Response to Reviewer Comments for *GOBAI-O_2: temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly two decades*

Thanks again to the reviewers for carefully looking over our manuscript and providing constructive comments and suggestions. Below we have included detailed responses (in bold) to the editor's and reviewers' comments, which have led to further improvement of the GOBAI-O_2 product and associated manuscript. A revised version of the manuscript and a version with tracked changes will accompany this document. References to line numbers in this document refer to the revised manuscript, rather than the original submitted version.

Sincerely,
Jonathan D. Sharp and coauthors

Topical Editor – Anton Velo

Public justification (visible to the public if the article is accepted and published):

In this article, the authors provide a way relevant work and dataset for the field, which promises significant future use by a wide number of different stakeholders. That also implies an important responsibility to ensure the precision and accuracy.

Getting to the content, I believe that reviewer#1's comments (after 1st iteration) about uncertainties of the ARGO-O2s should be addressed before publication, I copy them below:

[see Reviewer 1 comments]

And just as a minor comment, I'd have preferred the usage of ML term instead of AI in the work, as the latter tends to be used for smart systems with take autonomous decisions, but as the term definition is very ambiguous and broad (and includes ML), I've no objections.

We thank the editor for taking care to ensure that uncertainty in GOBAI-O_2 is carefully quantified and properly communicated. We indeed hope to see widespread use of the product in the future by a variety of stakeholders, so we recognize this as an important and necessary responsibility. In response to the reviewer comments, we have updated the data product with new estimates for oxygen measurement uncertainty, as well as a set of newly quality-controlled float oxygen profiles.

The inclusion of "AI" in the data product title was partly to form a coherent acronym, with the recognition that ML is a specific subfield under the AI umbrella. We now address this choice in the manuscript (lines 95–98).

Reviewer 1 – Anonymous

My comment is on the revised manuscript by Sharp and co-authors, developing a global map of dissolved oxygen concentrations since 2004 using sensor O2 measurements from autonomous floats and the QCed shipboard measurements from GLODAP. I have read the original manuscript, review comments, and the revised manuscript, and this report is based on the latest version. Starting from the conclusion, I would like to enthusiastically encourage the publication of this manuscript and the dataset with one further (minor) revision.

This work is an important milestone in the biogeochemical oceanography. The methodology of the machine-learning based oxygen maps was initially developed by Giglio et al (2018), and the authors did an excellent job of extending it to the global map. One notable step is that the authors merged the float and shipboard O2 with a small offset determined from the overlapping profiles. The float O2 is given a small and uniform offset based on the apparent, underestimation of a few micro mol/kg. This is important because we are concerned about the long-term trend O(1%) per decade or less, such that small offset like this can change the conclusion significantly. Another noteworthy effort in this paper is the creative use of the ESM output (GFDL-ESM4) to assess the skill of the gap-fill and mapping, adding confidence to the effectiveness of this ML-based approach.

I read that there were many excellent comments and questions from the first two reviewers, and I appreciate that the authors took the time to address these comments. I believe this process improved the manuscript significantly. I do not feel the need to repeat any of the points raised by these review comments at this time.

Below are my comments about the uncertainties of the GOBAI-O2 product that I would like the authors to consider & discuss before finalizing this paper. I do not wish to further delay the publication of this paper, but I think the authors are responsible for raising the awareness for the data users about the limitations and potential deficiencies of the ARGO-O2 dataset upon which GOBAI-O2 dataset is built.

We thank the review for their thorough evaluation of our work and enthusiasm about its publication. We've dedicated a significant amount of effort to evaluate the quality of the datasets, quantify the skill of the machine learning model predictions, and assess confidence in the GOBAI-O_2 fields through model-based simulations, so we appreciate the reviewer's recognition of these important steps. Finally, we acknowledge the importance of providing the highest quality data product possible at this time and fully alerting potential users to its strengths and weaknesses, so we will do our best to respond to the reviewer's concerns.

1. The ARGO-O2 is an evolving technology with variable accuracy and uncertainties even in the delayed mode dataset. Can we confirm whether or not GOBAI-O2 uses only optode sensors calibrated by the known methods (either climatology or in-air O2 measurements)? If other types of sensors are used, it should be stated.

