

Response to Reviewer #2

for “Uncertain current and future ocean deoxygenation due to internal climate variability and observational gaps”, by Takano and Ilyina, 2025
(response is in blue)

General Response

We thank both reviewers for the constructive comments on the manuscript. All comments have been addressed, and a detailed response for each is provided below. We are aware that both reviewers have recognized the validity of the dissolved oxygen (O₂) metrics we used in this study to detect the forced response on top of the internal climate variability (ICV). As the reviewers mentioned, the signal-to-noise ratio is high at a single depth (here 300m) due to isopycnal heaving, interior ocean wave activity, and associated steep vertical gradient of thermocline O₂, etc., and is not necessarily suitable for detecting multi-decadal trends. In the revised manuscript, we have followed the approach of Bopp et al. (2013) and the more recent study by Kwiatkowski et al. (2020), and used the vertical mean of O₂ over the upper ocean (here, between 100 m and 600 m, following Kwiatkowski et al., 2020). We have updated all the analyses accordingly.

Reviewer 2

Overview

This paper seeks to evaluate the extent to which sampling bias and internal climate variability (ICV) contribute to uncertainty in our understanding of ocean deoxygenation and oxygen variability over the past few decades. The authors approach this issue by sub-sampling a set of Earth system model ensemble members according to historical sampling patterns and comparing metrics from the sub-sampled fields to the full fields as well as to observation-derived data products. They use oxygen concentration at 300 meters depth to illustrate their results.

The authors show that sparse data coverage can lead to the conclusion that ICV is much greater than it is reality. They also find differences in trends computed from sub-sampled versus full-field results with a tendency for overestimation of the trend in data sparse regions. The authors indicate that saturation changes alone are not the major driver of deoxygenation over recent decades. They also

point out that ICV contributes to trends in addition to the forced response, especially over short timescales.

Finally, the authors test the impacts of three different scenarios for distributions biogeochemical (BGC) Argo floats over the next 20 years. They find that near-term deoxygenation might be overestimated without expansion of the BGC float array. The scenarios of planned expansion and expansion followed by a sharp cut tend to more realistically capture near-term global oxygen trends.

General Comments

I generally agree with the outcomes and conclusions of this work, but I am somewhat skeptical of the methodology and approach. I have concerns about the appropriateness of the dataset selected to represent observational coverage, the choice of a fixed depth level to illustrate the results, and the future-looking analysis.

I believe there should be more detailed discussion regarding the Ito et al. (2017) dataset, what their gaps in data coverage actually represent, and how this translates to the main conclusions of the paper. Ito17 objectively mapped their oxygen anomaly data using a Gaussian weight function with relatively long zonal and meridional length scales. So the “coverage” of the Ito17 dataset is upscaled from the actual observational coverage, and subsampling full fields of model ensemble members based on the Ito17 coverage therefore presents a rather optimistic view of observational coverage that assumes the simple length scales employed by Ito17 are adequate. At the same time, the efforts made by the referenced fully gap-filled datasets (Ito, 2022; Roach and Bindoff, 2023; Ito et al., 2024) to account for those observational gaps should not be ignored either. The results presented in Figures 1–4 and discussed in Section 3 represent a kind of intermediate state, whereby model ensemble members are subsampled according to unrealistically extended observational coverage, but no attempt is made to account for the gaps between those areas of coverage. I’m not sure how well this intermediate state represents studies that evaluate ocean deoxygenation trends and ICV. For this reason, I’m not fully convinced of the value of the results, although I generally agree with the points of discussion.

I understand the choice to focus the analysis to ~300 meters, as this is a depth with strong oxygen variability and relevance to marine ecosystems. However, it

is also a depth with steep vertical oxygen gradients that depend strongly on ocean physics and the exact positions of water masses. This makes it challenging to compare between observational datasets and model output in depth space, and makes trends very sensitive to ICV and to changes in observational coverage. Even if the authors would like to keep the focus on this 300 meter depth level, some supplemental analysis in potential density space an/or on other levels would be helpful.

The BGC-Argo analysis reads like it was tacked on late and is not a coherent piece of the paper. For one, the methodological information in Section 4.1 should be introduced earlier in the methods section. There is also a bit of a disconnect between the first part of the paper, where the sampling is based on the Ito17 data from shipboard observations alone, and the second part, where the sampling is based on simulated floats alone. More discussion of how these two observational networks might interact, and better yet inclusion of that interaction in the analysis, would be beneficial.

(Response) We thank the reviewer for the summary and thorough comments on the manuscript. The critique and comments on the methodology and approach were also raised by Reviewer 1 (i.e., the use of a fixed depth level O₂ as a metric), and we have revised the O₂ metric accordingly. The following are detailed responses to comments on the methodology and approach.

- 1. Representation of data gaps based on Ito et al. (2017):** Thank you for the comment, and this is a great point. Indeed, the spatiotemporal gap information from Ito et al., (2017) (hereafter Ito17) is upscaled due to an optimal interpolation procedure (i.e., an optimistic view of data coverage as the reviewer pointed out). Ideally, the best approach is to subsample the model output based on “the location of observational sampling points” (i.e., specific latitudes, longitudes, and time), but this will require high-resolution simulations with more temporally frequent model output (which is unfeasible at this point). Considering our focus is on large-scale (multi) decadal changes in O₂, we have revised our subsampling approach based on the spatiotemporal coverage information from “1-degree binned data” from the World Ocean Database. Binning the observational data is a standard procedure for gridded dataset constructions (e.g., Ito, 2022; Roach and Bindoff, 2023; Ito et al., 2024, 2026) before interpolation (i.e., gap filling) and provides us with the large-scale spatiotemporal coverage information suitable for subsampling the output from (relatively coarse resolution)

Earth System Models (ESMs). This approach will also prevent subsampling based on the upscaled information. The updated O₂ metrics (please see the second point (2) for details on updated O₂ metrics) are constructed from subsampled model output, using spatiotemporal coverage information from “1-degree binned data”. The interpolation and gap-filling methods also introduce uncertainties in constructing gridded products, and the community (e.g., Du et al., 2026; GODIP-DO) is currently conducting an observational intercomparison project. Addressing uncertainty arising from various interpolation methods is an important topic, but it is beyond the scope of this study, as we focus more on how spatiotemporal gaps will increase uncertainty in internal climate variability (ICV) and obscure the detection of forced trends. We have added the relevant statement on uncertainties to the summary, and the methodological uncertainty will be addressed in future studies based on an ensemble of models.

