argopy: A Python library for Argo ocean data analysis

Guillaume Maze, Kevin Balem

To cite this version:

Guillaume Maze, Kevin Balem. argopy: A Python library for Argo ocean data analysis. Journal of Open Source Software, 2020, 5, 10.21105/joss.02425 . insu-03683221

The Journal of Open Source Software

argopy: A Python library for Argo ocean data analysis

Guillaume Maze[1] **and Kevin Balem**[1]

1 Univ Brest, Ifremer, CNRS, IRD, LOPS, F-29280 Plouzané, France

DOI: 10.21105/joss.02425

Software

- Review
- Repository
- Archive

Editor: Kristen Thyng

Reviewers:

- @dhruvbalwada
- @DamienIrving

Submitted: 25 June 2020
Published: 01 September 2020

Summary

Argo is a real-time global ocean *in situ* observing system. It provides thousands of highly accurate ocean measurements every day. The Argo dataset has now accumulated more than 2.3 million vertical ocean profiles and accessing it for scientific analysis remains a challenge.

The Argo expert community, focused on delivering a curated dataset of the best scientific quality possible, has never provided its user base with a Python software package to easily access and manipulate Argo measurements: the **argopy** software aims to fill this gap. The **argopy** software can be used to easily fetch and manipulate measurements from Argo floats. It is dedicated to scientists without knowledge of the Argo data management system but is also designed to accommodate expert requirements.

Introduction

The ocean is a key component of the Earth's climate system. It therefore needs continuous real-time monitoring to help scientists better understand its dynamics and to predict its evolution. All around the world, oceanographers have managed to join their efforts and set up a Global Ocean Observing System among which *Argo* is a key component.

Argo is a global network of nearly 4000 autonomous probes measuring pressure, temperature and salinity from the surface to 2000m depth every 10 days. The localisation of these probes is nearly random between the 60^o parallels (see live coverage here). Data from the probes are collected by satellite in real-time, processed by several data centers, merged in a single dataset (comprising of more than 2.3 million vertical profiles as of June 2020) and made freely available to anyone through an ftp server or monthly zip snapshots.

The Argo international observation array was initiated in 1999 and soon revolutionised our perspective on the large scale structure and variability of the ocean by providing seasonally and regionally unbiased in situ temperature/salinity measurements of the ocean interior, key information that satellites can't provide (Riser et al., 2016). The Argo array reached its full global coverage (of 1 profile per month and per 3x3 degree horizontal area) in 2007, and pursues its evolution to fulfil new scientific requirements (Roemmich et al., 2019). Argo data have been used in more than 4000 scientific publications.

This online figure shows the current coverage of the network. It now extends to higher latitudes than the original $\pm 60^o$ and some of the floats are able to profile down to 4000m and 6000m. New floats are also equipped with biogeochemical sensors, measuring oxygen and chlorophyll for instance. All these evolutions of the network increase the total number of floats to nearly 4000. Argo is thus providing a deluge of in situ data: more than 400 profiles per day.

Each Argo probe is an autonomous, free drifting, profiling float, i.e. a probe that can't control its trajectory but is able to control its buoyancy and thus to move up and down the water column as it wishes. Argo floats continuously operate the same program, or cycle, illustrated

Figure 1. After 9 to 10 days of free drift at a parking depth of about 1000m, a typical Argo float dives down to 2000m and then rises back to the surface while profiling - measuring pressure, temperature and salinity. Once it reaches the surface, the float sends by satellite its measurements to a data center, where they are processed in real time and made freely available on the web in less than 24 hours.

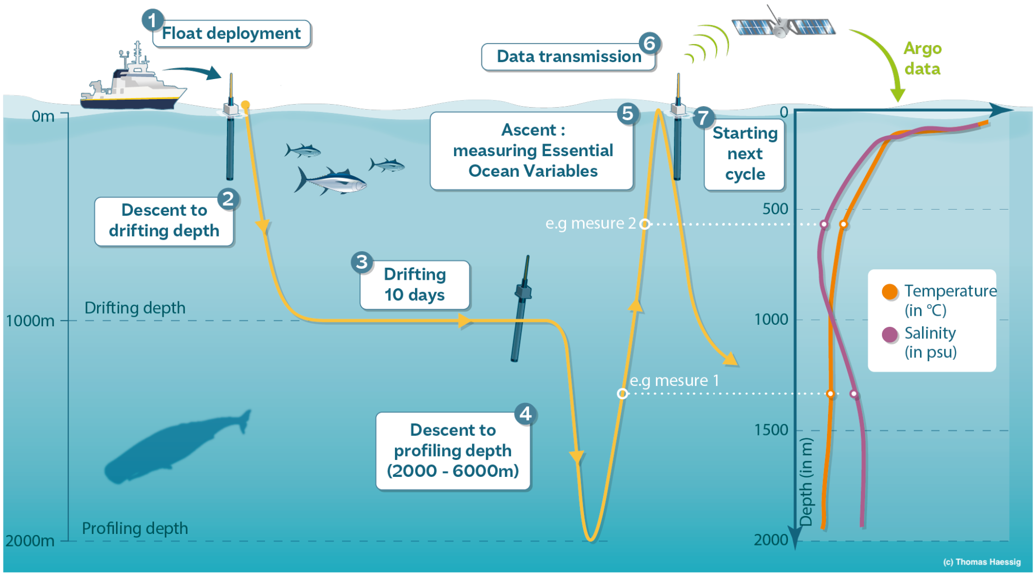

Figure 1: Typical 10 days program, cycle, of an Argo float.

