

Feedback from ADMT-18











M. Scanderbeg, S. Pouliquen
AST-19, Sidney, B.C., Canada, March 2018

Main topics

- Data monitoring, real time and near real time QC
- Format update, BGC merged multi-prof format
- DMQC monitoring and action items
- GDAC status update
- ARC activities at North Atlantic, Pacific, Indian, Southern Ocean

Data system monitored by AIC

Data Flow

Delivery Argo Global	96.39% 2/2018 	- Raw count	95% Target	# of registered units vs number of operational units (Global Argo)
Quality (DM Processing) Argo Global	74.38% 2/2018 	- Raw count	75% Target	# of DM obs vs # of DM eligible obs (> 12 months)
Quality (PSAL) Argo Global	88.25% 2/2018 	- Raw count	90% Target	# of monthly obs of best quality - PSAL
Quality (TEMP) Argo Global	91.92% 2/2018 	- Raw count	90% Target	# of monthly obs of best quality - TEMP
Timeliness (GDAC FR) Argo Global	94.35% 2/2018 	11281 Raw count	90% Target	% of monthly observations distributed within 24h (GDAC FR)
Timeliness (GDAC US) Argo Global	91.05% 2/2018 	11005 Raw count	90% Target	% of monthly observations distributed within 24h (GDAC US)
Timeliness (GTS FR) Argo Global	95.95% 2/2018 	6568 Raw count	90% Target	% of monthly observations distributed within 24h (GTS) (Global Argo)
Whitelist Argo Global	94.82% 2/2018 	3644 Raw count	95% Target	% of platforms whitelisted platforms vs operational platforms

Quarterly comparison with Altimetry

- **115 floats were extracted in most recent run :**
 - 59 'R', 54 'A', 2 'D'
 - 12 BLK (might need to be greylisted due to many problematic profiles being identified)
 - 103 CHK (need to be checked, but usually just one profile per float identified)
- **More and more floats on the list :**
 - More and more floats in the water
 - Some floats are on the list too long...
 - ➔ backlogs have to be corrected
 - ~50 % of the floats extracted show only one isolated very bad profile

DAC	Number of floats
aoml	75
bodc	3
coriolis	8
csio	2
incois	20
jma	1
kma	3
kordi	1
meds	2

Quarterly comparison with Altimetry

- Discussion at ADMT-18 on how to request feedback and prioritize removing floats with bad sensors from the data stream resulted in asking AIC to separate floats into two categories:

- Need to be checked
- May need to be greylisted

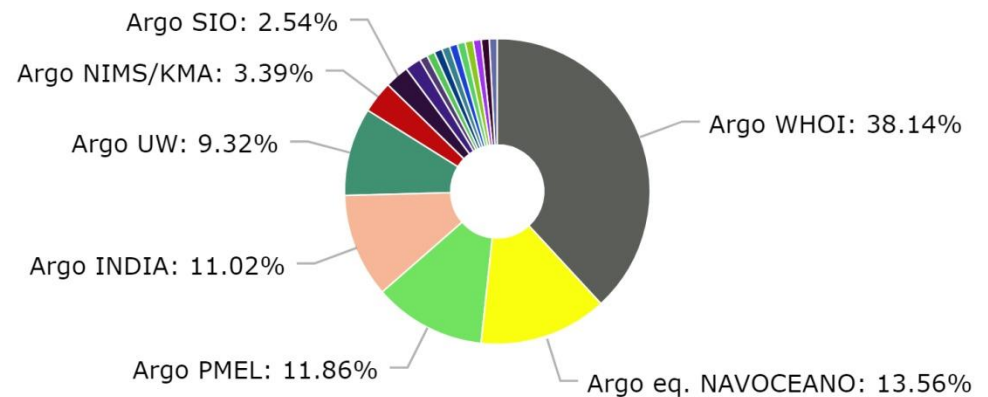
- AIC sends emails to DACs and DM-operators
- Can find this information on AIC and in S. Guinehut's list:

- ftp://ftp.ifremer.fr/ifremer/argo/etc/argo-ast9-item13-AltimeterComparison/QC_ARGO_ALTI_o82017.txt

1900438;chaotic,R/A,115-/,20170826
1900795;spike,R,134,20170826
1900830;chaotic,A,40-/,20170826
1900998;spikes,R,220,223,239,243,20170826
1901450;spike,R,138,20170826
1901481;spikes,R,102,147,20170826
1901501;spikes,R,110-124,20170826
1901531;spike,R,31,20170826
1901543;spike,R,32,20170826
1901554;spikes,R,7,19,20170826
1901587;spike,R,9,20170826
1901591;spike,R,47,98,20170826
1901627;spike,R,111,20170826

Altimetry QC floats pending feedback

- AST-18 Action Item 5 asked WHOI, UW, India, NAVO and PMEL to look at list and provide feedback
- Compared to last year:
- WHOI has provided feedback on 12 floats; continuing to work on this
- UW has for 10 floats
- India has for 7 floats
- PMEL and NAVO have not changed



Argo WHOI	45	Argo eq. NAVOCEANO	16
Argo PMEL	14	Argo INDIA	13
Argo UW	11	Argo NIMS/KMA	4
Argo SIO	3	Argo JAMSTEC	2
Argo AUSTRALIA	1	Argo KIOST	1
Argo UK	1	Argo UW-APL eq.	1
Coriolis-CONGAS	1	NAOS-France	1
Argo UW-SPURS eq.	1	Argo UK Bio	1
Argo BRAZIL	1	Argo CHINA	1

General quality of Argo dataset

- **All profiles, QC='1' = 1 232 891 profiles (as of August 2016)**

Floats	Correlation	Mean difference (cm)	Rms difference (cm)	Rms difference (%)	Nb profiles	% Total profiles
ALL	0.77	0.24	4.99	25.48	1 232 891	100
DATA_MODE='R'	0.72	0.70	6.19	34.86	160 120	13.0
DATA_MODE='A'	0.68	1.00	5.36	25.41	319 746	25.9
DATA_MODE='D'	0.88	-0.18	4.51	23.04	753 024	61.1

