Quality Controlled ARGO Profiling Oxygen Data: Open Source Sharing and Version Control Using iPython Notebooks on Github

Carole Sakamoto, Mike McCann, Josh Plant, Ken Johnson
Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing CA USA

OS13B
2038

Abstract

Profiling floats equipped with biogeochemical sensors are an ideal platform for observing the seasonal evolution of physical and chemical processes from the surface to the deep interior. As of August 2015, there have reportedly been 765 floats equipped with oxygen sensors with 363 currently active. Although new insights into oceanic biogeochemical processes have come from these data, the profiling oxygen data set has not been fully utilized because the dissolved oxygen data does not undergo QC procedures at the Argo data centers. The oxygen sensors on profiling floats have been demonstrated to produce highly stable and precise data over many months and years but only a small percentage have been verified with discrete samples taken at deployment to determine the sensor accuracy. Takeshita et al (2013) presented a climatology based quality control procedure utilizing the World Ocean Atlas 2009 (WOA09). We used a straightforward approach mentioned in their paper to calculate the sensor gain by comparing the surface %Sat(float) to the surface %Sat(WOA09) using > 1 year of data. We have developed Python scripts to provide QC'd data that are freely available and easily shared using IPython Notebooks hosted on GitHub. We will examine the consistency of the corrected oxygen data set, relative to existing climatologies, and its use in oceanographic studies.

Typical GitHub workflow. From HTTP://XKCD.COM/1597/

Why iPython Notebooks on GitHub?

- Free and open source (no licensing issues)
- Large community of users = help + many external libraries. stackoverflow.com is your friend.
- iPython Notebooks are an interactive environment that combines code execution, rich text, plots, and other media
- The GIT in GitHub = Version control system to keep versions straight and allow people to easily collaborate
- The HUB in GitHub = the center around which Git revolves is https://github.com where people can store their projects and network with likeminded people
- A platform that can support reproducible research
- The biofloat module and the example notebooks use Pandas. Pandas is a powerful software library written for Python specifically for data manipulation and analysis (especially time series data)

THE PROBLEM:

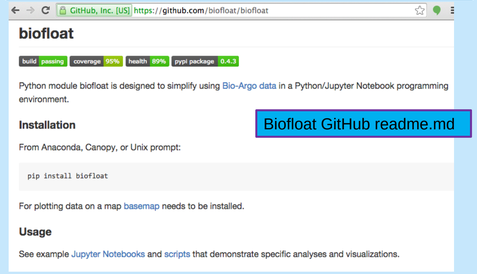
The profiling float oxygen data set is underutilized because it has not been quality controlled and tools for retrieving and visualizing, the data are not readily available

PROJECT OBJECTIVE:

- Develop software tools that the science community can use to do a first order calibration of oxygen data and visualization
- Develop software that is freely available on GitHub allowing for collaborating

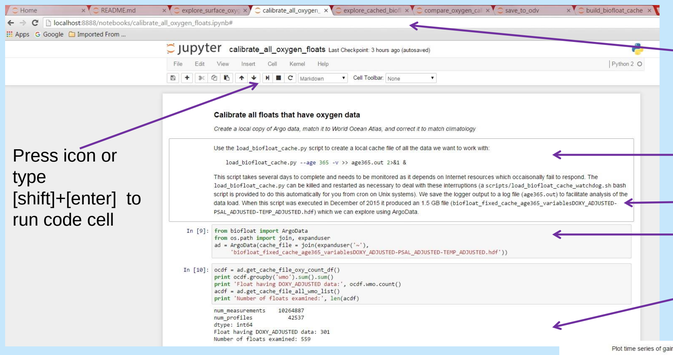
GREAT! HOW DO I USE IT?

- If new to Python, an easy way to install it is to install Anaconda at https:/www.continuum.io (Enthought Canopy is another option). These notebooks and the biofloat module are based on Python 2.7.
- At the Anaconda prompt, type- pip install biofloat The biofloat module contains the functions that do the heavy lifting for getting the data over the web, creating hdf files, and storing the data in Pandas dataframes
- Examples of ipython notebooks using biofloat are available on GitHub. Create a GitHub account at https://github.com and then search for biofloat. After installing biofloat you should be able to run the notebooks by typing ipython notebook at the Anaconda prompt
- After exploring the ipython notebooks, you can then create your own notebooks modified for your uses and you can contribute back to the repository.



