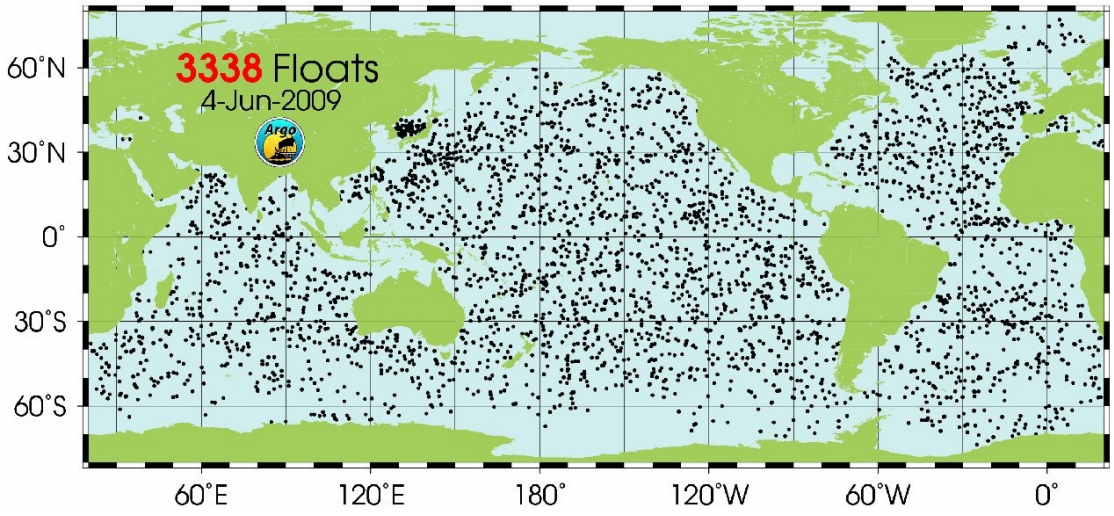
Argo data service: from data collection to final distribution". The respective roles of DACs and GDACs"



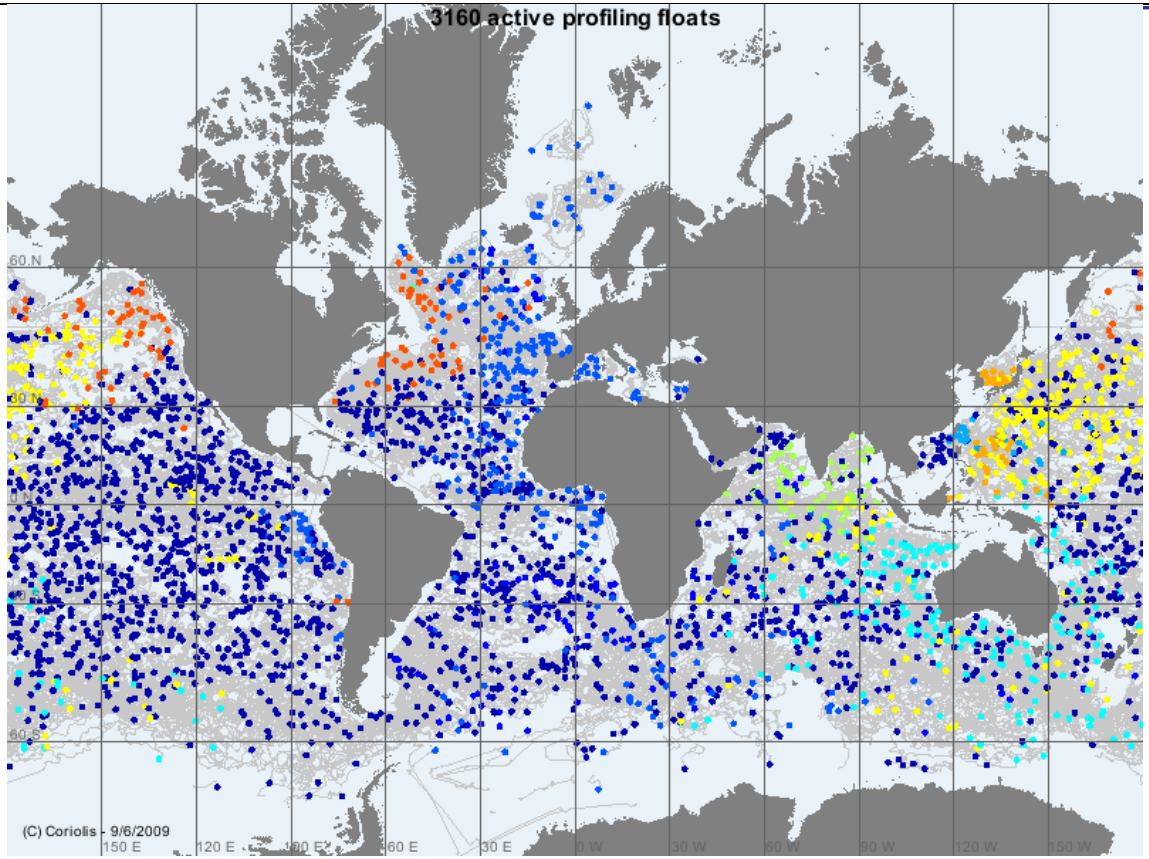
Loïc Petit de la Villéon
&
Sylvie Pouliquen
Ifremer

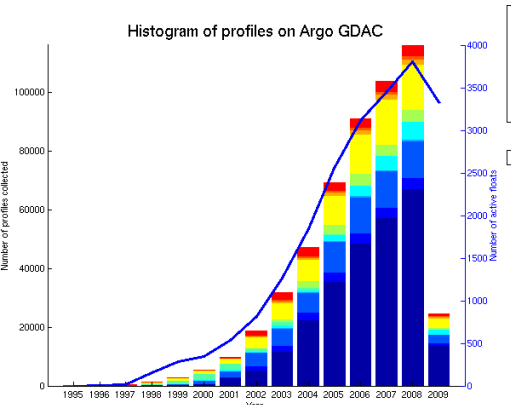
The missions of the Data Management structure and team

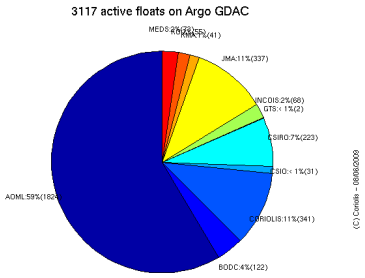
- Collect the data and process them
- Quality control the data within 24h using agreed procedures
- Distribute the Argo data on GTS and on Internet from a unique access point
- Correct in delayed mode the profiles using agreed procedures
- Inform Users on the Argo network
- Archive the data on the longer term
- Provide tools to monitor the network

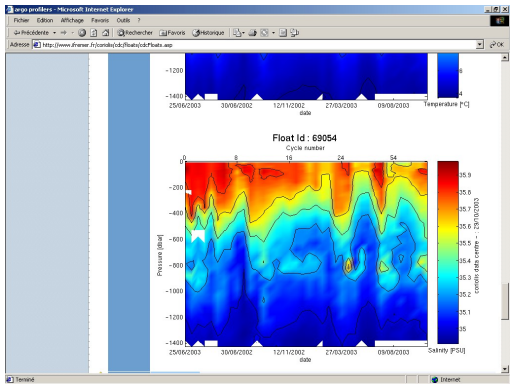
Argo data management

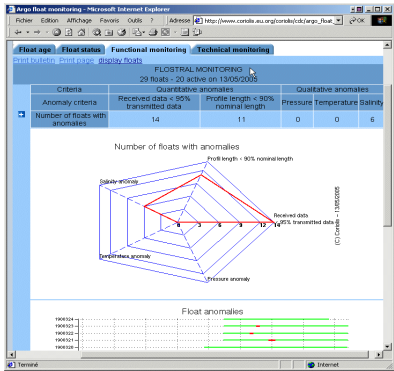
part of the integrated global observation strategy











Argo data management actors

PI : Principal Investigators

The scientists who deploy the floats, then carry out delayed mode QC and return data to DAC Centres within a few months of observations. Delayed mode operators may help Pis for Delayed mode QC

DAC: Data Assembly Centres

The DACs are facilities set up by many of the nations deploying floats.
Their role is crucial in acquiring the raw data from the floats and standardizing the handling. Their functions include :

- collecting data
- converting to standard exchange formats
- applying standardized real-time quality control
- delivering data to the GTS and GDACs within 24 hours of the surfacing and to PIs on a more relaxed schedule
- coordinating Argo data handling for the floats under their control

GDAC : Global Data Centres

The GDACs are the data servers where the master copies of the data reside.
The Argo global dataset is available from GDACs; central points for data distribution on Internet for all floats.
Located in Monterrey (US GODAE/FNMOC/USA) and Brest (Coriolis/Ifremer/France).
A synchronization between the 2 GDACs centres occurs daily.

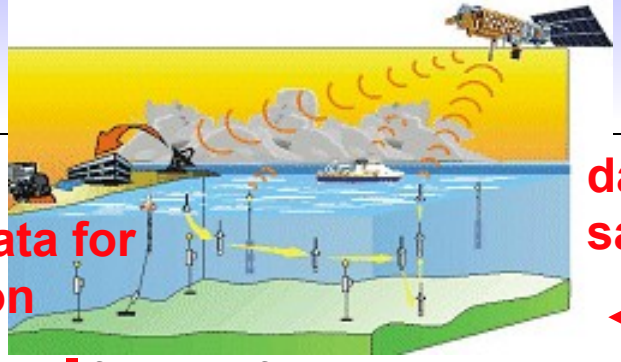
Argo data management actors

- ARC : Argo Regional Centres
  The Regional Centres look at data from ocean basins to verify float data consistency and generate products.
  They provide basin-wide synthesis of all floats data with other available data.
  They provide a feedback to Pis on delayed mode quality control.

- AIC : Argo Information Centre
  Centre located in Toulouse/France, in charge of informing on the Argo program status and to provide all necessary information to users.

- Argo long term archive
  Data centre located in NODC/USA in charge of the long term archive of all Argo data.

