Argo data management report 2018

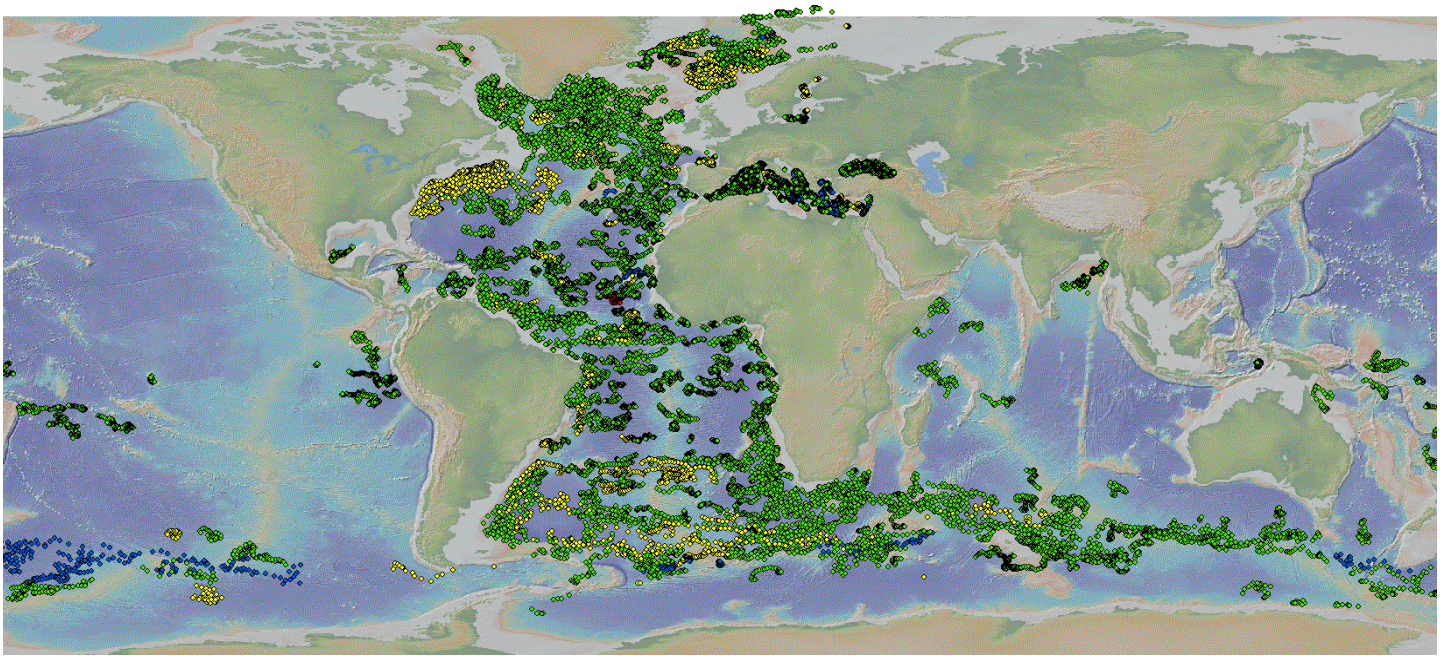
Coriolis DAC & GDAC

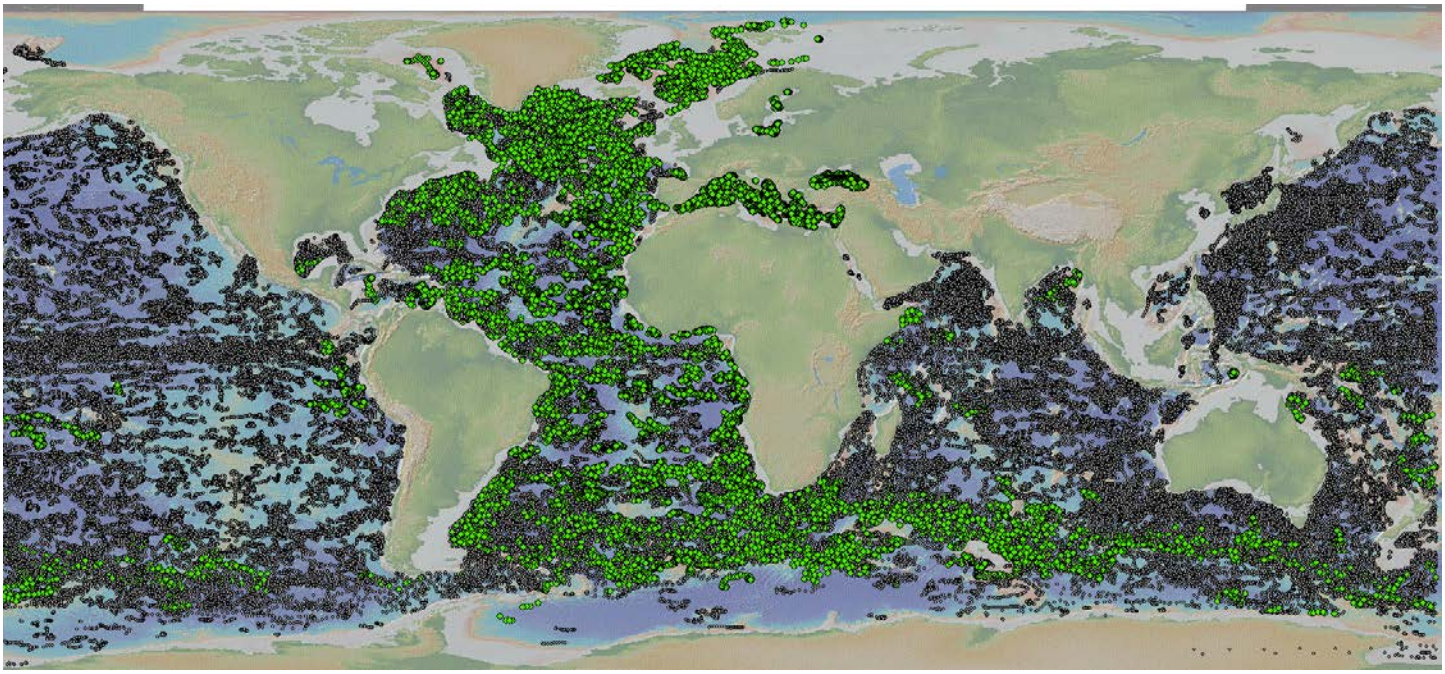
Data Assembly Centre and Global Data Assembly Centre

Annual report November 2018

Version 1.1, November 22nd, 2018

https://doi.org/10.13155/58109





1 DAC status

This report covers the activity of Coriolis data centre for a one-year period from November 1st 2017 to October 30th 2018.

1.1 Data acquired from floats

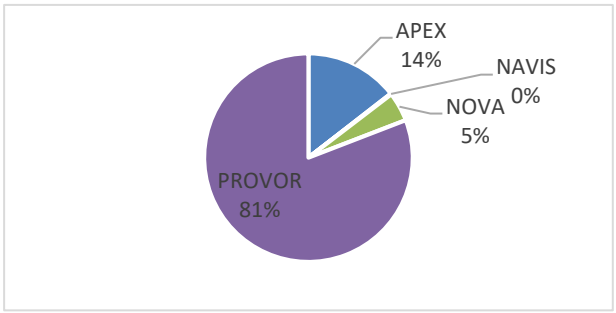
1.1.1 Active floats for the last 12 months

These last 12 months, **30 434 profiles from 794 active floats** were collected, controlled and distributed.

Compared to 2017, **the number of profiles is stable (+0.2%), the number of floats increased by 1%.**
These figures show a fair stability in Coriolis DAC activity.

The 794 floats managed during that period had 57 versions of data formats.

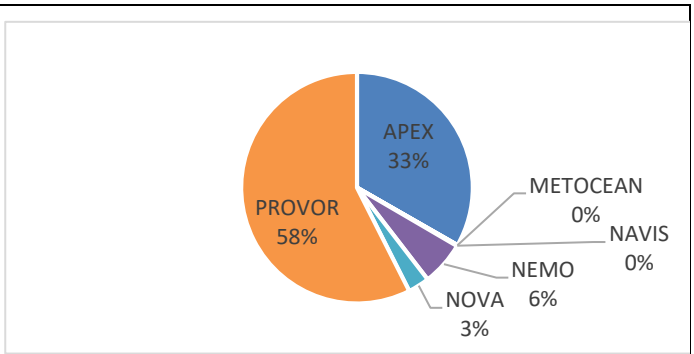
Coriolis DAC, active floats in 2018			
Float family	nb versions	nb floats	nb core profile files
APEX	23	115	3 630
NAVIS	1	1	35
NOVA	3	36	996
PROVOR	30	642	25 773
Total	**57**	**794**	**30 434**

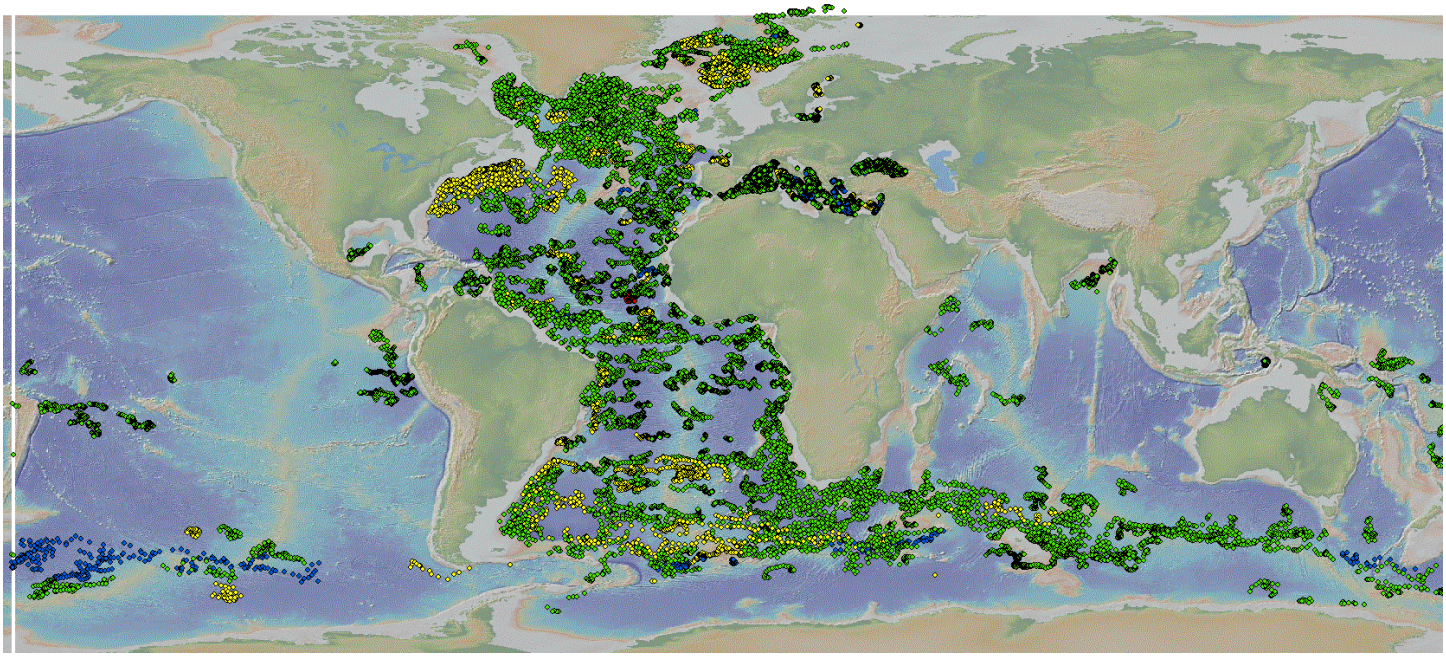


1.1.2 All floats managed by Coriolis DAC

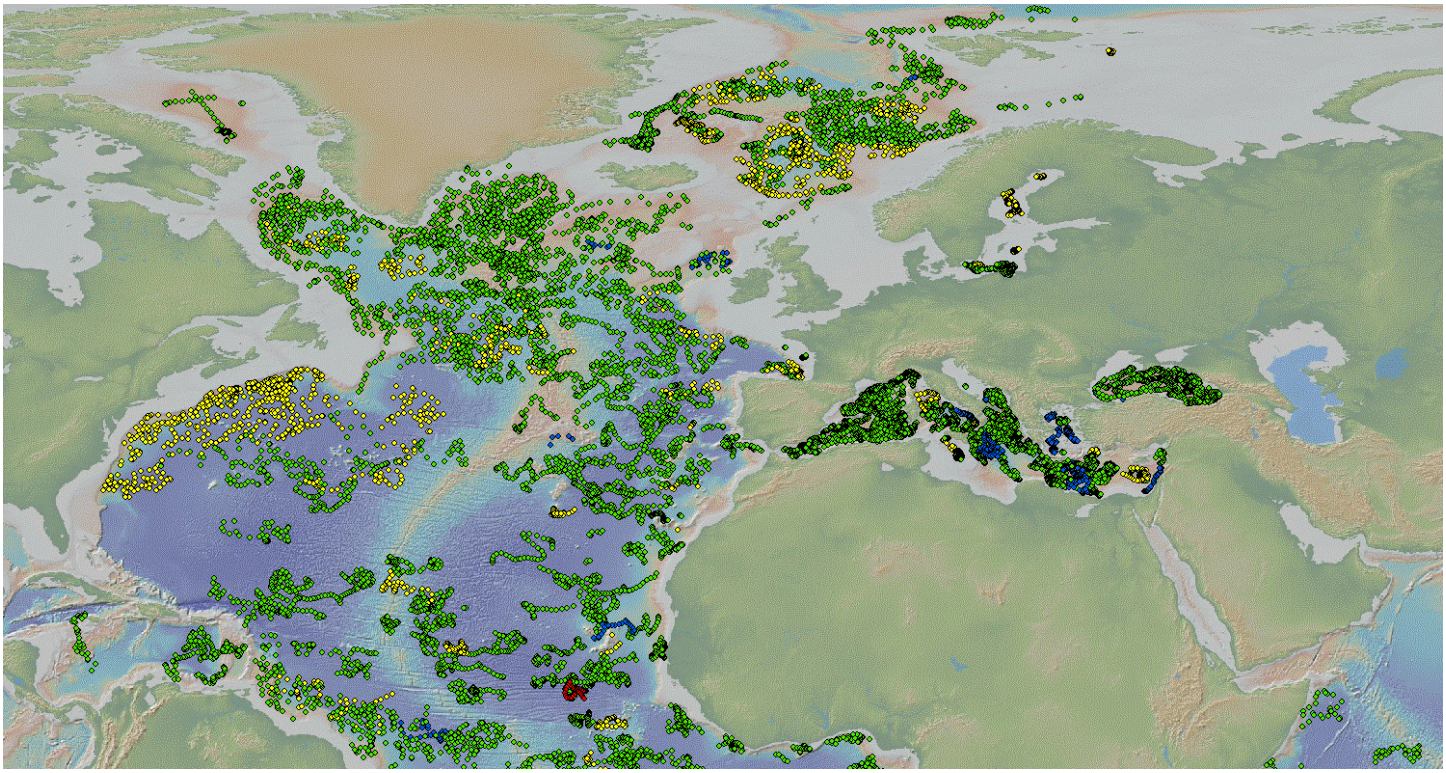
Coriolis DAC manages a total of 2 637 floats with 140 versions, from 6 families.
These floats reported 295 351 core Argo vertical profiles.

