A beginners' guide to accessing Argo data

John Gould – Argo Director Emeritus
Megan Scanderbeg – Argo Scientific Coordinator

Argo collects salinity/temperature profiles from a sparse (average 3° x 3° spacing) array of robotic floats that populate the ice-free oceans that are deeper than about 2000m. They also give information on the surface and subsurface currents. Most profiles are made up of about 200 data points, but floats with high speed communications may be sending many more data points given the higher bandwidth. The first Argo floats were deployed in late 1999 and the array contains more than 3600 active floats as of 2014. Argo data are made available to users quickly and free of restriction.

This document is intended to describe to Argo users:

(i) what they should expect from Argo data (real-time and delayed mode streams)
(ii) how to find Argo data

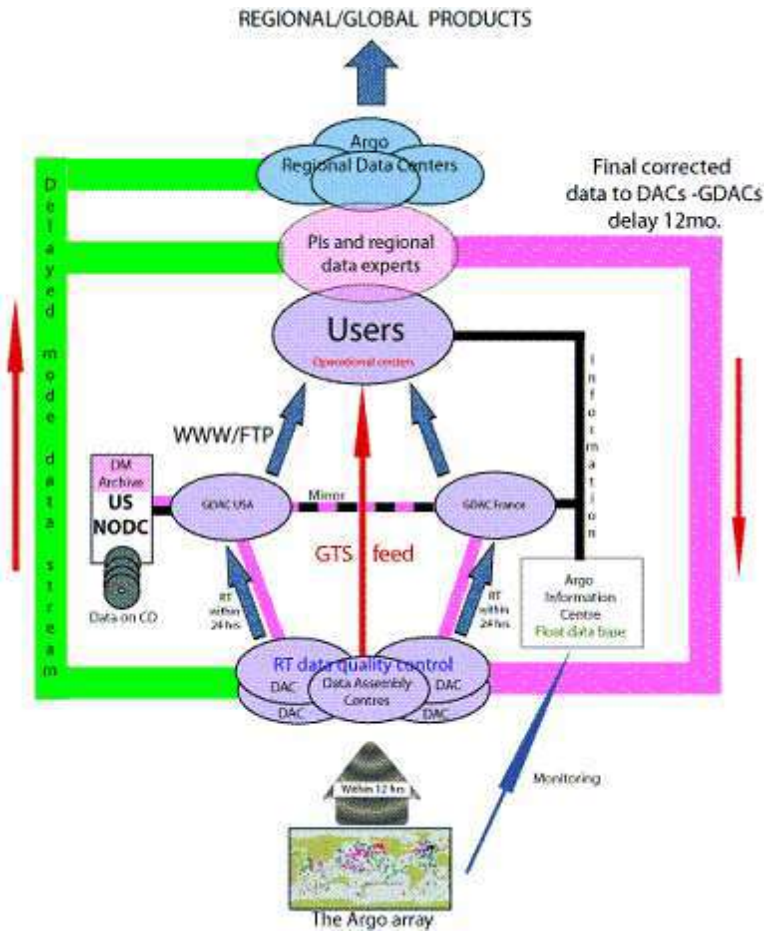
Complete documentation of the Argo data system is on the Documentation section of the ADMT website (http://www.argodatamgt.org). In particular, the "Argo user's manual" and the "Argo quality control manual" may be helpful.

Accessing Argo data

The data are available to users through several routes:

- via the GTS for operational centers (data are in TESAC or BUFR format)
- via DOIs on the GDACs
- via the two Global Data Assembly Centers (GDACs) in NetCDF format
- via Data selection tools on the GDACs
- via gridded fields and velocity products based on Argo NetCDF files from the GDACs
- via data viewers that incorporate Argo data
- via individual float data and plots at Coriolis GDAC
- via the Global Argo Data Repository (GADR) for archived and offline data

The pathways of data flow from the floats to the data centers and users are shown in the following diagram.



Schematic of Argo data flow

Preparation, availability and quality of Argo data

GTS data stream

Data that go out on the GTS are subject to a number of quality checks in national data assembly centers and only those measurements within a profile that pass all tests are inserted onto the GTS. The checks are also used to set quality flags in the data that go on to enter other data streams detailed below.