Argo data flow



data sent ashore via satellite

prepare data for distribution

floats deployment and data certification

float data from satellite

4-12 hours

24 hours, real-time QC

dac data assembly centres
aoml, coriolis, jma, meds, *cls*

gts : 24 hours

Delayed Mode Operators & PI
principal investigators
howard freeland, yves desaubies …

DM QC, few months

users
scientists, modelers, students…

24 hours

ftp : 24 hours
web : 12 hours

perform regional analyses

gdac global data centers
usgodae, coriolis

24 hours

1 day
to
several years

information about argo

long term archive

here is the master copy of argo data

1 year

regional centres
n. atlantic, pacific, indian

aic argo information center

long term archive
usnodc

Argo quality control process

Real-time (RT) data stream

Function: Apply agreed RT QC tests to float data. Assign quality flags.
Users: Operational centres, data assimilation, researchers needing timely data.
Timeframe: 24-48 hrs after transmission.
Who/Where: Performed by National Data Assembly Centres.

Delayed-mode (DM) data stream

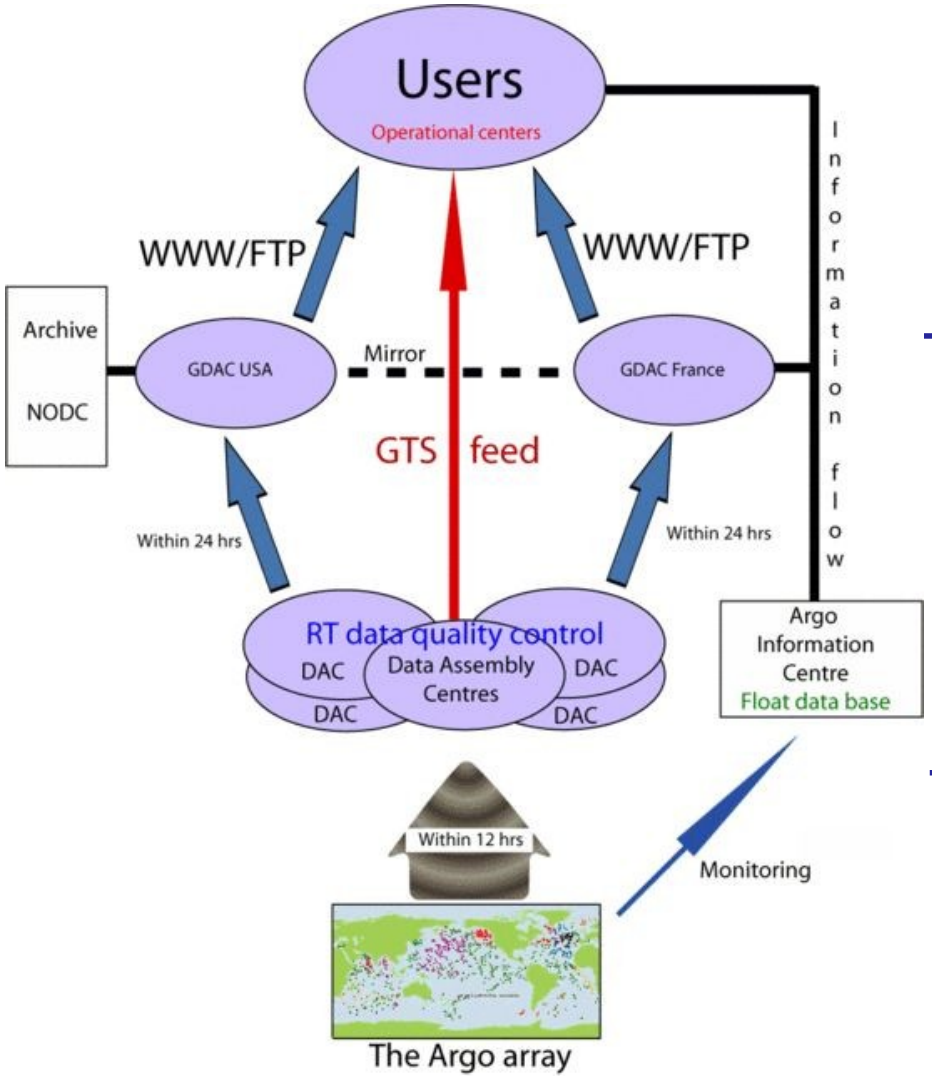
Function: Apply accepted DM procedures to float data. Provide statistically justified corrections using accepted methods. Provide feedback to RT system.
Users: All needing adjusted data with error estimates.
Timeframe: 12-18 months after transmission.
Who/Where: Perform by PIs with DM Agencies or Regional Centres.

Regional Analysis

Function: Provide basin-wide synthesis of all float data with other available data.
Users: Researchers of climate change and model validation.
Timeframe: Lifespan of float.
Who/Where: Perform by Regional Centres.

All data are available from the two Argo Global Data Assembly Centres: Coriolis at IFREMER, and USGODAE at Monterey.

Real Time Argo Data System



Two GDAC: Coriolis – US-Godae

Ten National Data Centers
AOML, BODC, CSIO, CSIRO, Coriolis, INCOIS, JMA, KMA,KORDI, MEDS

Argo real-time quality control

- Profiles : 17 automatic QC tests performed before gdac and gts distribution
  1. Platform Identification
  2. Impossible Date
  3. Impossible Location
  4. Position on Land
  5. Impossible Speed
  6. Global Range
  7. Regional Range
  8. Pressure Increasing
  9. Spike
  10. *Top and Bottom Spike : removed*
  11. Gradient
  12. Digit Rollover
  13. Stuck Value
  14. Density Inversion
  15. Grey List
  16. Gross salinity or temperature sensor drift
  17. *Visual QC (not mandatory)*
  18. Frozen profile
  19. Deepest pressure

- Trajectories : 7 automatic QC tests performed before gdac and gts distribution
  1. Platform Identification
  2. Impossible Date
  3. Impossible Location
  4. Position on Land
  5. Impossible Speed
  6. Global Range
  7. Regional Global Parameter

Argo real-time quality control

- Profiles : 17 automatic QC tests performed before gdac and gts distribution
  1. Platform Identification
  2. Impossible Date
  3. Impossible Location
  4. Position on Land
  5. Impossible Speed
  6. Global Range
  7. Regional Range
  8. Pressure Increasing
  9. Spike
  10. *Top and Bottom Spike : removed*
  11. Gradient
  12. Digit Rollover
  13. Stuck Value
  14. Density Inversion
  15. Grey List
  16. Gross salinity or temperature sensor drift
  17. *Visual QC (not mandatory)*
  18. Frozen profile
  19. Deepest pressure

- Trajectories : 7 automatic QC tests performed before gdac and gts distribution
  1. Platform Identification
  2. Impossible Date
  3. Impossible Location
  4. Position on Land
  5. Impossible Speed
  6. Global Range
  7. Regional Global Parameter

- QC flag scale
  0. No QC was performed

  1 Good data

  2. Probably good data
  3. Bad data that are potentially correctable

  4 Bad data

  5. Value changed
  6. Not used
  7. Not used
  8. Interpolated value
  9. Missing value

Argo real-time quality control

ALL measurements are provided to users. Use the flags to know if the data are good

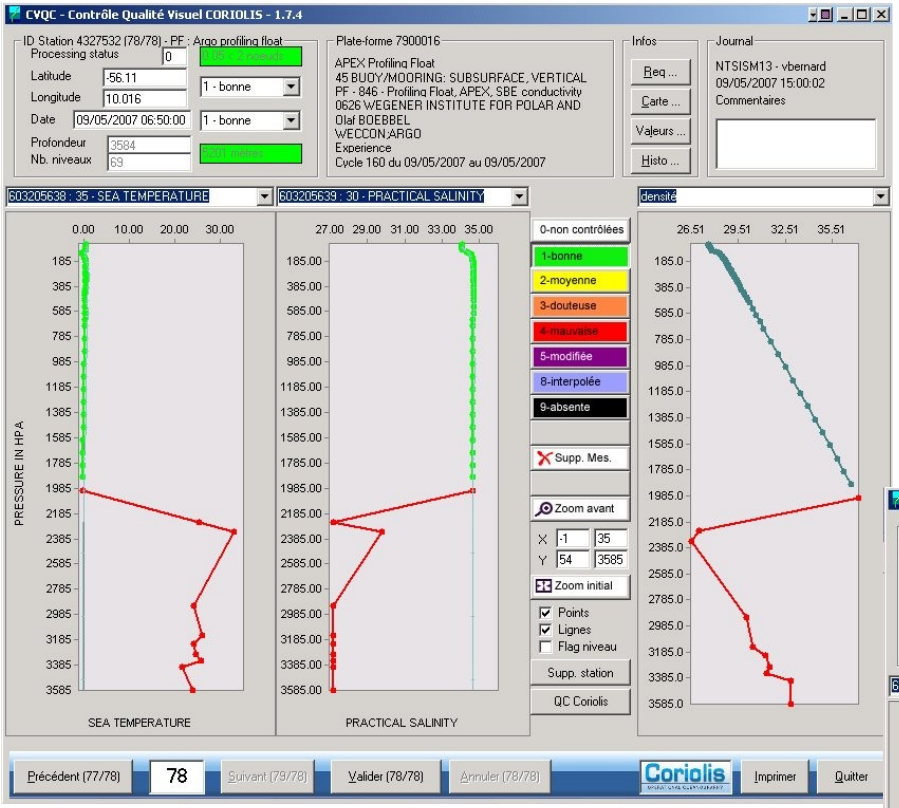
To be on the safe side only use data with flag=1

