**Response to reviewer 1 comments**

**I appreciate the constructive comments from the reviewer to improve the manuscript. Below, I respond to each comment with the original comments copied in black font and my responses in blue.**

Du et al. presents a worthwhile comparison of nine different oceanic dissolved oxygen datasets from differing methods. As a key biogeochemical property to understand current climate, including deoxygenation, patterns, this analysis is extremely worthwhile and valuable to the community. Despite some significant regionality, I am thoroughly impressed with how well these datasets align and only have minor questions for the authors.

Thank you for the positive comments on this study.

Major comments:

- I would like to see more information on how much of the data was from Winkler's vs. how much from sensors. That would be useful to understand some of the spatiotemporal biases.

Thank you for this comment. As the $O_2$ observations used to construct the eight climatologies are different in observation instruments/platforms and time coverage, but the major data sources (which constitute >95% of the data in WOD) are three: OSD, CTD, and Argo. Thus, we present the spatial distribution and annual number of profiles from these three instruments and added a new figure in the manuscript. The total numbers of OSD, CTD, and Argo profiles are 744,141, 215,345, and 188,497, respectively, for 1960-2024. It is clear that before 1990, OSD dominated; from 1990~2010, CTD dominated; and since ~2010, Argo has been widely used.

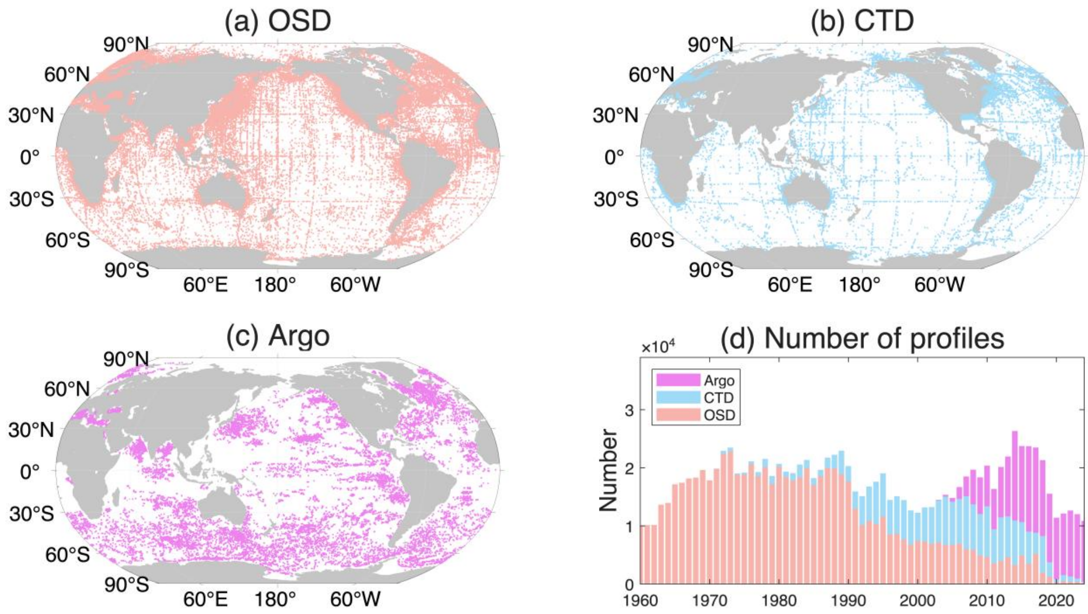


Figure 1 Spatial distribution (a: OSD, b: CTD, c: Argo) and annual number (d) of observation profiles.

Minor comments:

Line 65: Please define IPCC.

Thanks for the comment. The full name of IPCC, Intergovernmental Panel on Climate Change, has been added in Line 65.

Lines 66-68: Please keep numerical range formatting consistent.

Thanks for the comment. The numerical range in Line 66 has been changed from ‘0.3 %~2 %’ to ‘0.3 %-2 %’ and now the formatting are consistent for Lines 66-68.

Lines 75-77: Should be commas not semi colons.

Thanks for the comment. Semi colons in Lines 75-77 have been changed to commas as suggested.

Line 77: I would add “Since the late 19th century, oceanographers have measured ocean $O_2$ using many instruments with varying sampling resolutions.”

Thanks for help improve the manuscript. The content ‘with varying sampling resolutions’ has been added to Line 77.

Lines 78-85: Similar to my comment from above, I would expand on this by noting that Winklers are labour intensive, leading to lower sampling resolution, whereas sensor-based measurements have better spatiotemporal resolution, and the proliferation of the BGC-Argo program has dramatically increased observations.

Thanks for the suggestion. The extended discussion, 'The Winkler data are labour intensive thus leading to less sampling, whereas sensor-based measurements have better spatiotemporal resolution, and the proliferation of the BGC-Argo program has dramatically increased observational capability.', has been added in this section.

