

1 R2

The paper presents a data compilation of the freely and openly available BGC-Argo data with additional data processing steps performed by the authors. Their work is inspired by efforts like GLODAP or SOCAT, which add additional steps of quality control for a higher consistency dataset than that from the original observing network (repeat hydrography or SOCONET). Likewise, the authors performed additional visual quality control as well as automated checks that flag parts of the BGC-Argo data.

Those checks follow the intended use by the authors and the data compilation's potential user base, e.g., to remove density inversions above a certain threshold. This modifies the original BGC-Argo dataset beyond "just" quality control and should be clearly communicated as such. I.e., BGC-ArgoPlus is a cleaned version of the BGC-Argo data, but also one cleaned (somewhat) towards a certain science purpose. As long as the authors are transparent about this, this is fair and such choices up to their discretion. However, BGC-ArgoPlus should not simply be sold as a "better BGC-Argo" dataset.

In addition, the compilation contains additional, calculated parameters for convenience such as MLD and carbonate system parameters. The selection is again towards the anticipated user base. There is merit to having those calculations provided alongside the BGC-Argo data in a convenient way, as it promises to reduce spread across upstream studies due to different subtleties in the underlying calculations.

A key feature and positive aspect of this work is that the data compilation retains part of the original data structure, i.e., data from BGC-ArgoPlus can be straightforwardly linked back to the original BGC-Argo floats and profiles. This enormously helps the works transparency.

The manuscript is clearly and logically structured and can be followed easily.

Thank you for taking the time to review this manuscript and for your thoughtful contributions. We have tried to address each of your comments throughout (see specific responses below). For your point about the fact that this dataset has been QC'd / cleaned for a specific purpose and with specific biases, we totally agree. In response to your comments and those of another reviewer we have tried to make it clear that this is not a "best" or "final" dataset and in our "Practicalities and Caveats" section (p. 17) we try to accurately state our own background and interest and make it clear that others may make different decisions about what is signal and what is not.

Comments:

- The product's website and the manuscript differ in their statements about whether product data received additional adjustments that the authors deemed reasonable (e.g., DOXY; Bushinsky et al. 2025) or not. At present, there seems to be no adjustment done (following the manuscript). I think it would be important to preserve a version without additional data adjustments, so users can have a choice to follow the authors decision on such or not.

We agree that it is presently somewhat confusing about whether any correction for the deep oxygen biases has been made. As you note, for this first version of the dataset we do not make such a correction. It seemed to make the most sense to separate that effort from the current secondary QC approach as they are both valuable steps but have different purposes. We have changed the website to better reflect this and keep things clear. The offset correction will be submitted in a separate manuscript.

- It is unclear at present how the process will look like for updates of the BGC-ArgoPlus compilation. Will they include visual screening of all new data, as was done for the current compilation, or will they rely on automated checks on new data only? This will make a difference in the scientific utility judgement going forward and should be included in the manuscript.

Honestly, we do not know what updates will look like. We are in discussion about how to manage updates, but are not presently funded for continuation of this work. I think we will adopt a hybrid approach, where we manually check some floats but also try to improve our automatic routines to catch more and more of our previously identified outliers. It is a challenge, as the data gets updated at the DACs and we also have to decide when to re-assess previously checked floats. We are not including any specific comments about

how this dataset will be updated in the manuscript, as we would purely be speculating and the answer will depend on our ability to invest time and effort in the future.

- Some of the additional checks can be automated and efforts should be made to have those enter into the original processing systems, the Argo DACs (e.g., through the Argo Data Management Team meetings). A prime example seems the surface DOXY removal. Another one the bottom oxygen filtering.

We agree with this point. We are working to communicate our findings to people who are part of the ADMT and have begun work to refine / improve our automated detection process.

- It is less clear why there are surface removals on NITRATE with SUNA sensors: As they are typically placed at the bottom of a float (for buoyancy reasons), under no circumstances can they deliver data 'in air'. Any insights?