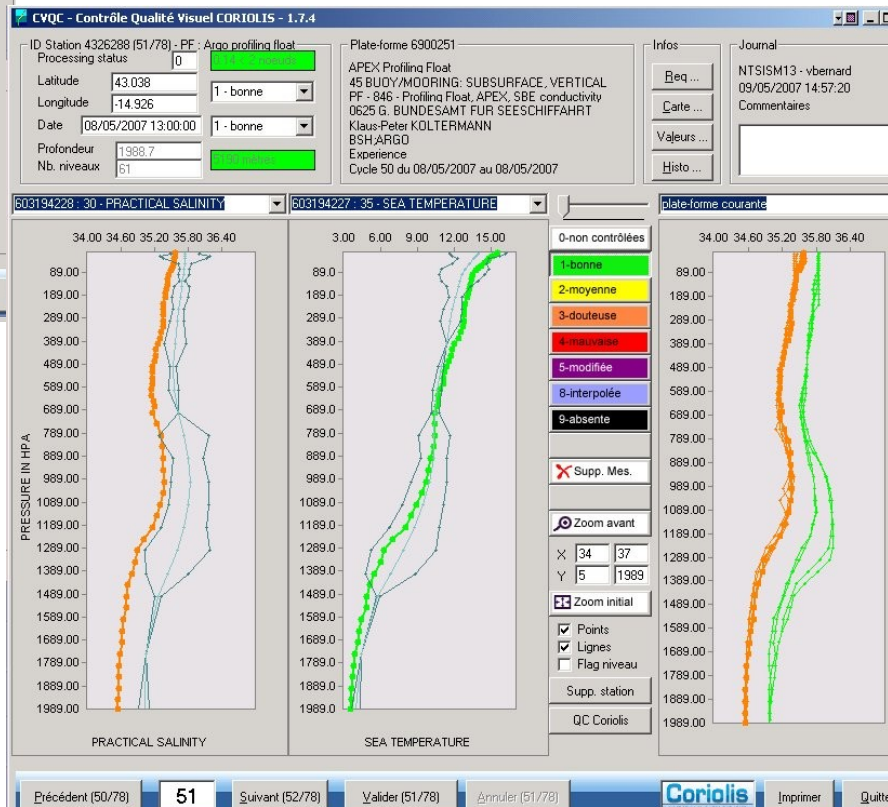
7 Regional Range
8 Pressure Increasing
9 Spike
10 Top and Bottom Spike : removed
11 Gradient
12 Digit Rollover
13 Stuck Value
14 Density Inversion
15 Grey List
16 Gross salinity or temperature sensor drift
17 Visual QC (not mandatory)
18 Frozen profile
19 Deepest pressure

7 Regional Global Parameter

- QC flag scale
  - 0 No QC was performed
  - **1 Good data**
  - 2 Probably good data
  - 3 Bad data that are potentially correctable
  - **4 Bad data**
  - 5 Value changed
  - 6 Not used
  - 7 Not used
  - 8 Interpolated value
  - 9 Missing value

Example of anomalies detected by RT QC





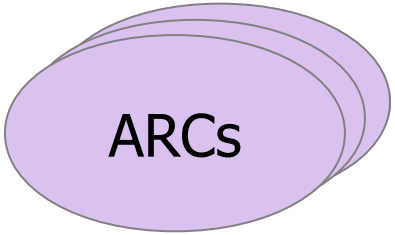
Real-Time QC can't detect all anomalies !!

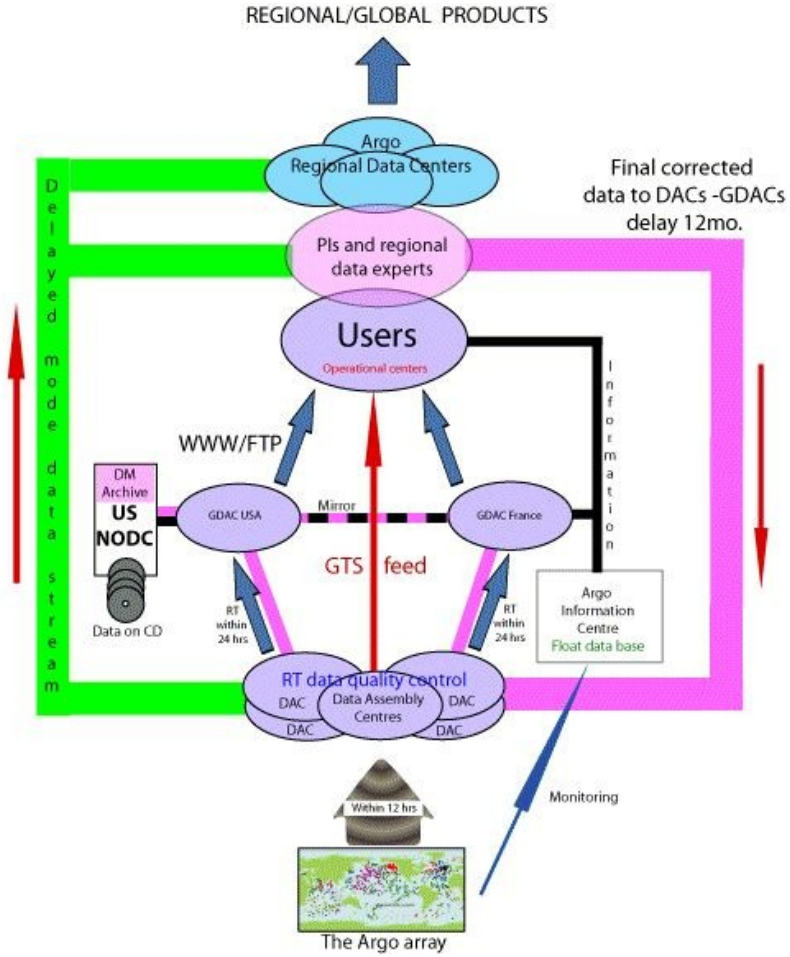
- Realtime QC automatic tests thresholds are a **compromise between letting bad data going through and stopping good data**

- Profiles are mainly qualified one by one and the experience has proven that **individual profiles may look good while they are not coherent with the neighbours**

Delayed Mode Argo Data System

Five Arcs coordinated by AOML- BODC Coriolis – INCOIS – IPRC





12 Delayed Mode Operators

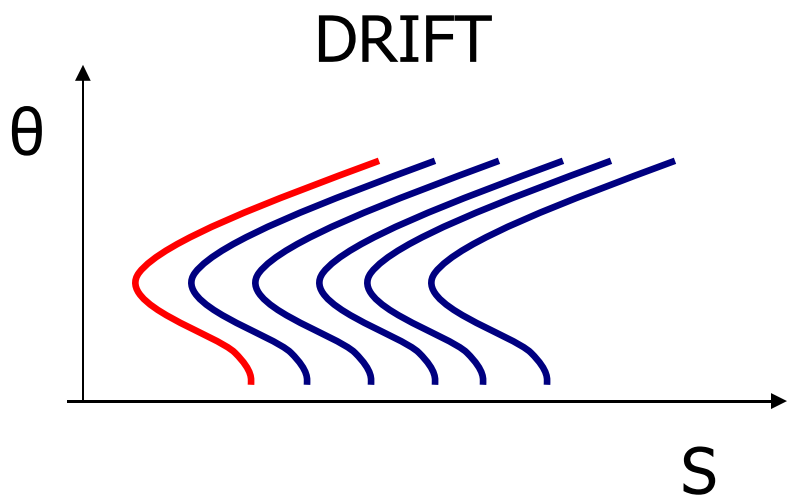
BODC–BSH- Coriolis-CSIO- CSIRO- INCOIS- JAMSTEC - OGS- PMEL- UCSD –WHOI - UW

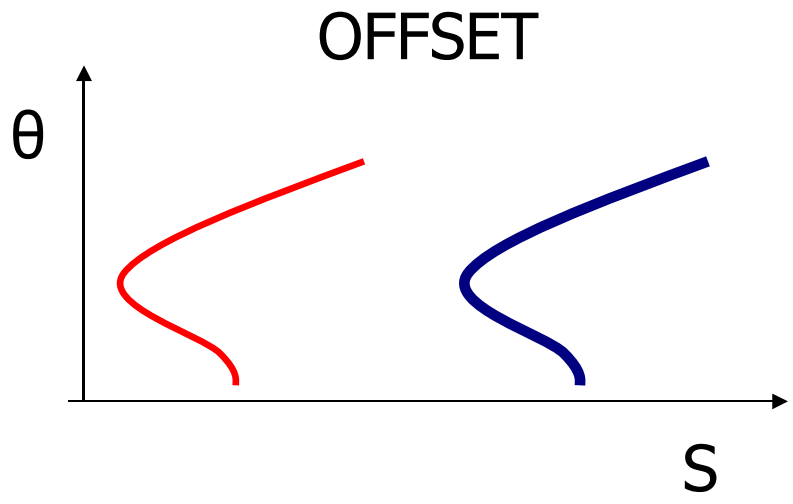
Why do we need delayed-mode processing?

Salinity is one of the main measurements made by Argo profiling floats.

However ...

... over time, the conductivity sensor can experience instrument drift that gives salinity measurements an artificial trend.

