https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Uncertain current and future ocean deoxygenation due to internal climate variability and observational gaps

Yohei Takano[1,4], and Tatiana Ilyina[2,3,4]

[1]British Antarctic Survey, NERC, UKRI, Cambridge, UK
[2]Department of Earth System Sciences, Universität Hamburg, Hamburg, Germany
[3]Helmholtz-Zentrum Hereon, Hamburg, Germany
[4]Max Planck Institute for Meteorology, Hamburg, Germany

Correspondence to: Yohei Takano (yokano@bas.ac.uk)

Abstract. Observed declines in oceanic oxygen (O_2) over recent decades are subject to substantial uncertainty due to internal climate variability (ICV) and limited observational coverage. Here, we quantify how observational uncertainty affects the assessment of both historical and future ocean deoxygenation by combining multiple observational datasets with a large ensemble simulation of the Max Planck Institute Earth System Model (MPI-ESM). We find that observational biases in ICV can amplify global and regional O_2 variability by 150%–500% in annual time series over the past 50 years. The combined effect of ICV and sampling bias can also introduce deviations of 5%–25% in estimated multi-decadal O_2 trends. Moreover, time-dependent changes in observational coverage complicate the interpretation of historical O_2 trends. Our results underscore the crucial need for a sustained, globally uniform ocean observing system to monitor long-term deoxygenation, assess its impact on marine ecosystems, and detect the anthropogenic signal in O_2 trends. We further show that near-future trend detection will remain sensitive to ICV, and observational gaps may distort the detection of scenario-based projections of O_2 trends, especially in the context of climate mitigation efforts.

1 Introduction

Multiple, independent lines of evidence show that the ocean has lost oxygen (O_2) in recent decades (Ito et al., 2017, 2024b; Ito, 2021; Helm et al., 2011; Roach and Bindoff, 2023; Schmidtko et al., 2017; Sharp et al., 2024). This observed deoxygenation is projected to accelerate with the ongoing ocean warming (Bopp et al., 2013, 2017; Cocco et al., 2013; Gruber, 2011; Keeling et al., 2010; Kwiatkowski et al., 2020; Laffoley and Baxter, 2019). With many marine organisms already living at their aerobic threshold (Deutsch et al., 2024), further loss of O_2 stands to precipitate species loss across wide regions, with profound implications for marine ecosystems and global fisheries (Breitburg et al., 2018; Levin, 2018). Accurate projection of the spatial distribution, magnitude, and timing of deoxygenation will be crucial for managing living marine resources, as well as for potentially motivating mitigation efforts (Bindoff et al., 2020). Confidence in such projections is gained, in part, through their faithful representation of observed changes over the historical era, a benchmark

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

rarely met by state-of-the-art Earth System Models (ESMs), most of which do not reproduce the observed changes in O_2 concentrations over the past 50 years (Oschlies et al., 2017, 2018; Stramma et al., 2012). However, there is only medium confidence in observed O_2 trends because the magnitudes of the global and regional O_2 loss diverge substantially between observationally based estimates (Bindoff et al., 2020; IPCC Special Report on the Ocean and Cryosphere, Chapter 5, page 471), which reduces the ability to accurately assess model fidelity and improvement.

Disagreement in observational estimates of ocean O_2 arises due to substantial uncertainties associated with spatiotemporal gaps in observational coverage and statistical methods used to fill the data gaps (Bindoff et al., 2020). In part, this uncertainty is also due to natural, background internal climate variability (hereafter ICV) (Bindoff et al., 2020; Long et al., 2016), which acts to obscure forced, anthropogenic signals in the climate system that require decades of observations for detection of a trend (Henson et al., 2017; Long et al., 2016; Rodgers et al., 2015; Schlunegger et al., 2020). It remains an open question as to what extent the disagreement between observational estimates and historical projections of ESMs is due to sampling artifacts, ICV, and the interplay between these two sources of uncertainty.

Sampling bias and ICV are sources of uncertainty that will continue to confound the detection of trends, even with the anticipated expansion of the global observing system (Moltmann et al., 2019). The presence of such uncertainty poses a challenge to assessing the oceans' evolving aerobic capacity, a crucial task for monitoring ecosystem health, and discerning the impacts of climate mitigation efforts on oxygen levels (Hoegh-Guldberg et al., 2018; Gulev et al., 2021). It remains an open question as to what extent the expansion of the global observing system will enhance the monitoring of trends in ocean O_2.

