION, PREVIOUS_VALUE)

// History row for each profile retreived using HISTORY_INDEX and HISTORY COUNT

// mN_HISTORY may be unlimited

```
char HISTORY_QC_INSTITUTION (mN_HISTORY, mN_STRING4) ;
HISTORY_QC_INSTITUTION:long_name = "Institution which performed QC";
HISTORY_ QC_INSTITUTION:Conventions = "GTSPP institution code" ;
HISTORY_ QC_INSTITUTION:_FillValue = " ";
char HISTORY_ QC_SOFTWARE(mN_HISTORY, mN_STRING4) ;
HISTORY_ QC_SOFTWARE:long_name = "Software which performed QC" ;
```

```
HISTORY_ QC_SOFTWARE:Conventions = "Institution dependent" ;
HISTORY_ QC_SOFTWARE:_FillValue = " " ;
char HISTORY_ QC_SOFTWARE_RELEASE(mN_HISTORY, mN_STRING4) ;
HISTORY_ QC_SOFTWARE_RELEASE:long_name = "Version/release of software which
performed QC" ;
HISTORY_ QC_SOFTWARE_RELEASE:Conventions = "Institution dependent" ;
HISTORY_ QC_SOFTWARE_RELEASE:_FillValue = " " ;
char HISTORY_ACTION(mN_HISTORY, mN_STRING4) ;
HISTORY_ACTION:long_name = " QC action performed on data" ;
HISTORY_ACTION:Conventions = "GTSPP (MEDS) action code" ;
HISTORY_ACTION:_FillValue = " " ;
char HISTORY_PARAMETER(mN_HISTORY, mN_STRING4) ;
HISTORY_PARAMETER:long_name = "Parameter  QC action is performed on" ;
HISTORY_PARAMETER:Conventions = "GF3 parameter code" ;
HISTORY_PARAMETER:_FillValue = " " ;
float HISTORY_IMMERSION(mN_HISTORY) ;
HISTORY_PRES:long_name = "Immersion QC action applied on" ;
HISTORY_PRES:_FillValue = 9999.f ;
HISTORY_PRES:units = "dbars (?)" ;
float HISTORY_PREVIOUS_VALUE(mN_HISTORY) ;
HISTORY_PREVIOUS_VALUE:long_name = "Parameter/Flag previous value before QC
action" ;
HISTORY_PREVIOUS_VALUE:_FillValue = 9999.f ;
```

Surface data

// Surface trajectory data

// One surface row = same index for (DATE, LATITUDE, LONGITUDE, TEMP, PSAL)

