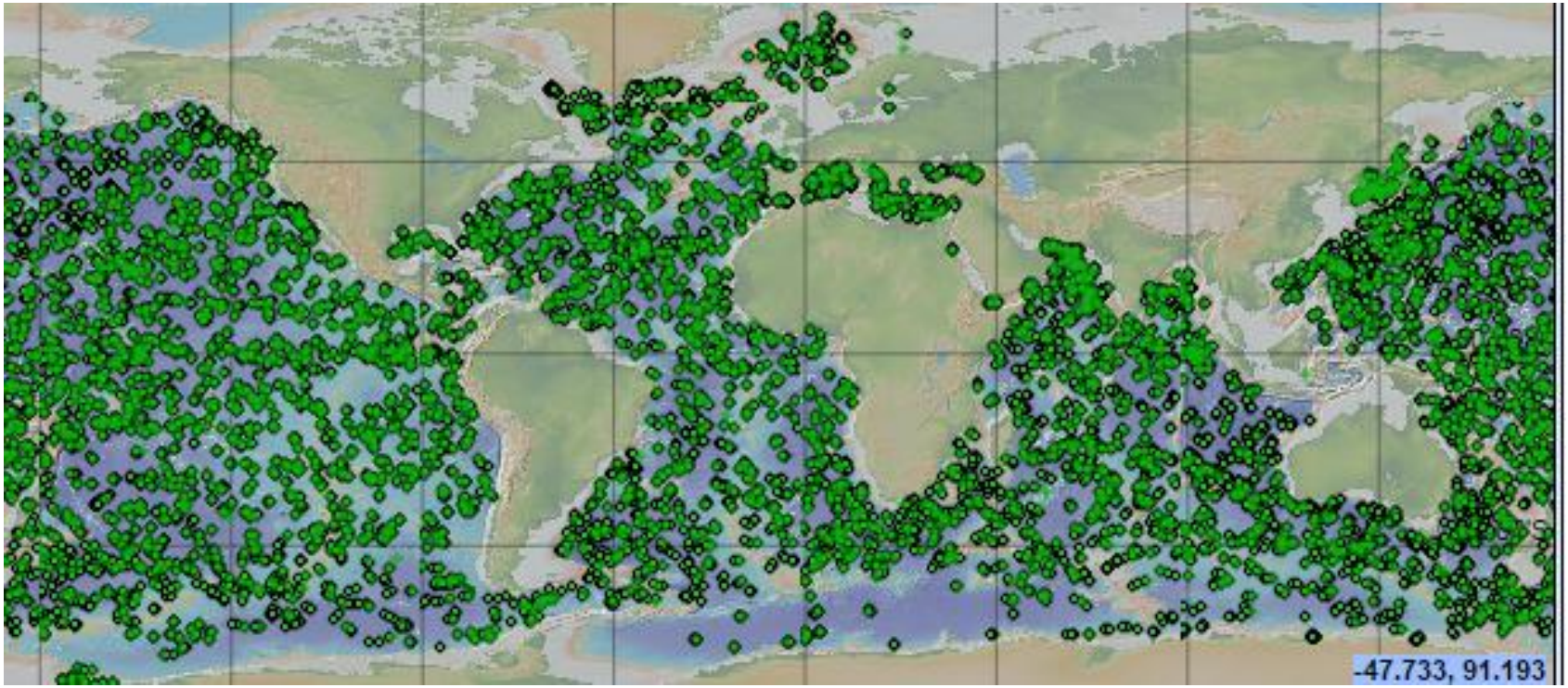
Feedback from ADMT15

S Pouliquen / Ann Gronell-Thresher
AST16

18-20 March 2015



Status of the Argo Data System

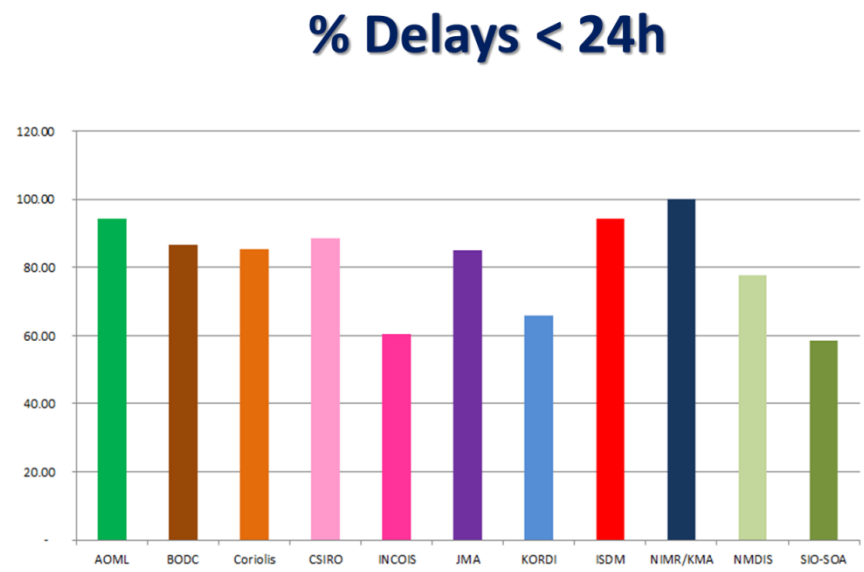
- Current ADMT efforts focus on
  - Monitoring the Real Time system and the delays
  - Updating Data Format to V3.1 to be able to handle:
    - Change in mission for Iridium and Argos3 floats
    - Separate surface and sub-surface profiles
    - Enhancing trajectory data
    - Have the capability to include new variables especially Bio-Argo new parameters
  - Improving data consistency and completeness
    - Density inversion tests
    - Monitoring data quality through OA and altimetry analyses
    - Mandatory metadata, standardization of Format-ID
    - Implementation of new format checker at GDAC
  - Monitoring DMQC processing
  - Feedback on ARC activity
  - Towards DMQC for Trajectory
  - Ramping-up of the Bio-Argo data management activity