My hunch is that the bad nitrate data is not due to the SUNA measurement itself, but from the calculation of nitrate using incorrect salinity or temperature. I think this is also probably one of the reasons for the spikes we see in nitrate that appear to coincide with sharp density or temperature gradients, for instance at the base of the mixed layer.

- The visual quality inspection on every float of the current dataset, which probably was the most time consuming part, is one of the most valuable pieces to feed back to the original Argo data system. The authors provide a file that traces all their additional flagging, which is great. Some more detail on whether only point-wise flags were/should be applied (i.e., at the given pressure/N_LEVELS index) or whether contiguous parts of the profile (e.g., starting at the given pressure for N_LEVELS) would be helpful, e.g., to help automate such a transfer back to Argo.

Unfortunately we did not record the outliers in this way. The tool we developed allows selection of areas that can either encompass a single point or groups of points. Similarly we can remove bottom points for one sensor for a given set of profiles or for all profiles for a given float. This is something to consider for the future though. I think that generally if more than a few points were selected and removed it can be assumed that they were all selected at once, but this is incomplete, I agree. The difficulty is in balancing the need for detail on why/how something was removed with the time it takes to do the outlier removal. Adding even 1 minute to the time it takes to examine /mark each float adds hours to the time needed to go through them all.

- On the slow response "goo" oxygen profiles observed and filtered out in the compilation: Again, this is a case where it should be made clear that data are flagged out based on the authors interpretation of good data. (I have not checked whether such "goo" cases are separately marked in the accompagnied file tracing the additional flagging? Would be good if they were seperate, because:)

See our response above. Additionally, we did not recognize the presence of these types of profiles until our analysis was well underway. Rather than add a different type of marking, we chose to keep these as "manual removals". However, going forward I think we can create a few new categories to label removals. Additionally, I think we can find many of these profiles with some careful filtering, which is something we can also try in the future.

- Did the authors had the chance to check some measurement timing information associated with those "goo" profiles? In my experience, floats sometimes happen to perform their buoyancy increase actions at a location that results in them getting a stronger buoyancy boost than expected, having them "shoot" upward through the water column compared to their typical ascent speed. With the sensor's time response still being the same, this can cause such seemingly delayed profiles, without any additional "goo" but just because the float profiles faster. The authors would have to look into the corresponding core profile file for such timing information ("MTIME" or "NB_SAMPLE_CTD"). It is not available in the s-profiles as data in the s-profiles can be interpolated.

Looking at the ascent rate as a potential culprit is a good idea. Based on your suggestion we looked at the "NB_SAMPLE_CTD" field for one float with multiple instances of a slow oxygen response in the ML and did not see an obvious correlation between the number of CTD samples within each averaging bin and the

presence of slow ML oxygen response. However, we changed the description to simply focus on the slow response and added this as a possible reason.

Additional comments:

- On the example figures given, please make sure to state the WMO and cycle number(s) somewhere within the figure labels/titles or the figure caption for traceability.

Thanks, all added now.

- Data: Please add a proper Argo reference, selecting the monthly snapshot that is closest to your data access. (e.g. <https://doi.org/10.17882/42182#117069> or <https://doi.org/10.17882/42182#117068> ?)

Fixed with the January 2025 monthly snapshot.

- l. 74: Replace "performed" by "implementation", since all should follow the same steps, i.e., perform the same calculation?

Changed.

- l. 89: Maybe add "and are subject to future updates when a float's data are revisited" to indicate the non-static nature of Argo data?

Added.

- only ascending profiles where used, no descending ones?

Most floats only sample on the ascent, but we looked at all profiles contained in the Sprof files and did not specifically look for ascending / descending profiles.

Typos:

- l. 53: data 2x

Fixed, thanks.

- Table 1: superscript "1" appears only in the footnotes, does it?

Fixed