To address these two open questions, we utilize global compilations of gridded observational datasets (Ito et al., 2017, 2024b; Ito, 2022; Roach and Bindoff, 2023) and the Max Planck Institute for Meteorology Grand Ensemble (MPI-GE), comprising 100-member ensemble simulations with MPI-ESM 1.1 (Maher et al., 2019). Using the large ensemble simulation as a testbed, we then quantify the influence of historical sampling bias on the reconstructed O_2 variability and trend. Second, we examine how future observing plans may influence the detection of ocean deoxygenation and its response to climate mitigation efforts. In this ensemble modeling study, we focus on the annual mean dataset to analyze interannual to decadal variability of O_2 from both the gridded observational dataset and the ensemble simulations. The gridded observational dataset is based on limited ocean measurements. The early gridded observational dataset (Ito et al., 2017) contains spatiotemporal gaps (i.e., no complete global gap filling has been conducted, which retains non-data regions as they are). More recent gridded observational dataset is based on advanced gap-filling methods, including machine learning methods with complete spatiotemporal gap fillings (Ito et al., 2024b; Roach and Bindoff, 2023; Sharp et al., 2024). Here, we focus on the uncertainty due to spatiotemporal gaps based on the early gridded observational dataset (Ito et al., 2017). This will enable us to address the combined uncertainty associated with spatiotemporal gaps and ICV using an ensemble modeling approach.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

The potential biases related to complete global gap filling, such as those introduced by interpolation methods (e.g. Ito et al., 2024a), are not addressed in this ensemble modeling study.

2 Methods

2.1 Large Ensemble Simulations and Observational Dataset, Comparison Considering Data Gaps

The large ensemble simulation is based on the Max Planck Institute Earth System Model version 1.1 (MPI-ESM 1.1) in low resolution (LR) configuration (Giorgetta et al., 2013). The atmospheric model, ECHAM6, has a T63/1.9 degree horizontal resolution (Stevens et al., 2013), and the ocean model, MPI-OM, has an approximately 1.5 degree horizontal resolution (Jungclaus et al., 2013). The 100 ensemble members were generated under CMIP5 historical and RCP scenario forcing (including RCP2.6, RCP4.5, and RCP8.5), extending from 1850 to 2099 (after 2005, RCP scenario forcing is applied). Each member starts from slightly different initial conditions generated by the pre-industrial control simulation. The MPI-OM incorporates the Hamburg Ocean Carbon Cycle model (HAMOCC 5.2) for ocean biogeochemistry (Ilyina et al., 2013). The land biogeography is represented by the model JSBACH (Reick et al., 2013). Details on the MPI's large ensemble simulations, the Grand Ensemble Simulations (MPI-GE), are on the project website (https://mpimet.mpg.de/en/research/modeling/grand-ensemble) and in the description paper (Maher et al., 2019). The ensemble modeling approach enables us to separate the signals emerging from the external forcing (e.g., greenhouse gases, volcanic eruptions, etc.) and ICV (Deser et al., 2020; Lehner et al., 2020; Maher et al., 2019). The signal that emerges from the external forcing is called the "forced response", defined as the ensemble mean, and the ICV is represented as the spread of ensemble members. Thereby, we can discern the role of ICV on decadal to multi-decadal trends.

We use four observation-based global O_2 datasets that have been publicly available. We use the gridded annual O_2 dataset from Ito et al., (2017) (hereafter, Ito17) as the basis for sub-sampling. The Ito17 dataset is based on the World Ocean Database 2013 (Boyer et al., 2013). The O_2 data from the World Ocean Database 2013 is binned to annual mean bases, quality control is applied, and the data is interpolated using a Gaussian weight onto a one-degree grid. The long-term climatology (1950-2015) has been subtracted from the data. Therefore, the Ito17 dataset consists of O_2 anomalies (ΔO_2). Note that, despite the interpolation procedures, the Ito17 data have not been fully gap-filled; therefore, regions with no data in 1-degree bins remain as missing values (Fig. 1b). We utilize this information on the observational data gaps for sub-sampling the model output. We also include a more recent gridded observational dataset from three different publications (Ito 2022; Ito et al., 2024b; Roach and Bindoff, 2023). Hereafter, we refer to the three datasets as Ito22, Ito24, and RB23. These gridded observational datasets (Ito22, Ito24, and RB23) are fully gap-filled using spatiotemporal interpolation and machine learning techniques. The gridded dataset of Ito22 is provided at a 1-degree resolution, excludes data from the Arctic Ocean, and covers the period from 1965 to 2020, based on annual O_2 anomalies (ΔO_2). RB23 has a native resolution of 0.5 degrees, which we interpolated to 1 degree for this analysis. It spans the period from 1955 to 2018 with annual mean values.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Ito24 is available at a 1-degree resolution and covers the period from 1965 to 2020 with monthly data. For consistency, we calculated annual means from the monthly data for our analysis. Long-term mean calculations for all datasets were performed from the earliest available period (1955 or 1965) up to 2015. All analyses in this study are based on the annual O_2 anomaly data (ΔO_2), which is calculated by subtracting the long-term climatology. No detrending has been applied, as we focus on both including signals from external forcing (e.g., greenhouse gases, volcanic eruptions) and ICV.