Lines 118-122: Personally, I don't think this type of paper outline is necessary, but that is up to you.

Thanks for the comment. After consideration, the paper outline has been retained here to follow the journal style.

Lines 188-190: Have you trimmed both GOBAI and IAP so that the exact years match up (i.e., 2004-2022)? That should correct for any bias specifically due to the dataset age.

We haven't done this in our study because we aim to provide a direct comparison of the existing datasets. We believe that climatologies constructed from observation profiles with exactly the same data age will be of great importance and more consistent in understanding the mapping method. Such exercises will be conducted in an internationally coordinated group (GO-DIP) (Ito et al. 2025).

Line 307: Gradients, plural.

Thanks. Corrected.

Lines 338-349: I appreciate the discussion of biological and physical controls on the annual cycle, but this feels like the first time underlying mechanistic drivers are being discussed. Can you similarly discuss biological and physical controls on spatial patterns or zonal structures?

Thanks for the comment. Some discussions have been added for the zonal mean annual cycle analysis (Fig. 8) in the manuscript: 'For both 0-100 m and 100-600 m, the phase transitions of the zonal mean temperature anomaly occur around June and January for all datasets, which are consistent with the annual variation analysis of $O_2$

anomaly in Fig. 5. The phase change of $O_2$ lags about one month behind the temperature change (June and December as shown in Fig. S12), reflecting that some biological and ventilation processes also impact the $O_2$ seasonal variations besides the solubility'. The figure of zonal mean annual cycles of temperature anomaly is added in Supplementary material as Fig. S12.

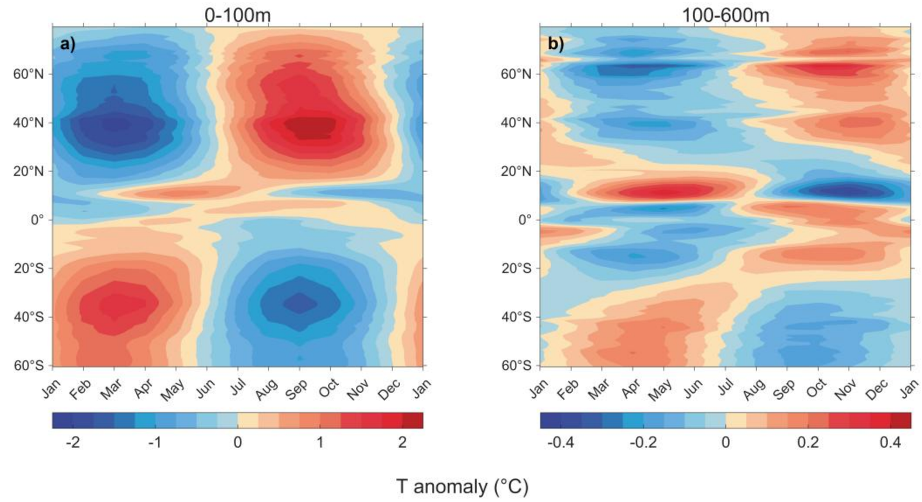


Figure S12 Zonal mean annual cycles of temperature anomaly for 0-100 m (a) and 100-600 m (b) for IAP temperature data product (unit:°C).

We also added some discussions about the zonal mean annual cycles of solubility and Apparent Oxygen Utilization (AOU) in the manuscript with a new figure as following:

The zonal mean annual cycles of solubility and Apparent Oxygen Utilization (AOU) of different data products for 0-100 m and 100-600 m are shown in Fig. 10. The solubility is calculated using the gridded temperature and salinity products of IAP (Cheng et al., 2024) following the Garcia and Gordon (1992) method. Generally, the zonal 0-100 m $O_2$ seasonal cycle is mainly dominated by the solubility, whereas the AOU dominates the 100-600 m $O_2$ pattern.

For 0-100 m, the AOU seasonal cycle in the mid-latitude is about two to three months lagging that of the temperature/solubility seasonal variation (Figs. 6, 10, S12). Within 20° N-60° N, the zonal mean AOU anomaly of 0-100 m is mostly negative from March to August. The phase transition of AOU occurs in March, which corresponds to the temperature minimum/solubility maximum. And AOU reaches a minimum in about May/June, corresponding to the phase transition of the zonal mean temperature/solubility anomaly (Fig. S12, 10). A possible explanation is from the combined impacts of biological and physical processes. The spring phytoplankton bloom lags the onset of temperature increase (starting in Mar.) because phytoplankton proliferation requires sufficient light from increasing spring insolation and nutrient entrainment due to a shallower mixed layer (Martin, 2012). And the net community production (NCP, the difference between gross community photosynthesis and community respiration) reaches a maximum in about May (Wang et al., 2022). The solubility reaches a minimum in September, which corresponds with the phase change of AOU. And the maximum of AOU lags about two months behind the temperature maximum/solubility minimum, which is also a combined effect of the prolonged response time of biological processes and the physical processes possibly induced by the deepening autumn mixed layer. The situation in 20° S-60° S is similar to 20° N-60° N for 0-100 m, mainly with the signs reversed. For 100-600 m, the annual variation in temperature/solubility is relatively small, and the pattern of AOU variation dominates, with the underlying processes requiring further investigation.