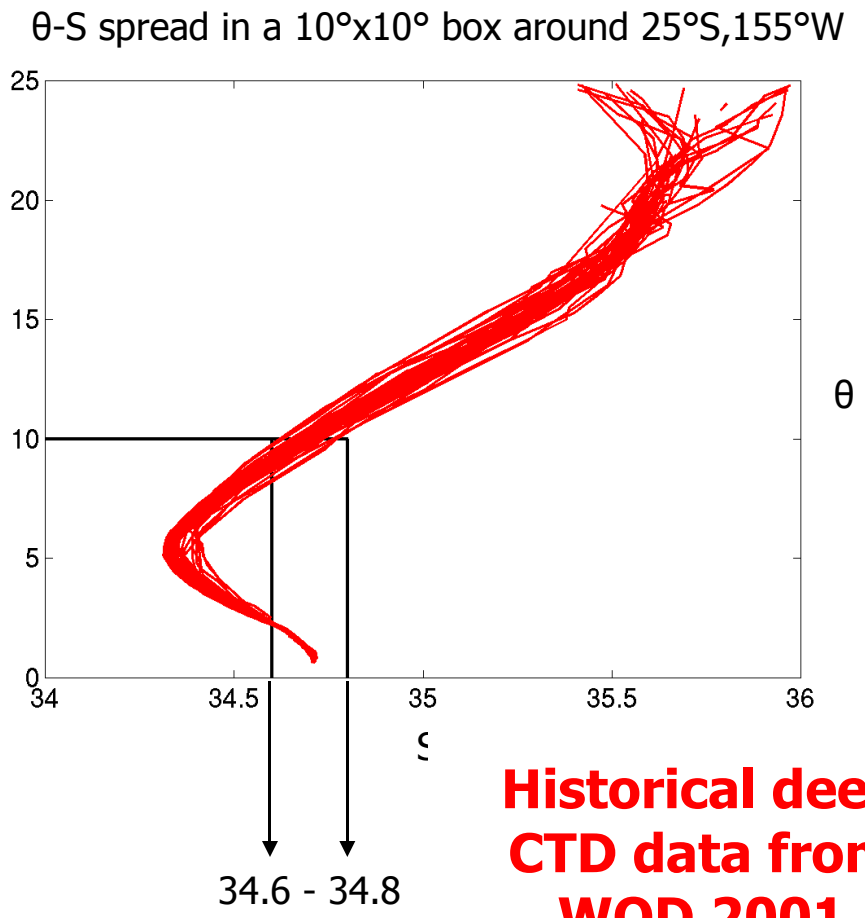




What is done in delayed-mode for salinity data?

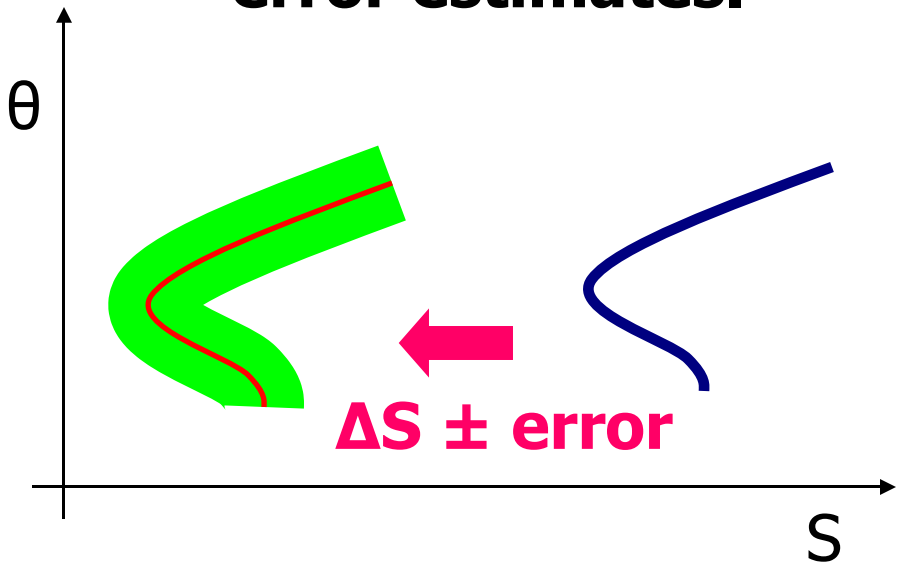
Basic theory …

By using reference deep CTD data and objective analysis, we can estimate what salinity should be at float locations.



Historical deep CTD data from WOD 2001

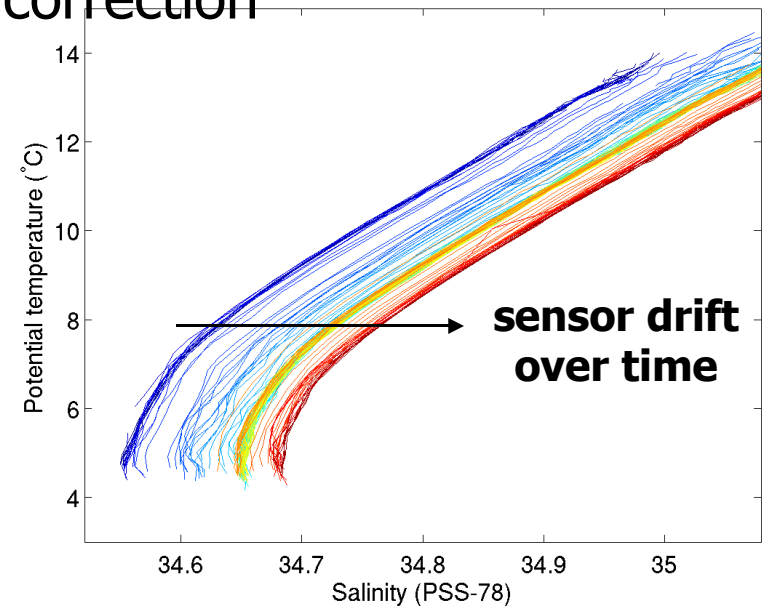
Weighted least squares fit between float-measured salinity and objectively-estimated salinity gives an adjustment factor ΔS, and with error estimates.

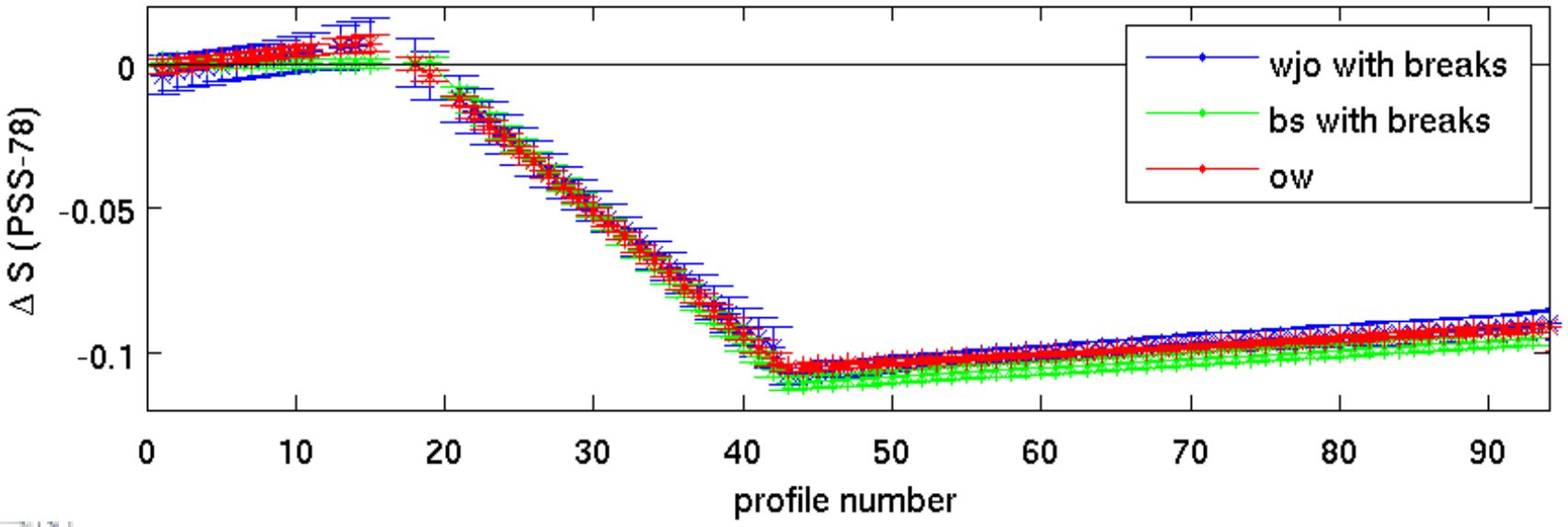


A good example of salinity drift correction

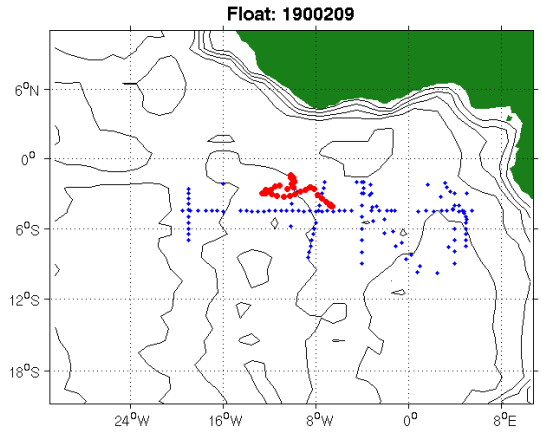
Example: WMO ID 3900161

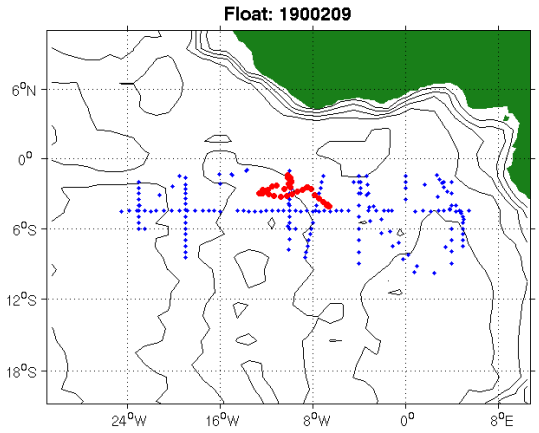
Salinity drifted salty over time (blue drifting to red).