- **All profiles, QC='1' = 1 405 441 profiles (as of August 2017)**

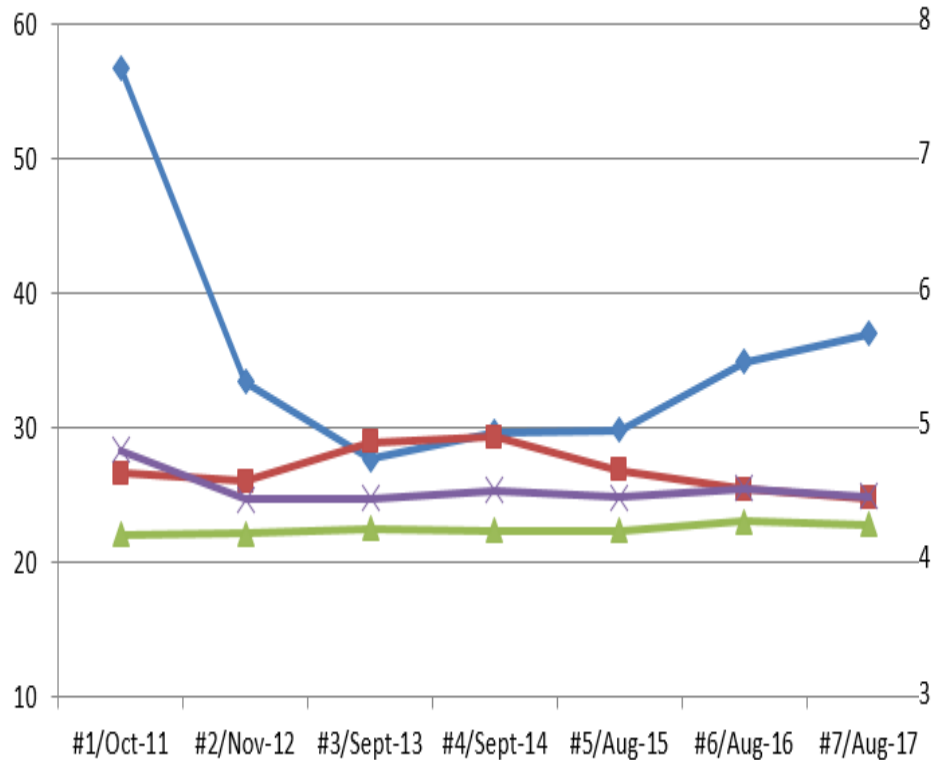
Floats	Correlation	Mean difference (cm)	Rms difference (cm)	Rms difference (%)	Nb profiles	% Total profiles
ALL	0.86	0.46	5.01	24.92	1 405 441	100
DATA_MODE='R'	0.79	1.00	6.08	37.00	153 846	11.0
DATA_MODE='A'	0.84	1.24	5.46	24.76	347 512	24.7
DATA_MODE='D'	0.88	0.07	4.61	22.79	904 083	64.3

➔ **Statistics are again degraded for 'R' observations, Number of 'D' increases**

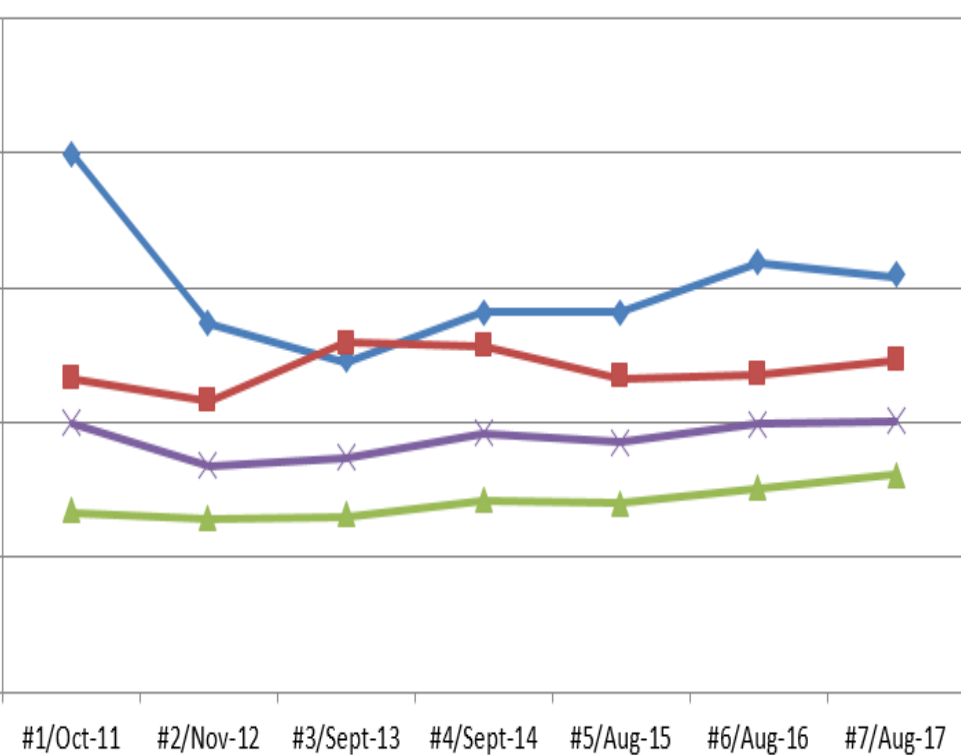
General quality of Argo dataset

- DATA_MODE='R'
- DATA_MODE='A'
- DATA_MODE='D'
- ALL

Rms difference % var(SLA)