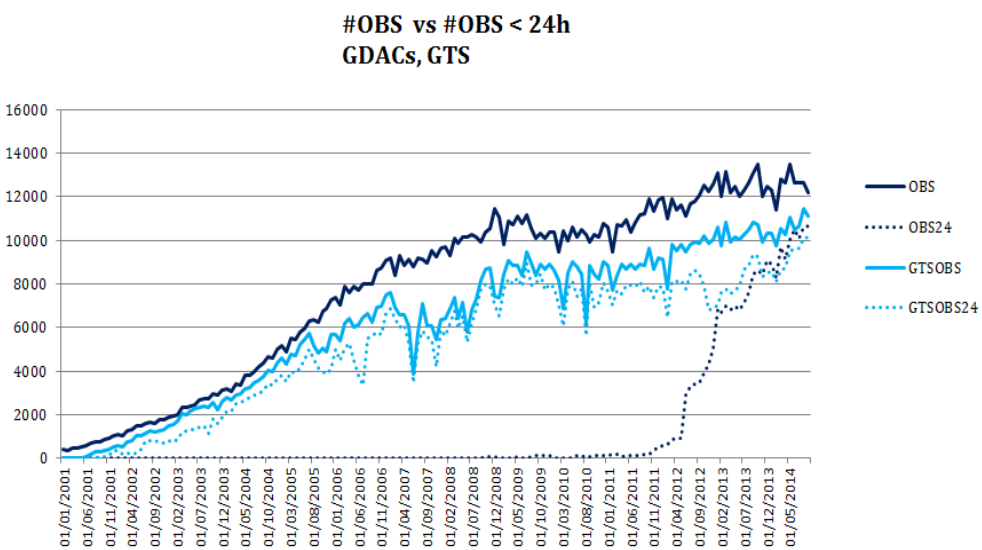
    ⇒ Bio-Argo workshop

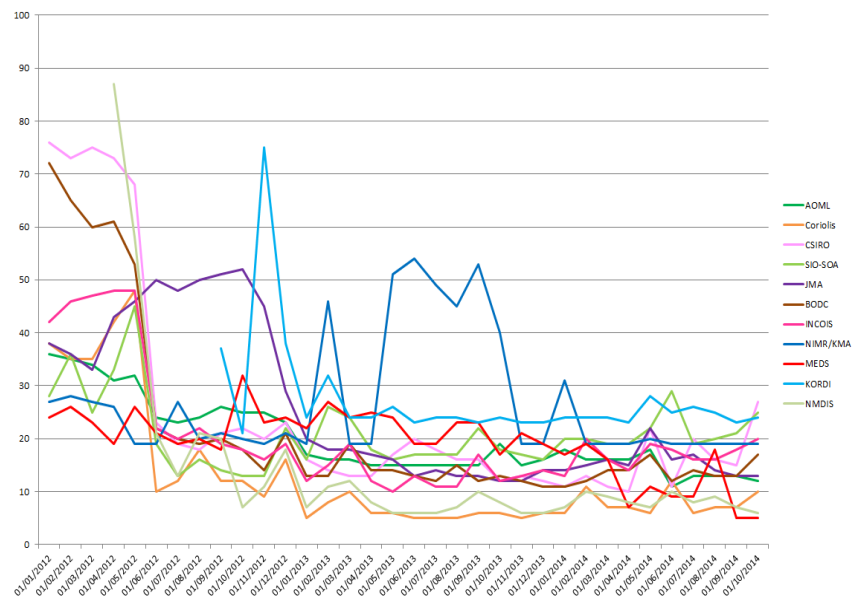
Real Time Data Stream - GDAC

g Team meeting 2015

- Delay monitoring has been continued by AIC using French GDAC (still to be implemented with US-GADC)
- Large improvement with past three years and main delays due to GDAC have been corrected
- Most data are delivered within 24h and more data in volume than on GTS