More than 90% of the BGC Argo floats used in the creation of GOBAI-O_2 are equipped with optode oxygen sensors. The rest are equipped with electrochemical oxygen sensors. Quality control measures for the oxygen sensors are primarily based on climatology comparisons and in-air measurements (see answer to next question).

2. The uncertainty range should be re-considered. Of the O(900) floats that passed the QC step, there is a diversity of sensors and calibration methods. Only relatively new ARGO-O2 profiles are

calibrated with in-air O2 measurement, and the older data (essentially all floats deployed 2015) had to be calibrated using climatology. Because GOBAI-O2 blends all kinds of O2 sensors, and uncertainties should be re-assessed in section 2.5. My suggestion would be no less than 3% for measurement uncertainty considering the climatological calibration from older O2 sensor profiles.

The reviewer raises an important question regarding the reliability of BGC Argo oxygen measurements, as methods of quality control for these measurements are being actively developed and refined. Common methods include calibration via in-air measurements, comparisons to surface climatologies, comparisons to subsurface measurements, and in situ optode calibrations. A statement has been added (lines 128–132) detailing the proportion of float profiles used in the development of GOBAI-O_2 that fall into each of these categories of quality control. These proportions are based on a thorough analysis of the calibration comments provided with the float data. We revise our measurement uncertainty estimate based on this analysis (see answer to next question).

3. In section 2.5, the combined measurement uncertainty is stated as 1.5%. I believe this is an average of 1% in GLODAPv2 and 2% in ARGO-O2. As stated above, ARGO-O2 should have 3% uncertainty at least, and the combined uncertainty should consider the number of profiles from each dataset. My reading is that the authors used O(21k) profiles from GLODAPv2 and O(133k) profiles from ARGO-O2. Therefore, measurement uncertainty should be dominated by the ARGO-O2, and due to the blend of unaccounted sensor/calibration type, it should be 3% in my opinion.

Measurement uncertainty has been amended to 3%, as suggested by the reviewer. We now discuss this choice as a consequence of multiple factors: GLODAP uncertainty, BGC Argo uncertainty, the lack of response time corrections to BGC Argo data, and the relative proportion of float profiles compared to ship profiles (lines 271–284). We have added a statement to section 3.2.4 indicating the implications of measurement uncertainty in GOBAI-O_2 for continued progress in sensor development and quality-control (lines 581–583).

4. The issue of sensor response time is briefly discussed in section 2.5. The community has not yet implemented the response time correction (RTC) of optode sensors. In section 2.5 the authors stated that 2% uncertainty of ARGO-O2 is still optimistic, but I don't think this statement adequately describe the problem. The lack of RTC in ARGO-O2 data not only increases the measurement uncertainty but also causes systemic bias in the vicinity of oxycline. This is a larger problem beyond the scope of this paper, however, it should be discussed and the data users must be warned. Much stronger statement of caution should be provided, perhaps in the discussion/conclusion. Current ARGO-O2 dataset can include significant bias and uncertainty in the oxycline region beyond the level that is characterized in the uncertainty analysis. Simply adding a constant offset or blending GLODAPv2 data using machine learning will NOT correct this issue. Progress in this area will be much needed for future research.

We thank the reviewer for highlighting this opportunity to emphasize the importance of considering optode sensor response time. We have expanded the discussion of sensor response time in section 2.5 and now highlight the potential for systematic biases (lines 280–284). We have also added a statement in the Conclusions to caution potential users about the

impact of the lack of response-time-corrections to float [O_2] data, especially when floats cross steep gradients (lines 676–677).

Reviewer 2 – Hernan Garcia

The authors have addressed my major concerns with the previous version. This is an interesting approach with potential for other applications. Thank you.

I would like the encourage the authors to include a statement of the new use of AI (GOBAI-O2) to map O2 in addition to other gap filling mapping techniques in the abstract or summary. This is the main take home message of the paper.

Sampling, integrating data of known quality, and understanding ocean O2 variability is difficult. I am curious about the potential use of AI to recognize similar S2D meso-scale (or larger scale) O2 distribution patterns? Are there repetitive temporal/spatial patterns and/or higher frequency in the observations and/or model output?

The authors thank the reviewer for evaluating the manuscript again. We have added a statement in the abstract indicating the novelty of machine learning for gap-filling ocean interior biogeochemical observations, and advocating for continued development alongside other mapping techniques (lines 22–23).