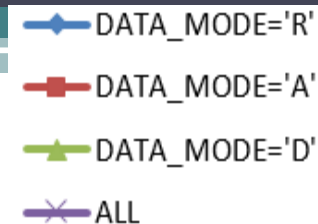


Rms difference cm

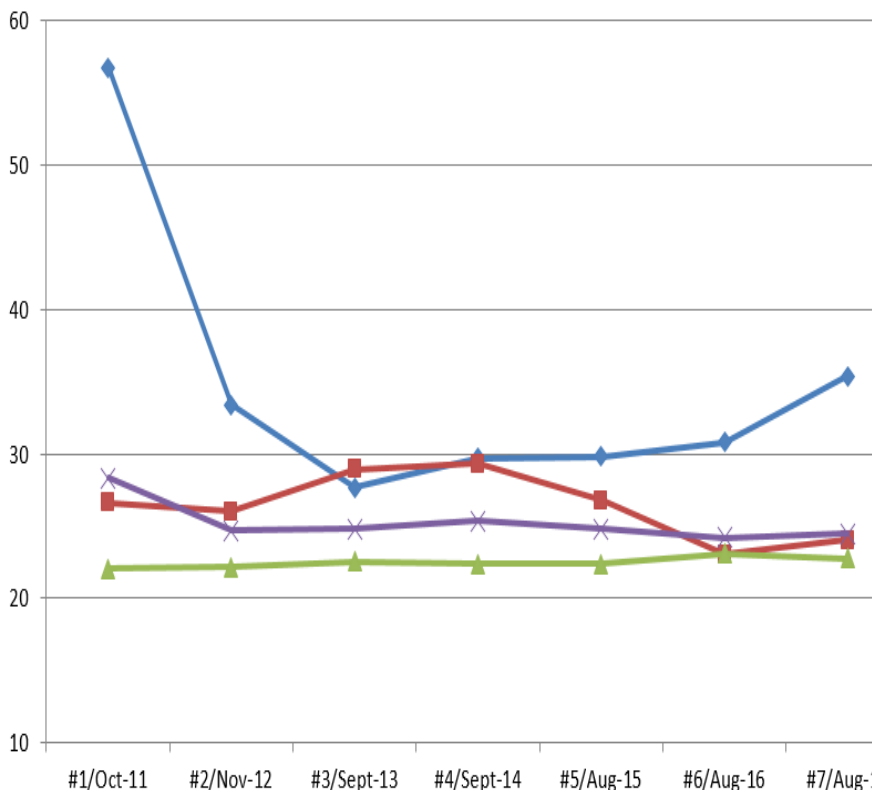


→ Statistics are degraded for 'R' observations

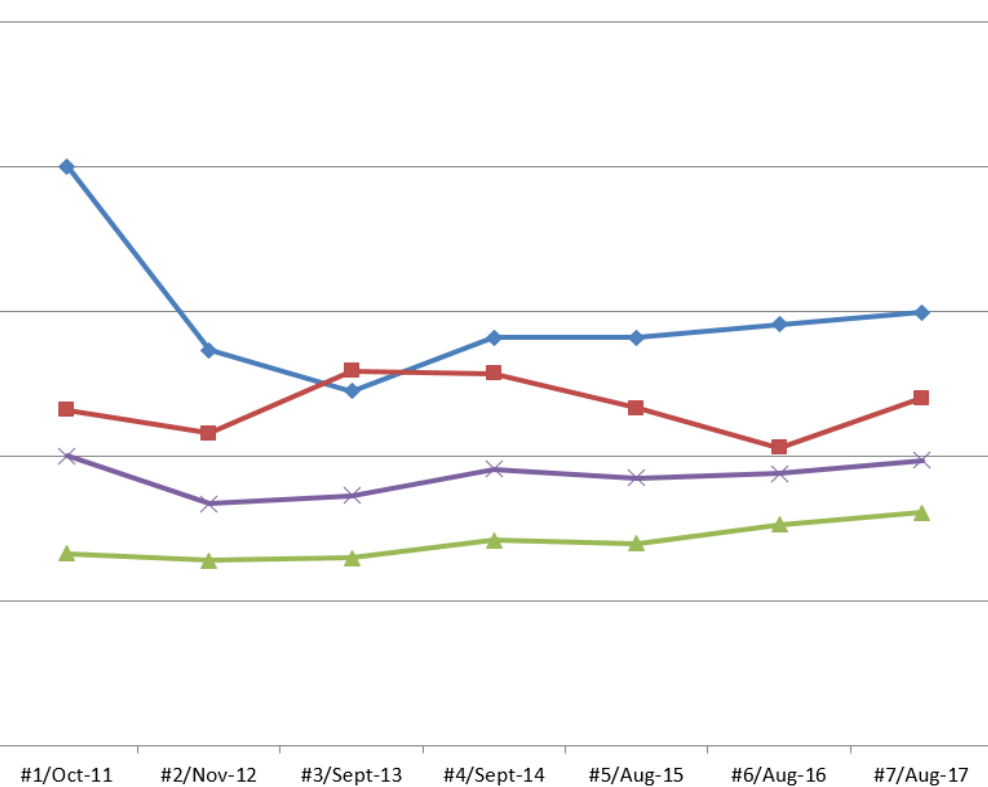
General quality of Argo dataset



Rms difference % var(SLA)



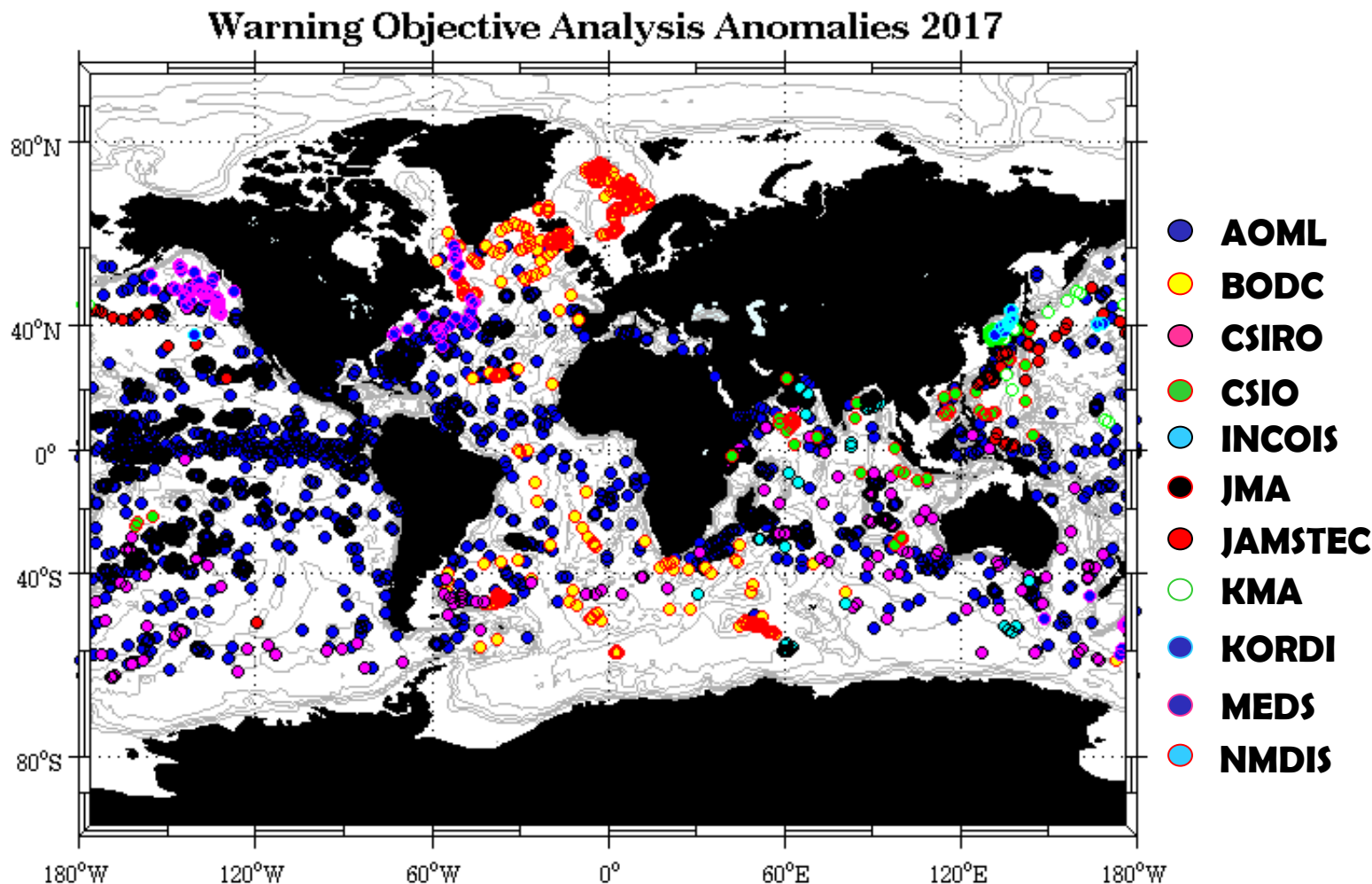
Rms difference cm



➔ Statistics are improved when data from the 115 floats present in the Altimetry QC list are rejected