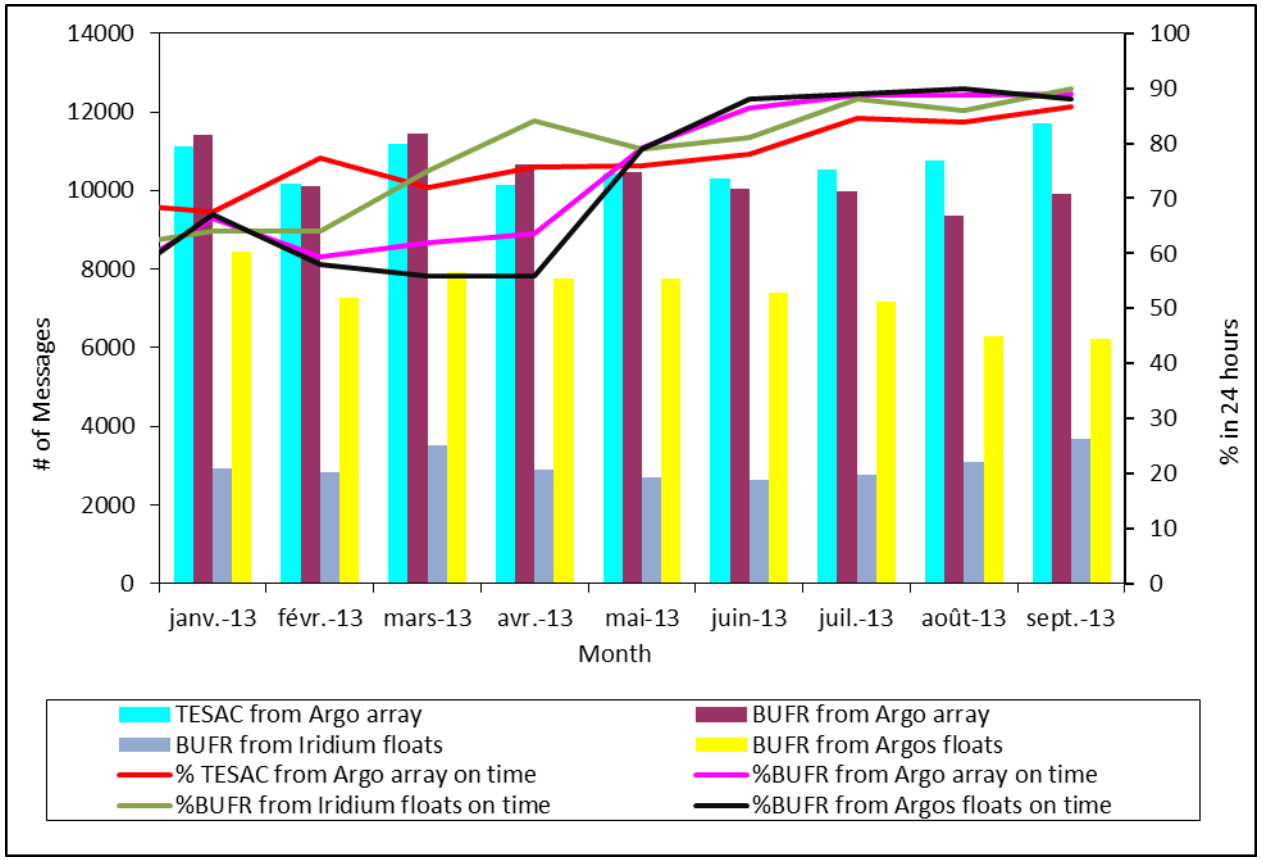






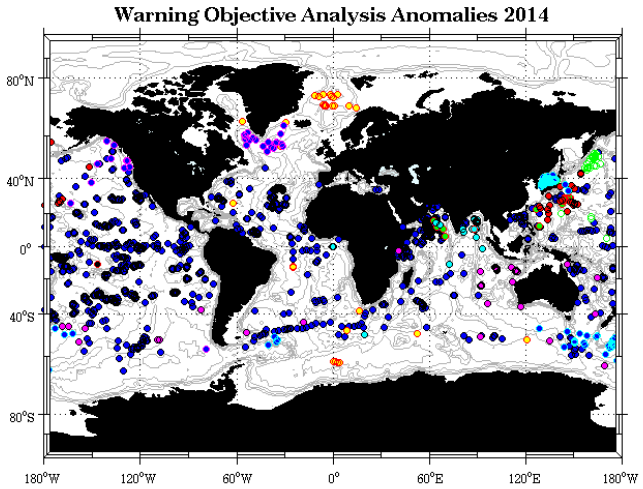
Real Time Data Stream - GTS

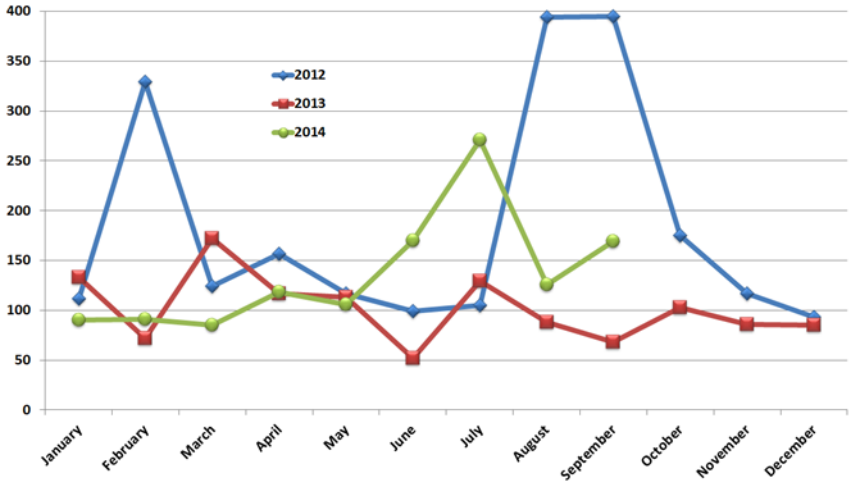
- On average 85 % TESAC data on GTS , and 87% BUFR data within 24 hours . About 75% at the end of 2013
- For most DACS there are similar BUFR than TESAC messages on GTS. Need to check that all measurements are transmitted (comparison between GTS and GDAC).
- Need to update the BUFR encoder from V3.1 format



Data monitoring

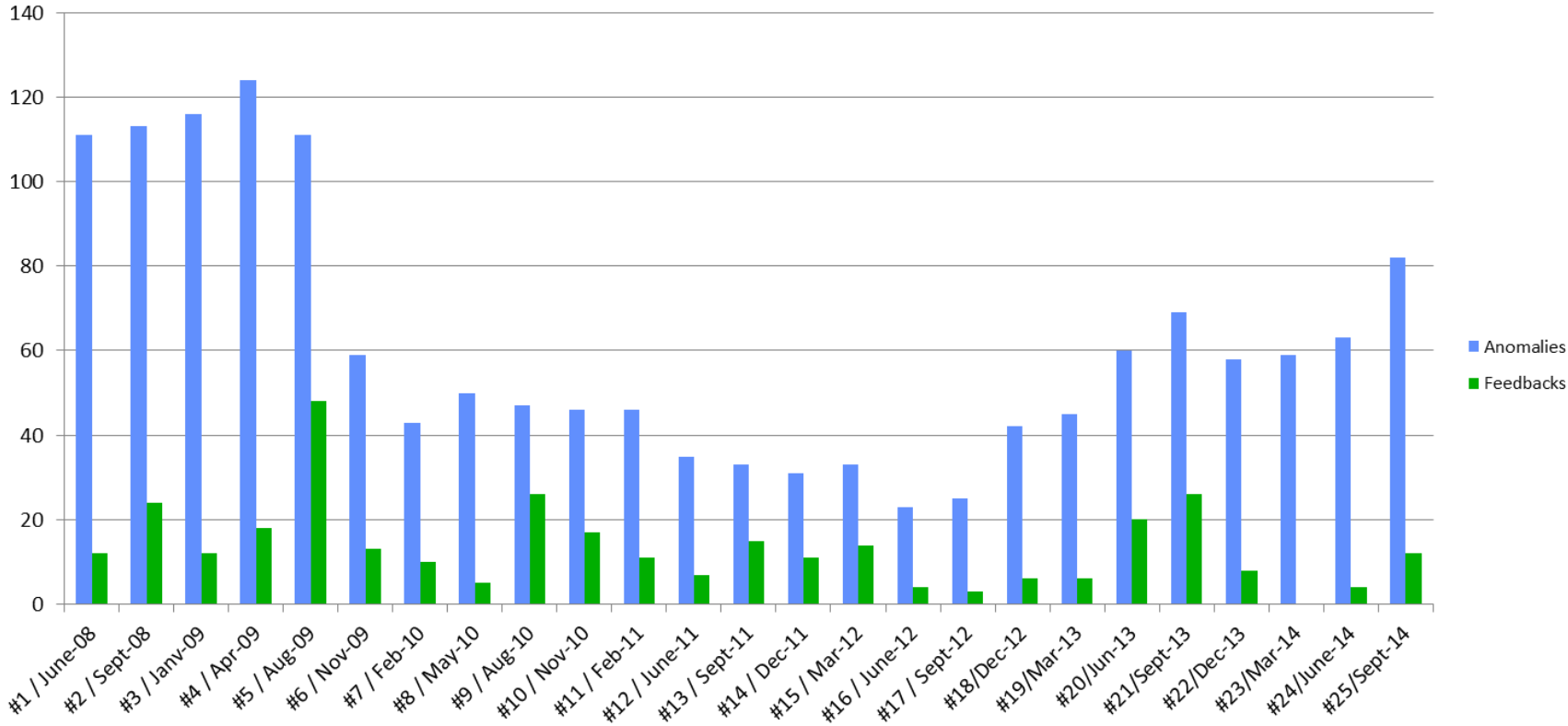
- Continuing manual QC checks
  - Daily: Coriolis performing an OA to detect anomalies and send "standardized" feedback to DACs. Dacs then provide feedback to Coriolis (is the data truly good or bad?) to improve the system and decrease false alerts
  - Monthly: Coriolis sends a report to the DACs summarizing all OA anomalies detected
    - Results are good with a mean of 125 failures per month
    - DACs are correcting profiles and providing feedback
  - Some DACS not populating adjusted fields correctly
    - Only P Adjusted is filled : under correction in concerned DACs
    - One Dac processed floats in delayed mode but kept the file name with R instead of D
    - No adjusted fields when float requires correction for SPO for one DAC





