The Journal of Open Source Software

ar-bgc-argo: Jupyter Notebook templates for searching, downloading, and post-processing biogeochemical Argo float time series

Haruto Fujishima[1*] and **Hakase Hayashida**[2*¶]

1 Department of Geophysics, Graduate School of Science, Tohoku University, Sendai, Japan **2** Application Laboratory, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals
Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

ar-bgc-argo is a set of Jupyter Notebook templates that transform raw profiles of biogeochemical Argo (BGC-Argo) floats into "analysis-ready" time series of ocean temperature, salinity, and biogeochemical variables. Users can search for floats based the dates, geographic region, and biogeochemical variables of their interests. After downloading the profile time series of a selected float, `ar-bgc-argo` can visualize, filter, interpolate, and save the post-processed time series as a netCDF file. In addition, `ar-bgc-argo` applies variable-specific treatments and derive additional oceanographic variables using empirical equations. `ar-bgc-argo` is designed to help expand the end users of the growing BGC-Argo float data.

Statement of need

Biogeochemical Argo (BGC-Argo) is a global network of autonomous profiling floats in the ocean that has brought dramatic advances in our understanding of ocean biogeochemistry and marine ecosystems in recent years (Thierry et al., 2025). The global BGC-Argo community aims to cover the global ocean with 1,000 floats that operate every 10 days or so, monitoring biogeochemical properties from the sea surface to 2,000 m deep (Claustre et al., 2020). These profiling floats are equipped with sensors that can measure up to six key variables: chlorophyll-a, pH, oxygen, nitrate, irradiance, and suspended particles (Bittig et al., 2019). As of December 2025, 861 BGC-Argo floats are already in operation, but only 111 of these are equipped with the six full sensors (Figure 1). The collected profiles are made publicly available in the netCDF format within a day or so (Wong et al., 2020).

Despite the growing application of BGC-Argo data, the raw profiles include technical errors and doubtful values because of poor sensor calibration and high sensitivity to noise and artifacts. Furthermore, the data are unfiltered (containing both good- and bad-quality samples) and have inconsistent sampling depths among profiles from a given float. These issues necessitate post-processing prior to scientific analysis, which requires technical knowledge hence, becomes a time-consuming task.