// Surface row for each profile retreived using SURFACE_INDEX and SURFACE_COUNT

// mN_SURFACE may be unlimited

```
        char SURFACE_DATE(mN_SURFACE,N_DATE_TIME) ;
                SURFACE_DATE:long_name = "Dates of each surface location" ;
                SURFACE_DATE:Conventions = "YYYYMMDDHH24MISS" ;
                SURFACE_DATE:_FillValue = " " ;
                SURFACE_DATE:epic_code = ??? ;
        double SURFACE_JULD(mN_SURFACE) ;
                SURFACE_JULD:long_name = "Julian day of each surface location
relative to REFERENCE_DATE" ;
                SURFACE_JULD:units = "julian days" ;
                SURFACE_JULD:Conventions = "Relative julian days with decimal
part (as parts of day)" ;
                SURFACE_JULD:_FillValue = -99999. ;
                SURFACE_JULD:epic_code = ??? ;
        double SURFACE_LATITUDE(mN_SURFACE) ;
                SURFACE_LATITUDE:long_name = "Latitude of each surface location"
;
                SURFACE_LATITUDE:units = "degrees_north" ;
                SURFACE_LATITUDE:_FillValue = -99999.f ;
                SURFACE_LATITUDE:valid_min = -90.f ;
                SURFACE_LATITUDE:valid_max = 90.f ;
                SURFACE_LATITUDE:epic_code = ???;
        double SURFACE_LONGITUDE(mN_SURFACE) ;
                SURFACE_LONGITUDE:long_name = "Longitude of each surface
location" ;
```

```
                SURFACE_LONGITUDE:units = "degrees_east" ;
                SURFACE_LONGITUDE:_FillValue = -99999.f ;
                SURFACE_LONGITUDE:valid_min = -180.f ;
                SURFACE_LONGITUDE:valid_max = 180.f ;
                SURFACE_LONGITUDE:epic_code = ???;
        char SURFACE_Q_DATE(mN_SURFACE) ;
                SURFACE_Q_DATE:long_name = "Quality on Surface location Date and
Time" ;
                SURFACE_Q_DATE:Conventions = "to be defined" ;
                SURFACE_Q_DATE:_FillValue = " depends on positionning system " ;
        char SURFACE_Q_POSITION(mN_SURFACE) ;
                SURFACE_Q_POSITION:long_name = "Quality on position (latitude and
longitude)" ;
                SURFACE_Q_POSITION:Conventions = "Positionning system convention
(argos class)" ;
                SURFACE_Q_POSITION:_FillValue = "depends on positionning system"
;
        float SURFACE_TEMP(mN_SURFACE) ;
                SURFACE_TEMP:long_name = "Sea surface temperature in situ T90
scale (corrected value)" ;
                SURFACE_TEMP:_FillValue = 9999.f ;
                SURFACE_TEMP:units = "degres Celsius" ;
                SURFACE_TEMP:valid_min = -3. ;
                SURFACE_TEMP:valid_max = 40. ;
                SURFACE_TEMP:comment = "not mandatory, present only if measured";
                SURFACE_TEMP:epic_code = ??? ;
        char SURFACE_QC_TEMP(mN_SURFACE) ;
                SURFACE_QC_TEMP:long_name = "Quality flag on sea surface
temperature" ;
                SURFACE_QC_TEMP:Conventions = "Q where Q =[0,9]" ;
                SURFACE_QC_TEMP:_FillValue = "9" ;
                SURFACE_QC_TEMP:comment = "not mandatory, present only if
measured";
```

Quality flags	
Value	***Description (to be precised)***
0	Unqualified
1	Correct value (All ckecks passed)
2	Probably good but value inconsistent with statistics (Differ from climatology)
3	Probably bad (spike, gradient, ... if other tests passed)
4	Bad value, Impossible value (out of scale, vertical instability, constant profile, ...)
5	Value modified during Quality Control
6-7	Not used (available)
8	Profile interpolated at standard depth
9	Missing value

Appendix 5: Argo Automatic Tests

Introduction

At the Argo data management meeting in Brest I was asked to propose the suite of tests to apply automatically to float data before the data are sent to the GTS. The agreed starting point was the suite of GTSPP test procedures. For your information, you can see the document that describes the test procedures in general (http://www.meds-sdmm.dfo- mpo.gc.ca/ALPHAPRO/gtspp/qcmans/MG22/guide22_e.htm) and the tests explicitly (http://www.meds-sdmm.dfo-mpo.gc.ca/ALPHAPRO/gtspp/qcmans/mg22/annexb.htm).

The following is the first draft of a document that will explain the automatic procedures applied to float data. As it is a draft, everything is up for discussion. The document summarizes the QC tests that will be applied automatically to profiling float data. The tests are presented in two groups. Failing a test in the first group means that nothing is sent to the GTS. Failing a test from the second group (a value is flagged as wrong) means that only the observation/level that fails is excluded from the GTS message.

Tests that cause data from a float to be flagged as wrong, or results that stop an entire message from being sent to the GTS are highlighted in bold. In these cases an explanation is provided why the data are stopped. Note that data values flagged as wrong should not be encoded into the TESAC message. No matter what flag an observation receives from the automated procedure, the original value and flag should be encoded into the netCDF file and sent to both the Argo servers and to the PI.