Data monitoring

- Quarterly: Comparison to altimetry : A summary is produced and sent to all DACs. Individual messages are sent by AIC when the same anomaly is present in two consecutive runs without any action from DACs/Pis. Most of the DM operator are now responding to these reports and fixing their anomalies
  - 82 floats: : : 34 'R', 44 'A', 4 'D' (43 new floats in 2014 at ADMT15)
- More floats recently extracted, thanks to the new test developed 2012r :
  - Float extracted if SLA/DHA compared to SLA/DHA_Adj is 10 % greater
  - Bad adjustment in RT, bad adjustment in DT
- Some old floats have been corrected but some still need urgent correction
- ~40 % of the floats extracted shows only one isolated very bad profile …



General quality of Argo dataset

- **All profiles, QC='1' = 840 594 profiles** (as of September 2013)

Floats	Correlation	Mean difference(cm)	Rms difference (cm)	Rms difference (%)	Nb profiles	% Total profiles
ALL	0.86	0.02	4.73	24.79	840 594	100
DATA_MODE='R'	0.83	0.13	5.45	27.69	110 666	13.2
DATA_MODE='A'	0.83	0.11	5.59	28.93	158 321	18.8
DATA_MODE='D'	0.88	-0.02	4.30	22.55	571 607	68

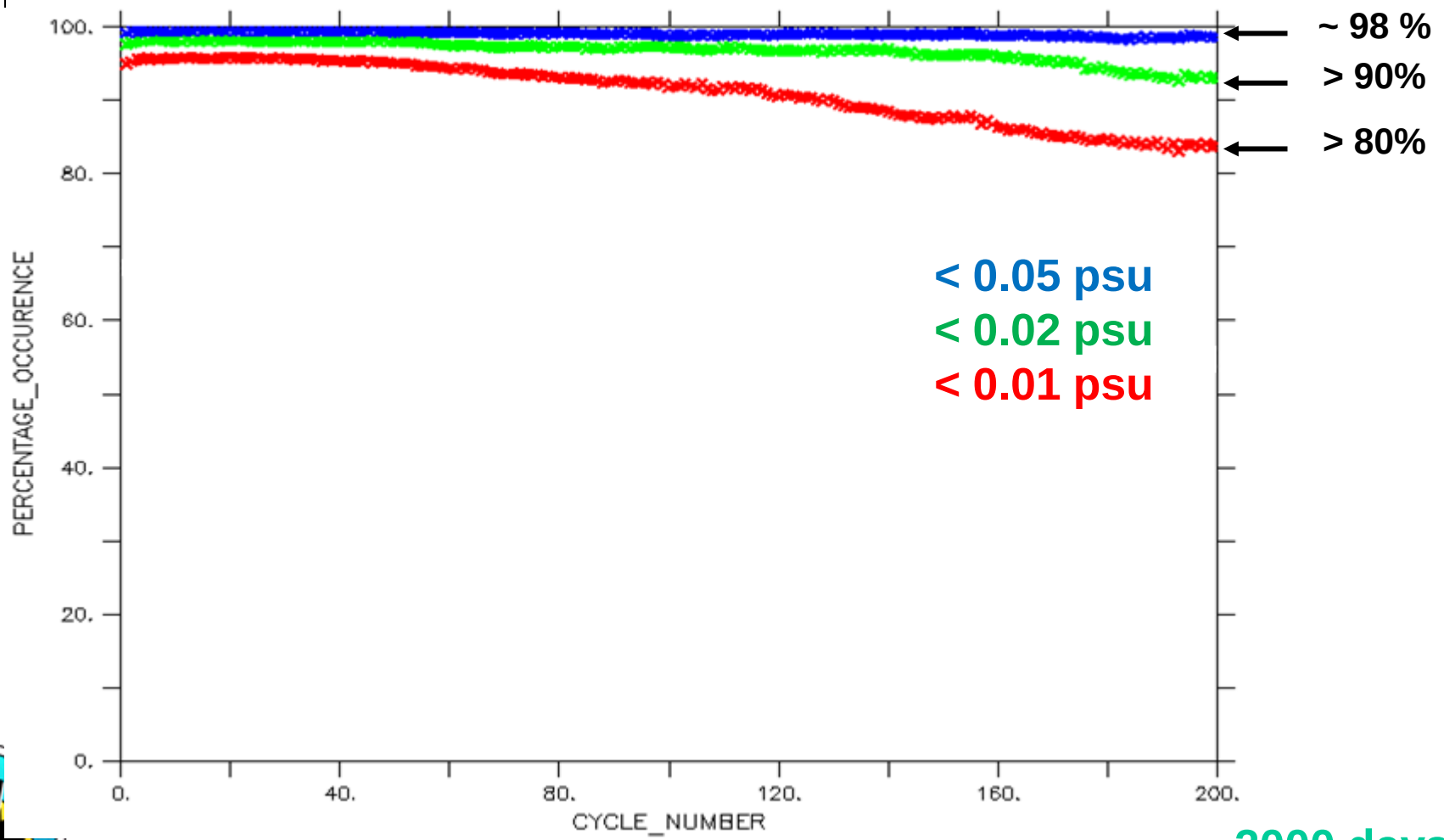
- **All profiles, QC='1' = 971 242 profiles (**as of September 2014**)**

Floats	Correlation	Mean difference(cm)	Rms difference (cm)	Rms difference (%)	Nb profiles	% Total profiles
ALL	0.85	0.26	4.91	25.33	971 242	100
DATA_MODE='R'	0.80	0.64	5.82	29.70	123 529	12.7
DATA_MODE='A'	0.83	0.62	5.57	29.30	231 131	23.8
DATA_MODE='D'	0.88	0.05	4.42	22.38	616 582	63.5