Coriolis DAC, all floats			
Float family	nb versions	nb floats	nb core profile files
APEX	67	877	101 062
METOCEAN	1	1	52
NAVIS	1	3	644
NEMO	8	162	8 902
NOVA	3	79	5 583
PROVOR	60	1 515	179 108
Total	**140**	**2 637**	**295 351**

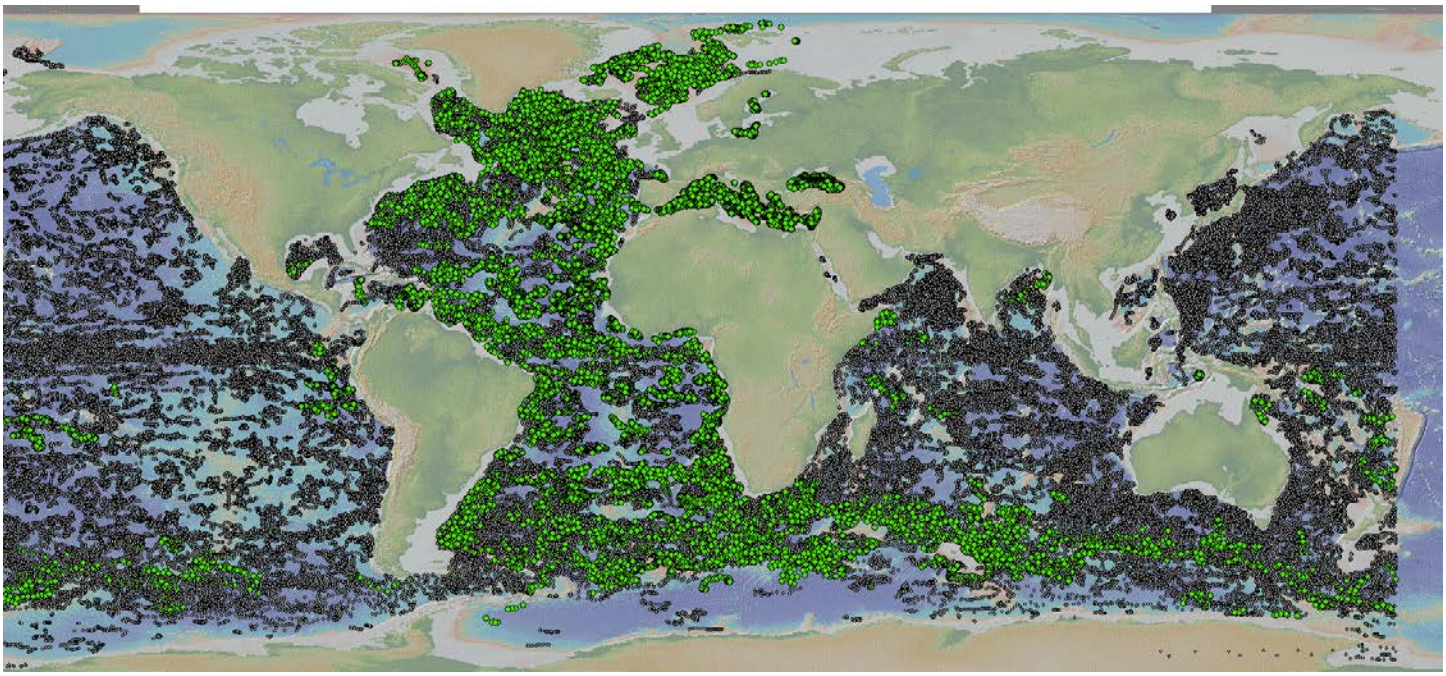




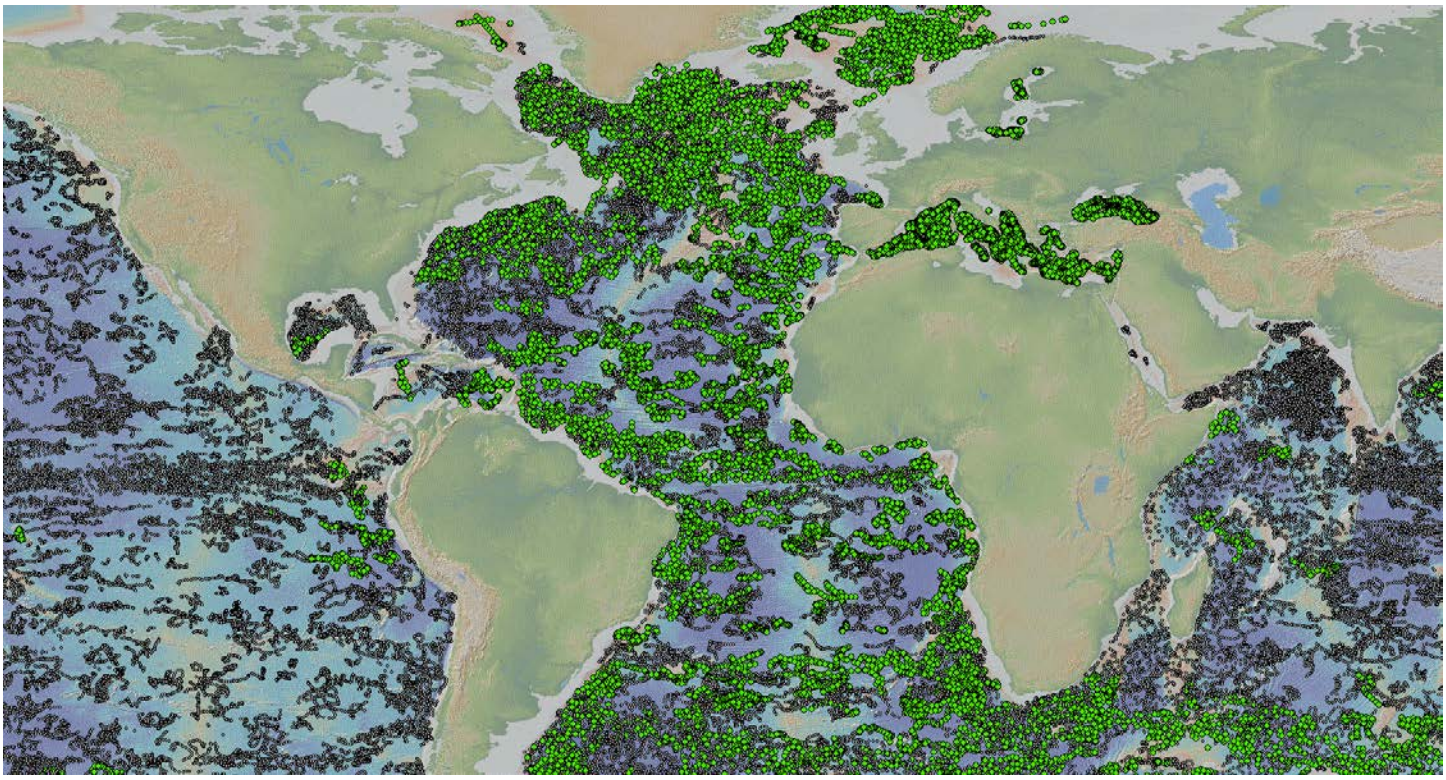
Map of the 30.434 profiles from 794 active floats decoded by Coriolis DAC this current year
Apex Navis Nova Provor



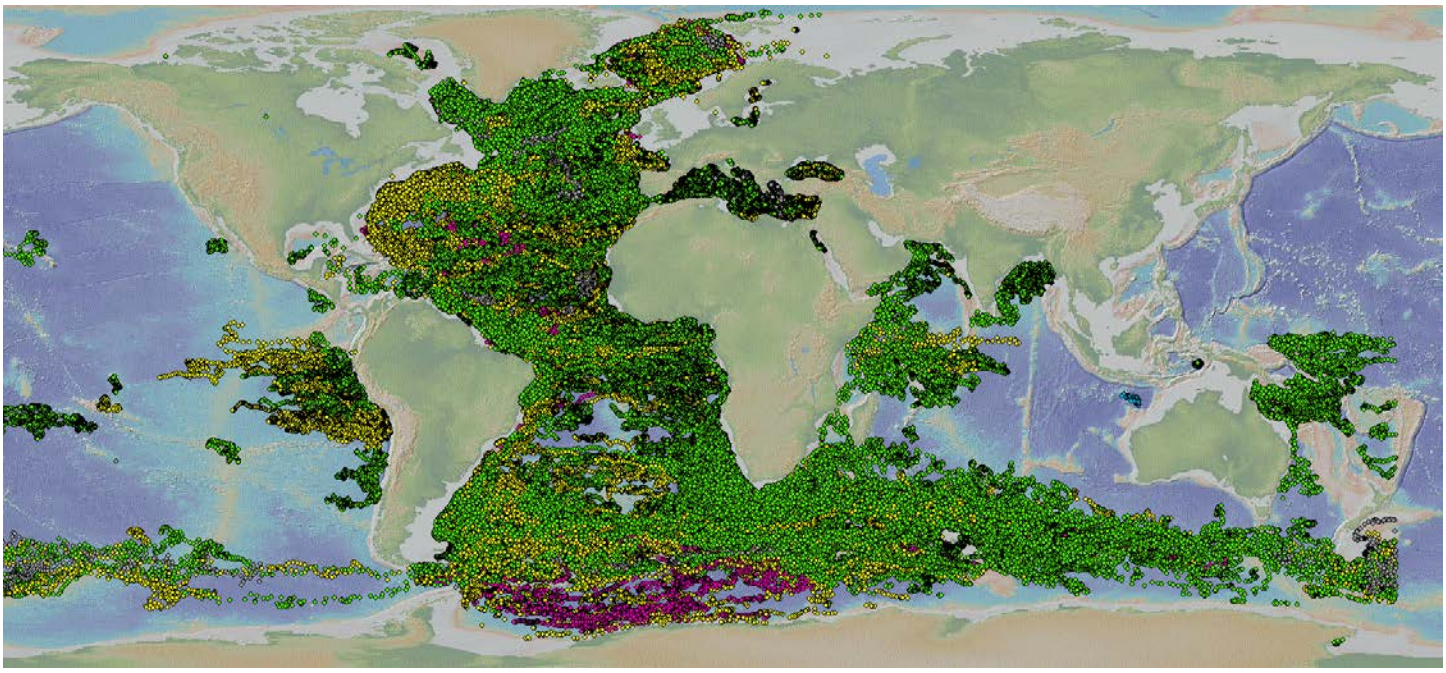
Map of active floats managed by Coriolis this current year, zoom on north Atlantic area



Map of the profiles from active floats decoded by Coriolis DAC this current year, among the other DAC's profiles (Coriolis: green, other DACs: grey)



Atlantic map active floats profiles from Coriolis DAC this current year, among the other DAC's profiles (Coriolis: green, other DACs: grey)



Map of the 295.351 profiles from 2.637 floats managed by Coriolis DAC
Apex Metocean Navis Nemo Nova Provor

1.1.3 BGC-Argo sensors on Coriolis floats

The data processing chain based on Matlab to manage data and metadata from Coriolis BGC-floats is continuously improved. These are advanced types of floats performing bio-geo-chemical (BGC) measurements.

Coriolis DAC manages 409 BGC-Argo floats from 5 families and 57 instrument versions. They performed 53.509 cycles.

The data processing chain is freely available:

- Coriolis Argo floats data processing chain, http://doi.org/10.17882/45589

BBP data reprocessing

In 2018, the BBP manual was updated: “BGC-Argo processing particle backscattering at the DAC level” http://dx.doi.org/10.13155/39459

To implement the updates, all BBP profiles were reprocessed during summer 2018. More than 28 000 files containing BBP data were resubmitted on the GDAC ftp server.

Chlorophyll data reprocessing

In 2018, the Quality control Chlorophyll-A manual was updated: “BGC-Argo quality control manual for Chlorophyll-A concentration” http://dx.doi.org/10.13155/35385

To implement the updates, all chlorophyll profiles were reprocessed during summer 2018. More than 28000 files containing chlorophyll data were resubmitted on the GDAC ftp server.

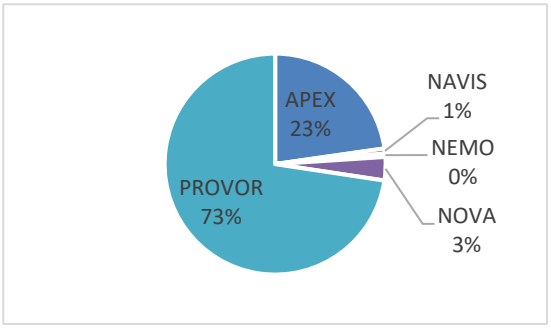
Oxygen data reprocessing

In 2018, the Oxygen manual was updated: “Processing Argo oxygen data at the DAC level cookbook” http://doi.org/10.13155/39795

To implement the updates, all oxygen profiles were reprocessed during summer 2018. More than 42000 files containing chlorophyll data were resubmitted on the GDAC ftp server.

Three PROVOR CTS5 floats deployed in 2017 in the framework of the greenedge project (http://www.greenedgeproject.info/) equipped with BGC sensors reappeared in july 2018 and transferred their data to the Coriolis DAC. They sent the first chlorophyll-A profiles acquired under ice for this type of floats (one per month under ice). All the data are available on the Coriolis DAC.

Bio-Argo floats processed by Coriolis DAC				
Familly	nb versions	nb floats	nb profiles	nb cycles
APEX	23	93	12 391	12 110
NAVIS	1	3	551	551
NEMO	1	2	297	297
NOVA	1	14	942	918
PROVOR	31	297	119 499	39 633
Total	**57**	**409**	**133 680**	**53 509**



General characteristics

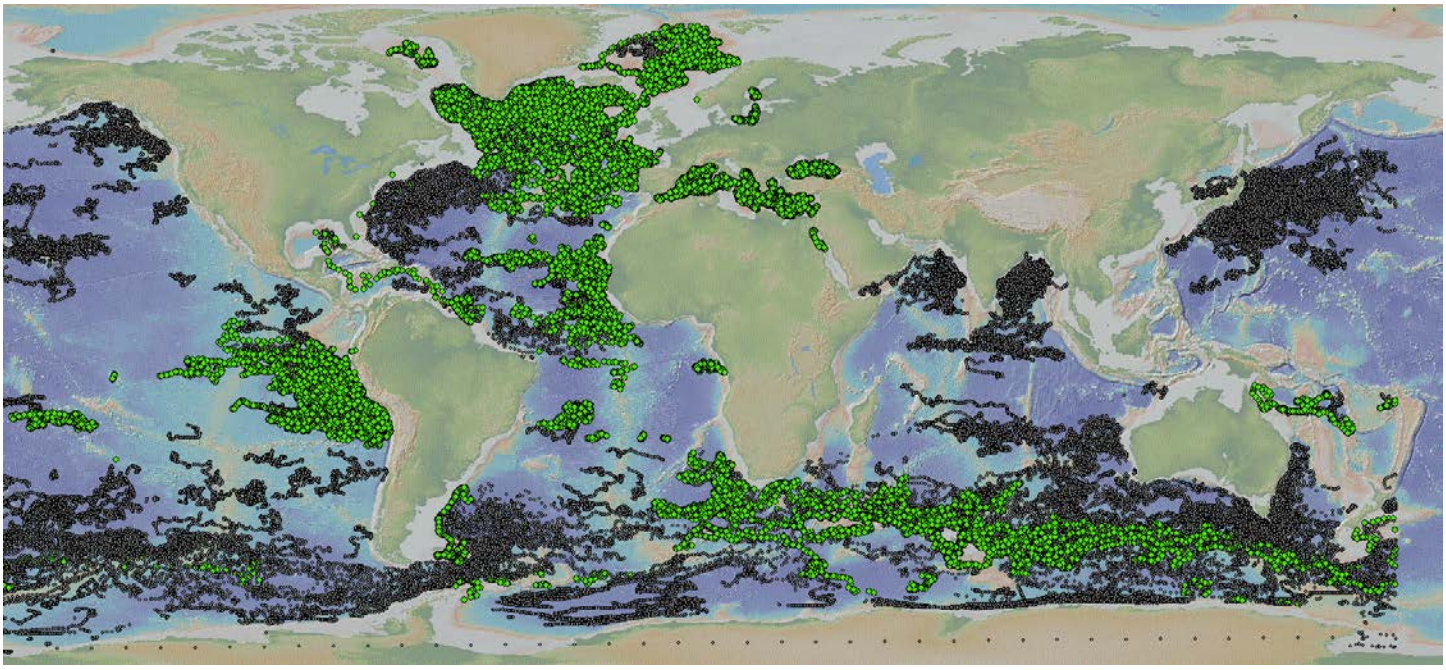
- Iridium sbd or rudics bi-directional communication or Argos
- Fourteen sensors are fitted on the floats
- Eleven BGC parameters reported

Coriolis BGC-Argo floats sensor	nb floats	nb profiles
AANDERAA_OPTODE_4330	265	34194
SATLANTIC_OCR504_ICSW	144	101752
ECO_FLBBCD	141	75852
DRUCK_2900PSIA	102	12349
SUNA_V2	59	8456
AANDERAA_OPTODE_3830	45	6081
SBE63_OPTODE	19	1775
C_ROVER	15	4356
SBE43F_IDO	10	1273
ECO_FLBB_AP2	8	1078
ECO_FLBB2	4	2016
ECO_FLNTU	4	1808
SEAFET	4	164
FLBB	2	616

The 14 types of sensors mounted on Coriolis BGC-Argo floats

parameter	nb profiles
DOXY	42606
CHLA	28278
BBP700	27374
DOWN_IRRADIANCE380	24446
CDOM	24029
NITRATE	7973
CP660	4330
TURBIDITY	904
BBP532	672
BISULFIDE	255
PH_IN_SITU_FREE	162

The 11 BGC parameters reported by Coriolis BGC-Argo floats



Map of the 409 BGC-Argo floats managed by Coriolis DAC (grey dots: the others DACs bio-Argo floats). They measure parameters such as oxygen, chlorophyll, turbidity, CDOM, back-scattering, UV, nitrate, bisulfide, pH, radiance, irradiance, PAR.