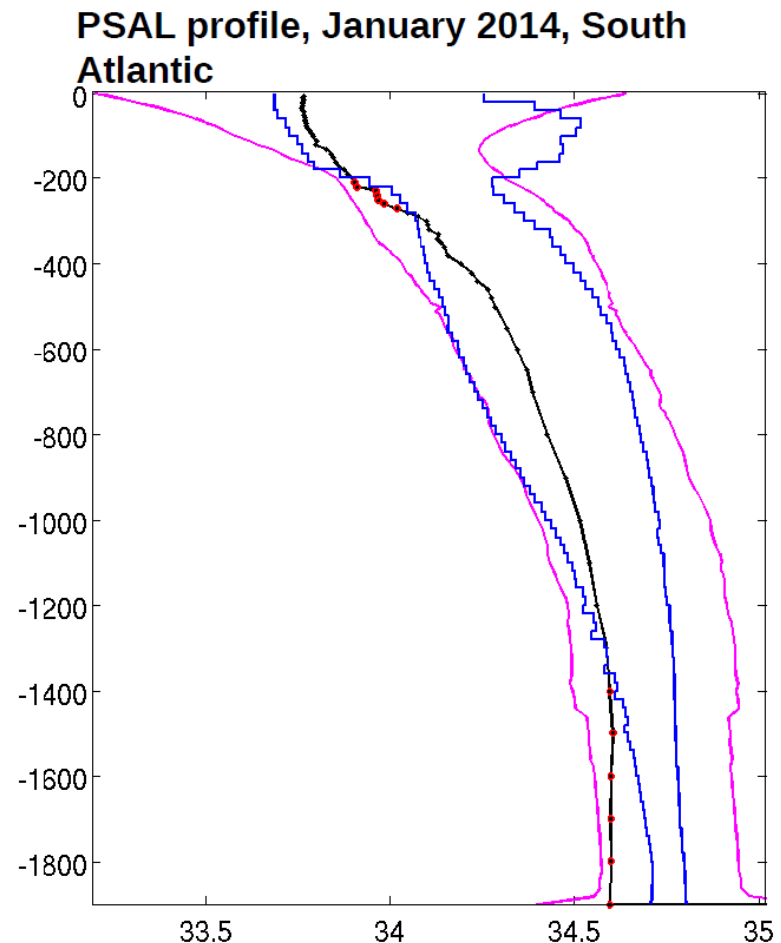
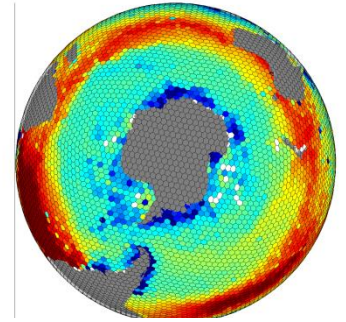
Daily Objective Analysis done at Coriolis

- Report sent monthly, list available online
- DACs, PIs mostly providing feedback
- Less than 400 warnings each month



Can we do more in real time?

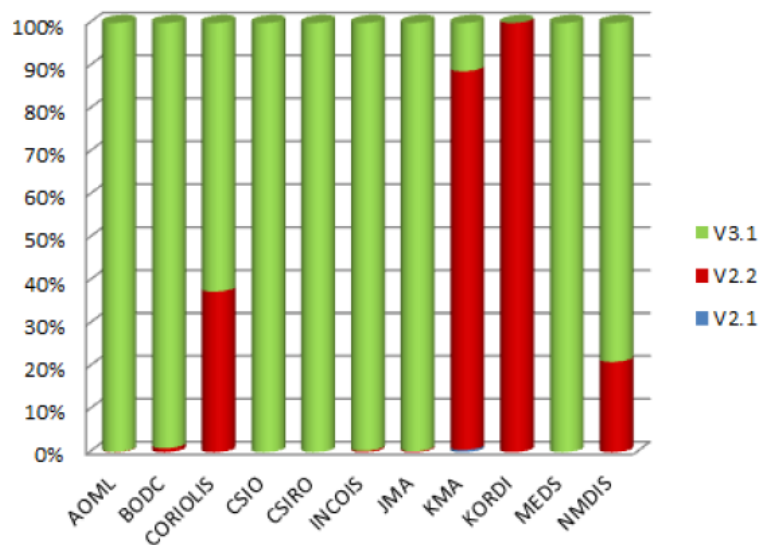
- New automatic MinMax tool developed by J. Gourrion to check quality of temp and salinity data
- Used reference field of minimum and maximum observed temperature in 120km wide hexagonal cells
 - Based on Argo, CTD and sea mammal profiles
 - Very efficient to find biased profiles
 - Lower wrong detection rate
 - Accurate in the deep ocean
- ADMT-18 Action Item asks working group to study building a gradient climatology from Argo with good, uncorrected data in a similar manner to test in MinMax tool



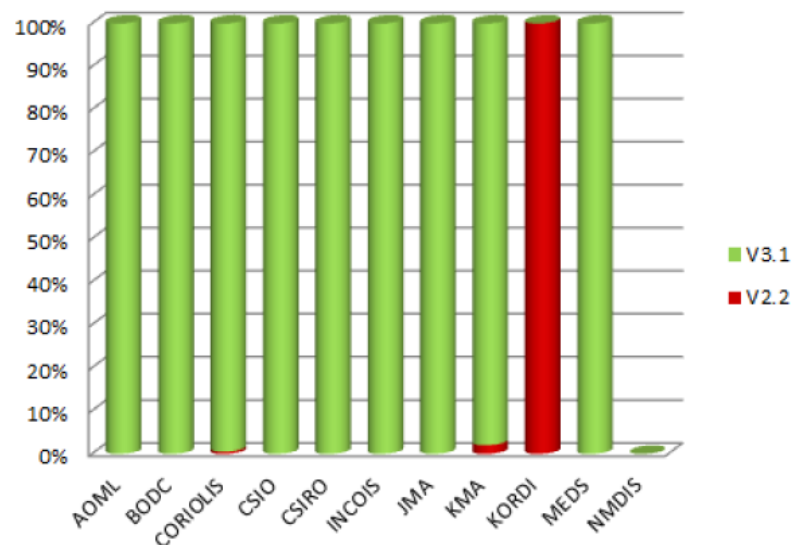
Format updates

- Big improvements in moving active floats to v3.1 for all file format types
- Dead floats still need work at some DACs
- Overall, lagging DACs are catching up
- Can change focus to improving content of v3.1 files
 - Planning on running Format Checker on all v3.1 files on GDACs with priority on delayed mode files to check for anomalies
 - Trajectory files improving as well; see talk on DAC Trajectory Workshop
 - Meta-file contents to have increased monitoring as well; see meta and tech file talk
 - BGC merged multi-profile files