➔ **Statistics are stable**

General quality of Argo dataset

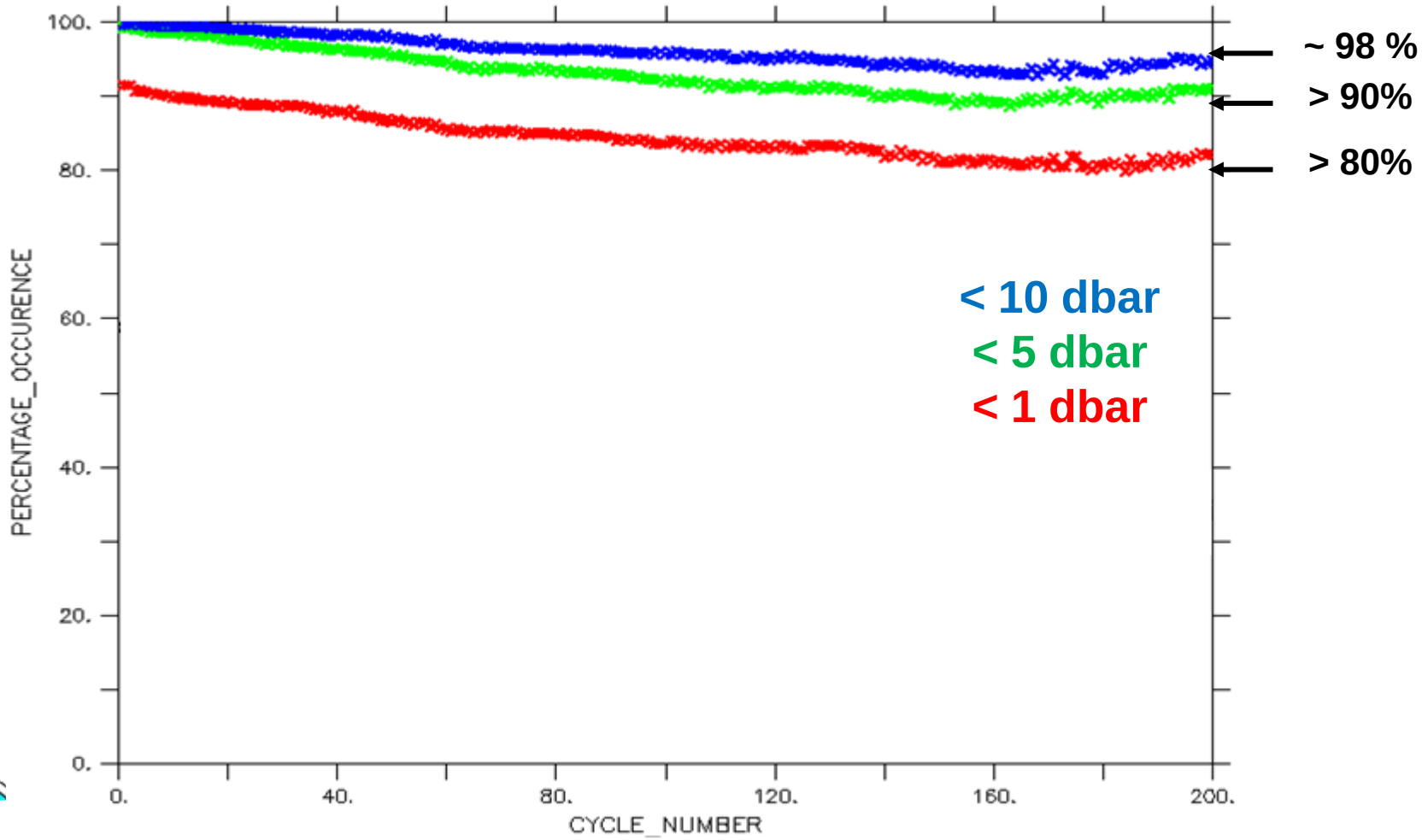
- **PSAL adjustments (613 165 profiles – 5404 floats – QC='1')**



~ 2000 days

General quality of Argo dataset

- **PRES adjustments**



~ 2000 days

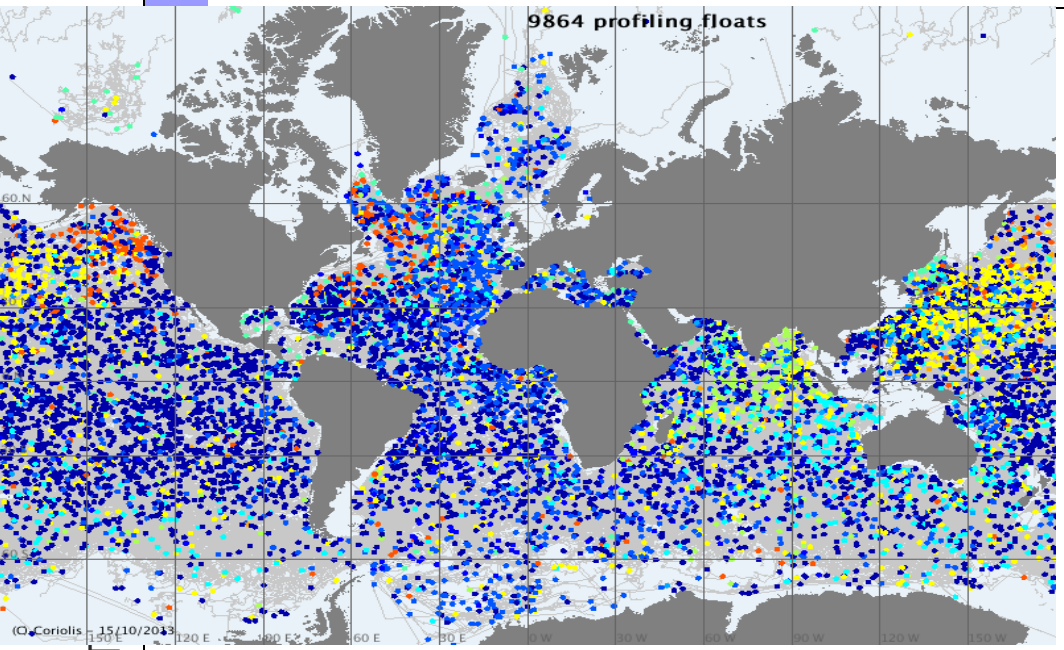
Real Time Data Stream

- Improvement in procedures
  - RTQC for umpumped SST defined and implemented at BODC, Coriolis, MEDS underway at AOML,INCOIS and JMA. Will be transmitted to GDAC in V3.1 format
  - Status on density test implementation:
    - the audit appeared to be catching data errors that had slipped through DMQC but more DM operators need to check their results.
    - RT errors have dropped but 30 files still contain density inversions larger than allowed.