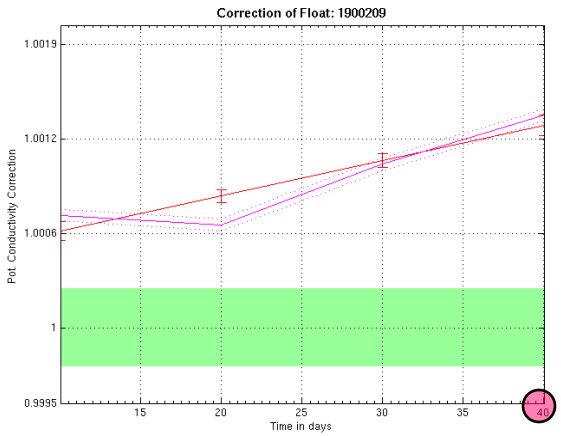




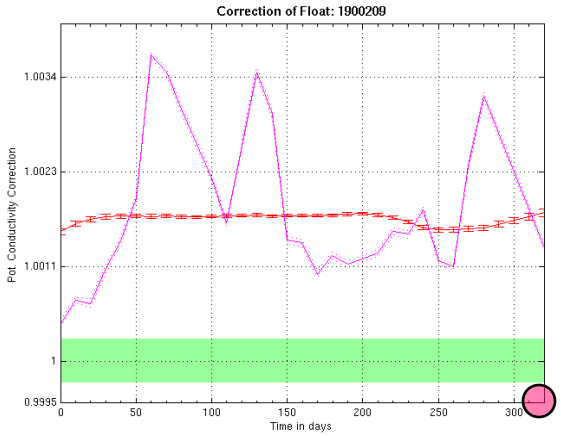
Delayed Mode – Reference database







Cycle 4

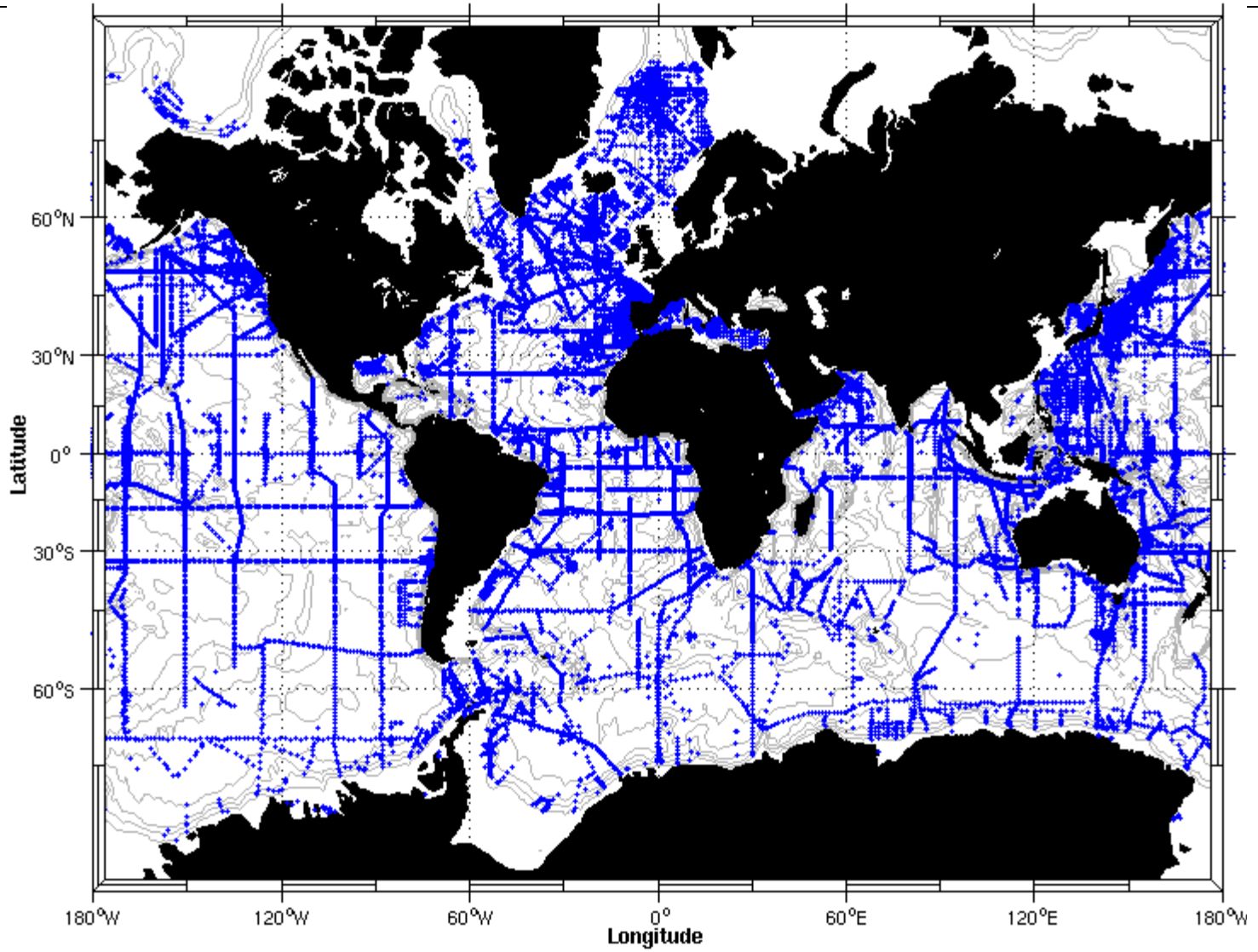


Cycle 32

The delayed mode QC process is based on the available CTD reference database. If we do not have enough reference data, the DMQC method can not work.

It is very important to collect all the CTD data from all the oceanographic cruises.

Reference data base for delayed mode QC : A need for more recent CTD

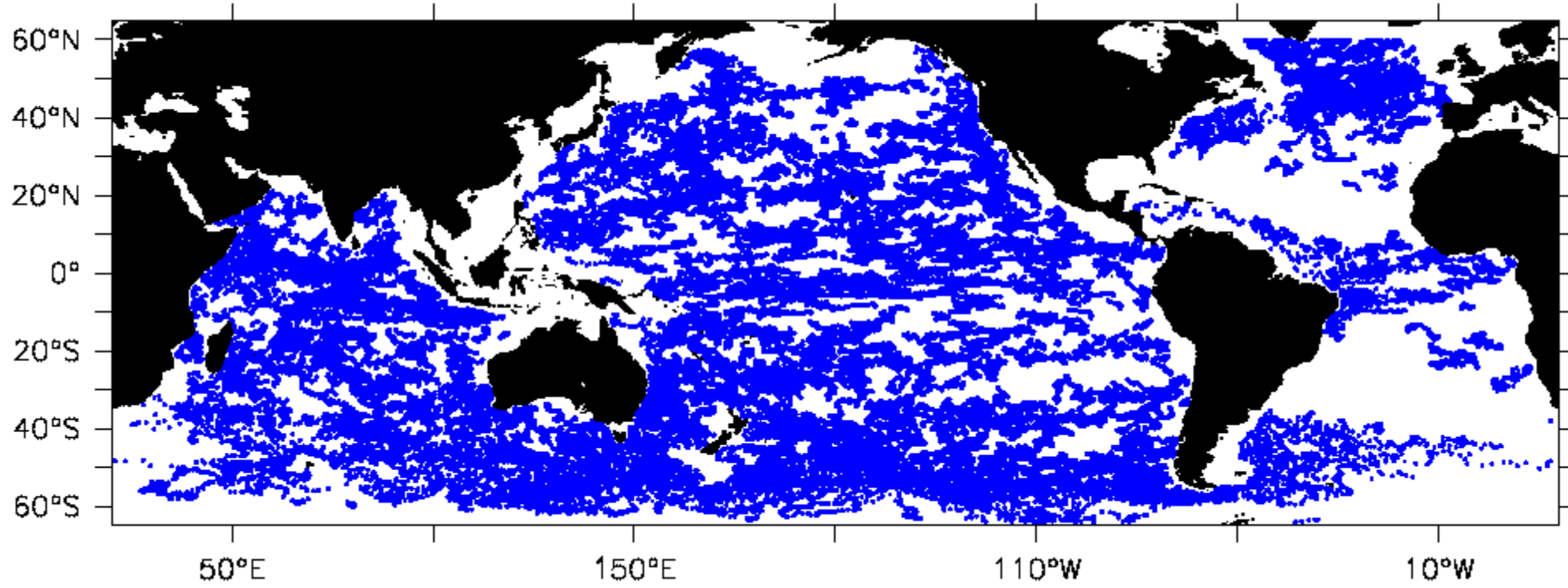


Coordinated by Coriolis (C Coatanoan)

Collaboration with Clivar/CCHDO NODC/USA and PIs

Reference data base for delayed mode QC : Argo good profile data base

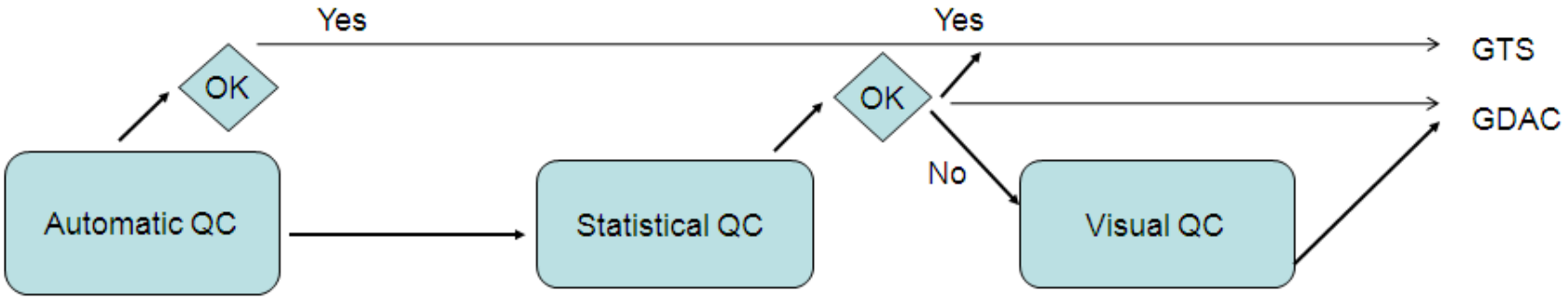
For poorly CTD sampled, high variability or complex area



