Response to Reviewer #1
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 (O_2) 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 O_2, 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 O_2 over the upper ocean (here, between 100 m and 600 m, following Kwiatkowski et al., 2020). We have updated all the analyses accordingly.

Reviewer 1

The authors have presented what is largely a methodological study focused on challenges in trend detection for ocean interior oxygen. Their analysis tools are clearly chosen, and their knowledge and judgment in choosing an ensemble and a range of observational resources is sound, as is their interpretation. So with that I believe that the manuscript is well-organized and well-presented, and overall I think it can meet the standard of Biogeosciences with major revisions. The reason for recommending major revisions is given below, but stems largely from what I think is an ill-advised choice of 300m as a horizon on which to assess observing system design.

(Response) We thank the reviewer for the positive comment and feedback on our work. As mentioned in the general response (and detailed response for both reviewers), we have revisited the O_2 metric we have used in our study and focused on the vertical mean of O_2 between 100m and 600m in the revised manuscript to address the major concern raised by the reviewers.

Main Points:

There is what I see as a highly problematic decision at the core of the framing of the study in choosing a fixed depth horizon (300m) for evaluating trends in the ocean interior. Though it is technically fine to do this, it should be expected to be a horizon that is characterized by elevated noise due to climate variations and wave activity in the ocean interior. Previous studies have tended to choose two different approaches to deal with the associated signal-to-noise ratio for this horizon being very high: (a) Long et al. (2016) considered changes on isopycnal surfaces in the ocean interior, and (b) Bopp et al. (2013) considered layer integrals for the upper ocean. These choices were made precisely because the signal-to-noise ratio at 300m is unnecessarily high. In short, no one would choose a 300m horizon for detecting anthropogenic trends, so it's perplexing that this fixed horizon was chosen for this study. Additionally, the authors do not provide any evidence of any particular importance of this depth horizon for any species or ecosystem community structure.

The core idea is that even with a perfect and gapless observing system for oxygen, natural variability at a fixed depth of 300m will be so large that it would take hundreds of years to detect a trend.

I believe that for this paper to be of interest to and to find broader applications within the broader community, including the ocean observations community, this problem must be addressed. One way to do this would be to choose a vertical layer spanning the upper ocean (again the approach of Bopp, 2013), if the problem is that the MPI model doesn't include monthly mean output. A much more challenging but robust approach might be to do a more comprehensive analysis for density layers and then to map back to depth horizons based on the mean depth of isopycnal layers.

(Response) We appreciate the constructive comments on the O_2 metric. We agree that, for detecting forced responses (i.e., externally forced multi-decadal and centennial O_2 trends), the single-depth O_2 metric is not necessarily suitable due to the high signal-to-noise ratio (also discussed in Long et al., 2016 with respect to isopycnal horizons). We also agree that, although upper-ocean O_2 is important for the marine ecosystem, there is no specific argument (or evidence) that a fixed depth of 300m is particularly important. We have updated the O_2 metric in the revised manuscript and chose to use the vertical mean of O_2 between 100m and 600m following Kwiatkowski et al. (2020). As the reviewer mentioned, the Max Planck Institute for Meteorology Grand Ensemble (MPI-

GE) [https://mpimet.mpg.de/en/research/modeling/grand-ensemble] does not include monthly-mean model output for ocean biogeochemistry, so it will be challenging to conduct a more comprehensive isopycnal O_2 analysis at this point. We hope to focus on this aspect for a future study.

More detailed points:

In the Abstract, the authors refer to "observational bias in ICV", do they mean "undersampling of ICV"? This also applies to the first sentence of the second paragraph of the "Summary and Discussion" section.
(Response) Thank you for the comment. In this context, we meant under-sampling of ICV (not necessarily addressing the bias in the observational dataset, such as measurement error, quality control, etc.). We have revised the statement.

It would help if the equations were numbered, but for the second equation (for AOU) the authors should be clear about which terms are saved as annual means and which are not.
(Response) All terms (i.e., temperature, salinity, and O_2) are based on the annual-mean model output. We have revised the equation labels and the statement to clarify this.

For the 4th paragraph of the "Summary and Discussion" (lines 550-561) it would be good if the authors could describe what distinguishes this study from previous studies.
(Response) Thank you for the comment on this. This study distinguishes itself from the previous studies in the following ways.

1) Based on one of the largest ensemble simulations based on an Earth System Model (ESM), we have quantified the uncertainty due to ICV, considering the spatiotemporal gaps in the observations. We explored whether the observed historical multi-decadal trends of O_2 could be explained solely by ICV, which has not been done in the CMIP6 multi-model analysis.
2) We have extended the context of historical analysis for ICV, considering spatiotemporal gaps in the observations, to address the consequences of detecting near-future trends of O_2 under different climate scenarios.

We believe that the approach we have presented in this study will be a helpful guideline for evaluating and analyzing simulations from CMIP6 and the future CMIP7 effort. We have included these statements in the revised "Summary and Discussion".

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] Long, M. C., C. Deutsch, and T. Ito. 2016: Finding forced trends in oceanic oxygen, *Global Biogeochem. Cycles*, 30, 381–397, doi:10.1002/2015GB005310.