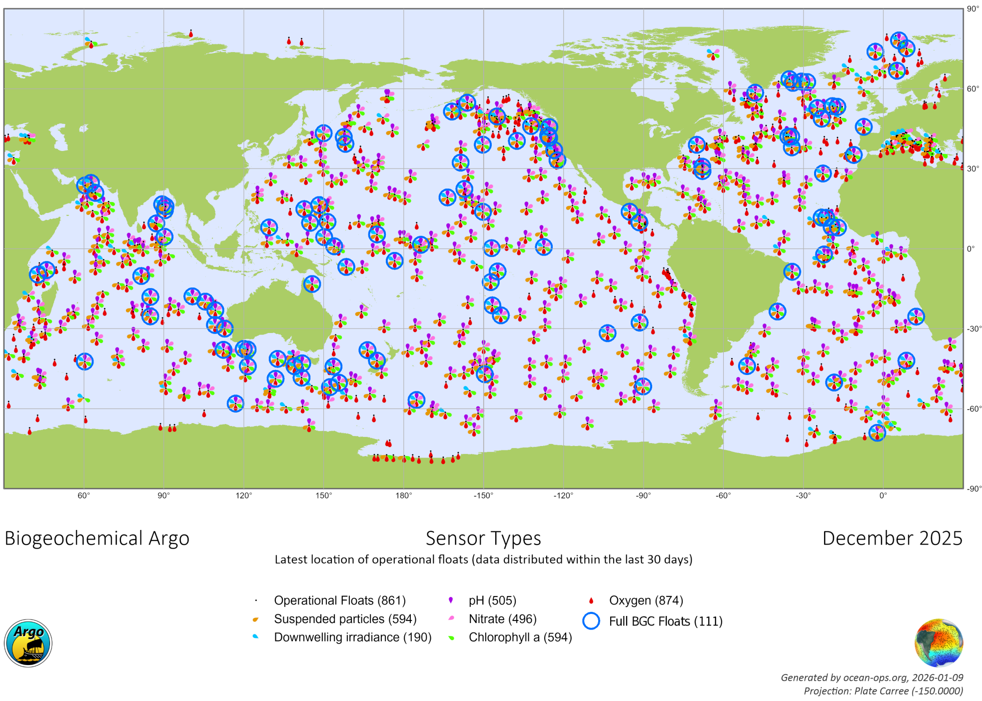

Figure 1: Figure 1: Global coverage of operational BGC-Argo floats as of December 2025 (https://www.ocean-ops.org/share/Argo/Maps/bgc.png; accessed on January 16, 2026).

Overview of ar-bgc-argo

`ar-bgc-argo` consists of three Jupyter Notebook templates: `search.ipynb`; `download.ipynb`; and `generate.ipynb`. To use these templates, users create a copy of the template of their interest, modify its inputs, and run it on their Jupyter environment.

search.ipynb

`search.ipynb` searches for BGC-Argo floats from the synthetic-profile index file (argo_synthetic-profile_index.txt) of the Global Data Assembly Center (GDAC; (Bittig et al., 2019)) and based on the user inputs, including the temporal and spatial coverages and the biogeochemical variables of interest. In addition, `search.ipynb` allows users to narrow down the float selection based on three key criteria:

- mindays: the minimum duration of the data record to ensure sufficient temporal coverage (e.g., at least 365 days).
- minfreq: the minimum sampling frequency required to capture temporal variability (e.g., at least every 14 days).
- maxdrift: the float's maximum drift speed (e.g., 0.05 m/s), which is particularly useful for identifying "quasi-Eulerian" floats suitable for one-dimensional modelling (e.g., (Bruggeman et al., 2024)).

The trajectories of all qualified floats based on the search criteria are drawn on a map and their temporal coverages are visualized on a time series. These visualizations provide an intuitive overview, enabling users to identify potential spatial and temporal biases in observational coverage within the study region of interest prior to data retrieval (Hayashida et al., 2025).

download.ipynb

`download.ipynb` retrieves the concatenated synthetic-profiles time series of a selected float ([wmoid]_Sprof.nc, where [wmoid] is the seven-digit World Meteorological Organization Identifier or WMO ID). While `download.ipynb` naturally follows the selection made in `search.ipynb`, it can also be used independently if the WMO ID of the target float is already known. Upon execution, `download.ipynb` creates a directory named after the WMO ID to store the data. It then identifies the correct file path from the synthetic-profile index file and downloads the data from one of the two GDACs.

generate.ipynb

`generate.ipynb` is the core component of `ar-bgc-argo`, designed to transform raw BGC-Argo profiles into "analysis-ready" time series suitable for immediate scientific application. The data processing workflow consists of eight steps (Figure 2). At every step, diagnostic plots are generated and enable users to visually verify the reliability of the post-processing, which prevents a "black-box" approach and ensures the production of high-quality datasets.