Why argopy ?

For non-expert users of the Argo dataset, it is rather complicated to get access to Argo measurements. Even though data are made freely available on the web, the Argo dataset consists of thousands of files organised using jargon, tens of different variables and many reference tables. The exhaustive Argo user manual is more than 100 pages long, which can be rather intimidating to go through for new users.

This complexity arises from the fact that Argo operates many different models of floats and sensors, quality control of *in situ* measurements from autonomous platforms requires a lot of complementary information (meta-data), and the Argo data management workflow is distributed between more than 10 Data Assembly Centers all around the world. The Argo data management is a model for other ocean observing systems and constantly ensures the highest quality of scientific measurements for the community (Wong & coauthors, 2020).

The result of this tremendous success in data management – in developing good practices and well calibrated procedures (see all the Argo Data Management Team documentation here) – is a very complex Argo dataset: the **argopy** software aims to help users navigate this complex realm.

Since the Argo community focuses on delivering a curated dataset for science, software packages exist for Argo data operators to decode and quality control the data (e.g. Detoc, Thepault, Carval, Mahoudo, & Garo, 2017). However, no open source softwares are available for scientists, who therefore must develop their own machinery to download and manipulate the data.

Python is becoming widely used by the scientific community and beyond: worldwide, and is the most popular and fastest growing language in the last 5 years (20%, source: http://pypl.github.io/PYPL.html). It offers a modern, powerful and open source framework to work with. Since, up to this point, no Python based software has been dedicated to the Argo dataset, it made sense to develop **argopy**.

Key features of argopy

argopy is a python software package that simplifies the process of accessing and manipulating Argo data. The two key features of **argopy** are its trivial fetching API of Argo data and its ability to provide data formatted for both beginner and expert users of Argo.

Data fetching

argopy provides a trivial fetching API of Argo data through a simple call to one of the 3 different ways to look at Argo data: over a space/time domain (with the *region* access point), for one or a list of specific floats (given their unique WMO number with the *float* access point) or for one or a list of float profiles (with the *profile* access point). This is as simple as:

```
from argopy import DataFetcher as ArgoDataFetcher
fetcher = ArgoDataFetcher().region([-75, -45, 20, 30, 0, 100, '2011', '2012'])
ds = fetcher.to_xarray()
```

Here we used **argopy** to fetch data between 75/45W, 20/30N, from 0 to 100db and for the entire year 2011. Once the user has defined what they need (the `fetcher` class instance in the example above), **argopy** will fetch data online and manage internally all the complicated processing of formatting the web request and creating a workable in memory data structure (the `to_xarray()` call above). By default, **argopy** uses the xarray data model; *xarray* is an open source Python package to easily work with labelled multi-dimensional arrays.

Data formatting

argopy aims to thrive in providing Argo data to non-experts. One key feature of **argopy** is the option for selecting a *user mode* that is either `standard` or `expert`. Standard users are those who want to focus on the measurements for scientific analysis; those who do not know, or don't want to be bothered with all the Argo jargon and multitude of variables and parameters.

For standard users (the default mode), **argopy** internally runs a series of processes that curate the raw data and provide a simplified and science focused dataset. For expert users, **argopy** will apply its data model to raw fetched data and return Argo variables that experts users are already used to.

And more

argopy has features to manipulate Argo data, for instance:

- the possibility to transform data from a collection of measurements to a collection of vertical profiles, and vice-versa;
- the possibility to interpolate irregularly sampled measurements onto standard pressure levels.

Another feature is the ability to cache fetched data, so that requests provide users with data much more rapidly, saving bandwidth and time.

Two last important features of **argopy** to mention here are:

- the possibility to fetch data locally, from a user copy of the entire or subset of the Argo database,
- the possibility to fetch only meta data (organised in *index* lookup tables), which allows the user to determine the regional Argo sampling, for instance. These more advanced features may be more of interest for `expert` users, since they require more knowledge of the Argo dataset.

Conclusion

argopy is filling an important gap in the ocean science community by providing an easy way to access a large and complex dataset that has proved to be very important in oceanographic studies. For information on all the features available with **argopy**, the reader is referred to the complete software documentation at https://argopy.readthedocs.io.

Acknowledgements

We acknowledge support from the Euro-Argo ERIC community during the genesis of this project. This software was created with support from the EARISE project, a European Union's Horizon 2020 research and innovation programme under grant agreement no 824131. Call INFRADEV-03-2018-2019: Individual support to ESFRI and other world-class research infrastructures.

References

Detoc, J., Thepault, B., Carval, T., Mahoudo, P., & Garo, M. (2017). Scoop-Argo : Visual quality control for Argo NetCDF data files. Dataset. doi:10.17882/48531

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbeoch, M., Gilbert, D., et al. (2016). Fifteen years of ocean observations with the global Argo array. *Nature Clim. Change*, *6*(2), 145–153. doi:10.1038/nclimate2872

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., et al. (2019). On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array. *Frontiers in Marine Science*, *6*, 439. doi:10.3389/fmars.2019.00439

Wong, A., & coauthors. (2020). Argo 1999-2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Frontiers in Marine Science, section Ocean Observation*, *in press*. doi:10.3389/fmars.2020.00700