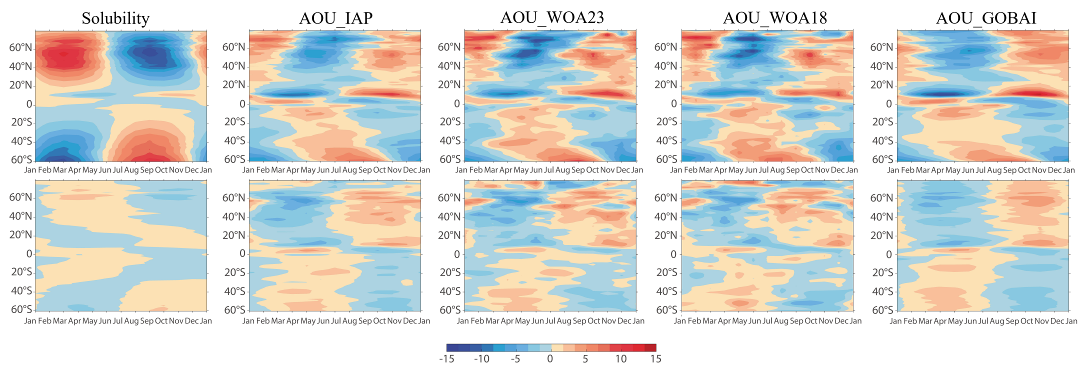


Figure 10 Zonal mean annual cycles of the solubility and AOU anomaly of different data products for 0-100 m (upper) and 100-600 m (lower) (unit: μmol $kg^{-1}$)

**Response to reviewer 2 comments**

**I appreciate the constructive comments from the reviewer to improve the manuscript. Below, I respond to each comment with the original comments copied in black font and my responses in blue.**

The paper by Du al. compares eight gridded dissolved oxygen products to assess global dissolved oxygen climatology, annual cycle and oxygen minimum zone (OMZ) representation. This is timely and useful contribution for the oxygen community. A multi-product comparison of this kind is much needed and valuable.

The focus on both climatology and annual cycle is worthwhile, and the effort to define common comparison metrics through GODIP-DO is a strong point. In my view, the manuscript is strongest as a descriptive intercomparison. It is less convincing when it moves toward explaining why products differ, or when it comes close to treating product spread as formal uncertainty estimate.

The main findings are clear: the global mean vertical oxygen structure is broadly similar across products, while local upper-ocean differences can be much larger.

My main concern is that the manuscript sometimes over-interprets what this study design can support. The products differ in their source observations, quality check, bias correction, mapping approaches, resolution, mask and climatological periods. The study is therefore well suited to attributing those differences to a specific cause. The manuscript acknowledges this point, but not consistently enough throughout.

Thank you for the constructive comments. Yes, this study aims to comprehensively assess the differences among eight O2 climatological data products within the framework of GODIP-DO. We can only provide some potential reasons for the differences, with the specific reasons requiring further study, as in Ito et al. 2025, where a common input experiment is conducted. The “uncertainty” can not be derived from this study because it is poorly defined by the ensemble spread of either product. According to this suggestion, we have refined the manuscript carefully and clarified which conclusions are directly supported by this study and which need further investigation.

Main issues

- The abstract states that the intercomparison "allows assessing the robustness of uncertainties through the spread of products", and in the conclusion presents the results as a starting point for resolving the uncertainty budget. That is reasonable at a broad level, but the wording needs to be more careful. What the paper really quantifies is inter-products spread or product disagreement, not total uncertainty in a formal sense.

Thank you for this comment. We have modified the last sentence of the abstract as ' Our results help to depict and understand the spread among the available $O_2$ gridded datasets' and also refined the wording in the conclusion section accordingly.

Some comments

- The annual cycle section seems weaker than the climatology section, and could be discussed more in the text.

Thank you for the comment. More discussions of the annual cycle, based on the zonal mean distribution of Apparent Oxygen Utilization (AOU), and a new figure for the four monthly climatological data products have been added to the manuscript following Fig. 9 (zonal mean annual cycles of $O_2$ anomaly).