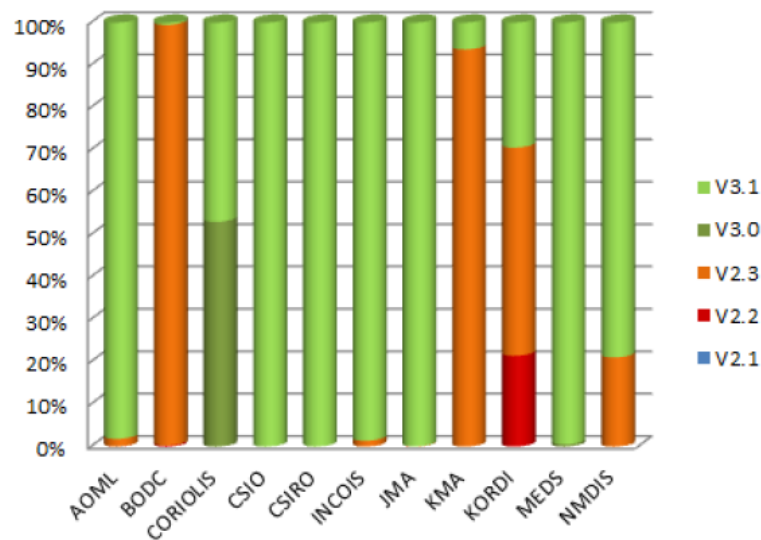
Metadata Files - Dead floats



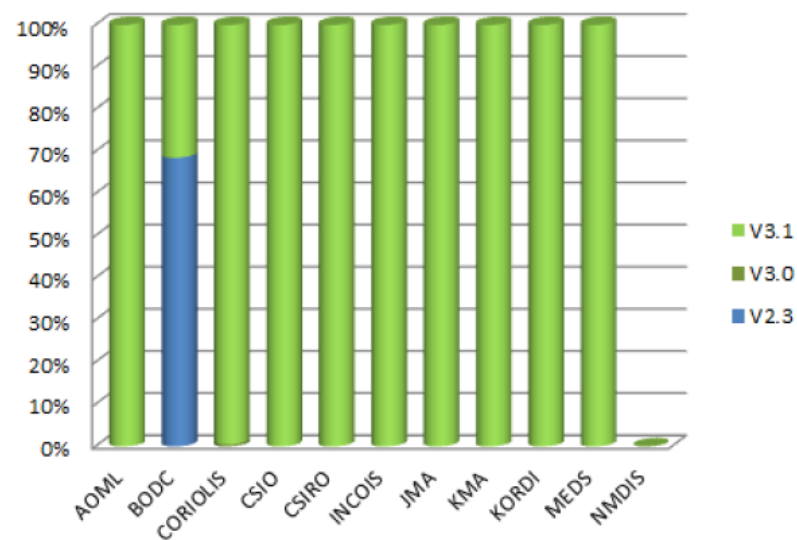
Metadata Files - Active floats



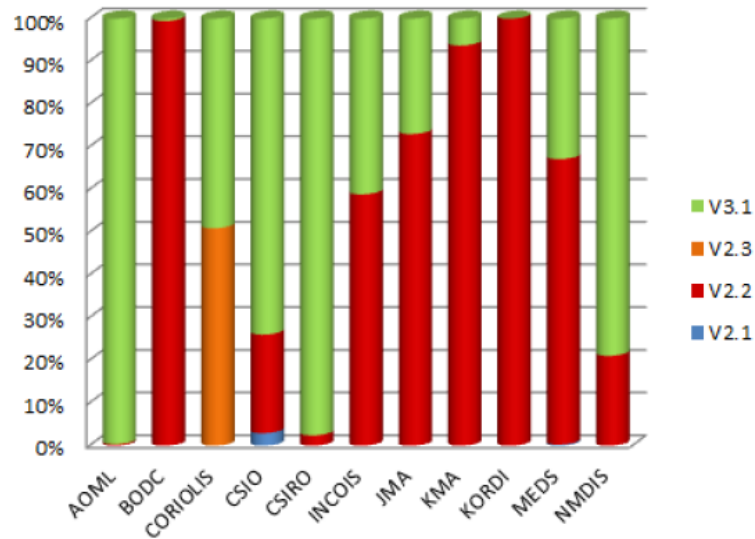
Technical Files - Dead floats



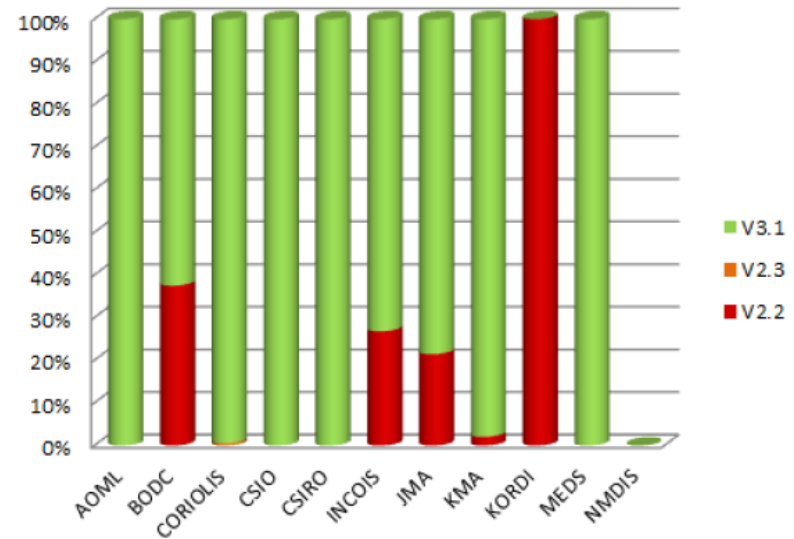
Technical Files - Active floats



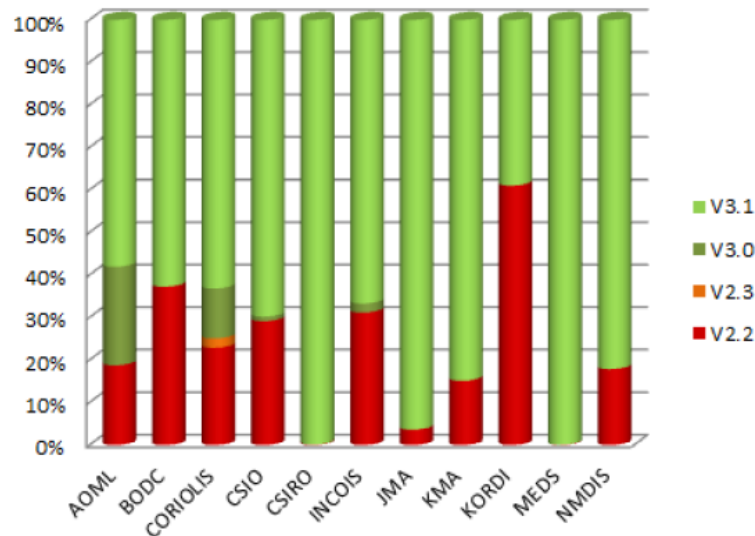
Trajectory Files - Dead floats



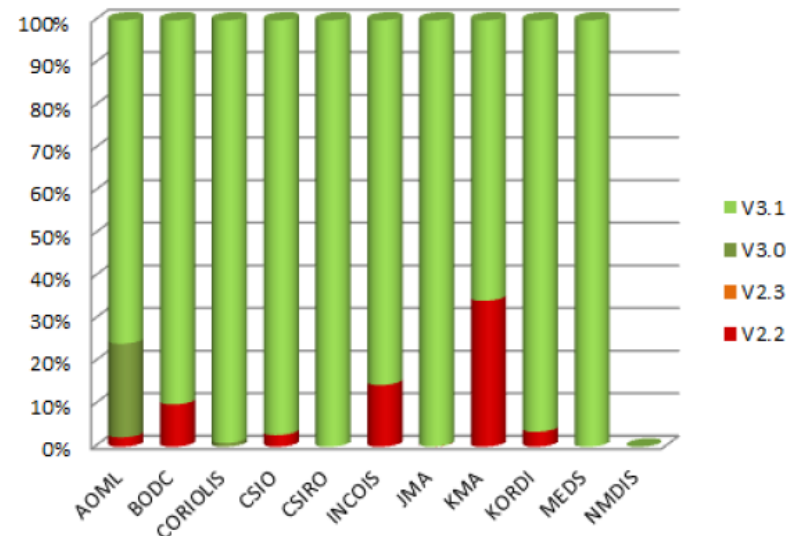
Trajectory Files - Active floats



Profile files - Dead floats



Profile Files - Active floats



BGC-Argo multi-profile files

1. Argo archives all data from floats, but size of merged multi-profile file for some BGC Argo floats is huge
 2. BGC-Argo scientific community is maturing and users are asking for more user friendly files
- To address 1 above, Argo is:
 - Investigating using NetCDF4 to compress data to reduce disk size (does not help with memory requirements when actually using data)
 - Developing a new merged multi-prof file format with unique pressure axis reporting all pressures and measurements
 - To address 2 above, BGC-Argo is:
 - Developing synthetic BGC profiles with co-location of BGC parameters (within constraints) on pressure levels
 - Aiming for a simpler structure when there are several vertical sampling schemes

DMQC topics

- Action Item 13 from AST-18 asked for investigating DMQC frequency and prioritizing floats to plan revisit times
- QC is costly to Argo in terms of manpower and money:
 - G. Maze studied ways of using machine learning to reduce QC burden on Argo; see his talk Thursday
