

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

Haruto Fujishima  <sup>1\*</sup> and Hakase Hayashida  <sup>2\*¶</sup>

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

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Open Journals](#) ↗

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#))

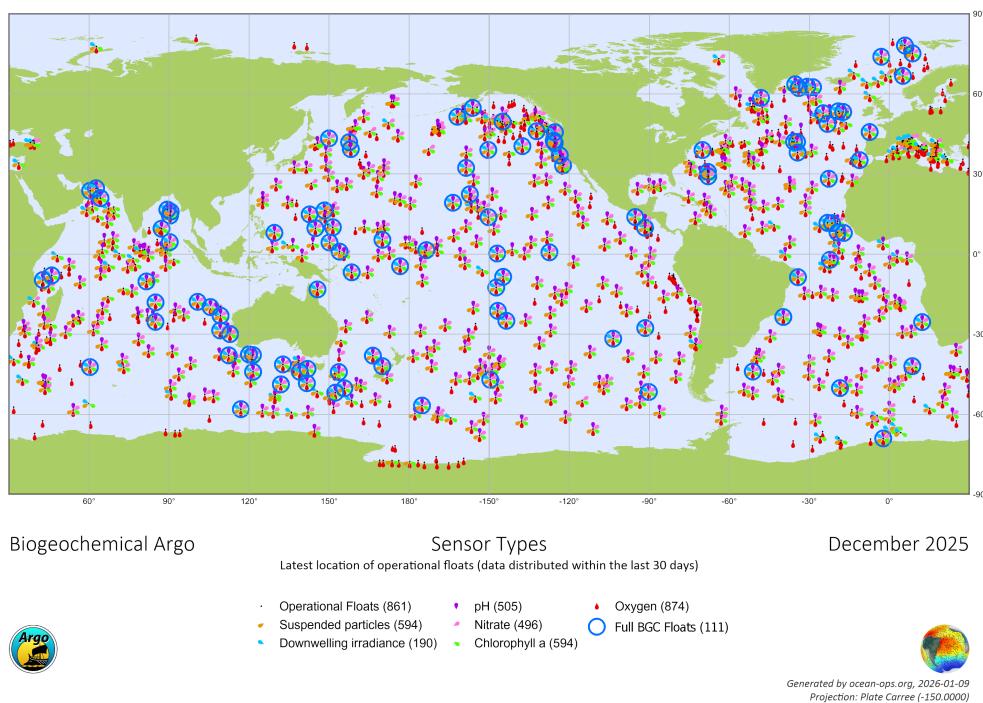
## Summary

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:** 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).

## 34      Overview of ar-bgc-argo

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

### 38      search.ipynb

39      search.ipynb searches for BGC-Argo floats from the synthetic-profile index file (argo\_synthetic-  
 40      profile\_index.txt) of the Global Data Assembly Center (GDAC; [Bittig et al., 2019](#)) and based  
 41      on the user inputs, including the temporal and spatial coverages and the biogeochemical  
 42      variables of interest. In addition, search.ipynb allows users to narrow down the float selection  
 43      based on three key criteria:

- 44      ■ mindays: the minimum duration of the data record to ensure sufficient temporal coverage  
   (e.g., at least 365 days).
- 45      ■ minfreq: the minimum sampling frequency required to capture temporal variability (e.g.,  
   at least every 14 days).
- 46      ■ 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](#)).