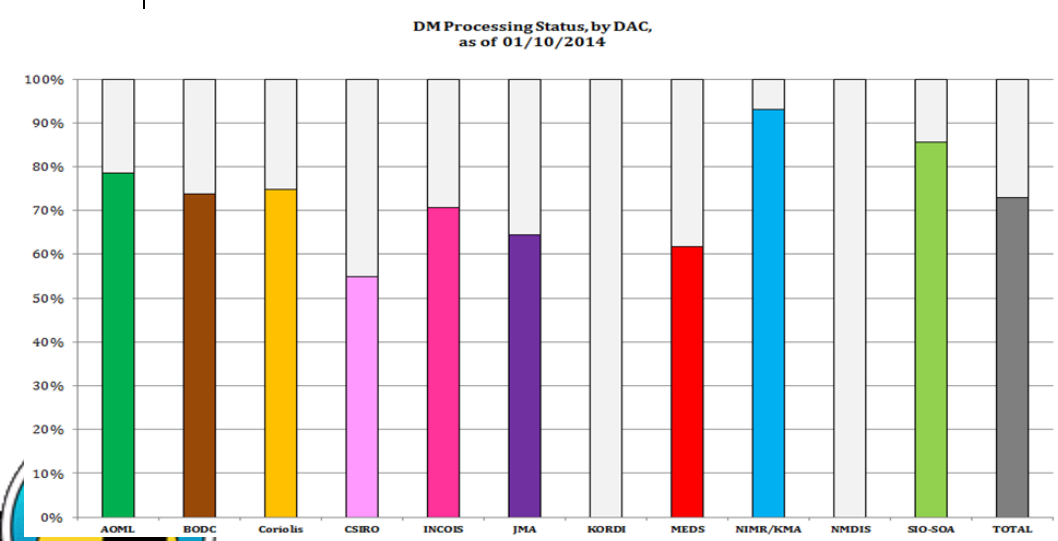
General Data Management

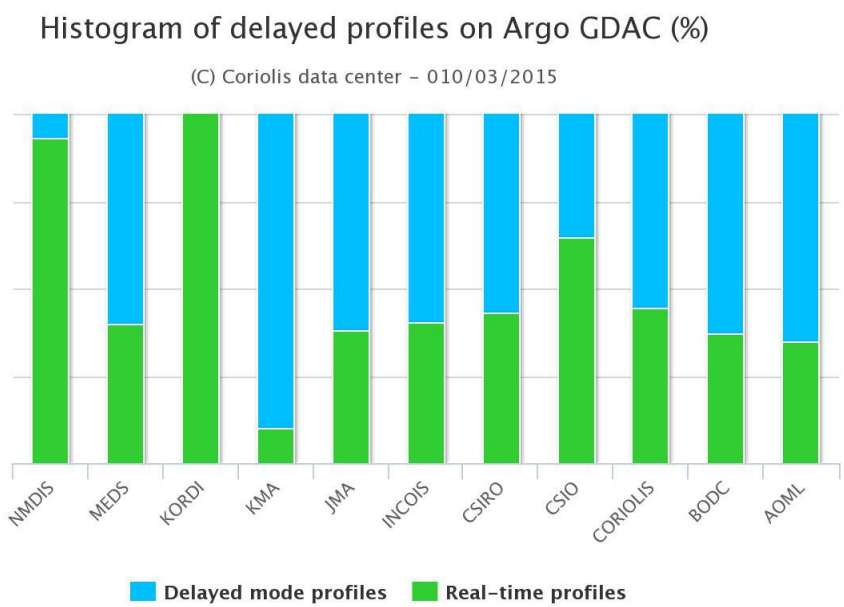
- Metadata files are still in flux
  - *new standard reference tables (SENSOR_MODEL and SENSOR_MAKER)*
  - *the updated core Argo configuration parameter table (Esmee van Wijk) and the new Bio Argo configuration parameter table (Catherine Schmechtig)*
  - *A new table for Standard-Decoders (Mathieu Belbeoch)*
  - *The maintenance of this table need to be defined (ADMT16)*

Delayed mode QC



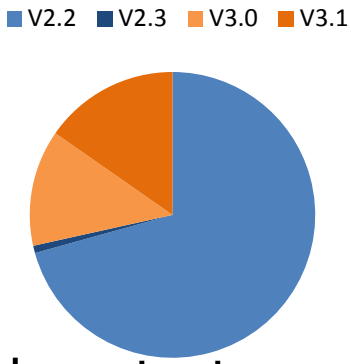
- Backlog as at ADMT – 27%
- about 1,6%(14000) D-files have anomalies (slow correction)
- Reducing the backlog has stalled for past 3 years
- Some DACS have no DFiles

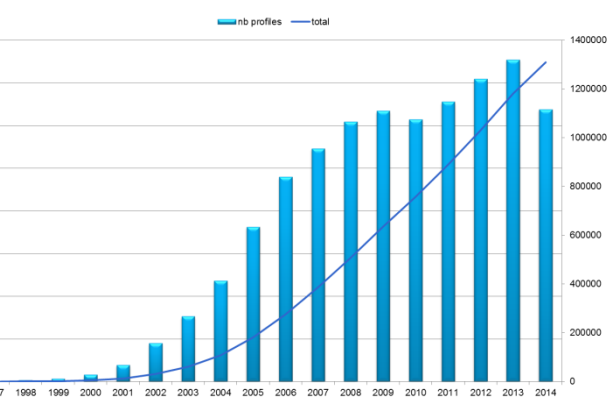


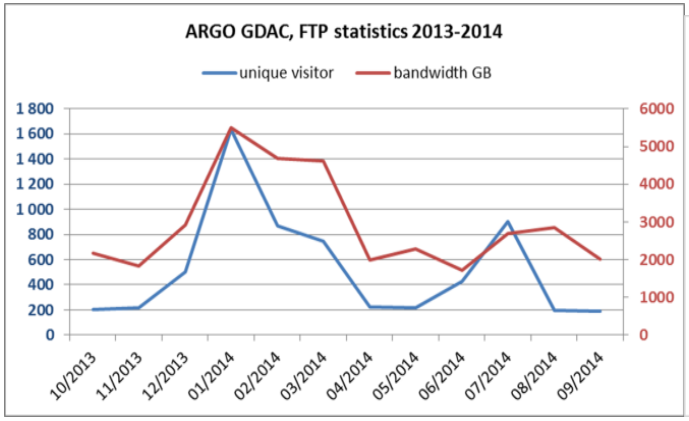


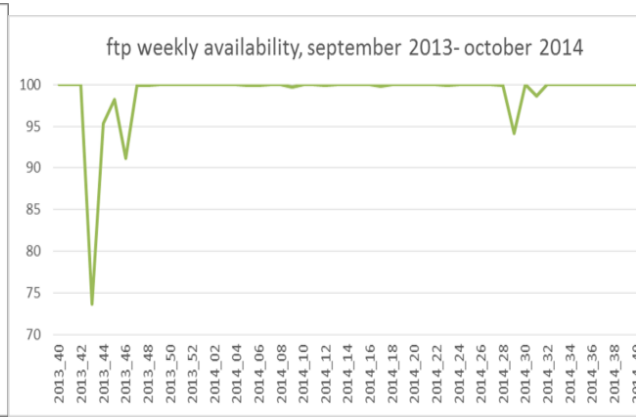
GDAC status –

- Monthly snapshot of the DAC directory with manuals are performed at Coriolis and a DOI is assigned (cf Justin Talk)
- At both GDACS
  - V3 format are accepted
  - Detailed index files are updated hourly
  - Compressed file submission
- The content format checker has been finalized to check contents including conformity of tech-file variable names and units with the agreed list