 - G. Maze recommended ways to improve meta data in Argo profile files to aid in this process

DMQC 1st visit and revisit time

B. King did analysis to answer the questions:

- 1) How often do we presently revisit DMQC ?
- 2) What would be implications of having the target changed to

First DMQC at $t = 12$ months

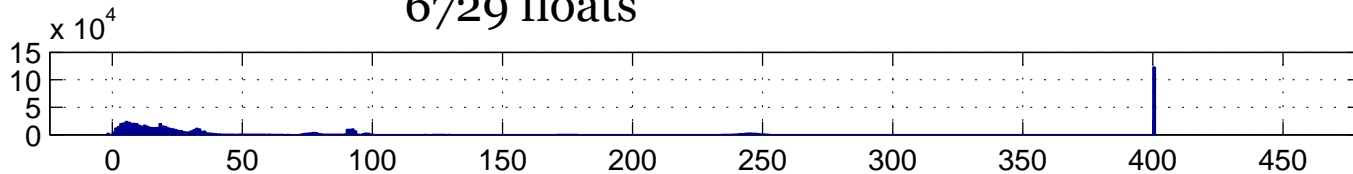
Second DMQC at $t = 36$ months instead of 24 months

- Audit performed on archived index files up to September 2017
- Focusing in on period of 2010 - 2018
- First and Second DMQC inferred from dates on which data mode is D and the file had a changed “update date” in the index files
- Have examined first appearance of D, and the elapsed time to the second appearance of D

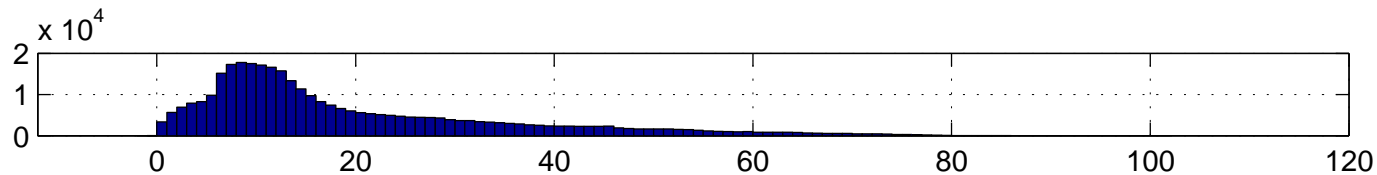
How often are floats dmoded?

2010 – 2018: All DACs

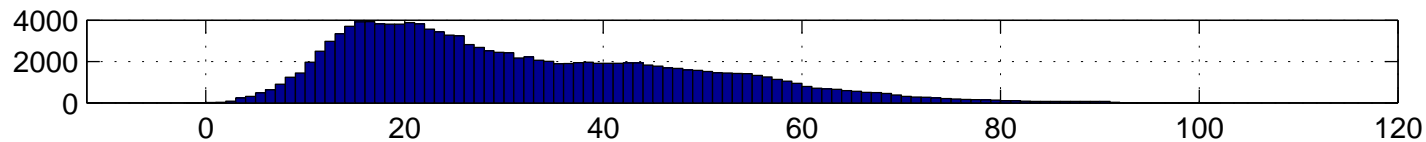
6729 floats



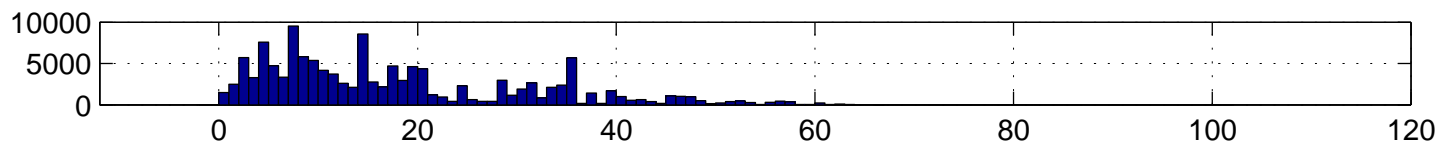
R file age (hours) 798,849 cycles



First D file age (months) 343,737 cycles



Second D file age (months) 130,745 cycles



Second minus first D file age (months) 130,745 cycles

DMQC frequency and size conclusions

- 1) At cycle 10, more adjustments are > 0 than < 0
- 2) As the floats age, the balance swings towards $\text{adj} < 0$

3) Floats deployed since 2010

118/2710 had $\text{adj} > 0.005$ at $\text{cyc} = 36$ (4.3%)

99/1961 had $\text{adj} > 0.005$ at $\text{cyc} = 72$ (5.0 %)

66/1285 had $\text{adj} > 0.005$ at $\text{cyc} = 108$ (5.1 %)

47/1918 had change of adj from $\text{cyc} = 36$ to 72 (2.5 %)

42/1257 had change of adj from $\text{cyc} = 36$ to 108 (3.3 %)