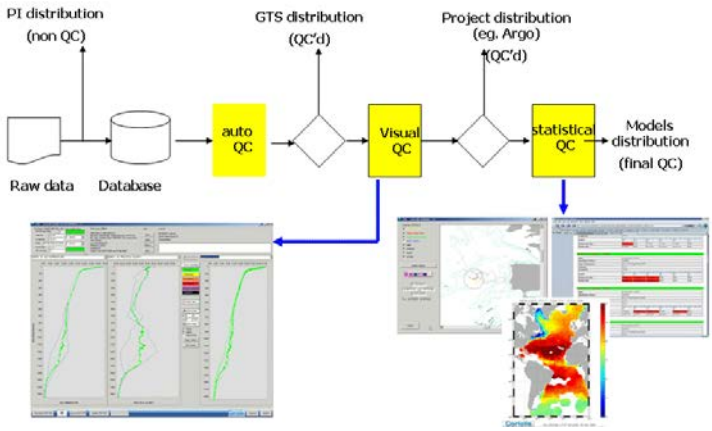
© Antoine Poteau, Observatoire Océanologique de Villefranche (CNRS/UPMC)
Deployments of a bio-argo Provor in Ligurian sea

1.2 Data issued to GTS

All profiles processed by Coriolis are distributed on the GTS by way of Meteo-France. This operation is automatically performed. After applying the automatic Argo QC procedure, the Argo profiles are inserted on the GTS every 2 hours. Argo profiles are inserted on the GTS 365 days per year, 24 hours a day.

The profile files are sent as TESAC and BUFR messages by way of Meteo-France. Meteo-France accepts Coriolis as valid BUFR messages and circulate them on neighbouring nodes.

Once a day, floats data that are less than 21 days old are checked in an objective analysis (ISAS) that triggers alert and visual inspection for suspicious observations.

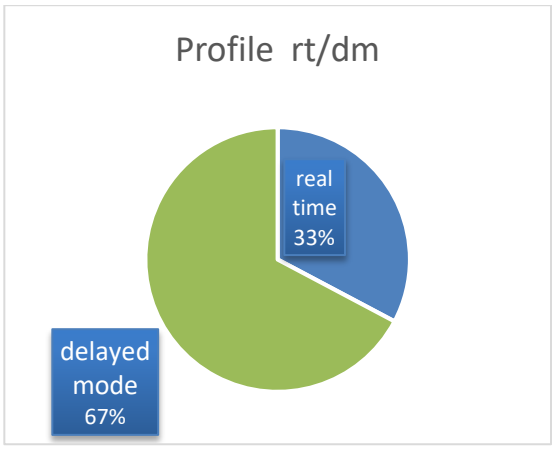


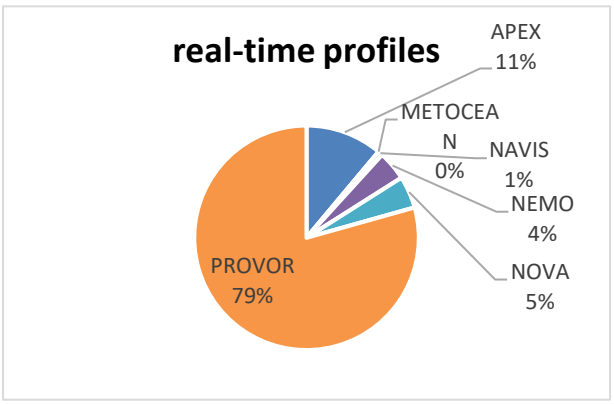
CORIOLIS DAC: Argo data flow

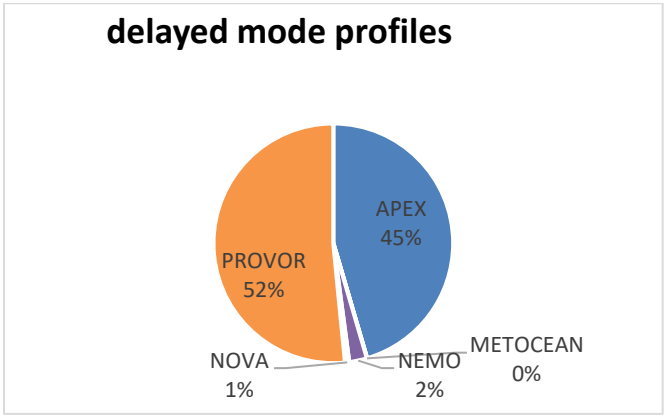
1.3 Data issued to GDACs after real-time QC

All meta-data, profiles, trajectory and technical data files are sent to Coriolis and US-GODAE GDACs. This distribution is automated.

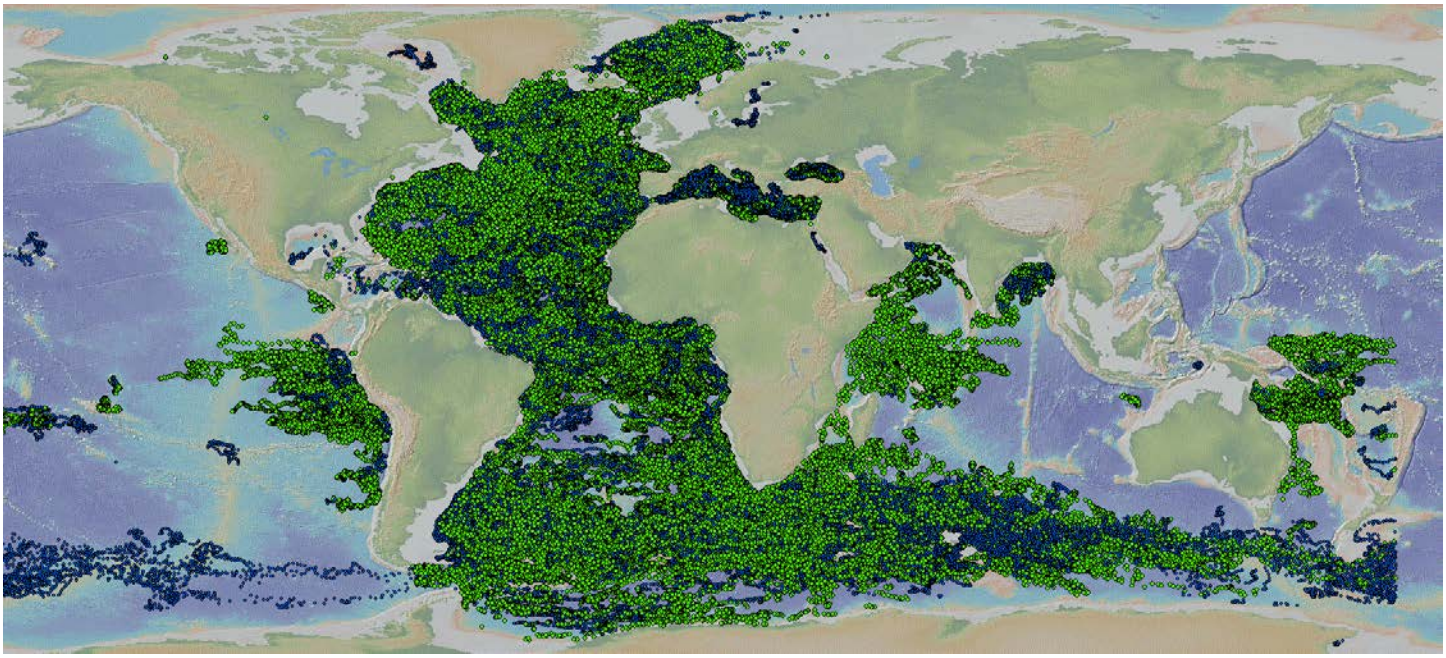
All Coriolis floats, number of profile files on GDAC				
Family	nb floats	nb profiles	RT profiles	DM profiles
APEX	877	101 078	10 735	90 343
METOCEAN	1	52	-	52
NAVIS	3	644	644	
NEMO	162	8 902	4 128	4 774
NOVA	79	5 592	4 477	1 115
PROVOR	1 515	179 303	76 818	102 485
Total	**2 637**	**295 571**	**96 802**	**198 769**







Distribution of Coriolis DAC real-time and delayed mode profiles (96 802 profiles – 198 769 profiles)



Map of real-time profiles and delayed mode profiles
Real time: green dots, delayed mode: blue dots

1.4 Data issued for delayed mode QC

Delayed mode profiles

All profile files are sent to PIs for delayed QC. Most of the Atlantic data handled by Coriolis are checked by the European project Euro-Argo.

Preparation of Argo delayed mode trajectories

The delayed mode trajectories derived from Andro trajectory product are available from:

- ftp://ftp.ifremer.fr/ifremer/argo/etc/coriolis-custom/argo-andro-data/data/dac/coriolis/

The Andro trajectory TRAJ3 files are available for most of the DACs. Each DAC may decide to use these files to provide delayed mode trajectory on GDAC.

Coriolis DAC will use these files as its delayed mode trajectories for old floats versions.

1.5 Delayed mode data sent to GDACs

An Argo delayed mode profile contains a calibrated salinity profile (psal_adjusted parameter).

- A total of **60.598 new or updated delayed mode profiles** was sent to GDACs this year.
- **A total of 198.769 delayed mode profiles** where sent to GDACs since 2005.
  The number of delayed mode profiles increased by 11% this year.

1.6 Web pages

The web site of the French DAC is available at:

- http://www.coriolis.eu.org/Observing-the-Ocean/ARGO

This web page describes all Argo floats:

- http://www.ifremer.fr/co-argoFloats/
    - Individual float description and status (meta-data, geographic map, graphics : section, overlaid, waterfall, t/s charts)
    - Individual float data (profiles, trajectories)
    - FTP access
    - Data selection tool
    - Global geographic maps, GoogleEarth maps
    - Weekly North Atlantic analyses (combines Argo data and other measurements from xbt, ctd, moorings, buoys)

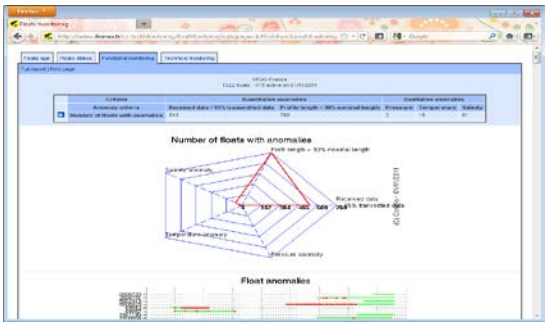
This web page describes all Argo floats interoperability services from Coriolis:

- http://www.coriolis.eu.org/Data-Products/Data-Delivery/Argo-floats-interoperability-services2
    - Display an individual float's data and metadata in HTML or XML format
    - Display all Argo floats, display a group of floats

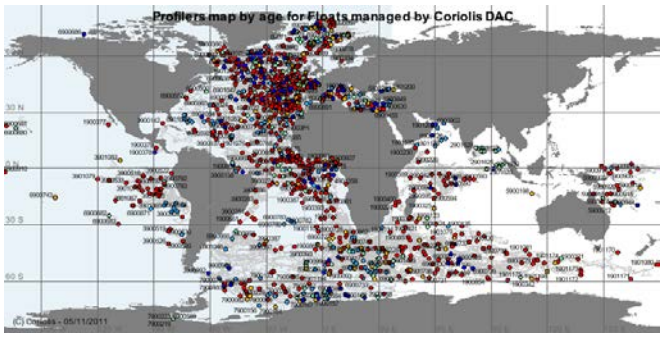
- Argo profiles and trajectories data selection (HTML or XML)
- All individual float's metadata, profile data, trajectory data and technical data
- Argo profiles data on OpenDAP, OGC-WCS and http
- Argo data through RDDAP data server (www.ifremer.fr/erddap)
- Argo data through Oceanotron data server
- Argo profiles data through GCMD-DIF protocol
- Argo data through RDF and OpenSearch protocols
- Display Argo profiles and trajectories with GoogleEarth

Some pages of Coriolis web site are dedicated to technical monitoring:

- http://www.coriolis.eu.org/Data-Products/At-sea-monitoring



Example 1: technical monitoring of Argo-France floats

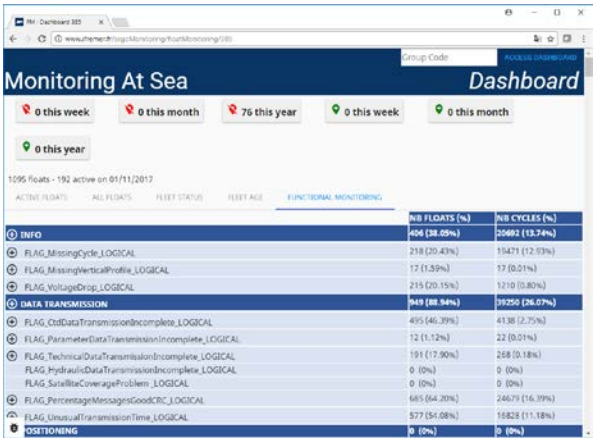


Example 2: age map of floats managed by Coriolis DAC.

Within Euro-Argo ERIC (European Research Infrastructure Consortium), a dashboard with pre-defined alerts on anomalies is operational: http://www.ifremer.fr/argoMonitoring/floatMonitoring/632

For better interactivity, in 2019, the backoffice of the dashboard will be transferred in big data solution : all Argo metadata indexed in Elasticsearch, all data pushed in Cassandra noSQL database.