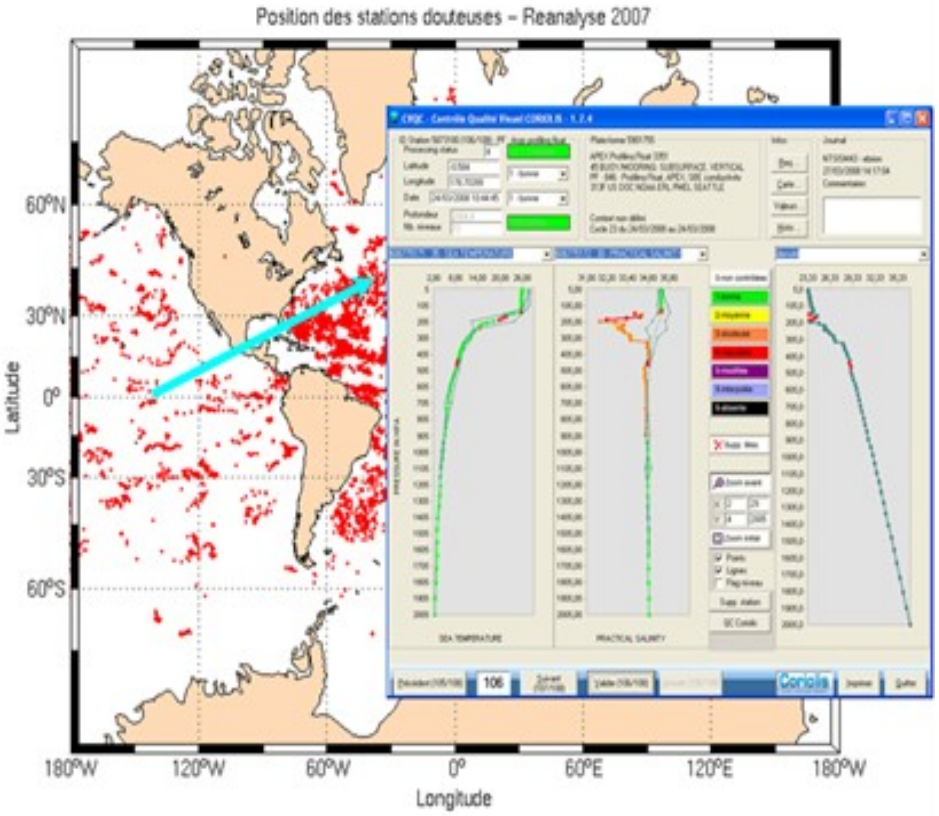
How improve the Argo Dataset between RT delivery and DM Processing !!!!!!!!!!!!!

- Between RT and DM => Improve by global analysis comparing Argo to other observations
  - Comparison with climatology
  - Comparison with altimetry
  - Comparison with Argo neighbours (on progress)

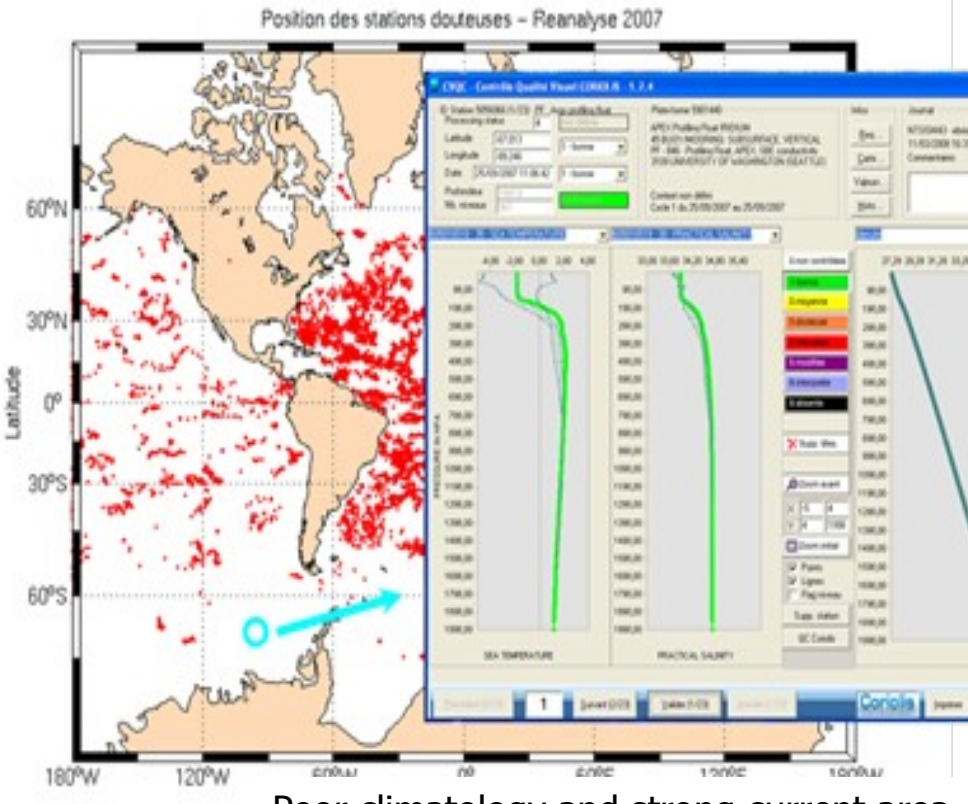
Data Consistency an Argo Regional Centre activity



Emilie Brion Coriolis France



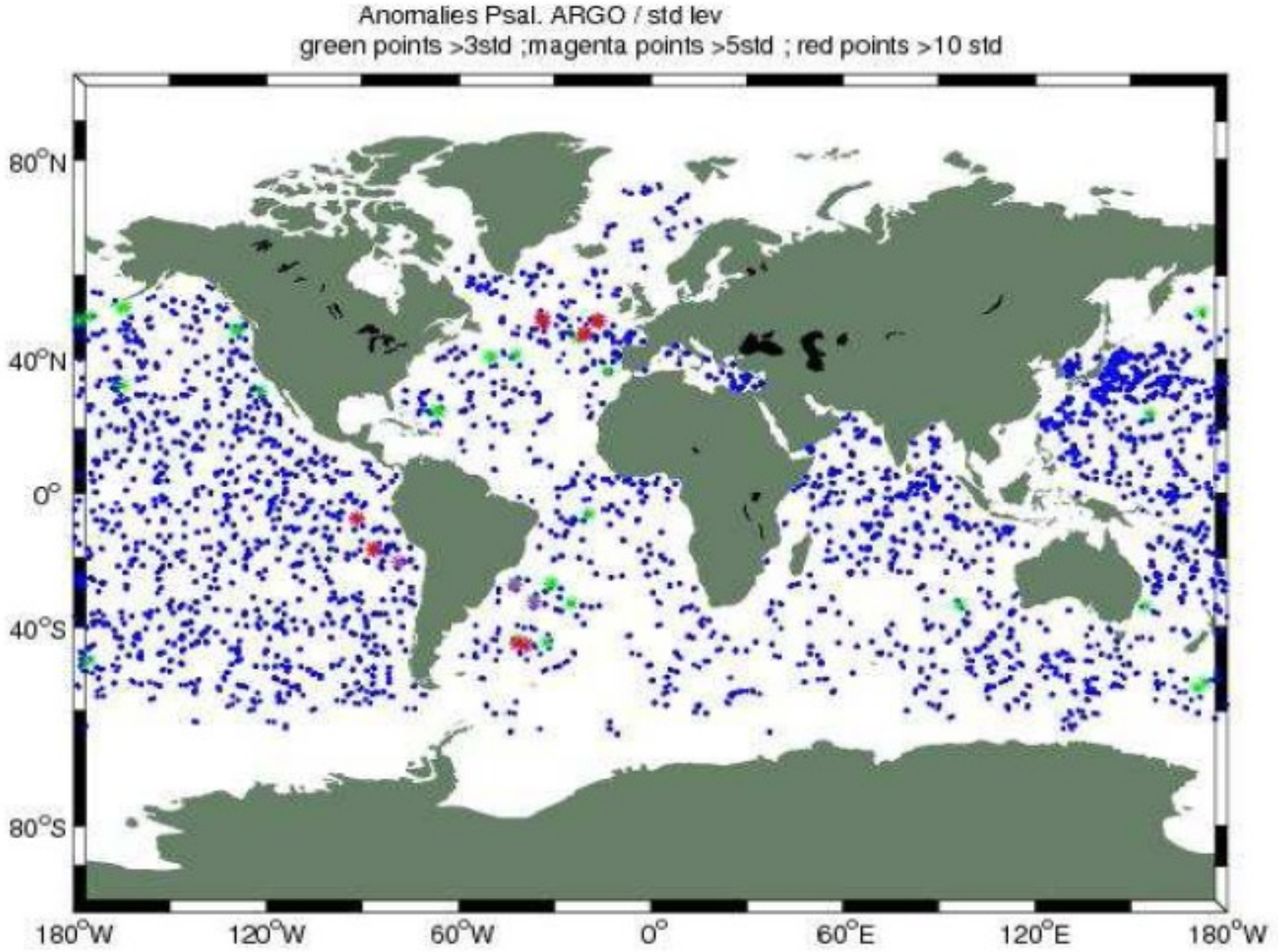
Spike too small to be detected automatically



Poor climatology and strong current area

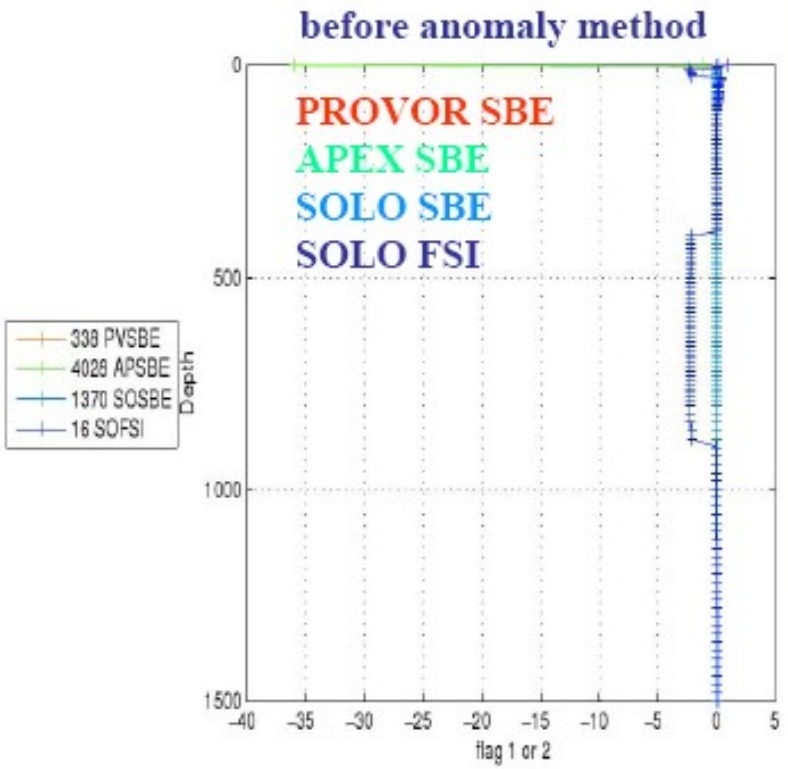
First results : Under test mode at Coriolis since beginning of 2009