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

## 55 `download.ipynb`

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

## 63 `generate.ipynb`

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

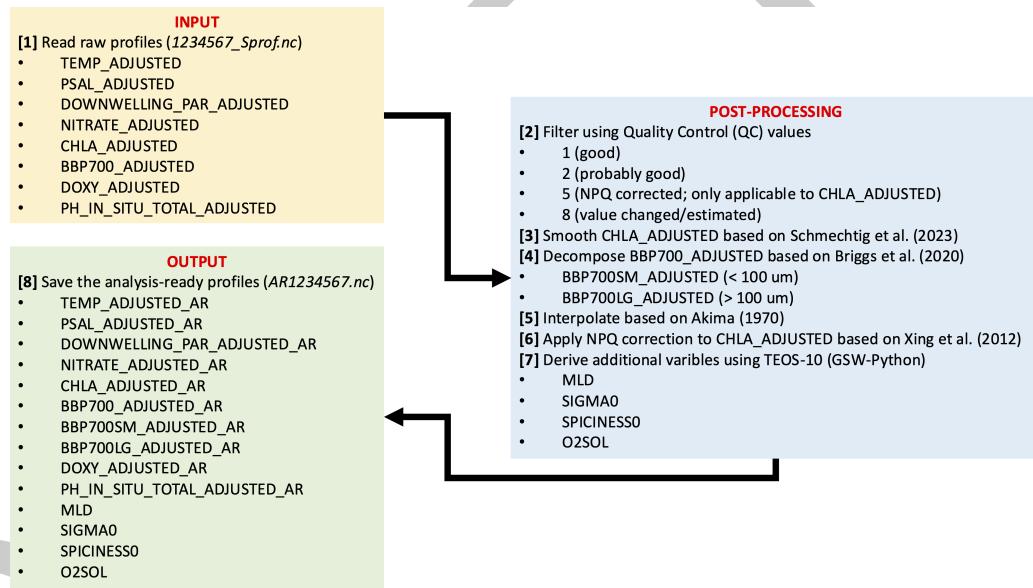


Figure 2: Figure 2: Schematic workflow of the data-processing pipeline implemented in `generate.ipynb`, which involves procedures based on Briggs et al. (2020), Schmechtig et al. (2023), and Xing et al. (2012). This example assumes a full-sensor float with an arbitrary WMO ID (1234567).

## 69 Software design

70 ar-bgc-argo was designed to help get started with the BGC-Argo data analysis, while allowing  
 71 the users to see and modify the source code as they run the Jupyter Notebook templates.  
 72 Doing so motivates the users to learn the code and customize it for their own application.  
 73 This approach is structurally different from library-based products (e.g., Maze & Balem, 2020),  
 74 which are designed to work without needing to see the source code. Furthermore, ar-bgc-argo  
 75 gives immediate access to post-processed data as soon as the data become available on GDACs.  
 76 This near-real-time capability is advantageous over data archiving such as Johnson et al.  
 77 (2023), which is updated approximately every six months.

## 78 Research impact statement

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

## 91 AI usage disclosure

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

## 93 Acknowledgements

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

## 99 References

- 100 Bittig, H. C., Maurer, T. L., Plant, J. N., Schmechtig, C., Wong, A. P. S., Claustre, H., Trull, T.  
101 W., Udaya Bhaskar, T. V. S., Boss, E., Dall'Olmo, G., Organelli, E., Poteau, A., Johnson,  
102 K. S., Hanstein, C., Leymarie, E., Le Reste, S., Riser, S. C., Rupan, A. R., Taillardier, V.,  
103 ... Xing, X. (2019). A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage.  
104 *Frontiers in Marine Science*, 6. <https://doi.org/10.3389/fmars.2019.00050>
- 105 Briggs, N., Dall'Olmo, G., & Claustre, H. (2020). Major role of particle fragmentation  
106 in regulating biological sequestration of CO<sub>2</sub> by the oceans. *Science*, 367, 791–793.  
107 <https://doi.org/10.1126/science.aay1790>
- 108 Bruggeman, J., Bolding, K., Nerger, L., Teruzzi, A., Spada, S., Skákala, J., & Ciavatta, S.  
109 (2024). EAT v1.0.0: A 1D test bed for physical–biogeochemical data assimilation in natural  
110 waters. *Geoscientific Model Development*, 17, 5619–5639. <https://doi.org/10.5194/gmd-17-5619-2024>
- 112 Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the Global Ocean with  
113 Biogeochemical-Argo. *Annual Review of Marine Science*, 12, 23–48. <https://doi.org/10.1146/annurev-marine-010419-010956>
- 115 Doi, T., Baba, Y., & Hayashida, H. (submitted). Seasonal chlorophyll-a prediction in the  
116 tropical pacific with a global climate model incorporating marine biogeochemistry (SINTEX-  
117 F2bio). *Journal of Climate*.
- 118 Hayashida, H., Strutton, P. G., Hashihama, F., Horimoto-Miyazaki, N., Katano, T., &  
119 Yasui-Tamura, S. (2025). Status and prospects of biogeochemical Argo observations of

- 120 chlorophyll-a in the northwest Pacific. *Journal of Oceanography*, 81, 289–299. <https://doi.org/10.1007/s10872-025-00750-2>
- 121
- 122 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>
- 123
- 124
- 125
- 126
- 127
- 128 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>
- 129
- 130 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/>
- 131
- 132
- 133
- 134 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., Bolland, 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>
- 135
- 136
- 137
- 138
- 139
- 140 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>
- 141
- 142
- 143
- 144
- 145
- 146 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.
- 147
- 148
- 149