The front office will adopt and customize the OceanWorks interface developed with NASA-JPL.



Within ENVRIPLUS EU project, an Argo big data demonstration: http://co-discovery-demo.ifremer.fr/euro-argo/
All Argo data is indexed and available through an Elasticsearch API.
More on https://www.youtube.com/watch?v=PKU_JcmSskw

Data centre activity monitoring: Coriolis operators perform an activity monitoring with an online control board.

Fonction	Description	Etat J	Etat J-1	Etat J-2	Etat J-3	Dernière exécution (TU)
CO-05-08-08	Archive GDAC Argo					UNDERWAY-LOCKED 2017-11-25T02:07:01Z
CO-03-07-01	Argo files controler					OK 2017-10-13T08:00:46Z
CO-05-08-11	Argo grey list diffuser					OK 2017-11-24T11:05:02Z
CO-03-07-01-02	Argo stat controler					OK 2017-11-22T01:02:21Z
CO-01-07-08	Collecte Argo Coriolis EDAC					OK 2017-11-25T09:45:04Z
CO-01-07-03	Collecte Argo DAC - FTP					OK 2017-11-25T10:09:04Z
CO-01-07-01-02	Collecte Argo DAC - Table argo index profiles					OK 2017-11-25T09:58:50Z
CO-01-07-01-aoml	Collecte Argo DAC - aoml					OK 2017-11-25T10:00:04Z
CO-01-07-01-bodc	Collecte Argo DAC - bodc					OK 2017-11-25T10:01:02Z
CO-01-07-01-coriolis	Collecte Argo DAC - coriolis					OK 2017-11-25T10:02:21Z
CO-01-07-01-csio	Collecte Argo DAC - csio					OK 2017-11-25T10:03:02Z
CO-01-07-01-csiro	Collecte Argo DAC - csiro					OK 2017-11-25T10:04:03Z
CO-01-07-01-incois	Collecte Argo DAC - incois					OK 2017-11-25T10:05:02Z
CO-01-07-01-jma	Collecte Argo DAC - jma					OK 2017-11-25T10:06:05Z
CO-01-07-01-kma	Collecte Argo DAC - kma					OK 2017-11-25T10:07:03Z
CO-01-07-01-kordi	Collecte Argo DAC - kordi					OK 2017-11-25T10:08:02Z
CO-01-07-01-meds	Collecte Argo DAC - meds					OK 2017-11-25T10:09:03Z
CO-01-07-01-nmdis	Collecte Argo DAC - nmdis					OK 2017-11-25T10:10:02Z
CO-01-07-06-aoml	Collecte Argo DAC BDD - aoml					OK 2017-11-25T09:42:07Z
CO-01-07-06-bodc	Collecte Argo DAC BDD - bodc					OK 2017-11-25T09:42:03Z

www.ifremer.fr/co-tbo/?theme=CORIOLIS&category=COARG&function=CO-01-07-08&date=22/11/2017

Argo GDAC operations monitoring: every working day, an operator performs diagnostics and take actions on anomalies (red or orange smileys)

1.7 Statistics of Argo data usage (operational models, scientific applications, number of National Pis…)

Operational oceanography models; all floats data are distributed to:

- French model Mercator (global operational model)
- French model MARC (regional operational model)
- French model Soap (navy operational model)
- EU Copernicus models (Foam, Topaz, Moon, Noos)
- EuroGoos projects

Argo projects: this year, Coriolis data centre performed float data management for **47 Argo scientific projects and 50 PIs (Principal Investigators).**

List of Coriolis scientific PIs and project names

project	nb floats
euro-argo	2218
coriolis	1126
bsh	493
goodhope	174
naos	146
argomed	136
argo italy	135
remocean	119
awi	84
gyroscope	84
mocca	70
ovide	70
dap	69
argo_spain	59
pirata	59
mocca-eu	55
argo_awi	40
wen	40
ifm-geomar	38
congas	32
flostral	30

List of projects with more than 30 active floats

List of project with less than 30 active floats: argo norway, gmmc, mfstep, argo_fin, argo greece, cirene, pomme, shom, frontalis, ifm, eto_bb, argo spain, flops, cmgp, rrex asfar, tropat, egypt, argo poland, atlantos, sfb460, bioargo, eaims, ifm2, sagar, gmmc_cnes, mersea, rrex, argo_chile, geovide, mocca-germany, amop, asfar, aspex, gmmc ovide, bwr, narval, prosat, soclim, argo geomar, medargo_it, naos-france, outpace, ticmoc, argo bulgary, brazilian navy argo program, dekosim, ge moose, gmmc argomex, mouton, track, argn, cicio, cienperu, mocca-italy, moose, naos-canada, socib, track2010, argo-finland, cnes, gmmc moana maty, hymex, mafia, mocca-poland, norargo, previmer, sri_lanka, vsf, bioargo-italy, e-aims, euroargo, lefe, perseus, shackelton, upsen, weccon, argo brazil, argo_cr, asa, bide, capricorn, eu fp7 hypox, gmmc perle, heracles, medargo, mgoodhope, mocca-eu, dekosim (metu), mooxy, opportunity (sail), plumrho leg 1, proteusmed, argo_lebanon, argo_mexico, argo_poland, calypso, i-p-study, jerico, lov-atlantos, mocca-ned, mocca-netherlands, peacetime, physindien, protevs swot, sojana - antigua to azores

pi	nb floats
birgit klein	241
christine coatanoan	218
virginie thierry	209
sabrina speich	201
pierre-marie poulain	180
holger giese	136
bernard bourles	113
olaf klatt	90
birgit klein	81

rena czeschel	76
andreas sterl	73
fabrizio d'ortenzio	53
herve claustre	52
gerd rohardt	48
pedro velez belchi	42
klaus-peter koltermann	38
xavier andre	34
christophe maes	33
alain serpette	32
detlef quadfasel	31
rosemary morrow	30

List of Principal Investigators (PI) in charge of more than 30 floats

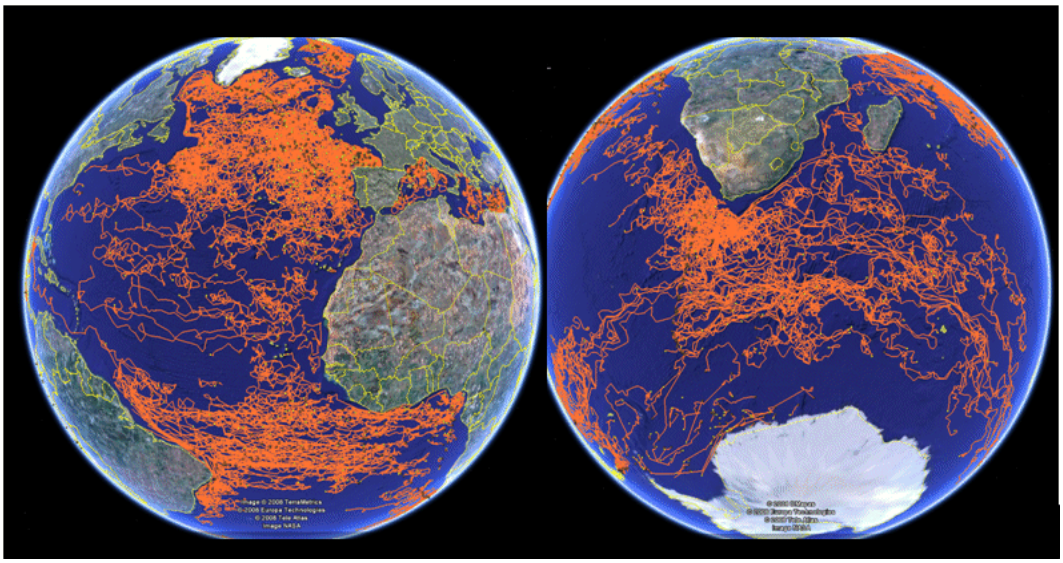
List of Principal Investigators (PI) in charge of less than 30 floats: walter zenk, christine provost, dimitris kassis, laurent coppola, jerome vialard, kjell arne mork, waldemar walczowski, romain cancouët, thierry delcroix, jose lluis pelegri, pedro velez, gerard eldin, fabien durand, antoine poteau, tero purokoski, wilmar van der zwet, isabelle taupier-lepage, franck dumas, jean-baptiste sallee, marcel babin, olaf boebbel, bert rudels, einar svendsen, gregorio parrilla, jens schimanski, osvaldo ulloa, sunke schmidtko, jens meincke, camille daubord, elodie martinez, louis prieur, peter brandt, serge le reste, violeta slabakova, cecile cabanes, fabien roquet, sophie cravatte, alban lazar, bettina fach, luis felipe silva santos, stephane blain, vincent echevin, xavier carton, yves morel, frederic vivier, guillaume maze, marek stawarz / birgit klein, pedro joaquin velez belchi, stephanie louazel, arne kortzinger, gilles reverdin, pascal conan, romain cancouet, sven petersen, thierry moutin, vincent dutreuil et serge le reste, yves gouriou, agus atmadipoera, brian king, christoph kihm, daniel ballestero, hubert loisel, jordi font, josep lluis pelegri, julia uitz, juliet hermes, katrin latarius, louis marie, serguey gladyshev, tobias ramalho dos santos ferreira, anja schneehorst, antje boetius, e. stanev, gerasimos korres, j. haapala, jose luis pelegri, l. tuomi, liu zenghong, louis marié, nathanaele lebreton, ochoa de la torre, sorin balan, velez belchi pedro

1.8 Products generated from Argo data

Sub-surface currents ANDRO Atlas

Based on Argo trajectory data, Michel Ollitrault and the Ifremer team are regularly improving the "Andro" atlas of deep ocean currents. The ANDRO project provides a world sub-surface displacement data set based on Argo floats data. The description of each processing step applied on float data can be found in:

- Ollitrault Michel, Rannou Philippe (2013). **ANDRO: An Argo-based deep displacement dataset**. SEANOE. http://doi.org/10.17882/47077



Argo trajectories from Coriolis DAC are carefully scrutinized to produce the "Andro" atlas of deep ocean currents.

2 Delayed Mode QC

(Please report on the progress made towards providing delayed mode Argo data, how it's organized and the difficulties encountered and estimate when you expect to be pre-operational.)

2.1 Delayed mode operations

At the Coriolis data centre, we process the delayed mode quality control following four steps. Before running the OW method, we check carefully the metadata files, the pressure offset, the quality control done in real time and we compare with neighbor profiles to check if a drift or offset could be easily detected. As each year, we have worked on this way with PIs to strengthen the delayed mode quality control.

Some floats have been deployed from some projects, meaning a lot of PIs and a lot of time for explaining the DM procedure to all of them. A few PIs are totally able to work on DMQC following the four steps but this is not the case for most of them. Since the unavailability of the PIs leads to work by intermittence and then extend the period of work on the floats, we did the work with a private organism (Glazeo) to improve the realization of the DMQC, exchanging only with the PIs to validate results and discuss about physical oceanography in studied area. Working in this way, we largely improve the amount of delayed mode profiles.

For a few projects, there are still no identified operators to do DMQC, for instance the first run has been done by students which have now left institutes or are not available to carry on with this work. We have made a lot of progress with BSH (Birgit Klein) taking into account also floats from other German institutes and OGS (Giulio Notarstefano) for the MedSea.

Some DM files have been updated to format version 3.1 taking into account a new decoder (matlab) developed at Coriolis. This work has been done for some Provor and Apex, few files need to be manually updated.

Regular DM files submission is performed each year but an effort has been done since the year 2017 and following in 2018 to increase the DM files number.

2.1.1 A focus on MOCCA project delayed mode activity

Within the Euro-Argo **MOCCA** project (deployment and processing of 150 floats between 2016 and 2018), RT and DM processing has been organized between the involved partners. Half of the fleet is processed in RT by Coriolis, the other half by BODC using the same processing chain developed by Coriolis.

DMQC is performed by DM operators according to their expertise and the deployment locations. BSH (Birgit Klein, Katrin Latarius), OGS (Giulio Notarstefano), BODC (Matt Donelly) and Ifremer (Gaëlle Herbert, Christine Coatanoan) are highly involved.

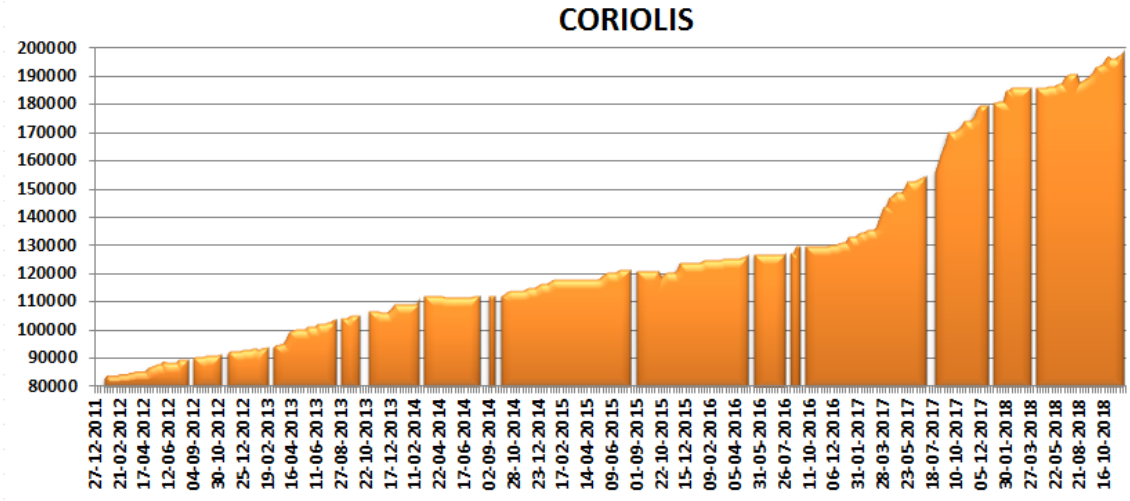
Substantial resources and manpower have been made available through the project to effectively process the floats in both RT and DM, but also to carry on additional activities targeting the improvement of the overall data quality of the Argo dataset, either working on the development of new techniques, improving the reference database or performing pilot/case studies etc. Training of potential new DM operators is also an objective within Euro-Argo and the work has started with the organization of a first European DMQC workshop in April 2018 where 30 people participated.