ADMT-18 action items on DMQC

- First DMQC needs to be done at 1 year
- Second DMQC visit for drift might be relaxed to 3 years for groups struggling with DMQC
- Add deepest profile sampled to index file to facilitate monitoring
- DMQC files will no longer be accepted in anything other than v3.1 by July 2018
- DMQC Orphan floats:
 - Ask AIC to manage DM-operator by parameter instead of by float
 - Ask M. Belbéoch to provide list of DMQC orphan floats at ADMT to assign volunteers for each parameter on a yearly basis

ADMT-18 action items for machine learning

- Study how to better capture DM-operator and institution for each parameter in profile
- Review list of software reported in HISTORY_SOFTWARE variable to distinguish between human made actions and non-actions to improve machine learning studies
- Improve HISTORY section in User Manual and propose FileChecker tests to ensure HISTORY variables are compliant with ref tables and each other

GDAC status

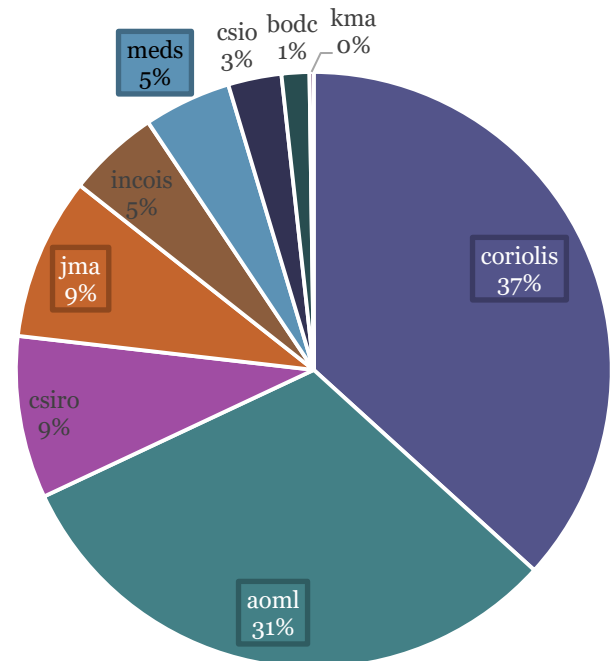
DAC	metadata files 2017	increase from 2016	profile files 2017	increase from 2016	delayed mode profile files 2017	increase from 2016	increase from trajectory files 2017	increase from 2016
AOML	6 601	10%	996 019	13%	686 144	20%	8 064	12%
BODC	636	18%	70 637	23%	33 015	5%	479	14%
Coriolis	2 554	11%	263 894	17%	179 361	38%	2 465	11%
CSIO	370	8%	44 934	15%	10 221	0%	365	7%
CSIRO	806	8%	140 870	14%	120 680	24%	781	10%
INCOIS	422	7%	58 538	14%	28 008	1%	379	2%
JMA	1 550	7%	188 716	11%	115 044	20%	1 484	4%
KMA	227	5%	29 005	11%	23 094	11%	208	0%
KORDI	119	0%	16 578	2%	6 986		119	0%
MEDS	472	9%	47 983	8%	31 687	23%	456	8%
NMDIS	19	0%	2 460	0%	0		19	0%
Total	13 776	9,35%	1 859 634	13,75%	1 234 240	22,02%	14 819	9,93%

BGC-Argo at GDAC level

- In November 2017, 131,308 BGC Argo cycle files from 863 floats were available on Argo GDAC. This is a big increase compared to 2016:
+65% more profiles and +54% more floats.

DAC	Number BGC floats	Number B files
coriolis	317	41 147
aoml	270	36 917
csiro	76	20 149
jma	76	14 602
incois	43	5 201
meds	41	3 821
csio	25	6 344
bodc	13	3 041
kma	2	86
Total	863	131 308

DAC distribution of Argo-bgc floats

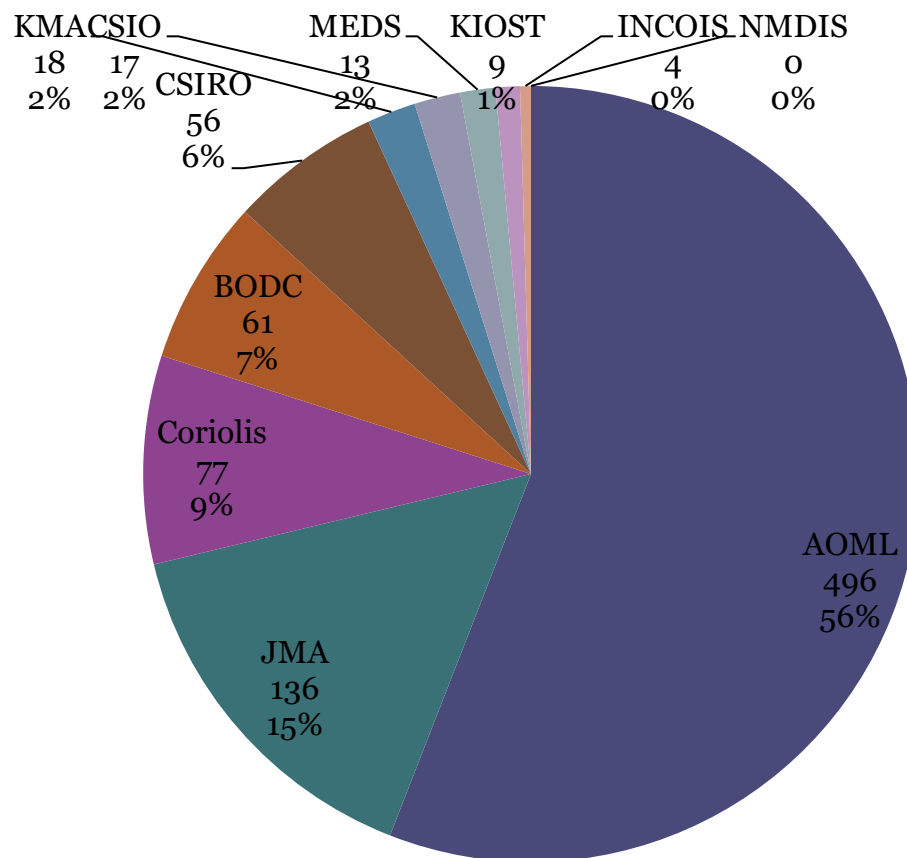


GDAC servers

- FTP weekly average performance was 99.870%
- Daily GDAC synchronization takes about 100 minutes; failed 5 times in past year
- Submitted files are automatically collected from DACs **every 10 minutes**
- At Coriolis:
 - monthly average of 449 unique visitors, performing 4552 sessions and downloading 3.3 terabytes of data files
 - Majority of users are located in USA, China, Australia and Europe
- Continue to offer rysnc and ERRDAP services

Grey list

- GDAC hosts a grey list of the floats which are automatically flagged by automatic QC
- Grey list has 887 entries (November 25th 2017) compared to 1000 entries one year ago