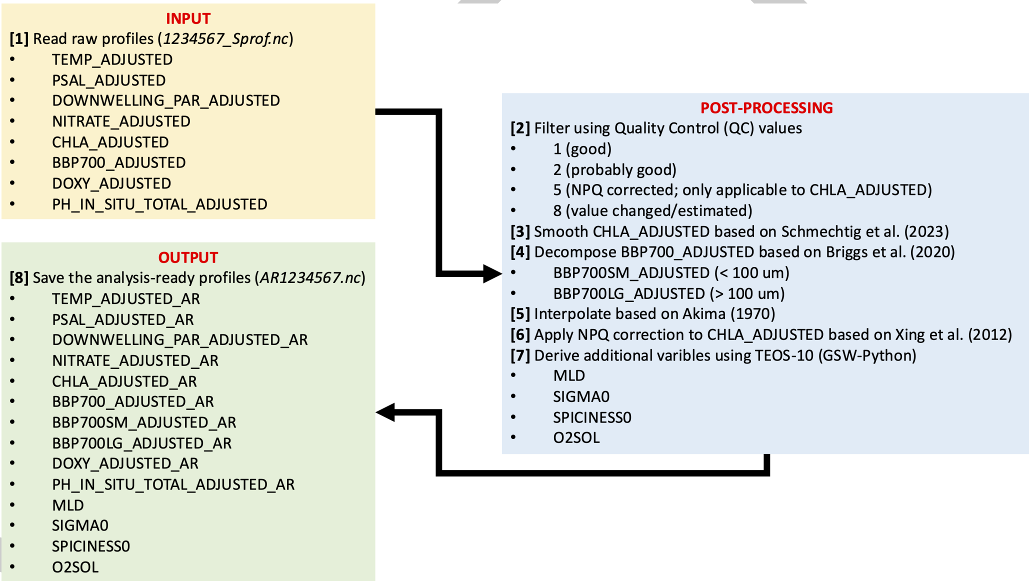

Figure 2: Figure 2: Schematic workflow of the data-processing pipeline implemented in `generate.ipynb`, which involves procedures based on Briggs et al. (2020), Schmechtig et al. (2023), and Xing et al. (2012). This example assumes a full-sensor float with an arbitrary WMO ID (1234567).

Software design

`ar-bgc-argo` was designed to help get started with the BGC-Argo data analysis, while allowing the users to see and modify the source code as they run the Jupyter Notebook templates. Doing so motivates the users to learn the code and customize it for their own application. This approach is structurally different from library-based products such as `argopy` (Maze & Balem, 2020) that are designed to work without needing to see the source code. Furthermore, `ar-bgc-argo` gives immediate access to post-processed data as soon as the data become available on GDACs. This near-real-time capability is advantageous over data archiving such as Johnson et al. (2023), which is updated approximately every six months.

The Journal of Open Source Software

Research impact statement

There are various ways in which `ar-bgc-argo` can be used for oceanographic research. First, `ar-bgc-argo` may be used stand alone to monitor and understand the vertical structure of physical and biogeochemical properties in a specific region of interest on a near-real-time basis, which may help in research expedition planning. Second, `ar-bgc-argo` may be adapted to serve as initial and boundary conditions for one-dimensional ocean models as mentioned above (Bruggeman et al., 2024). Third, `ar-bgc-argo` is a quick and easy tool for assessing the performance of large-scale ocean biogeochemical model simulations as it provides refined float time series that has uniform vertical grids ready for immediate comparison. In fact, `ar-bgc-argo` has been used for assessment of seasonal ocean prediction of chlorophyll-a and nitrate in the equatorial Pacific (Doi et al., submitted). Lastly, `ar-bgc-argo` is intended to contribute to attracting new end users of BGC-Argo data globally, but particularly in Japan, by providing the Japanese language support.

AI usage disclosure

Generative AI tools were used for debugging the Jupyter Notebook templates.

Acknowledgements

BGC-Argo data are freely available through one of the two GDACs. We thank Hidehiro Fujio for computational assistance and Pete Strutton, Clara Vives, Kanako Sato, Yoshimi Kawai for insightful discussion on BGC-Argo floats and feedback on earlier versions of `ar-bgc-argo`. This work was supported by JSPS KAKENHI Grant Number JP24H02226 and the Nakajima Foundation.