Here is a non-exhaustive list of Data Management activities performed and financed under the MOCCA umbrella:

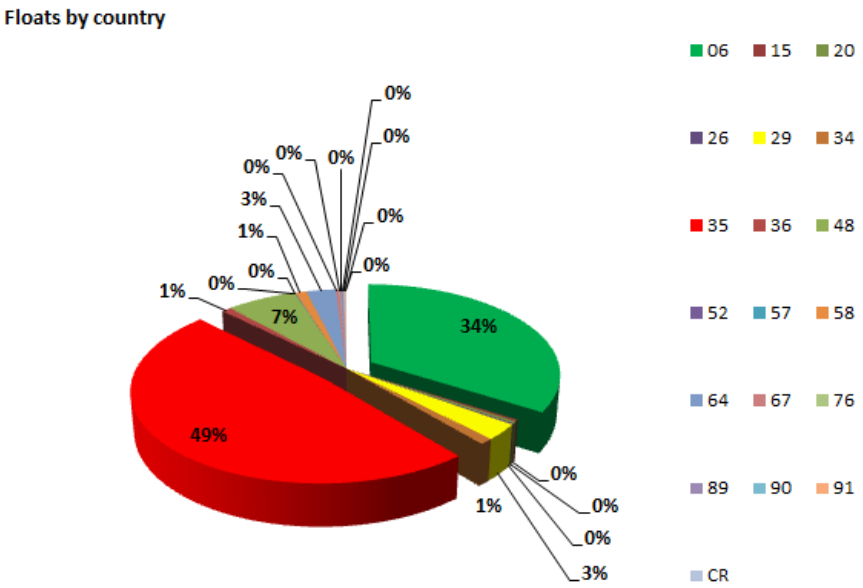
- DMQC of the 150 MOCCA floats
- Update of the reference dataset for the DMQC activity in the Med and Black Sea
- DMQC training for the activity in the Black Sea
- Work on development of some software that can help in the DMQC activity
- Inter-comparison of floats in the Nordic Seas (regional study on data quality)

- Organisation of DMQC in the Baltic sea
- Work on the tuning of ICE-SENSING algorithm for the Nordic Seas
- Work on New RTQC method using MinMax climatology
- Review and development of DMQC training and resources
- Improving under-ice positioning methods in the high-latitude Southern Ocean
- Improving availability of Southern Ocean specific DMQC resources
- Performance assessment of new DMQC method based on machine learning + development of the associated infrastructure

2.1.2 Coriolis delayed mode activity in charts and numbers

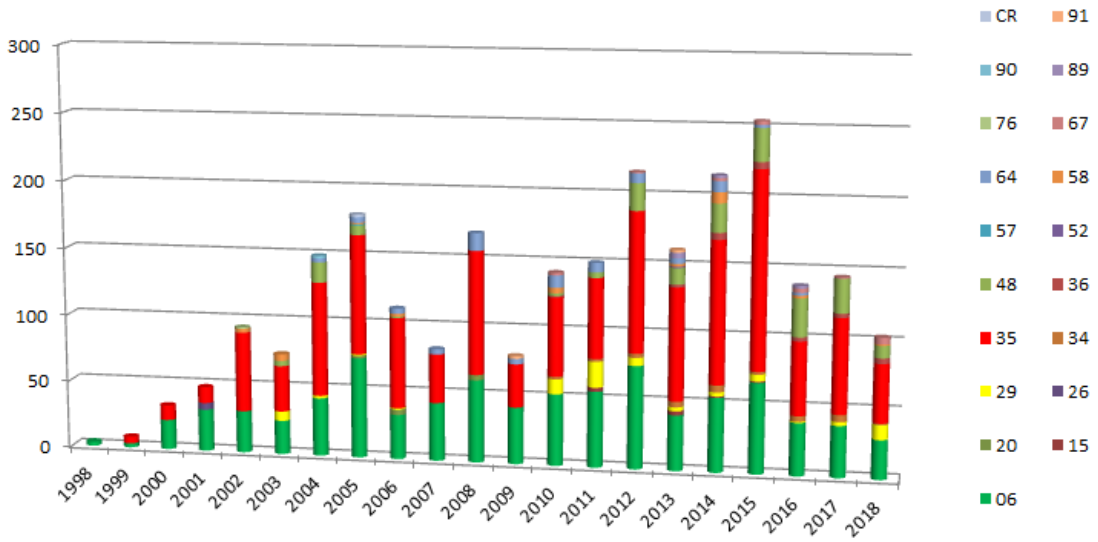


Evolution of the DM profiles' submission versus dates



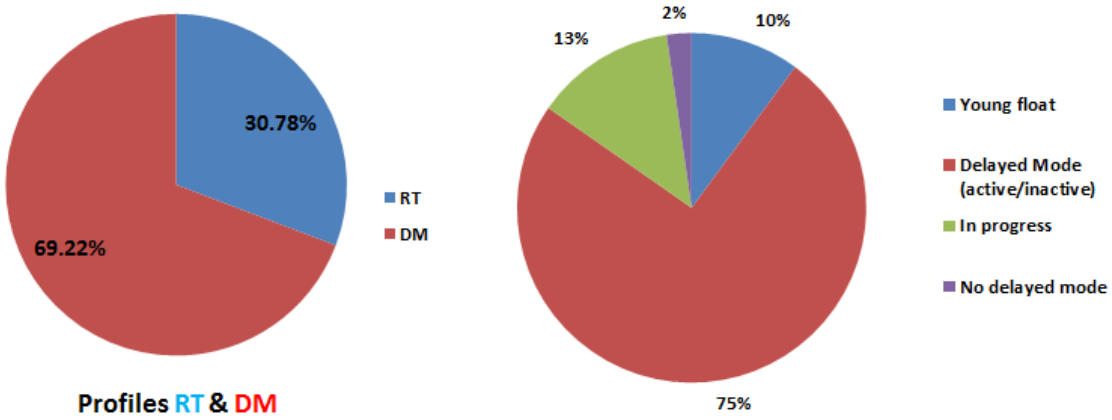
Percentage of floats by country in the Coriolis DAC.

Codes for the countries: 06 : Germany - 15 : Bulgaria - 20 : Chili – 26 : Denmark – 29 : Spain – 34 : Finland - 35 : France – 36 : Greece - 48 : Italy – 52 : Lebanon - 57 : Mexico - 58 : Norway – 64 : Netherlands – 67 : Poland – 76 : China – 89: Turkey - 90 : Russia – 91 : - South Africa - CR : Costa Rica



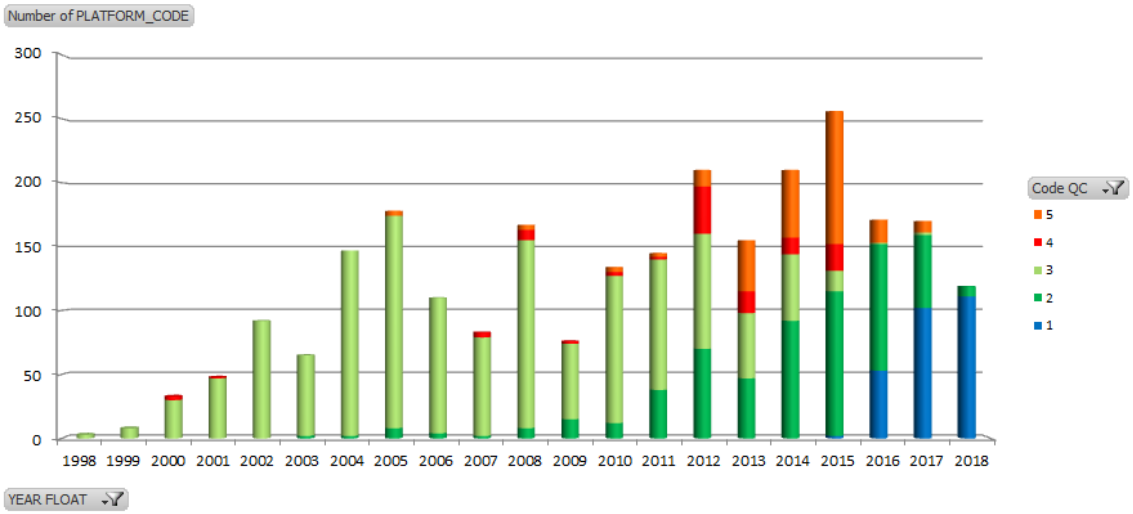
Number of floats by country and by launch's year in the Coriolis DAC

During the last year (from October 2017 to November 2018), 28682 new delayed mode profiles where produced and validated by PIs. A total of 199133 delayed mode profiles where produced and validated since 2005.



Status of the floats processed by Coriolis DAC.
Left: in terms of profile percent and right: in terms of float percent (DM : delayed mode – RT : real time).

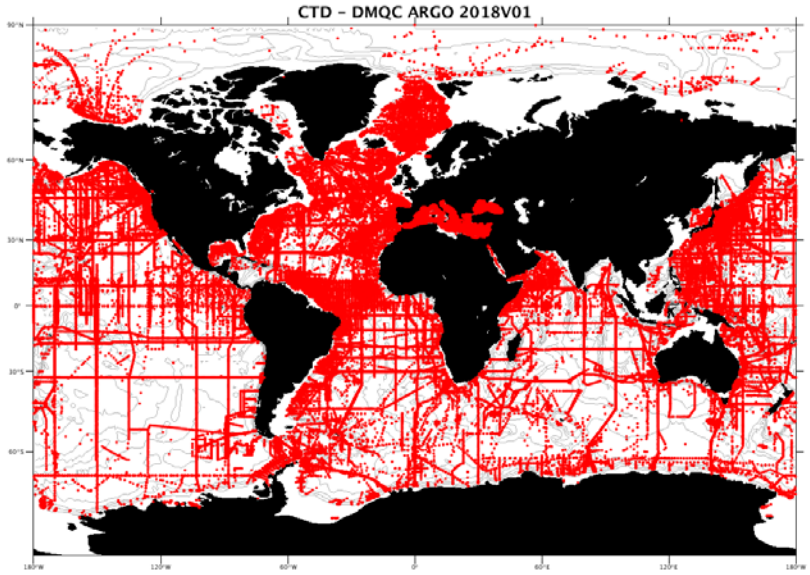
The status of the quality control done on the Coriolis floats is presented in the following plot. For the two last years (2017-2018), most of the floats are still too young (code 1) to be performed in delayed mode. For the years 2012-2013-2014, we are still working on the DMQC of some floats. The codes 2 and 3 show the delayed mode profiles for respectively active and dead floats.



Status of the quality control done on profiles sorted by launch's year, code 1: young float, code 2: active float, DM done, code 3 : dead float, DM done; code 4 : DM in progress, code 5 : waiting for DM, code 6 : problems with float.

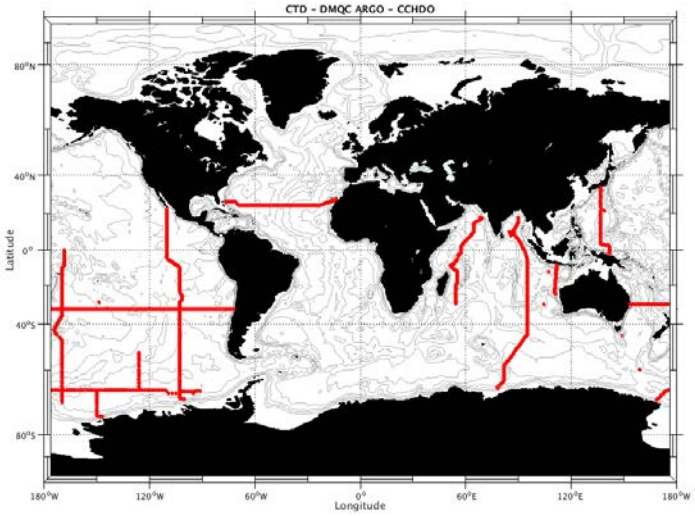
2.2 Reference database

In beginning of 2017, *a* new version 2017V01 has been provided with some updates on a few boxes, following the feedback sent by some scientists. Since March 2018, a new version 2018V01 including OCL updates, CTD from PI, correction from feedbacks is available on the ftp site.



This version is divided in smaller tar balls, one by wmo box area (1-3-5-7): for instance, CTD_for_DMQC_2018V01_1.tar.gz for all boxes starting with wmo 1, then we will have 4 tar files.

New works are in progress and a new version (2018V02) should be delivered by the end of this year. This version will take into account CTD from the GO-SHIP program (data from 2016 to 2018) and downloaded from the CCHDO Website, as well as a few CTD from scientists.


CTD - DMQC ARGO - CCHDO
80°N
40°N
0°
40°S
80°S
Latitude
180°W
120°W
60°W
0°
60°E
120°E
180°W
Longitude

3 GDAC Functions

(If your centre operates a GDAC, report the progress made on the following tasks and if not yet complete, estimate when you expect them to be complete)

- National centres reporting to you
- Operations of the ftp server
- Operations of the www server
- Data synchronization
- Statistics of Argo data usage : Ftp and WWW access, characterization of users (countries, field of interest : operational models, scientific applications) …

3.1 National centres reporting to you

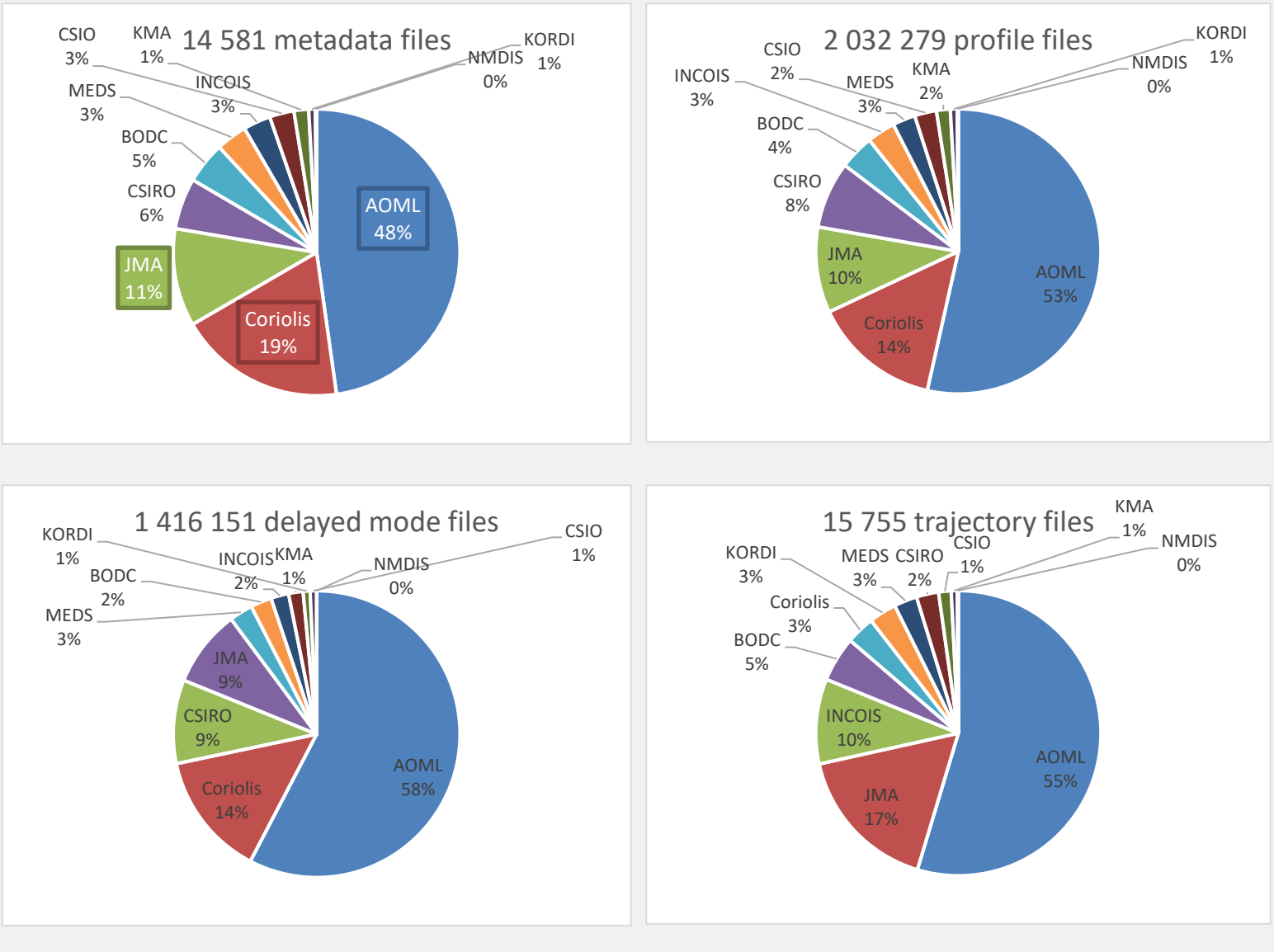
Currently, 11 national DACs submit regularly data to Coriolis GDAC.