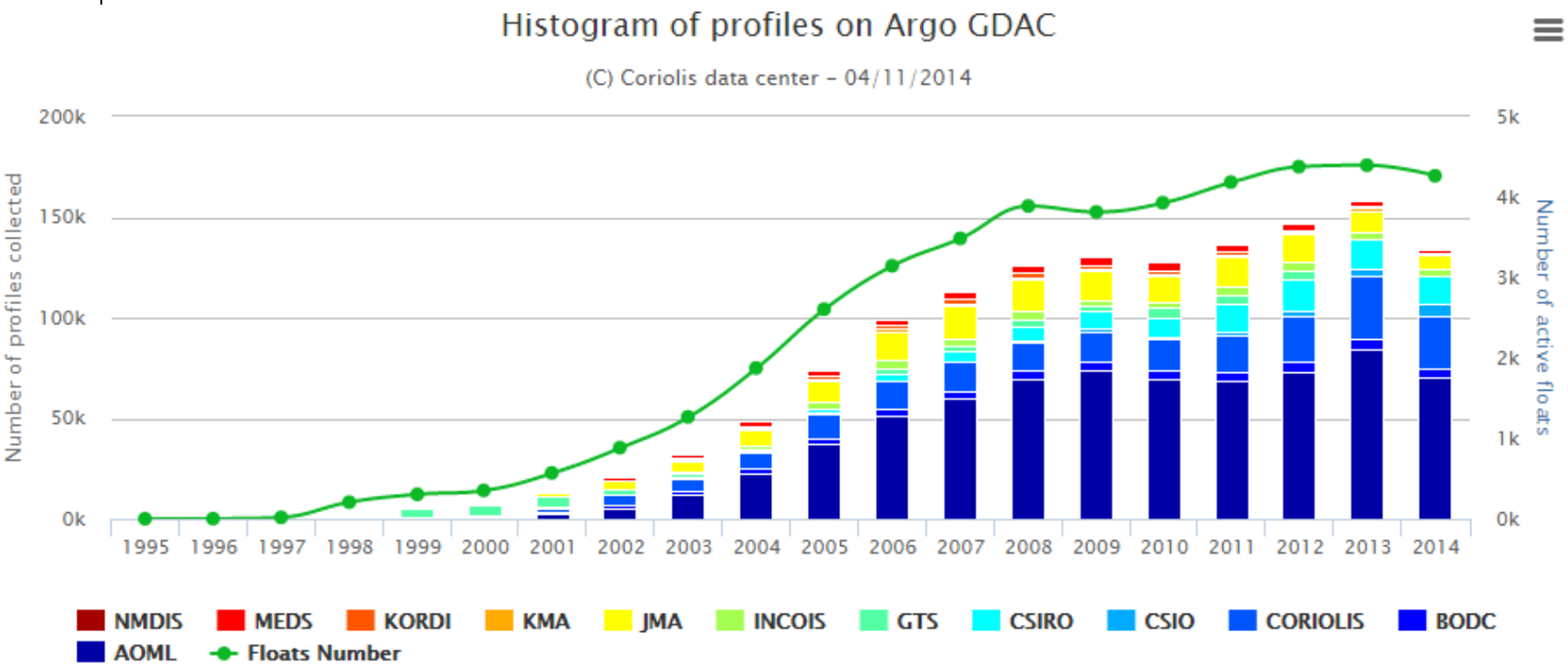








GDAC operations status



16th Argo

- Status available on
- http://www.argodatamgt.org/Monitoring-at-GDAC

GDAC status

- Currently, 11 DACs submit regularly data to GDACs.
- On November 27th, 1 303 072 files were available from the GDACs

DAC	metadata files 2014	metadata files 2013	increase from last year	profile files 2014	profile files 2013	increase from last year2	delayed mode profile files 2014	delayed mode profile files 2013	increase from last year3	trajectory files 2014	trajectory files 2013	increase from last year4
AOML	5 191	4 750	9%	701 226	611 161	15%	485 436	445 834	9%	5 817	4 617	26%
BODC	472	435	9%	47 329	42 136	12%	31 221	31 221	0%	420	415	1%
Coriolis	1 884	1 693	11%	168 971	145 718	16%	111 454	104 902	6%	1 795	1 579	14%
CSIO	276	140	97%	18 325	11 623	58%	10 141	9 201	10%	201	137	47%
CSIRO	621	596	4%	96 450	79 427	21%	44 076	37 324	18%	566	566	0%
INCOIS	339	302	12%	41 529	37 007	12%	26 410	26 409	0%	335	299	12%
JMA	1 339	1 229	9%	150 463	138 226	9%	91 672	85 536	7%	1 325	1 215	9%
KMA	184	168	10%	20 925	18 358	14%	17 180	13 970	23%	176	160	10%
KORDI	119	119	0%	15 459	14 849	4%	0	0	#DIV/0!	113	119	-5%
MEDS	379	368	3%	40 432	37 911	7%	23 481	23 449	0%	371	362	2%
NMDIS	19	19	0%	1 963	1 622	21%	0	0		19	19	0%
Total	**10 823**	**9 819**	**10%**	**1 303 072**	**1 138 038**	**8%**	**841 071**	**777 846**	**8%**	**11 138**	**9 488**	**17%**

New service : Mirroring GDAC data rsync service

- In July 2014 we opened a dedicated rsync server
  - http://www.argodatamgt.org/Access-to-data/Argo-GDAC-synchronization-service
- It provides a synchronization service between the "dac" directory of the GDAC with a user mirror. From the user side, the rysnc service:
  - Downloads the new files
  - Downloads the updated files
  - Removes the files that have been removed from the GDAC
  - Compresses/uncompresses the files during the transfer
  - Preserves the files creation/update dates
  - Lists all the files that have been transferred (easy to use for a user side post-processing)
- Mirroring GDAC using rsync . Examples
  - Synchronization of a particular float
    rsync -avzh --delete vdmzrs.ifremer.fr::argo/coriolis/69001 /home/mydirectory/...
  - Synchronization of the whole dac directory of Argo GDAC
    rsync -avzh --delete vdmzrs.ifremer.fr::argo/ /home/mydirectory/..