Currently the data are in both the TESAC and the BUFR format with the expectation to switch entirely to BUFR in the future. Temperature and salinity data are truncated to two decimal places. The vertical co-ordinate is depth not pressure (as measured). No corrections are made to salinities in the GTS data stream. The BUFR format is less restrictive than TESAC and allows for quality flags to be included.

An Argo Grey list of floats is maintained at the GDACs (http://www.usgodae.org/ftp/outgoing/argo/ar_greylist.txt and ftp://ftp.ifremer.fr/ifremer/argo/ar_greylist.txt). The Grey list contains floats which may have problems with one or more sensors. Therefore, data from floats on the Grey list is not sent out on the GTS and should be treated with caution.

Real-time data quality checks (see the Argo Quality Control Manual on the http://www.argodatamgt.org/Documentation page for more details)

These checks will be reordered and all DACs will apply tests in same order to the profile data.

1. Platform ID *
2 Impossible date *
3 Impossible location *
4 Position on land *
5 Impossible speed *
6 Global range test *
7 Regional parameter range*
8 Pressure increasing
9 Spike test
10 Top – bottom spike - obsolete
11 Gradient test
12 Digit rollover
13 Stuck value
14 Density inversion
15 Grey list
16 Gross salinity or temperature drift
17 Visual QC – not mandatory in real time
18 Frozen profile
19 Pressure not greater than Deepest_Pressure +10%

Test 17 is not mandatory in real time. Tests marked * are also applied to trajectory data (see below).

An Argo “Grey list” of floats on which some sensors may have problems is available at the GDACS. Data from floats on the “Grey list” is not sent out on the GTS and should be treated with caution.

Profile data, data flags (see below) and probably trajectory data will at some stage become available using the BUFR format that is less restricted than TESAC.

DOIs on the GDACs

For users interested in using actual Argo NetCDF files in a scientific paper, an Argo DOI is the best option to ensure reproducibility of data and scientific results. There are several **DOIs** associated with Argo data, all of which are described on the Argo Data Management page. Each month, a snapshot is taken of the GDAC (learn more about the **GDACs**) and assigned a DOI. When it is time to write a paper, obtain the most recent monthly DOI, a zipped tarball, and use that as the starting place for all Argo data analysis. There is no need to create a mirror of Argo data via ftp. Instead, the mirror comes in one big file.

Data from GDACs

For users interested in manipulating the actual Argo NetCDF files, the GDACs should be the route to access Argo data. Both GDACs offer access to the complete Argo data collection, including float metadata, detailed trajectory data, profile data and technical data all in NetCDF

format. The http and ftp sites are identical and are organized into three main folders: a "dac" folder which sorts the data by Data Assembly Centre (DAC), a "geo" folder which sorts the data by ocean basin, and a "latest_data" folder which includes the most recent data. There are also several index files in the top directory containing a list of metadata on each type of Argo data file (meta, prof, tech and traj) contained in the "dac" and "geo" folders. It is possible to download these lists and search them for floats in specific regions, times, DACs, etc. There is also a grelist which contains a list of floats that likely have sensor problems. It is important to understand the naming system of the files as well as the variable names and quality control flags within each data file. See the the Argo User's Manual and the Argo quality control manual on the http://www.argodatamgt.org/Documentation page for more information.

Data are available at the GDACs in two versions:

1) **Real time data** is the first form of Argo data available to the public. Because of the requirement for delivering data to users within 24 hours of the float transmitting its profile data, the real time quality control tests are limited and automatic. If data from a float fails these tests, the data will not be distributed onto the GTS but will be on the GDACs as NetCDF files. Real time files on the GDACs all start with "R" (e.g. R5900400_001.nc).

   The real time data should be free from gross errors in position, temperature and pressure. The uncalibrated salinity values are available on both the GTS and at the GDACs. If a salinity offset is known, it may appear as an "adjusted salinity" (PSAL_ADJUSTED) variable on the GDACs. These data are identified with 'R' in the "Data mode" variable if no adjustments were made and are identified with 'A' if an adjustment was made.

   In general these data should be consistent with ocean climatologies even though no climatology tests have been performed. For science applications sensitive to small pressure biases (e.g. calculations of global ocean heat content or mixed layer depth), it is not recommended to use "R" files.