The additional GTS DAC contains all the vertical profiles from floats that are not managed by a national DAC. These data come from GTS and GTSPP projects. The GTS profiles are quality controlled by the French DAC (Coriolis).

On November 25th 2018, the following files were available from the GDAC FTP site.

3.1.1 GDAC files distribution

DAC	metadata files 2018	increase	profile files 2018	increase2	delayed mode profile files 2018	increase3	trajectory files 2018	increase4
AOML	6 967	6%	1 087 404	9%	816 536	19%	8 609	7%
BODC	683	7%	79 921	13%	33 680	2%	509	6%
Coriolis	2 748	8%	295 350	12%	198 769	11%	2 661	8%
CSIO	403	9%	50 576	13%	10 221	0%	397	9%
CSIRO	841	4%	153 793	9%	133 816	11%	805	3%
INCOIS	450	7%	65 141	11%	28 418	1%	413	9%
JMA	1 610	4%	197 647	5%	123 006	7%	1 519	2%
KMA	241	6%	31 711	9%	23 094	0%	224	8%
KORDI	110	-8%	15 878	-4%	11 156		107	-10%
MEDS	509	8%	52 398	9%	37 455	18%	492	8%
NMDIS	19	0%	2 460	0%	0		19	0%
Total	14 581	6%	2 032 279	9%	1 416 151	15%	15 755	6%



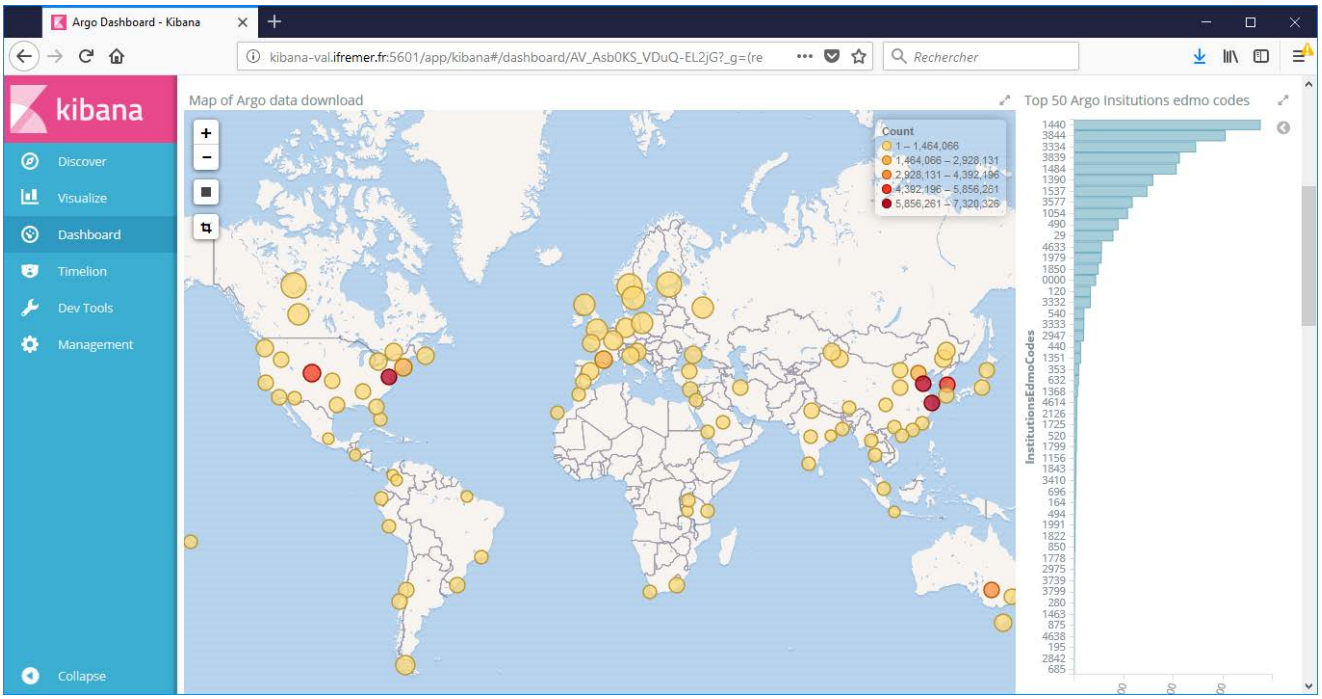
Number of files available on GDAC, November 2018

3.1.2 FTP dashboard: give credit to data providers

Within EU AtlantOS project, Ifremer is setting up a dashboard to monitor data distribution and give credit to data providers such as Argo floats.

FTP downloads log files are ingested in an Elsaticsearch index. A link between downloaded files, download originators, floats included in the downloaded files and institution owners of the floats is performed. These links are displayed in a Kibana dashboard.

This dashboard will offer the possibility to give credit to Floats owner institutions such as how many data from one particular institution was downloaded, by whose data users.



Geographical distribution of GDAC ftp downloads in 2017

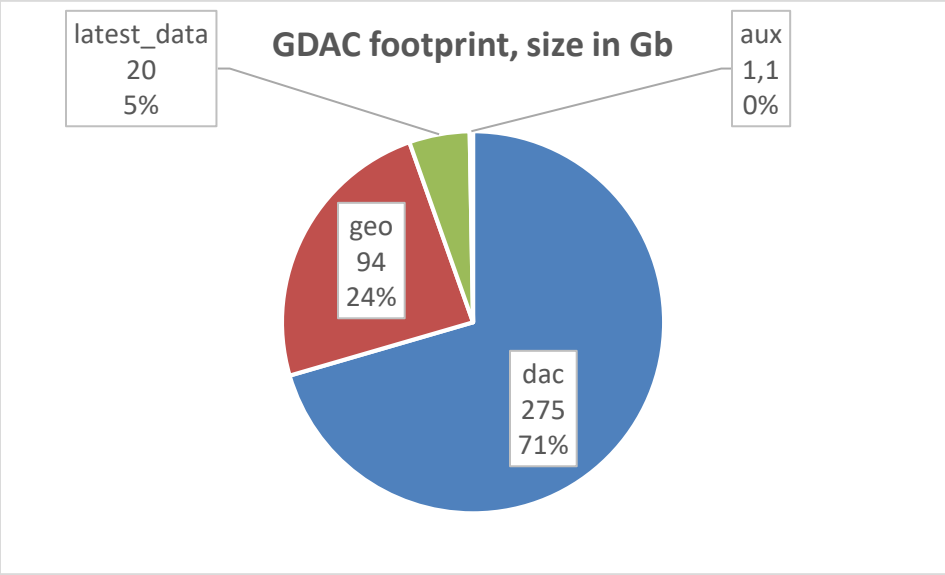
The majority of users (red dots) are located in USA, China, Australia and of course Europe. The right side histogram sorts the floats institution code (1440: PMEL, 3844: WHOI, 3334: INCOIS, 3839: UWA, 1484: CSIRO, …).

The top 50 of floats institutions downloads and the top 50 of data user's

3.1.3 GDAC files size

- The total number of NetCDF files on the GDAC/dac directory was 2 420 372
- The size of GDAC/dac directory was 245 G (+95%)
- The size of the GDAC directory was 553G

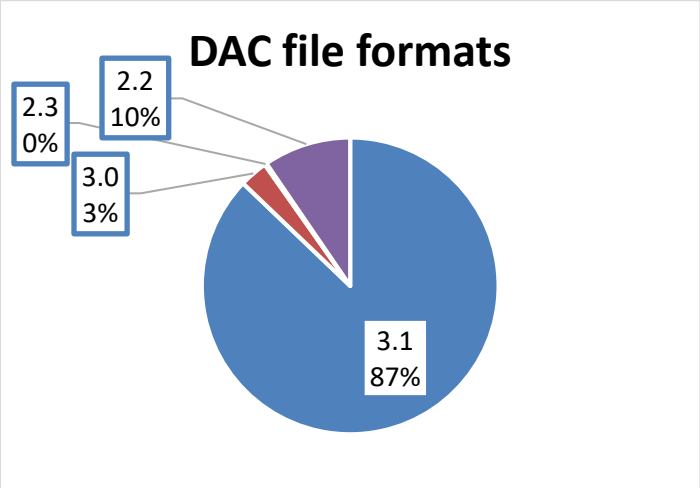
branch	GDAC size in G	since 2016
dac	275	12%
geo	94	11%
latest_data	20	43%
aux	1,1	57%

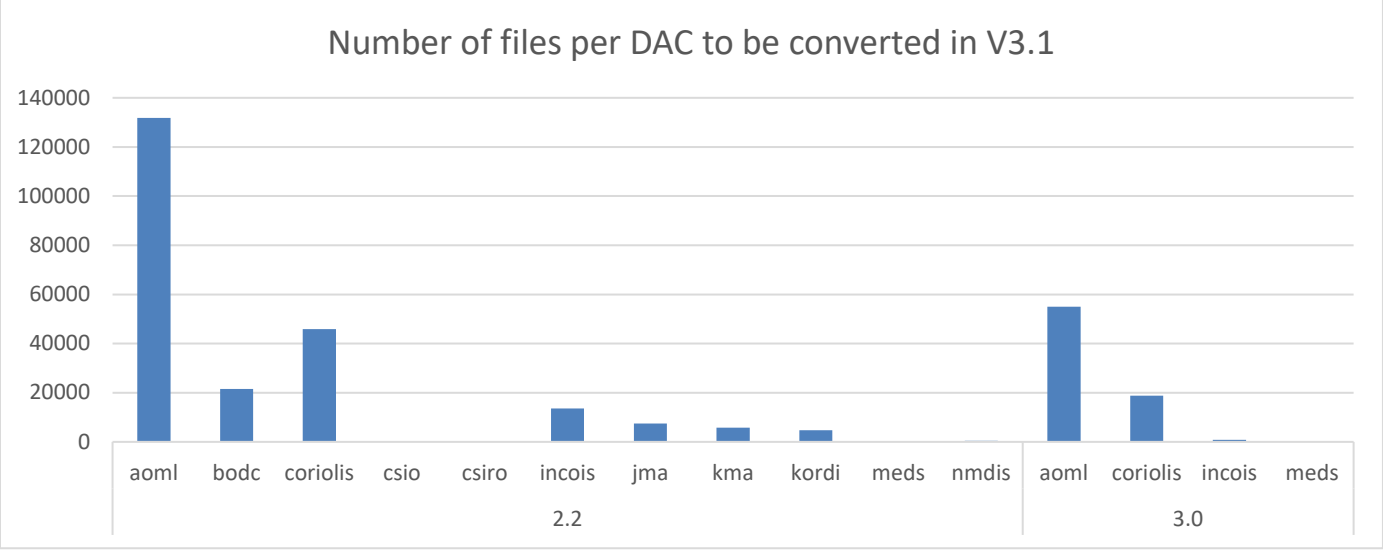


3.1.4 Argo NetCDF transition to format V3.1

The transition from Argo format 2.* and 3.0 toward Argo NetCDF format 3.1 is underway. In 2018, the number of files in format version 3.1 is heading toward 90%.

format version	nb files	percentage
3.1	2 113 833	87%
3.0	74 670	3%
2.3	5 422	0%
2.2	231 497	10%
2.1	12	0%
Total	**2 425 434**	**100%**





File format	number of files
2.2	**231497**
aoml	131785
bodc	21554
coriolis	45916
csio	63
csiro	10
incois	13562
jma	7452
kma	5747
kordi	4674
meds	286
nmdis	448
3.0	**74670**

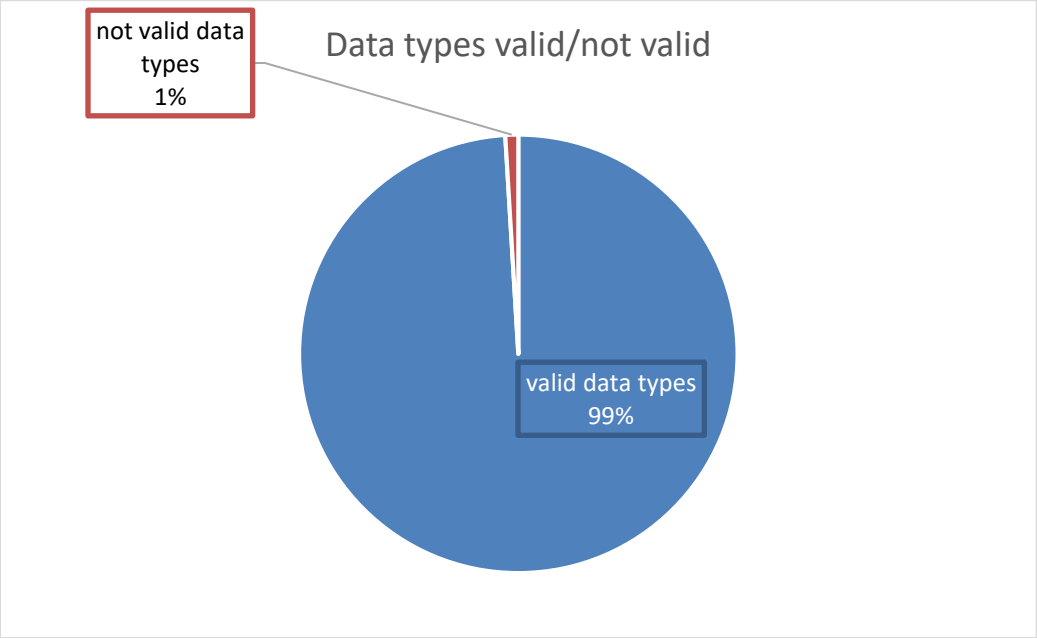
aoml	55061
coriolis	18741
incois	864
meds	4
Total général	**306167**

The files in format version V3.1 are much more homogeneous than their previous versions. The controls applied by the format checker on V3.1 is much more exhaustive. The controlled vocabulary listed in the 27 reference tables is used for V3.1 format checks. A non-valid content is automatically rejected. Only valid V3.1 content appears on GDAC.

Example of valid content checked by the format checker on V3.1 files

There are 8 valid DATA_FORMAT variables listed in reference table 1 (there are 26 more tables…).
A survey on GDAC files shows that 24 779 files (1% of the total) do not have a valid DATA_FORMAT.
The V3.1 files are not affected by this kind of problem.