AUX directory

- Both GDACs have set up the AUX directory and several DACs have subfolders on it containing
- Tech files with additional timing information:
ftp://usgodae.org/pub/outgoing/argo/aux/coriolis/3901089/3901089_tech_aux.nc
ftp://usgodae.org/pub/outgoing/argo/aux/aoml/5902485/5902485_tech_aux.html
- Fragment under-ice detection profiles:
ftp://usgodae.org/pub/outgoing/argo/aux/aoml/7900678/profiles/7900678_006_aux.nc
- Fragment test cycle start up profiles:
ftp://usgodae.org/pub/outgoing/argo/aux/aoml/7900678/profiles/7900678_000D_aux.nc
- Formats are not consistent, but data is being distributed
- Some data may never be QC'd

ARC activities

- Advanced automatic real time QC (AQC) at PARC
- DMQC consistency checks at NA-ARC
- Adoption of orphan floats for DMQC at IARC, SOARC
- Research into other DMQC approaches at IARC
- Ref DB analysis and search for additional Ref DB CTD profiles at NA-ARC, IARC and SOARC
- Studying under-ice positioning at SOARC
- Development of products
 - ISAS2015 at NA-ARC
 - Gridded products and visualization tools at IARC
 - Monthly AQC dataset at PARC

Near real time automatic QC at PARC

The aim is to produce useful QC information for PIs based on process of making AQC v2.0 dataset.

- Applied to Argo R, A, and D mode files from GDACs
- Processing parameters
 - PRES, TEMP, PSAL
- Conducting QC and adding flags for...
 - P,T,S profiles
 - Position_QC
 - 1 (good), 2 (probably good), 8 (interpolated value)
 - Juld_QC
 - 1(good),2(probablye good),8(interpolated value)
 - * Exclude when real time QC flags of PTS profiles are all 4 and 9 except for unpumped layer.
- Frequency of delivery from the web site
(http://www.jamstec.go.jp/ARGO/argo_web/argo/?page_id=100&lang=en)
 - Monthly, releasing in previous month

Consistency of DM salinity data in the NAARC region

- Based on GDAC snapshot from may 2017
 - 1653 floats in delayed mode
 - 1345 unbiased floats, according to PI's decision
 - Using four sets of configuration parameters:
-
- Only 24 floats for which the estimated offset is larger 0.01 PSU and larger than two times the statistical error, for all the four runs.
-
- Results obtained are mostly in agreement with the PI's decision

Conclusions

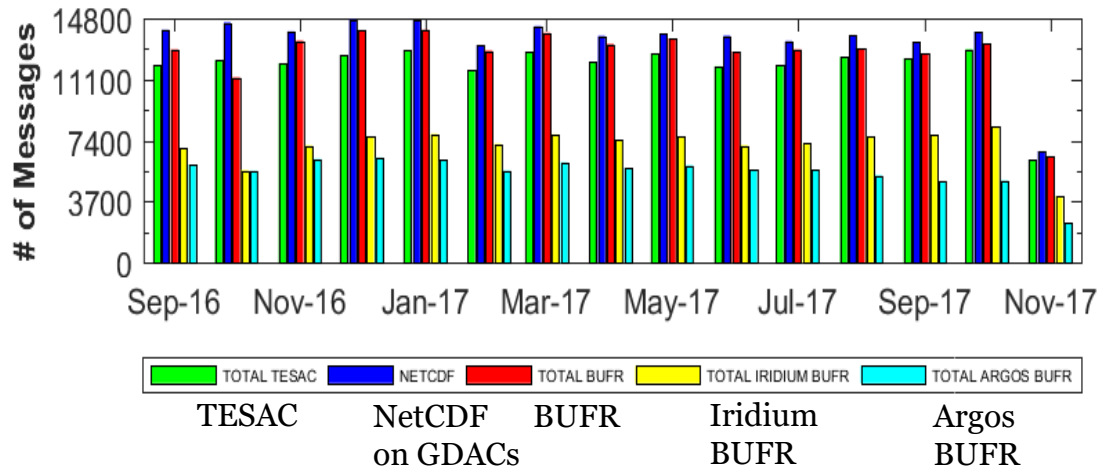
- Real time processing is going well
- DACs and PIs are working to finalize conversion to v3.1 file formats; format checker improves quality and consistency of the Argo dataset
- DMQC still high priority after one year; would like to do it yearly and some groups are catching up now that format conversion is winding down
- More feedback being sent on near real time QC tests
- Exploring additional real time QC tests
- Should consider machine learning possibilities for both real time and DM QC

Conclusions part two

- GDACs continue to grow in size with good availability
- BGC dataset is growing quickly; BGC group developing more user friendly scientific files and trying to reduce file size and complexity
- AUX directory is set up and ready to receive pilot or experimental sensor data as needed
- Most ARCs continue to function and provide needed work to improve Argo data set quality and consistency

Real Time: GTS report

* ADMT-18 Action Item to stop TESAC transmission on 1, July 2018. More on this in Fiona's BUFR format presentation



Monthly Average between
Sep 2016 - Oct

2017:

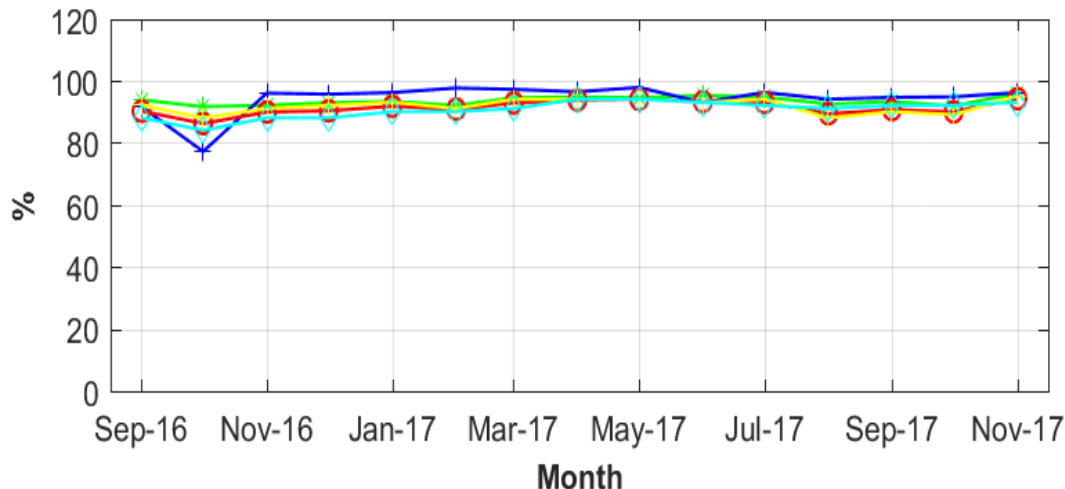
TESAC: 12,328 msg

BUFR: 13,145 msg

Iridium BUFR: 7,374 msg

Argos BUFR: 5,729 msg

Netcdf files: 13,971 files



Monthly Average between
Sep 2016 - Oct 2017:

TESAC on GTS in 24 hr: 93%

BUFR on GTS in 24 hr: 91%

Iridium BUFR on GTS in 24 hr: 91%

Argos BUFR on GTS in 24 hr: 91%

BUFR/NetCDF files at GDAC: 94%