Biofloat GitHub readme.md

Example Jupyter Notebooks that have been developed include:
build_biofloat_cache.ipynb
compare_oxygen_calibrations.ipynb
explore_cached_biofloat_data.ipynb
explore_surface_oxygen_and_WOA.ipynb
calibrate_all_oxygen_floats.ipynb
save_to_odv.ipynb



Press icon or type [shift]+[enter] to run code cell

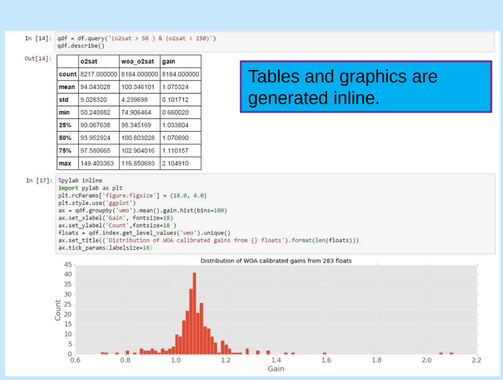
Notebooks run in web browser locally

Markdown cell

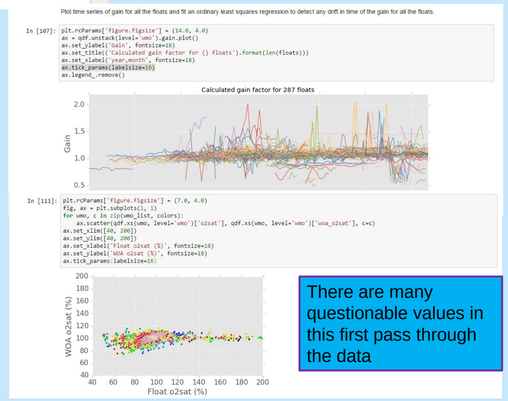
DOXY_ADJUSTED data was used

Code cell

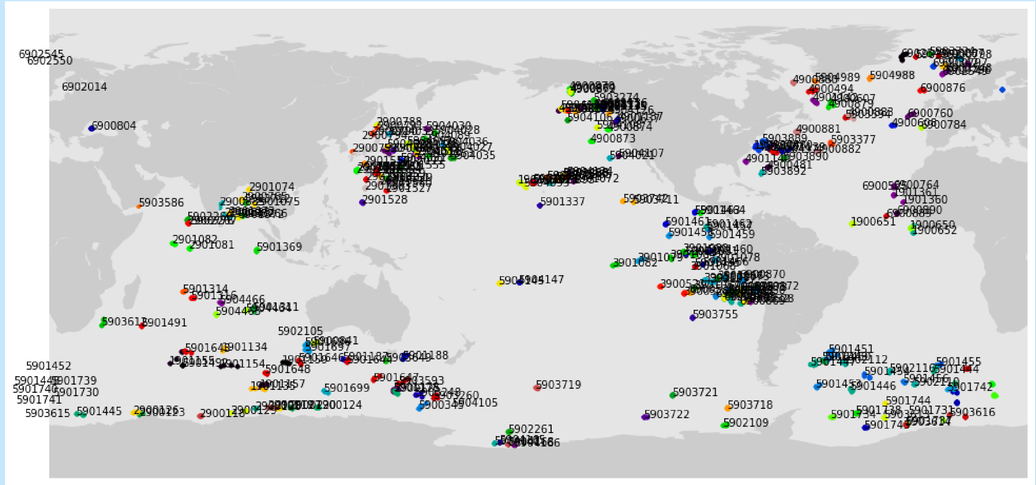
Output from code in above cell



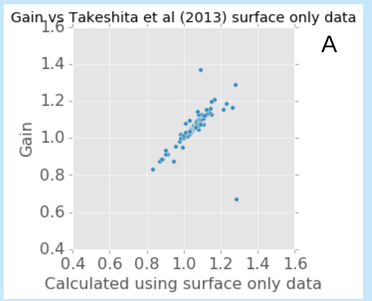
Tables and graphics are generated inline.

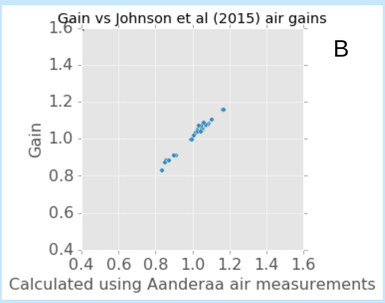


There are many questionable values in this first pass through the data



Map of the locations of the last 4 profiles from each float with DOXY_ADJUSTED oxygen data. Yes, we know that you can't read the float numbers when they overlap but it serves to illustrate the density of the data.





(A) Gain using monthly WOA2013 surface saturation vs gain computed using surface only data calculated by Takeshita et al (2013), and (B) gain using monthly WOA2013 surface saturation vs gain computed using air oxygen measurements with Aanderaa optodes. Wherever possible, air oxygen measurements substantially improve the accuracy of the optode oxygen sensor (Johnson et al. 2015)

Future work

- Can extend the notebooks for use with other bioArgo data
- Develop a community of users contributing notebooks

References

Johnson, K., J. Plant, S. Riser, and D. Gilbert (2015), Air oxygen calibration of oxygen optodes on a profiling float array, *J. Atmos. Oceanic Technol.*, *32*, 2160-2172.

Takeshita, Y., T. R. Martz, K. S. Johnson, J. N. Plant, D. Gilbert, S. C. Riser, C. Neill, and B. Tilbrook (2013), A climatology-based quality control procedure for profiling float oxygen data, *J. Geophys. Res. Oceans*, *118*, 5640-5650.

Acknowledgements

These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System.

the David & Lucile Packard FOUNDATION

Argo