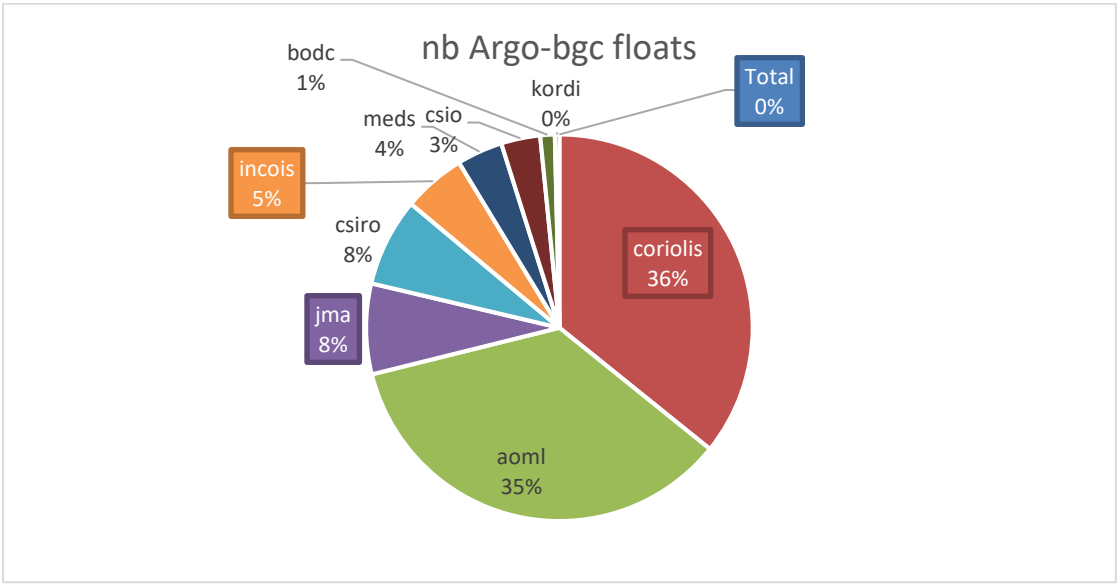
data_type	nb files	valid type
Argo profile	2027964	yes
B-Argo profile	166019	yes
Argo profile merged	165099	yes
ARGO profile	21336	no
Argo meta-data	14581	yes
Argo trajectory	14556	yes
Argo technical data	13639	yes
ARGO trajectory	1093	no
Argo technical	572	no
B-Argo trajectory	445	yes
Argo Trajectory	110	no
ARGO technical data	20	no

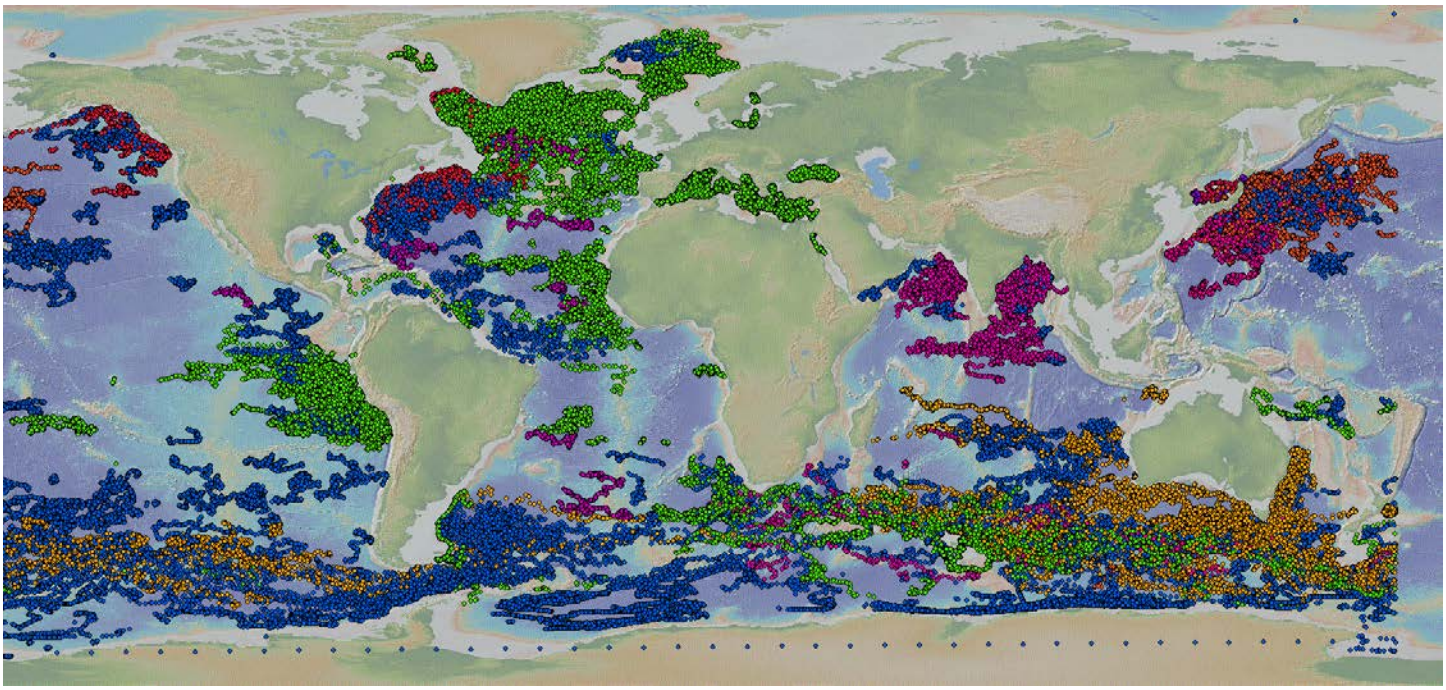

Data types valid/not valid
not valid data types 1%
valid data types 99%

3.1.5 BGC-Argo floats

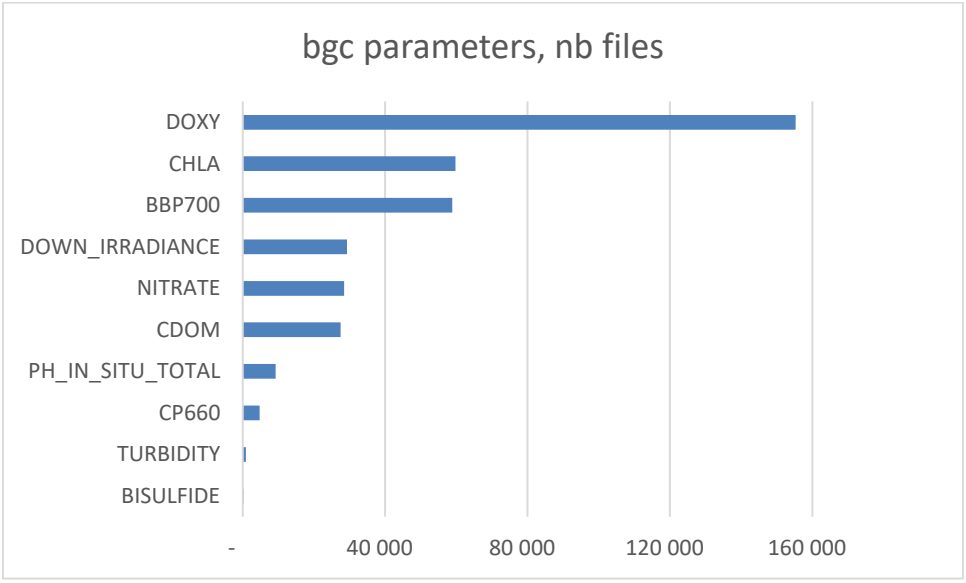
In November 2018, 165 639 BGC-Argo profiles from 1073 floats were available on Argo GDAC. This is a strong increase compared to 2017 : +26% more profiles and +24% more floats.

DAC	nb bio floats	nb bio files
aoml	378	55 176
bodc	13	3 674
coriolis	383	51 981
csio	35	6 784
csiro	79	21 067
incois	56	7 352
jma	81	14 865
kma	3	419
kordi	4	240
meds	41	4 081
Total	**1073**	**165 639**





BGC-Argo profiles, colored by DACs



Main BGC-Argo physical parameters, number of profiles

parameter	nb files
BISULFIDE	255
TURBIDITY	904
CP660	4 722
PH_IN_SITU_TOTAL	9 209
CDOM	27 455
NITRATE	28 475
DOWN_IRRADIANCE	29 259
BBP700	58 852

CHLA	59 756
DOXY	155 309

3.2 Operations of the ftp server

For each individual DAC, every 30 minutes, meta-data, profile, trajectory and technical data files are automatically collected from the national DACs. The 11 DACs are processed in parallel (one process launched every 3 minutes).

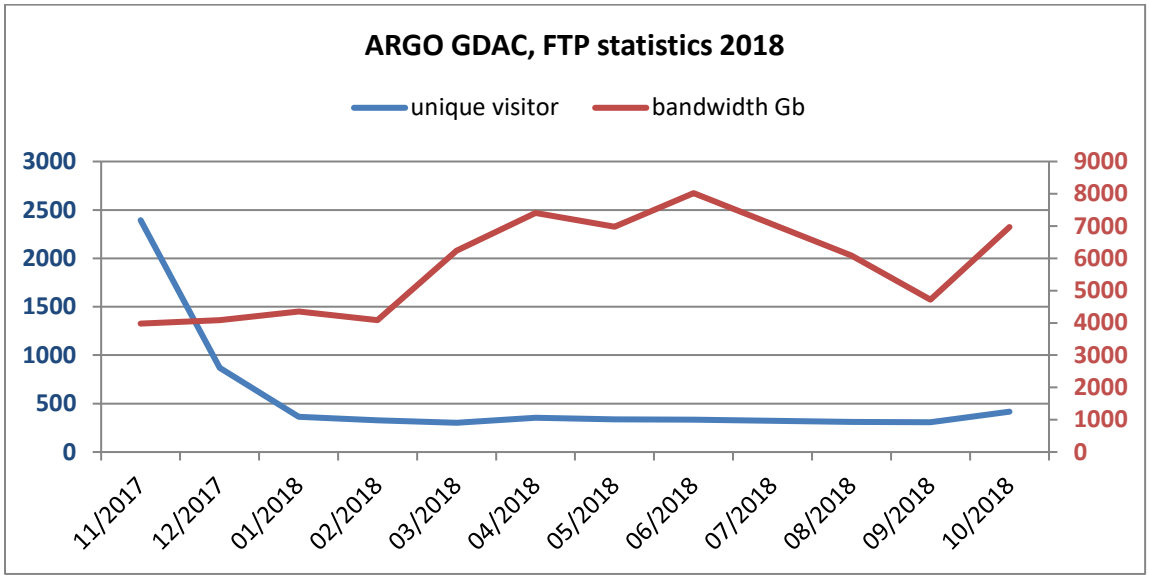
Index files of metadata, profiles, trajectories, technical and auxiliary data are hourly updated.

GDAC ftp address: ftp://ftp.ifremer.fr/ifremer/argo

Statistics on the Argo GDAC FTP server: ftp://ftp.ifremer.fr/ifremer/argo

There is a monthly average of 561 unique visitors, performing 4302 sessions and downloading 5.9 terabytes of data files.

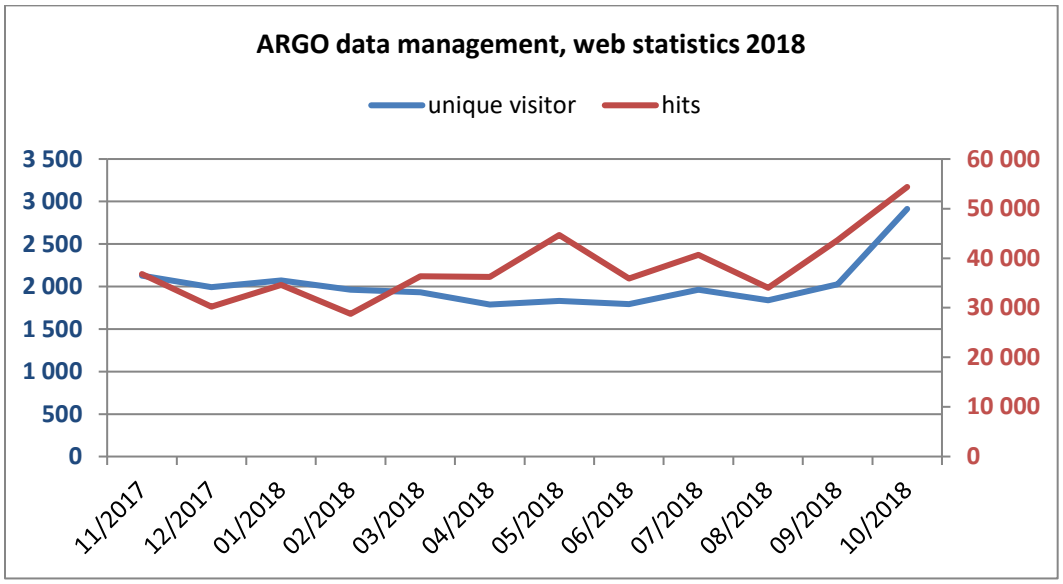
The table below shows an unsusual of visitors in November and December 2017 on GDAC FTP; we do not have a specific explanation.



ARGO GDAC FTP statistics				
month	**unique visitor**	**number of visits**	**hits**	**bandwidth Gb**
11/2017	2394	6 600	2 815 582	3978,83
12/2017	868	5 090	2 933 822	4081,45
01/2018	365	4 246	4 523 428	4352,77
02/2018	328	3 896	2 261 206	4088
03/2018	302	3 551	11 768 691	6238
04/2018	354	4 200	3 943 831	7405
05/2018	339	3 853	4 509 977	6977
06/2018	335	4 281	2 695 216	8020
07/2018	408	4 414	2 932 583	8388
08/2018	312	3 767	3 943 330	6088
09/2018	308	3 519	3 737 580	4718
10/2018	417	4 202	10 758 854	6972
Average	**561**	**4 302**	**4 735 342**	**5 942**

Statistics on the Argo data management web site: http://www.argodatamgt.org

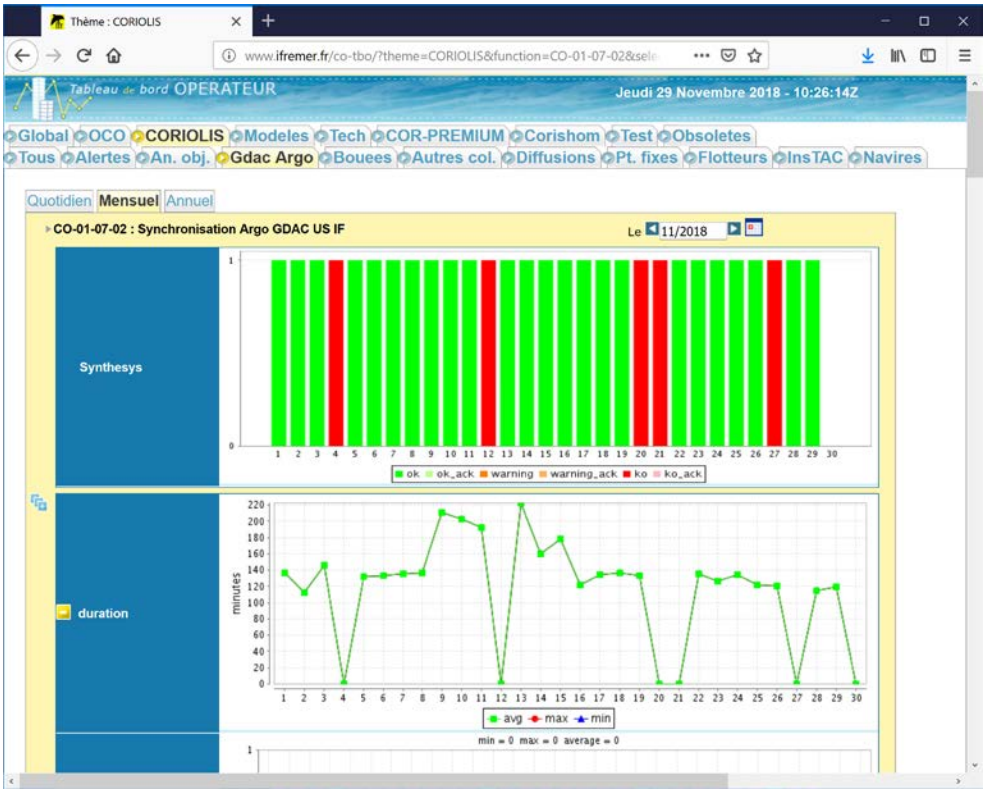
There is a monthly average of 2020 unique visitors, performing 2861 visits and 38029 hits. The graphics shows a slightly increasing number of unique visitors.