References

Bittig, H. C., Maurer, T. L., Plant, J. N., Schmechtig, C., Wong, A. P. S., Claustre, H., Trull, T. W., Udaya Bhaskar, T. V. S., Boss, E., Dall'Olmo, G., Organelli, E., Poteau, A., Johnson, K. S., Hanstein, C., Leymarie, E., Le Reste, S., Riser, S. C., Rupan, A. R., Taillandier, V., … Xing, X. (2019). A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.0050

Briggs, N., Dall'Olmo, G., & Claustre, H. (2020). Major role of particle fragmentation in regulating biological sequestration of CO2 by the oceans. *Science*, *367*, 791–793. https://doi.org/10.1126/science.aay1790

Bruggeman, J., Bolding, K., Nerger, L., Teruzzi, A., Spada, S., Skákala, J., & Ciavatta, S. (2024). EAT v1.0.0: A 1D test bed for physical–biogeochemical data assimilation in natural waters. *Geoscientific Model Development*, *17*, 5619–5639. https://doi.org/10.5194/gmd-17-5619-2024

Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the Global Ocean with Biogeochemical-Argo. *Annual Review of Marine Science*, *12*, 23–48. https://doi.org/10.1146/annurev-marine-010419-010956

Doi, T., Baba, Y., & Hayashida, H. (submitted). Seasonal chlorophyll-a prediction in the tropical pacific with a global climate model incorporating marine biogeochemistry (SINTEX-F2bio). *Journal of Climate*.

Hayashida, H., Strutton, P. G., Hashihama, F., Horimoto-Miyazaki, N., Katano, T., & Yasui-Tamura, S. (2025). Status and prospects of biogeochemical Argo observations of chlorophyll-a in the northwest Pacific. *Journal of Oceanography*, *81*, 289–299. https://doi.org/10.1007/s10872-025-00750-2

Johnson, K. S., Fassbender, A., Gray, A., Nicholson, D., Purkey, S., Riser, S. C., Takeshita, Y., Talley, L. D., Wijffels, S. E., Gilson, J., Grady, L. A., Guisewhite, N., Maurer, T. L., Parise, K., Plant, J. N., Robbins, P., Rupan, R. A., & Swift, D. D. (2023). *Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) and Global Ocean Biogeochemistry (GO-BGC) Biogeochemical-Argo Float Data Archive*. UC San Diego Library Digital Collections. https://doi.org/10.6075/J0SJ1KT8

Maze, G., & Balem, K. (2020). Argopy: A Python library for Argo ocean data analysis. *Journal of Open Source Software*, *5*(53), 2425. https://doi.org/10.21105/joss.02425

Schmechtig, C., Claustre, H., Poteau, A., D'Ortenzio, F., Schallenberg, C., Trull, T., Xing, X., Poteau, A., D'Ortenzio, F., Schallenberg, C., Trull, T., & Xing, X. (2023). *Biogeochemical-Argo quality control manual for chlorophyll-a concentration and chl-fluorescence, Version 3.0* (Version 3.0). IFREMER. https://archimer.ifremer.fr/doc/00243/35385/

Thierry, V., Claustre, H., Pasqueron de Fommervault, O., Zilberman, N., Johnson, K. S., King, B. A., Wijffels, S. E., Bhaskar, U. T. V. S., Balmaseda, M. A., Belbeoch, M., Bollard, M., Boutin, J., Boyd, P., Cancouët, R., Chai, F., Ciavatta, S., Crane, R., Cravatte, S., Dall'Olmo, G., … Zuo, H. (2025). Advancing ocean monitoring and knowledge for societal benefit: The urgency to expand Argo to OneArgo by 2030. *Frontiers in Marine Science*, *12*. https://doi.org/10.3389/fmars.2025.1593904

Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., … Park, H.-M. (2020). Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats. *Frontiers in Marine Science*, *7*, 700. https://doi.org/10.3389/fmars.2020.00700

Xing, X., Claustre, H., Blain, S., D'Ortenzio, F., Antoine, D., Ras, J., & Guinet, C. (2012). Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the Kerguelen region (Southern Ocean). *Limnology and Oceanography: Methods*, *10*(7), 483–495.