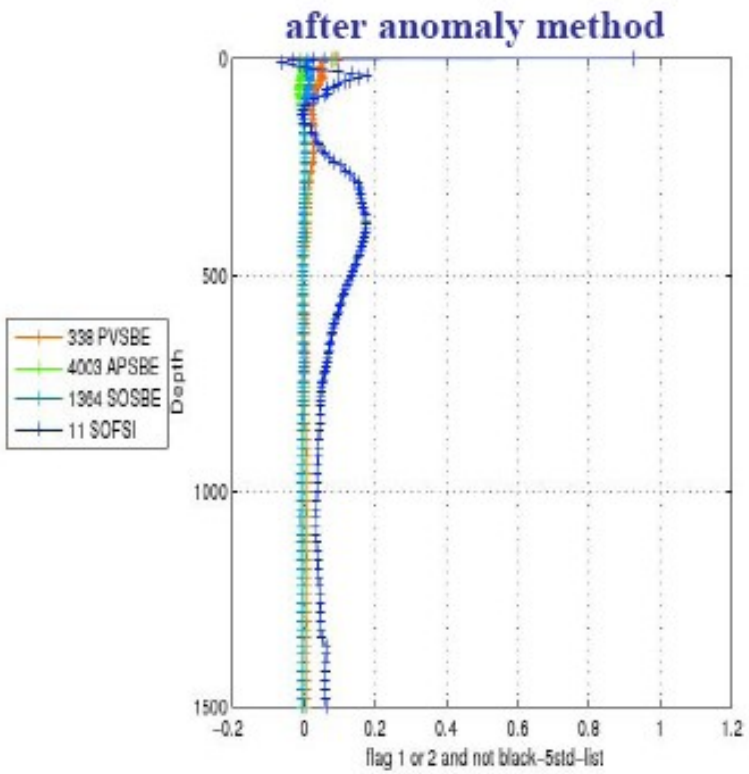
- Comparison with a Climatology computed from Argo (K. von Schuckmann & al)



Global comparison between different instrument types

Global average of salinity bias





Carole Le Saout /Glazeo Cecile Cabanes /Coriolis

Comparaison to Altimetry

- As, Sea Level Anomalies (**SLA**) from altimeter measurements and Dynamic Height Anomalies (**DHA**) calculated from in-situ T and S profiles are complementary but also **strongly correlated**
  - ➔ **Satellite altimeter measurements are used to check the quality of the Argo profiling floats time series**

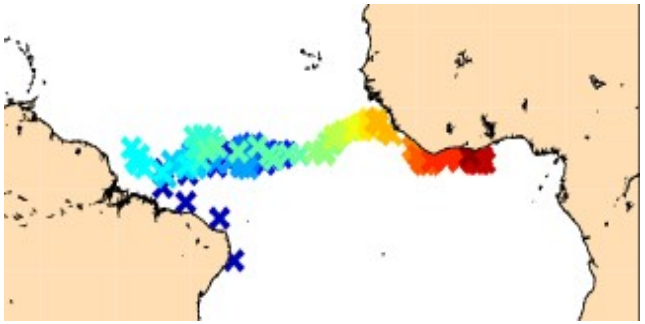
➔ Altimeter measurements represent the mesoscale and the interannual variability – more efficient than the use of climatological fields

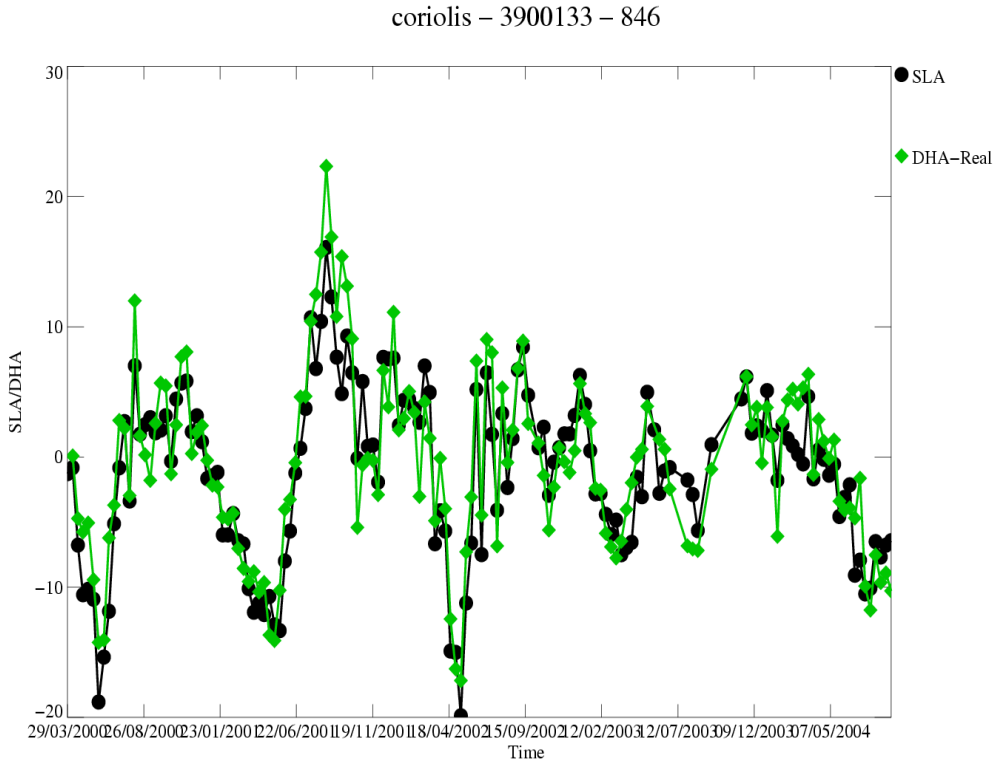
- The main idea is **to compare** co-located **SLA** and **DHA** to detect systematic or punctual errors in the Argo data sets time series

- Analysis are performed for **each** float time series

*S. Guinehut, C. Coatanoan, A.-L. Dhomps, P.-Y. Le Traon and G. Larnicol, 2008: On the use of satellite altimeter data in Argo quality control, accepted in JAOT.

Validate the consistency by comparing to altimetry



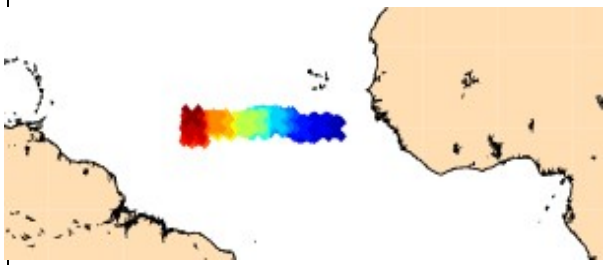


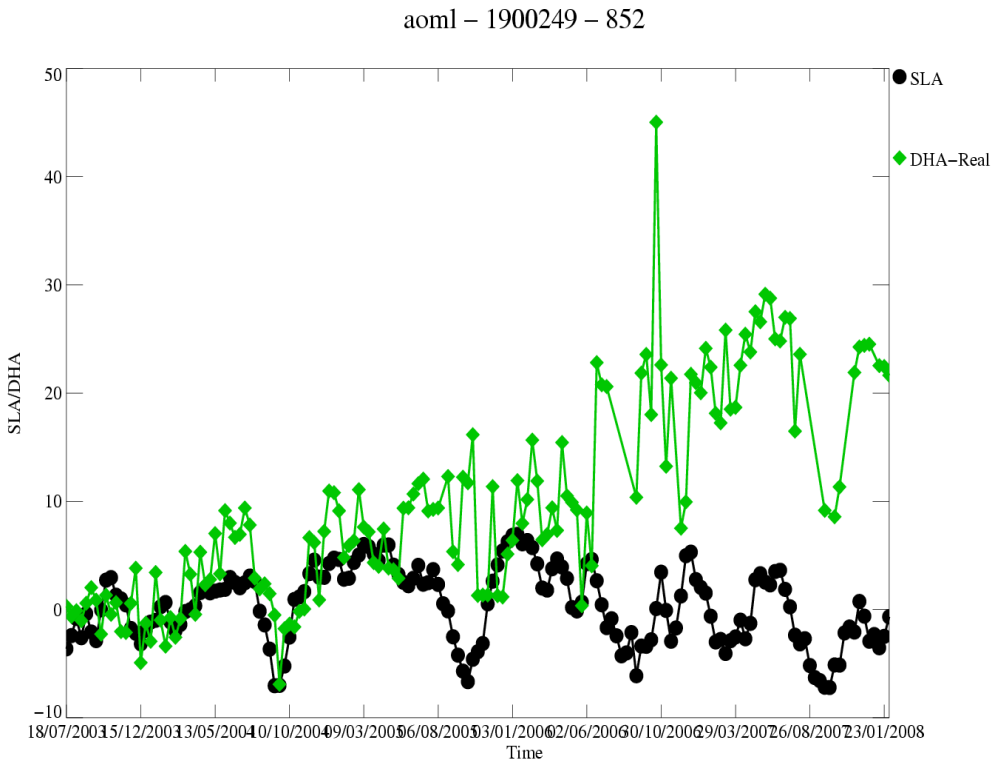
Float : 3900133
r : 0.91
rms-diff : 20.44 %
mean-diff : -0.73 cm
samples : 147