2. **A single-depth O₂ metric:** Thank you for the comment on the metric. As Reviewer 1 also mentioned, we agree that detecting forced response (i.e., anthropogenic O₂ trends) is not necessarily feasible using the single-depth O₂ metric due to the high signal-to-noise ratio (i.e., strong O₂ variability driven by isopycnal heaving associated with steep vertical O₂ gradients and ocean physical processes). Considering the limitations of the ocean biogeochemistry model output from MPI-GE and the approach used in previous CMIP-based studies (e.g., Bopp et al., 2013; Kwiatkowski et al., 2020), we have decided to use the annual vertical-mean O₂ metric, following Kwiatkowski et al. (2020). More specifically, we calculated the vertical mean of O₂ between 100m and 600m from both the full and newly subsampled annual mean O₂ from MPI-GE. We then revisited the analysis of forced response and ICV.
3. **The BGC-Argo-like analysis and future subsampling:** Thank you for the comment on this. The motivation for this analysis is to quantify uncertainty arising from the spatiotemporal gap, considering the recent BGC-Argo effort for the near-term detection (and monitoring) of multi-decadal deoxygenation in a somewhat idealized manner. We agree that there are some discontinuities in the introduction of the method, and we have revised the manuscript, moving the methodological information from section 4 to section 2. One important point we should highlight is that, according to the

recent World Ocean Database (and the recent review by Ito et al., 2026), the BGC-Argo data have been steadily increasing since the mid-2000s, while the amount of bottle data has been declining since the 1990s (please see Fig. 1 in Ito et al., 2026). This has implications for transitions not only in the observational platform but also in observational sampling regions (please see Fig. 1 in Ito et al., 2026). We have included this background information and discussion in the revised manuscript associated with our analysis in sections 3 and 4.

Line-by-line comments

152: Vertical oxygen gradients are strong in the thermocline, so the 10-meter gap between observational and model comparisons is not trivial.

(Response) Thank you for the comment. We have revised the O₂ metric to vertical mean O₂. This will make a more equal comparison of depth between observations and models.

225: Was a consistent mask applied to compare the fully gap-filled observational products and model ensemble members? If not, discrepancies may come from different areas of the global ocean being included in the global means.

(Response) The Ito22 dataset does not include the Arctic Ocean, and comparing it with other observational products and model ensemble members leads to discrepancies. We revised all the global mean O₂ timeseries based on “quasi-global mean” excluding the Arctic Ocean for all calculations for consistency following the recent review from Ito et al., (2026).

454: Given the average lifetime of a BGC Argo float, the “cut” scenario is not realistic. Even if deployments stopped immediately, the reduction would be a more gradual one.

(Response) We have revised the analysis and introduced a “gradual decrease” scenario for a more realistic evaluation.

References

[1] Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.

2013: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, *Biogeosciences*, 10, 6225–6245, <https://doi.org/10.5194/bg-10-6225-2013>

[2] Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T. 2020: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, *Biogeosciences*, 17, 3439–3470, <https://doi.org/10.5194/bg-17-3439-2020>

[3] Ito, T., S. Minobe, M. C. Long, and C. Deutsch. 2017: Upper ocean O₂ trends: 1958–2015, *Geophys. Res. Lett.*, 44, 4214–4223, doi:10.1002/2017GL073613.

[4] Ito, T. 2022: Optimal interpolation of global dissolved oxygen: 1965–2015. *Geoscience Data Journal*, 9, 167–176. <https://doi.org/10.1002/gdj3.130>

[5] Roach, C. J., and Bindoff, N. L. 2023: Developing a new oxygen atlas of the World's oceans using data interpolating variational analysis. *Journal of Atmospheric and Oceanic Technology*, 40(11), 1475–1491. <https://doi.org/10.1175/jtech-d-23-0007.1>

[6] Ito, T., Cervania, A., Cross, K., Ainchwar, S., & Delawalla, S. 2024: Mapping dissolved oxygen concentrations by combining shipboard and Argo observations using machine learning algorithms. *Journal of Geophysical Research: Machine Learning and Computation*, 1, e2024JH000272. <https://doi.org/10.1029/2024JH000272>

[7] Du, J., Cheng, L., Ito, T., Garcia, H. E., Wang, Z., Sharp, J. D., Roach, C. J., Minobe, S., Zhou, Y., Lu, B., Navarra, G. G., and Bushinsky, S. M. 2026: Climatology and annual cycle of global ocean dissolved oxygen represented by multiple observational gridded products, *Ocean Sci.*, 22, 1483–1499, <https://doi.org/10.5194/os-22-1483-2026>

[8] Takamitsu Ito, Yohei Takano, Yassir A. Eddebbar, Jerry F. Tiputra, Zhankun Wang, Shoshiro Minobe, Lijing Cheng, Juan Du, Yumi Abe. 2026. Are Simulated Ocean Deoxygenation Rates Consistent with the Observational Reconstructions?. *Annual Review of Earth and Planetary Sciences*. 54: In press.

<https://doi.org/10.1146/annurev-earth-032524-123111>