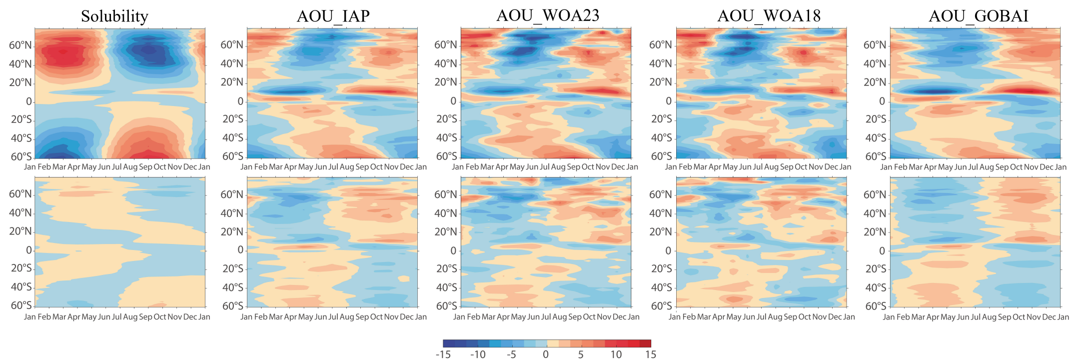


Figure 10 Zonal mean annual cycles of the solubility and AOU anomaly of different data products for 0-100 m (upper) and 100-600 m (lower) (unit: μmol $kg^{-1}$) .

- There is also a coverage mismatch across products. Some extend only to 2000 m, while others reach deeper; some provide annual climatology only, whereas others include monthly fields. The authors do point this out, but the implications for comparability should be discussed more explicitly, especially for the OMZ analysis.

Thanks for the comment. The monthly/annual climatologies for different data products are introduced in Lines 135-137. And only the four monthly climatologies (IAP, WOA18, WOA23, GOBAI) are included in the annual cycle intercomparison (Lines 326-327). For the OMZ analysis, we calculated the total volume of OMZ60 and OMZ90 globally and for each basin using the climatologies with depths of 5500/6000m, excluding GOBAI and SJTU because their maximum depth only reaches about 2000 m below the ocean surface, as explained in Lines 440-442. In the analysis section of OMZ areas versus depth levels, the following discussion is added as suggested to explicitly point out the depth coverage difference across products: 'So that all the climatologies, including GOBAI and SJTU with the maximum depth of 2000 m could be involved in the analysis of OMZ areas at depth levels.'.

- The handling uncertainty remains descriptive. I would encourage the authors to be more precise in how they frame product disagreement. As noted above, the paper measures inter-products spread, not uncertainty in a formal sense.

Thanks for the comment. We have removed or modified most of the descriptions about 'uncertainty' because we agree that inter-product spread is not a good quantification of the uncertainty.

- At times, especially in the conclusion, the manuscript moves too quickly from descriptive comparison to causal explanation. For example, the statement that discrepancies "...could give an insight into regions where the accuracy of gridded data reconstruction is relatively more sensitive to the mapping method and observation data distribution... " is stronger than the analysis allows. This comparison does not isolate the effect of mapping method or observation data distribution, since products differs in several ways at once. I would suggest softening this kind of language. More generally, the final paragraph is trying to look ahead, but it becomes somewhat generic. I think the paper would end more strongly by saying that it establishes a baseline descriptive benchmark, while controlled intercomparisons are still needed to partition the causes of spread, especially in regions with strong gradients and sparse observations.

Thanks for the comment. The statement of causal explanation in the conclusion part has been softened as suggested: 'Substantial local differences (25 μmol kg-1 for the upper 1000 m climatological mean) can be seen, which could influence the baseline from which anomalies and trends are calculated and could give an insight into regions that are relatively more sensitive in the process of gridded data reconstruction'.

And the last paragraph in the conclusion section has been modified as suggested: 'The overall spread across products results from all uncertainty sources (e.g., measurement errors, mapping errors, different time periods, etc.). Controlled intercomparisons that isolate each error source are needed in the future to understand the contribution of a single factor, especially in regions showing a large spread. It will eventually help the community to improve the methodologies and reduce the spread in the future. Another caveat is that only a limited number of products are included in this inter-comparison. We hope to maintain and extend this activity in the future and serve as a regular intercomparison exercise to provide critical information to the data users.'

Overall, the main revision need is not the dataset comparison itself, which is useful, but a tighter framing of what the study can and cannot infer.

We have carefully revised the expressions in this study to focus on establishing a description of inter-product differences, and we also expect a careful investigation of error sources in the future through controlled intercomparison experiments.

Some wording comments:

- Line 509:" by its depth-mean spatial distribution" is awkward and should be rephrased.

Thanks for the comment. It has been rephrased as 'Substantial local differences (25 μmol $kg^{-1}$ for the upper 1000 m climatological mean) could be observed between data products.'.

- Lines 527-528: "serve as a regular practice" would read better as "serve as a regular intercomparison exercise".

Thanks for the suggestion. The content here has been modified as suggested.