Group 1: Tests that stop complete profiles from going to the GTS

1.1 Platform Identification

This test will stop data going to the GTS. Every centre handling float data and posting them to the GTS will need to prepare a metadata file for each float. Since each float being handled by a centre should have a metadata file, there is no reason why, except a mistake, a float ID should be unknown. Letting mistaken IDs out to distribution causes confusion later on. This will be acute if the float ID is used to generate the unique data tag that we discussed in Brest.

1.2 Impossible Date/Time

This test will stop data going to the GTS. The test requires that the observation date and time from the float be earlier or the same as the processing date and time, that the month be 1-12, day be the appropriate number for the month, that the hour be 0 to 23 and minutes be 0 to 59. Data are not valuable if the date is wrong.

1.3 Impossible Location

This test will stop data going to the GTS. The test requires that the observation latitude and longitude from the float be located between +/- 90degrees N and +/- 180degrees E. Data are not valuable if the position is wrong.

1.4 Position on Land

This test will stop data going to the GTS. The test requires that the observation latitude and longitude from the float be located in an ocean. We suggest use of at least the 5 minute bathymetry file that is generally available. Data are not valuable if the position is wrong.

1.5 Impossible Speed

Drift speeds for floats can be generated given the positions and times of the floats when they are at the surface. In all cases we would not expect the drift speed to exceed 3 m/s. If it does, it means either a position or time is wrong, or a float is mislabeled. In any of these cases, there is something seriously wrong with the data.

1.6 Impossible Sounding

This test applies to TESACs for which a sounding field is provided. Since this is unlikely for TESACs from profiling floats, this test does not apply.

Group 2: Tests that remove selected values and levels from data going to the GTS

2.1 Global Impossible Parameter Values

The ranges set for these variables are quite wide. However, upper limits of 35 degrees or T and 40 PSU for salinity may not be high enough to accommodate certain bodies of water, such as the Red Sea. Before choosing to have this test failure stop data going to the GTS, we need to reconsider these limits.

2.2 Regional Impossible Parameter Values

The ranges set for these special regions need to be considered. I view this as the same way as test 2.1 Before choosing to have this test failure stop data going to the GTS, we need to reconsider these limits.

2.3 Increasing Depth

We have already seen reports from floats where there are multiple observed values at a given depth. I would accept the first occurring value at a depth and flag all of the rest of the observed values at the same depth as wrong.

2.4 Profile Envelop

This would not cause a flag of “wrong” to be set on any values.

2.5 Constant Profile

This would not cause a flag of “wrong” to be set on any values.

2.6 Freezing Point
This test is based on the UNESCO algorithm and so can be relied upon. If a value fails the test it should be flagged as wrong.

2.7 Spike
This test is relatively insensitive in that only serious spikes are caught. Any values failing this test should be flagged as wrong.

2.8 Top and Bottom Spike
This test is relatively insensitive such that it is not able to identify failures such as a spike at the surface of 0.5 degrees. We should perhaps tighten up the thresholds. I recommend failed observations be flagged as wrong.

2.9 Gradient
This test is relatively insensitive to all but the most extreme gradients. Points failing should be flagged as wrong.

2.10 Density Inversion
Experience has shown that small density inversions do appear in flat data presently circulating on the GTS. They do not appear to affect adjacent observations.

2.11 Bottom
This checks that the deepest observation depth is not deeper than the depth of water from a bathymetry file. Since the bathymetry file is only approximately correct, we can either allow the deepest reported depth to exceed the bathymetric depth by some factor (10%?) before a flag or wrong is set, or elliminate the test.

2.12 Temperature Inversion
This test is one that attempts to catch problems that are characteristic of temperature profiles from XBTs. Any observed value failing this test would not be flagged wrong.

3.x Climatology
Because climatologies use averaging techniques to arrive at their values, by their very nature they exclude extreme but real observations. We would not use a failure in a comparison of an observed value to a climatological vale to flag the observed value as wrong.