2) **Delayed mode**. These data profiles have been subjected to detailed scrutiny by oceanographic experts and the adjusted salinity has been estimated by comparison with high quality ship-based CTD data and climatologies using the process described by WJO, Böhme and Send or OW[1]. This process is carried out on a 1 year long “data window” and so no“Delayed Mode” observations can be less than 1 year old. These data are appropriate for applications sensitive to small pressure biases. Read below to

[1] Owens, W.B. and A.P.S. Wong, 2009: An improved calibration method for the drift of the conductivity sensor on autonomous CTD profiling floats by theta-S climatology. *Deep Sea Research Part I: Oceanographic Research Papers*, **56**, 450-457.

Böme, L. and U. Send, 2005: Objective analyses of hydrographic data for referencing profiling float salinities in highly variable environments. *Deep Sea Research Part II - Tropical Studies in Oceanography*, **52**, 651-664.

Wong, A.P.S., G.C. Johnson and W.B. Owens, 2003: Delayed-mode calibration of Autonomous CTD profiling float salinity data by Theta-S climatology. *Journal of Atmospheric and Oceanic Technology*, **20**, 308-318.

learn about which variables to use and how to interpret the quality control flags which are vital to pulling the best quality data from the files.

Variable names
It is important to understand the basic variable naming structure within the profile files. There are two versions of temperature, salinity and pressure in each file: a "real time" version (TEMP, PSAL, PRES) and a "delayed mode" version (TEMP_ADJUSTED, PSAL_ADJUSTED, PRES_ADJUSTED). It is recommended that users work with the delayed mode, or adjusted version, if it is filled.

In the "D" files, the adjusted variables are filled with the data that has been corrected after examination by the oceanographic experts, making it the highest quality data for that profile

The quality control flags are:

0	No QC tests have been performed
1	Observation good
2	Observation probably good (implies some uncertainty)
3	Observation thought to be bad but may be recoverable
4	Observation thought to be bad and irrecoverable

Accessing data

The data are held at the GDACs in NetCDF format and the GDACS provide a number of means of selecting data:

- By monthly snapshot – each month a snapshot of all the data at the GDACs is taken and assigned a DOI.
- By the GDAC synchronization service
- By DAC
- By ocean basin
- By data selection tools (http://www.usgodae.org/cgi-bin/argo_select.pl and http://www.argodatamgt.org/Access-to-data/Argo-data-selection)

Which GDAC should I use?

This depends on a number of factors but the first issue is location. North American users will likely use the Monterey site and European users the Coriolis site. Beyond this the choice is a matter of which GDAC has a mode of access that best suits the user's needs.
These access modes are described below:

Data selection tools on the GDACs

Both GDACs also provide data selection tools which allow users to enter different search criteria and then select different ways to display and receive the data. Regardless of the GDAC, users can search for data using the following means of selecting data:

- By selecting a latitude/longitude and time range

- By DAC (Data Assembly Centre)
- By Delayed Mode or Real Time quality

After providing selection criteria at the US GDAC Data Browser, one can expect the following:

- A profile location plot for all profiles returned by the query (may be plotted with or without float ID for queries returning many profiles)
- Download of selected profiles (in netCDFnetCDF Multi-Profile format) as a TAR file
- Plots of T-P and S-P for individual profiles
- Plots of float tracks for individual floats

Image of US GDAC webpage interface

The Coriolis GDAC interface allows one to search by a few additional criteria:

- profile or trajectory file
- measured parameter
- platform type (ie, access to additional non-Argo data (XBT, CTD, drifters, moorings, thermosalinograph,ADCP) is available from Coriolis through the same interface

The results can be downloaded as either NetCDF or ASCII data and a map is displayed with the locations of all the resultant floats.

via the GDAC synchronization service

The rsync server "vdmzrs.ifremer.fr" server provides a synchronization service between the "dac" directory of the GDAC with a user mirror. See the **ADMT website** for more details. 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/...

Other Portals

Argo float data and metadata are available a variety of web services suchs as OpenDAP, OGC-WCS, etc. Click**here** for more information.

Gridded files based on Argo NetCDF files at the GDACs

Argo's more than 3000 floats provide 100,000 plus temperature and salinity profiles each year, but often it is necessary to combine the profiles into a grid to look at different water properties. Additionally, trajectories from Argo floats drifting at depth can be used to calculate velocities which may or may not be gridded to examine the currents at 1000 dbars. There are now several gridded fields and velocity products available on the world wide web based on Argo data, as well as other data sources. The Argo website maintains a list of known files based on Argo data here: http://www.argo.ucsd.edu/Gridded_fields.html.