Enhanced file checker

- The main difference with this checker is that it will check for consistency amongst several of the parameters within the files in addition to the format itself.
- The enhanced file checker will operate on every file that is received from the DAC's prior to them being accepted into the GDAC. Files that have errors will not be inserted into the GDAC. Files that have errors will cause an email to be sent to the responsible DAC.
- In October of 2014, a test was performed where the US GDAC and results were distributed to the DAC's.
- Discussion at ADMT and some tuning were requested
- Additional tests performed early 2015
- Will be turned into operation 30 March 2015
  - run in a mode where pre-v3.1 files will only be subjected to format checks; no data-consistency checks will be performed on pre-v3.1 files. This essentially duplicates current format checking.
  - Version 3.1 files will be subjected to both format checks and data consistency checks.

Transition to V3.1 for historical profiles

- All Dacs have to develop the capability to generate V3.1 files according to the format defined in Liverpool (ADMT14), with the separation between core-Argo data and b-Argo data.
- The responsibility for the data and metadata distributed by the GDACs is on the DACs . The GDACs do not change or enhance the data received from DACs. Therefore the upgrade of historical data to V3.1 should be performed/managed by the DACs.
- The DACs have to provide the additional information to produce good V3.1 files
  - The detail of vertical sampling schemes (or decide to leave it empty)
  - The description of the float's mission (a majority of unique mission)
- Coriolis DAC is converting its 2.2 and 3.0 files to 3.1 with a Matlab patch. When validated, this patch will be available to the DACs who need it. (Provor is done. Apex is coming)

Status at ADM14

	RT	Historical
AOML	Reprocessed in 3.0 no plan in 3.1 Equivalent	Reprocessed by DM operator Annie : convert all D files to 3.1 John : done in 3.1 for Argos Plan To be done for Iridium Paul : plan to use the Matlab tool PMEL Plan in next 12 months
BODC	Reprocess in V3.1 from database	Will do it for their floats
Coriolis	Done for RT	Converter for Provor (checked by Ifremer)and APEX (checked by BSH) and OGS (for the med)
CSIO	RT in 2015	????
CSIRO	Done for Meta, Profile (including B-File but were rejected), and Tech files. To be done in Traj	In progress for D-Files by patch
INCOIS	Working on it before end Nov	Will use the patch
JMA, JAMSTEC	Profile and Metafile conversion planned by AST16 for Argos. For Tech, Traj and iridium by ADMT16.	Profile and Metafile conversion planned by AST16 for Argos . For Tech, Traj and iridium by ADMT16
KMA	Testing 3.0, plan 3.1 by AST	A bit later for Historical
KORDI	??	??
MEDS	Meta and Profile files done, and traj within 3 months	At the same time
NMDIS	December (will convert Profiles first, then 2-3 more months for trajectory and meta)	At the same time

Trajectory progress

- Trajectory format now able to hold all information from the ANDRO DEP files
- Conversion from Andro has been performed
  - All ANDRO trajectory information has been transferred to an Argo V3.1 delayed mode trajectory file
  - From GDAC profiles : pressure, temperature and salinity adjustments have been recorded in the adjusted parameters of the V3.1 trajectory file.
  - Available at : ftp://ftp.ifremer.fr/ifremer/argo/etc/coriolis-custom/argo-andro-data
  - Documentation at ftp://ftp.ifremer.fr/ifremer/argo/etc/coriolis-custom/argo-andro-data/argo-andro-data_20141016.pdf
- *Each DAC with PI/DM has to take the responsibility for the decision to use or not the ANDRO converted D-Traj files as a first version of their D-files. Each DAC/PI/D-Operator should do an assessment of some of their floats in order to define their strategy and report to ADMT. If they decide to use the Andro files, then they should retrieve the D-traj files from Coriolis FTP site and submit them to GDAC as usual.*

Trajectory progress : Need to define a strategy for DMQC . Some ideas at ADMT14

- The first step was to apply corrections found in delayed mode to the <PARAM> variables wherever they occur in the cycle. The qc flags on the <PARAM> variables should also be applied.
- The second step would be to quality control cycle number. Some floats do not send their cycle number and it must be estimated. Errors occur and these should be fixed in delayed mode.
- The final requirement could be quality control of surface times and positions.
  - Check that JULD measurements from Argos or Iridium were in chronological order. If not, the traj file can be sent back to DACs for reprocessing.
  - Next, apply the JAMSTEC position QC procedure when not done in real time by the DAC.
  - checking measurement codes were applied correctly
  - estimates of timing information
  - filling in measurement code 301 for the best calculated pres/temp during drift
  - apply grounded flag

ARC activities

- **South Atlantic ARC**: No new features reported .
- **MedArgo ARC:** Handles overall coordinating of float operations including deployments and production of Argo products. Also perform DMQC on the Argo data from this region. Have 245 floats with more than 70 floats active presently active and the involvement of 2 new countries(Lebanon and Turkey) .
- **North Atlantic ARC**: producing maps and gridded products as well as some time series plots. All available on the web. They also provide a data mining tool for the N Atl Argo data. Result of the assessment of DMQC in Northern Atlantic was performed. Proposal for modified OW method has been described in the following paper: http://www.coriolis.eu.org/News-Events/Newsletters/Coriolis-10 . The method was run on 578 floats and for 26 floats the salinity correction proposed by the PI differs significantly from Na-Arc results . PIs or DM operators of the 26 floats have been informed and the DM corrections have been revised or revisions are in process http://www.argodatamgt.org/Argo-regional-Centers/North-Atlantic-ARC/Overall-consistency-of-DM-corrections

ARC activities - contunue

- **Pacific ARC**: PARC has been providing float monitoring information in the Pacific region and the products developed at JAMSTEC, IPRC, and KIOST on the web site. JAMSTEC has released two new products.
- **Southern Ocean ARC**: no new features reported
- **Indian ARC**: working to archive IO data and making them available though a web-GIS, producing a DVD of Argo data for students and others with low bandwidth web connections, continuing DMQC on Indian floats, identifying and supplying CTD data from Indian research cruises for inclusion in the Argo reference database, held workshops to demonstrate the use of Argo data to researchers and generate value added products to be served on the web.