ARGO GDAC web statistics					
month	unique visitor	visits	pages	hits	bandwidth Go
11/2017	2 128	3 012	5 380	36 832	1,06
12/2017	1 992	2 693	4 511	30 195	1,12
01/2018	2 072	2 906	5 296	34 603	944,41
02/2018	1 961	2 459	4 006	28 775	1,71
03/2018	1 933	2 718	4 963	36 372	2,88
04/2018	1 788	2 563	4 768	36 218	1,49
05/2018	1 829	2 666	5 733	44 710	2,56
06/2018	1 795	2 585	4 847	35 920	1,52
07/2018	1 963	2 889	5 764	40 707	3,59
08/2018	1 839	2 618	5 278	34 029	1,47
09/2018	2 026	2 921	6 029	43 625	1,57
10/2018	2 913	4 303	7 947	54 367	1,50
Average	**2 020**	**2 861**	**5 377**	**38 029**	**80,41**

3.3 GDAC files synchronization

The synchronization with US-GODAE server is performed once a day at 03:55Z



The synchronization dashboard in November 2018: the daily synchronization time takes on average 2 hours.

You may notice on the dashboard that the synchronization process reported 5 errors in November (red bars):

- "Can't create the ftp connection to usgodae.org"
  There was an ftp connection problem between Coriolis and US GDACs

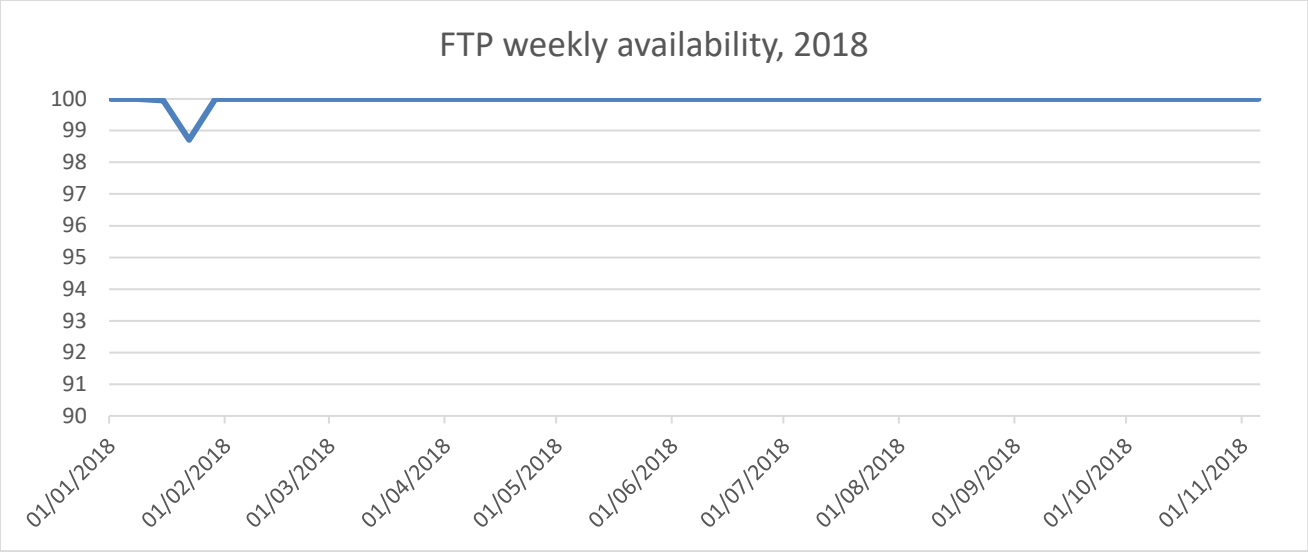
3.4 FTP server monitoring

The Argo GDAC ftp server is actively monitored by a Nagios agent (http://en.wikipedia.org/wiki/Nagios).

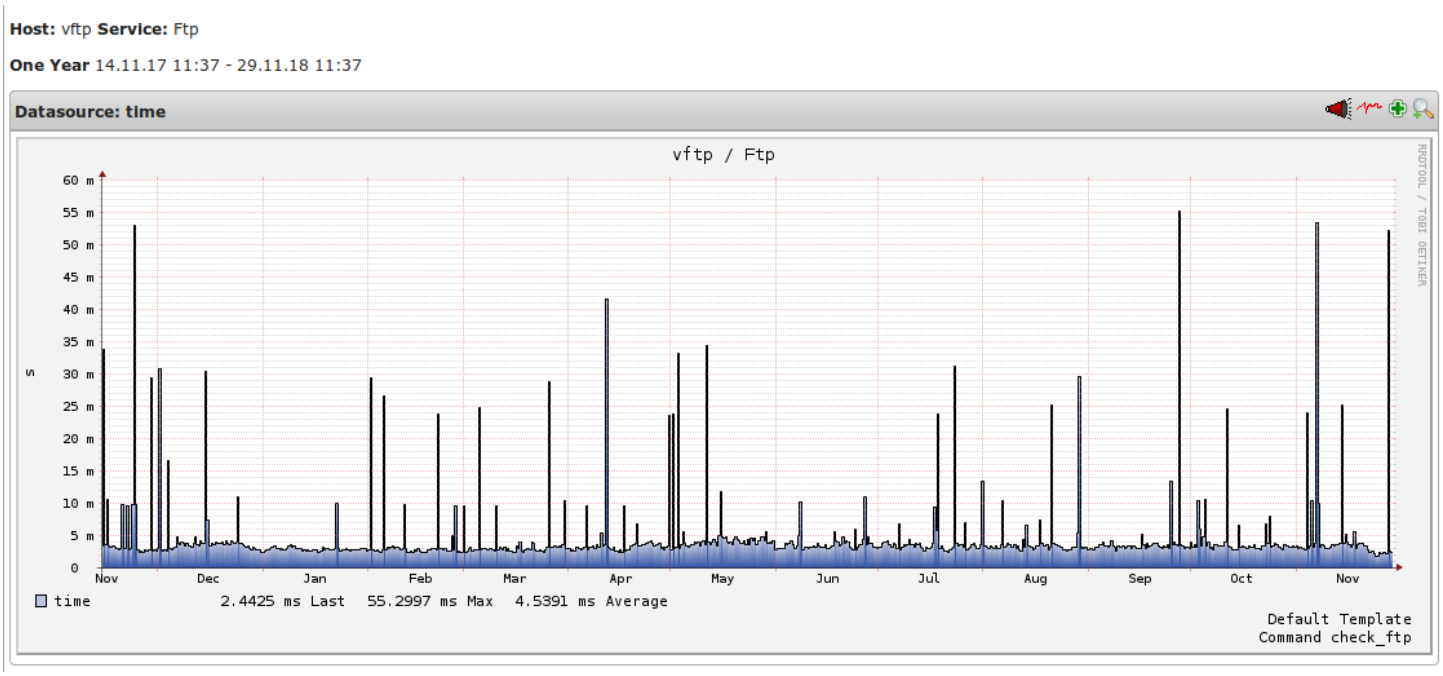
Every 5 minutes, an ftp download test and an Internet Google query are performed. The success/failure of the test and the response time are recorded. The FTP server is a virtual server on a linux cluster.

On the last 11 months, the FTP server was operational on 99.970% of time, non-operational during 14 minutes (0.003%).

FTP server monitoring 01/01/2018 - 29/11/2018			
Status	percentage	duration	comment
OK	99,970%	332d 9h 57s	operational
Warning	0,027%	0d 2h 10m 10s	poor performance
Unknown	0,000%	0d 0h 0m 0s	
Critical	0,003%	0d 0h 14m 50s	non operational

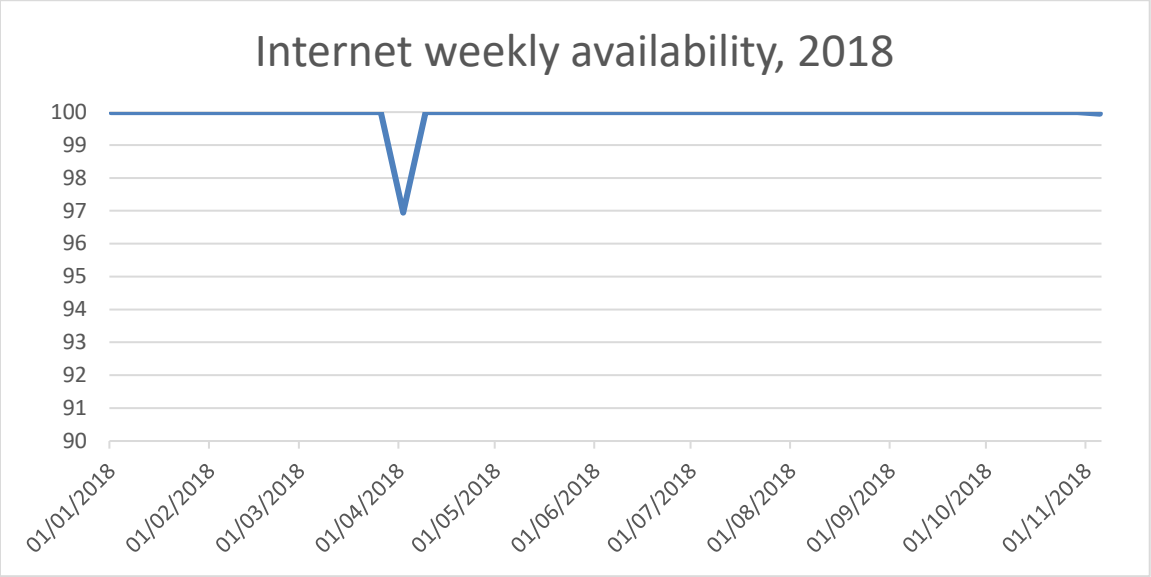


Nagios ftp monitoring: between January and November 2018



FTP server response time monitoring

Internet access monitoring 01/01/2018 - 29/11/2018			
Status	percentage	duration	comment
OK	99,934%	332d 6h 36s	operational
Warning	0,000%	0d 0h 0m 0s	poor performance
Unknown	0,000%	0d 0h 0m 0s	
Critical	0,066%	0d 5h 18m 1s	non operational

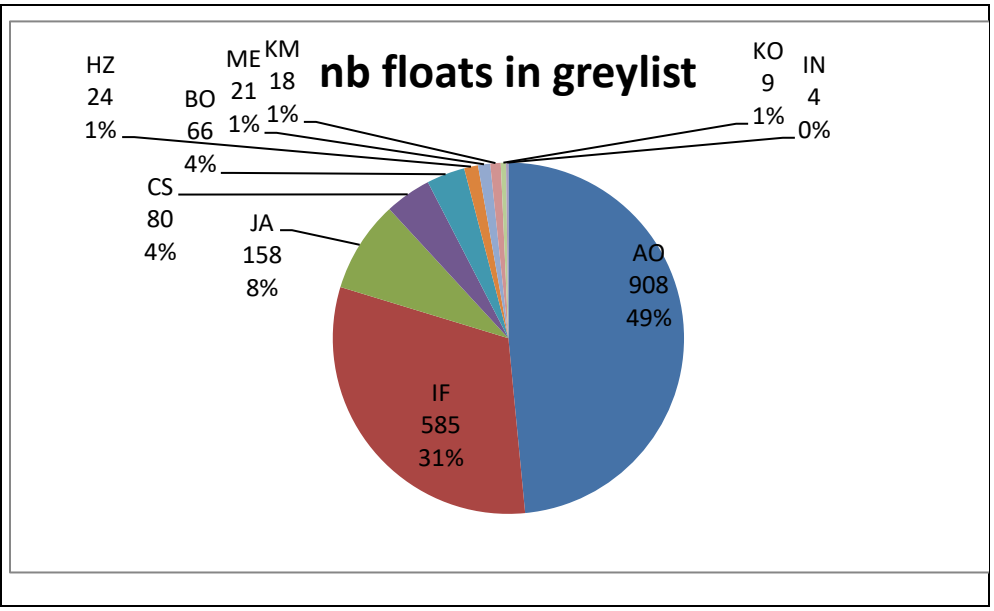


Nagios Internet monitoring: between January and November 2018

3.5 Grey list

According to the project requirements Coriolis GDAC hosts a grey list of the floats which are automatically flagged before any automatic or visual quality control. **The greylist has 1873 entries** (November 29th 2018), compared to 887 entries one year ago. **The 111% increase is noticeable; it can partly be attributed to BGC sensors in greylist.**

DAC	nb floats in greylist
AOML	908
Coriolis	585
JMA	158
CSIRO	80
BODC	66
NMDIS	24
MEDS	21
KMA	18
KIOST	9
INCOIS	4
Total	**1873**



Distribution of greylist entries per DAC and per parameter

AOML reports a high percentage of pressure and temperature in the greylist, compared to other DACs.

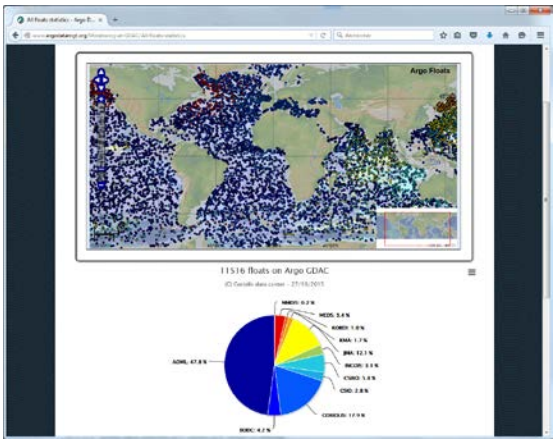
Coriolis reports many BGC greylist entries.

Greylist record	Nb floats
AOML	**908**
PRES	155
PSAL	618
TEMP	135
BODC	**66**
PRES	15
PSAL	34
TEMP	17
CSIRO	**80**
PRES	12
PSAL	52
PSAL	1
TEMP	15
NMDIS	**24**
PRES	3
PSAL	18
TEMP	3
Coriolis	**585**
DOXY	12
PRES	5
PSAL	121
TEMP	18
BBP700	119
CDOM	170
CP660	41
CHLA	84
DOWNWELLING_PAR	3
DOWN_IRRADIANCE380	3
DOWN_IRRADIANCE412	3
DOWN_IRRADIANCE490	3
NITRATE	2
BBP532	1
INCOIS	**4**
PRES	1
PSAL	2
TEMP	1
JMA	**158**
DOXY	1
PRES	24
PSAL	101
TEMP	32
KMA	**18**
PRES	6
PSAL	6
TEMP	6

KIOST	**9**
PRES	3
PSAL	3
TEMP	3
MEDS	**21**
PRES	2
PSAL	16
TEMP	3
Total général	**1873**

3.6 Statistics on GDAC content

The following graphics display the distribution of data available from GDAC, per float or DACs. These statistics are daily updated on: http://www.argodatamgt.org/Monitoring-at-GDAC



3.7 Mirroring data from GDAC: rsync service

In July 2014, we installed a dedicated rsync server called vdmzrs.ifremer.fr described on:

- http://www.argodatamgt.org/Access-to-data/Argo-GDAC-synchronization-service

This server 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)

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/...

3.8 Argo DOI, Digital Object Identifier on monthly snapshots

A digital object identifier (DOI) is a unique identifier for an electronic document or a dataset. Argo data-management assigns DOIs to its documents and datasets for two main objectives:

- Citation: in a publication the DOI is efficiently tracked by bibliographic surveys
- Traceability: the DOI is a direct and permanent link to the document or data set used in a publication
- More on: http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier

Argo documents DOIs

- Argo User's manual: http://dx.doi.org/10.13155/29825

Argo GDAC DOI

- Argo floats data and metadata from Global Data Assembly Centre (Argo GDAC) http://doi.org/10.17882/42182

Argo GDAC monthly snapshots DOIs

- Snapshot of 2018 November 8th http://doi.org/10.17882/42182#59903
- Snapshot of 2014 October 8th http://doi.org/10.17882/42182#42280
- Snapshot of 2012 December 1st http://doi.org/10.17882/42182#42250