While each producer makes every effort to remove errors from their product, some might still exist. Therefore, users are reminded that these products are to be used cautiously. Please notify the producer if an error is discovered. Please credit the product appropriately.

Data Viewers that incorporate Argo data

There are a variety of that allow users to look at Argo data (http://www.argo.ucsd.edu/Data_viewers.html). The Global Marine Argo Atlas is a graphical user interface (GUI) that helps users easily look at Argo data and compare it to other global gridded data sets like Reynolds SST and Aviso alitmetry. Users can choose from maps, sections, time series, line drawings and a few products and can enter geographic specifications, time limits, depths, etc. to customize the plot. Postscript and JPGs are available as well as an ascii or NetCDF output of the data plotted. The Atlas is updated monthly for the PC version and is static for the Mac version although users can manually update it each month. (http://www.argo.ucsd.edu/Marine_Atlas.html)

For users interested in the Indian Ocean, there is the Indian Ocean Argo Data Viewing Application. This GUI (http://www.argo.ucsd.edu/incois_ADV.html) offers the option to look at profile data and value added products like temperature, salinity, currents, etc. at various depths in the water column. This viewer works on Windows and covers 2001 - March 2014.

Ocean Data View (ODV) and Java Ocean Atlas (JOA) also support Argo profile data.

There is an Argo Google Earth layer (http://www.argo.ucsd.edu/Argo_GE.html) produced by the Argo Information Centre that shows the positions of all active and inactive floats and makes it possible to plot all the trajectories in a certain region and time range. The balloons that appear when a user clicks on the float location show profile data from the float and make it possbile to quickly look at profiles from floats in various areas of the world ocean.

Both GDACs offer data selection tools described in the GDAC section above.

Individual float data and plots at Coriolis GDAC

For users interested in quickly getting information about individual floats, the Coriolis GDAC offers a description of all floats by WMO ID#. The description includes some meta data about the float, plots of the float's trajectory and data, and links to download individual files in NetCDF or ASCII.

GADR data at US NODC

The US NODC maintains the Global Argo Data Repository (GADR) which operates as a long-term archive for Argo data. The GADR has the responsibility to manage updates to Argo data that are reanalyzed some time later and for which corrections may be applied.
While the GDACs are the main source of Argo data to users with high speed internet access, there will be some who cannot get data in this manner. The GADR can provide alternate means for users to get Argo data.

Argo Regional Centers (ARCs)

Argo has a number of regional centers whose functions include:

- Performing regional analysis of all the Argo data in the region to assess its internal consistency as well as its consistency with recent shipboard CTD data.
- Providing feedback to PIs about the results of the regional analysis and possible outliers.
- Facilitating development of a reference database for delayed mode quality control. This includes assembling the most recent CTD data in their region.
- Preparing and distributing Argo data products on a regular basis. The main data product
will be a consistent Argo delayed mode dataset for their region, but other products might include weekly analyses of temperature, salinity and currents calculated from floats. Documentation of these products will also be provided.

The centers are identified as follows:

Pacific ARC: http://apdrc.soest.hawaii.edu/argo/
North Atlantic ARC: http://www.argodatamgt.org/Argo-regional-Centers/North-Atlantic-ARC
South Atlantic ARC: http://www.aoml.noaa.gov/phod/sardac/index.php
Indian ARC: http://www.incois.gov.in/Incois/argo/argo_dataregional.jsp
Southern ARC: http://www.ukargo.net/southern_ocean_argo_regional_centre/
MedArgo ARC: http://nettuno.ogs.trieste.it/sire/medargo/active/index.php

Tools for assisting with Argo data handling

Some people have difficulty working with NetCDF format files on the Argo GDAC servers. To encourage wider use of Argo data the Argo Information Centre has made an inventory of the tools that are available and links to the sources of this information. The links can be found from UCAR website (http://www.ucar.edu/ucar) or NetCDF documentation (http://www.unidata.ucar.edu/packages/netcdf/index.html).

Here are some **simple Matlab programs** to read in a netCDF Argo profile and trajectory file There are also programs to plot the CTD data from the profile file and the trajectory locations.

Users are encouraged to share the tools they develop with the rest of the Argo Community.