Stephanie Guinehut – CLS - FRANCE

Validate the consistency by comparing to altimetry

- Progressive drift of the salinity/pressure sensors :



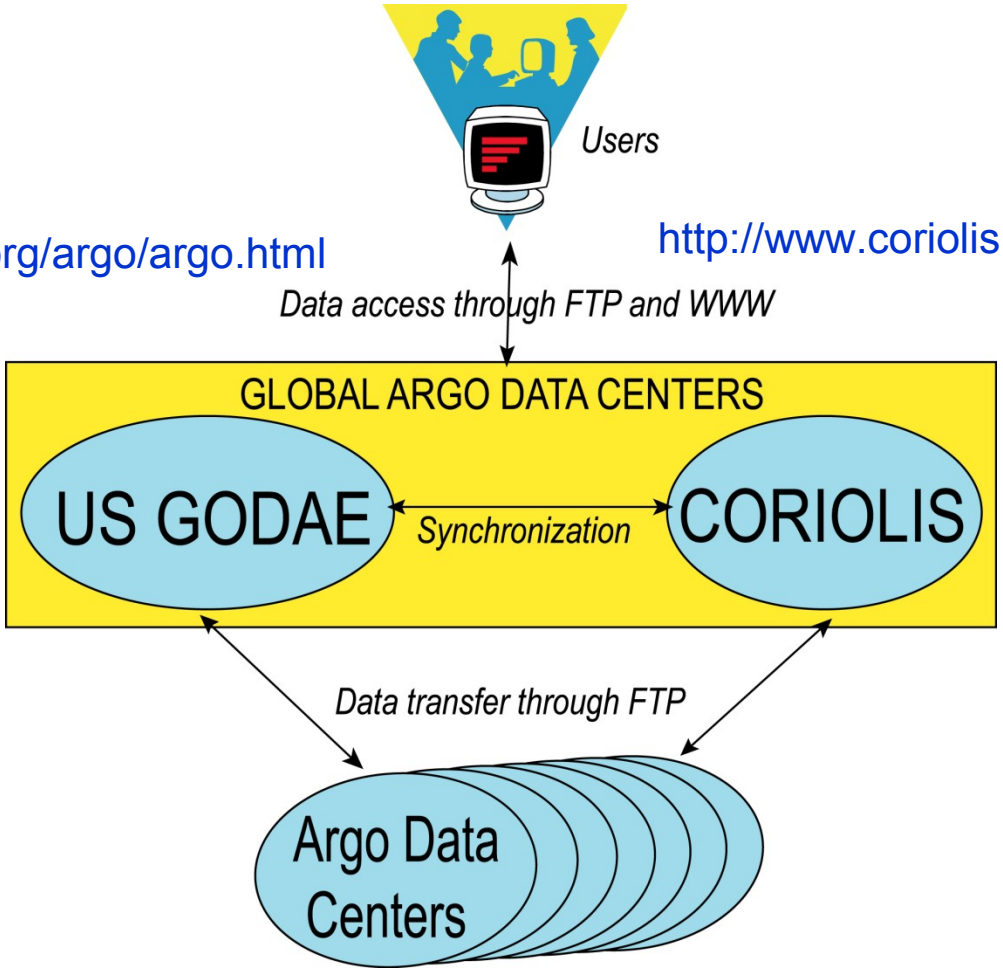


Float : 1900249
r : 0.00
rms-diff : 1538.0 %
mean-diff : -8.98 cm
samples : 152

Stephanie Guinehut – CLS - FRANCE

Data Access Argo: GDAC global FTP servers Monterrey & Brest

- The GDAC FTP sites provide the master copy of Argo data set (meta-data, profiles, trajectories and technical informations).
- The GDAC keeps the highest processing level of floats data.
  It is not in charge of archiving the different versions of float data.



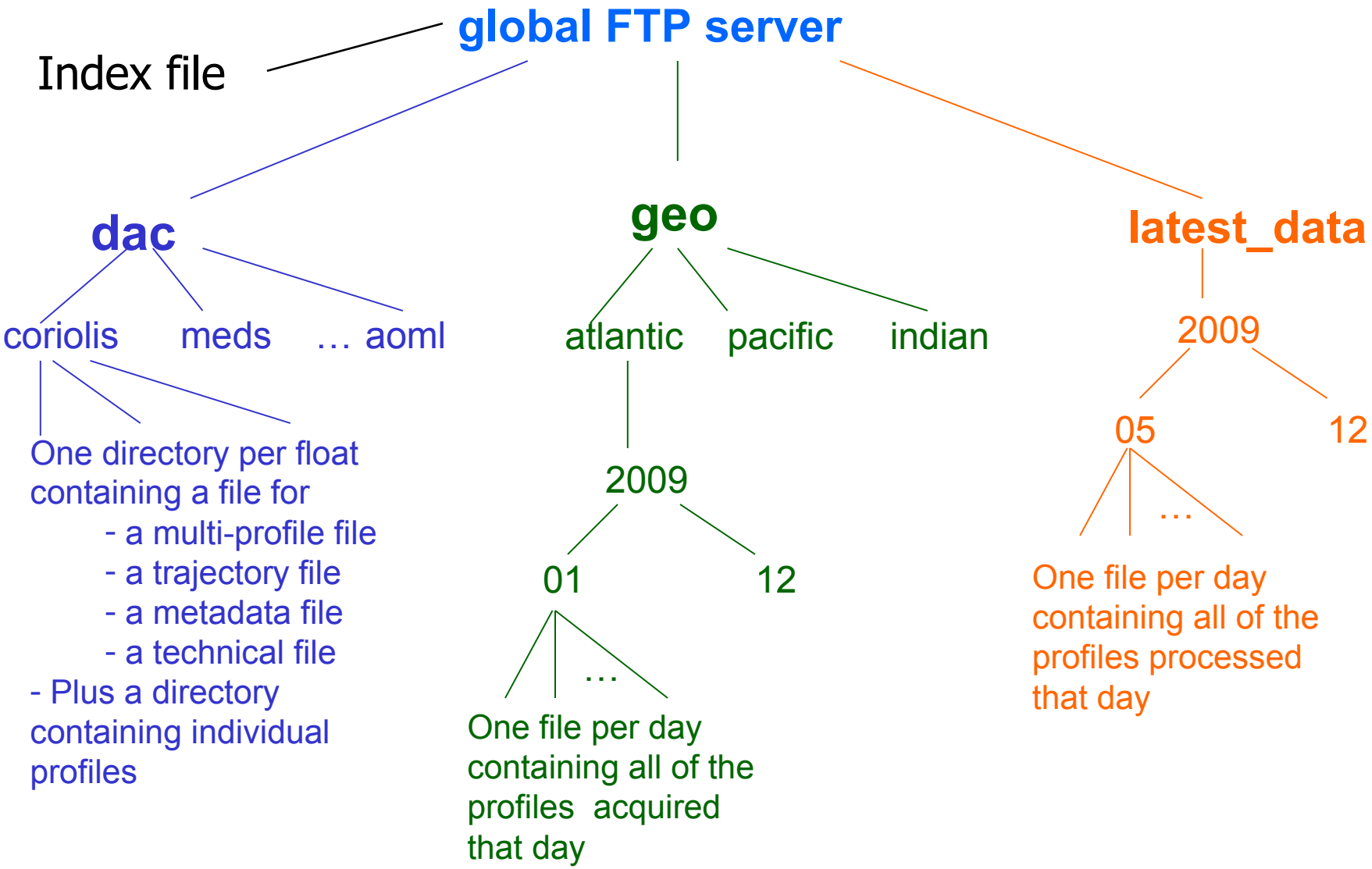
http://www.usgodae.org/argo/argo.html

http://www.coriolis.eu.org/cdc/argo.htm

Argo GDAC :content and data formats

- On GDAC FTP site, for each float, 4 types of information are handled.
  - Metadata file: general informations on the float life.
  - Profile file: one file per profile.
    It contains both original data acquired by the float and the best available profile together with quality flags.
  - Trajectory file: one file containing the complete trajectory of the float as well as the measurements collected while drifting.
  - Technical file: one file containing the technical information provided by the float.
- Naming convention for Profiles files: Starts with an **R** when Real-time, and **A** when adjusted in real-time, with a **D** for processed in delayed mode
- Argo NetCDF format is used for metadata, profile, trajectory and technical files.
  Argo NetCDF data format is described in « Argo users's manual »
  http://www.coriolis.eu.org/cdc/argo/argo-dm-user-manual.pdf

Argo GDAC FTP structure : will be detailled on Thursday



Lessons Learned

- Open data policy is an asset for an international observing program
- It's possible to process data in RT in a coherent way but need important coordination
- The DAC /GDAC architecture works and is reliable
- Delayed Mode processing in a coherent and subjective-less manner is challenging and requires a lot of coordination, training and cooperation
- Argo Data Management system is in progress and work is on going to improve data Quality between Real-time and Delayed Mode processing
- Feedback from users is welcomed and necessary to improve the system (support@argo.net)

Argo data management reference documents

- Reference documents http://www.coriolis.eu.org/cdc/argo_rfc.htm
  - Data management web page
  - Beginner's guide to Argo data
  - Data management policy
  - GDACs organization
  - User's manual; data formats
  - Quality control manual

- Data access
  - The FTP adresses are :
    - ftp://www.usgodae.org/pub/outgoing/argo
    - ftp://ftp.ifremer.fr/ifremer/argo
  - Data discovery : web interface
    - http://www.usgodae.org/cgi-bin/argo_select.pl
    - http://www.coriolis.eu.org/cdc/dataSelection/cdcDataSelections.asp
  - Data access : OpenDAP
    - http://www.ifremer.fr/cgi-bin/nph-dods/data/in-situ/argo

- User Desk : support@argo.net
- GDAC Coriolis : codac@ifremer.fr