To facilitate comparison between observed estimates of ΔO_2 and those of the large ensemble simulations, we sub-sampled the annual model output based on the time-varying data coverage information from the gridded annual dataset of Ito et al., (2017) (Fig. 1b, and d). This allows us to quantify the observational sampling bias in estimating the ICV and forced response. In this study, we focus on O_2 near the thermocline (at around 300m depth), which is commonly used in observational and modeling studies of ocean deoxygenation (Oschlies et al., 2017; Schmidtko et al., 2017; Stramma et al., 2012). The O_2 minimum begins to emerge near the thermocline, and the observed low-frequency variability and trend of thermocline O_2 are among the challenging ocean biogeochemical properties to simulate by current models (Oschlies et al., 2017; Stramma et al., 2012). We note that there is no consensus for a standard ocean deoxygenation metric from previous studies (Bopp et al., 2013; Cabré et al., 2015; Kwiatkowski et al., 2020; Oschlies et al., 2017, 2018; Stramma et al., 2012; Takano et al., 2023). The O_2 metric we use here focuses on the upper ocean, which may be more crucial for marine ecosystems. With spatiotemporal sub-sampling in mind for the large ensemble simulations, we construct weighted-area mean ΔO_2 time series from both observations and all 100 ensemble simulations. The sub-sampling analysis is extended to assess near-future deoxygenation under three different climate change scenarios (RCP2.6, RCP4.5, and RCP8.5) (Meinshausen et al., 2011). Further details on sub-sampling methods will be introduced in the latter sections.

The spatiotemporal data coverage based on the Ito17 dataset is summarized in a spatial map and a time series plot (Fig. 1b and the bar chart in Fig. 1d). Fig. 1b shows the temporal data coverage at 1-degree resolution, consistent with the methodology used in the supplementary information of Ito et al. (2017) and reproduced here using publicly available data. Temporal coverage is quantified by counting the number of available annual data points at each grid cell. The count is then divided by 58 (the number of years from 1958 to 2015) and multiplied by 100 to obtain the percentage coverage for each grid point. The global time series of spatial coverage represents the proportion of 1-degree ocean grid cells that contain data in each year (Fig. 1d). A value of 100% indicates that all ocean grid cells have data in that year. For each year from 1958 to 2015, we assess whether the spatial coverage exceeds 50%. The total number of such years (up to 58) is displayed in the upper right corner of Fig. 1d and is referred to as the *Spatiotemporal Coverage Metric* (STCM). The global STCM is 26. This metric can also be calculated regionally and will be used to rank ocean regions based on their spatiotemporal data coverage. Over the 1958–2015 historical period, the North Atlantic Ocean has the highest STCM (55), while the South Pacific Ocean has the lowest (8).

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

2.2 Observational Sub-sampling of the 100-Member MPI-GE

To address the gaps in observational data coverage, we developed a sub-sampling method. We applied it to annual model output from the MPI-GE, 100 ensemble simulations from an Earth System Model. We note that the month-to-month variability (including seasonal cycles) is not addressed in our study, and the focus is on annual to decadal variability, partly because of the limitation in output frequency of the three-dimensional ocean biogeochemical variables; only annual model output is available for most of the simulations (Maher et al., 2019). One potential bias could stem from the fact that more measurements are taken in a particular season (e.g., more observations in summer compared to winter in certain regions), which is not considered in this study. This could impact the observed signal, possibly underestimating the annual variability in the observations. We first subtracted the long-term climatology (1950-2015) from all annual mean model output (i.e., O_2, temperature, and salinity) to calculate the annual anomaly data. Temperature and salinity are needed for calculating oxygen saturation and Apparent Oxygen Utilization (AOU) (explained in an upcoming section). We utilize the annual anomaly data from the model at 310m (denoted as $\Delta O2\_300$) from 1958 to 2015 for analysis in this study. The annual time-dependent sub-sampling mask based on the Ito17 dataset had been projected onto MPI-OM's native grid (GR15, approximately 1.5 degrees) using the nearest neighborhood interpolation method. Finally, we applied the time-dependent sub-sampling mask to $\Delta O2\_300$ fields from the model output. It is important to note that the Ito17 dataset is based on global compilations of bottle and CTD data, which are then subject to a gap-filling procedure. Reconstruction of a gridded observational-like dataset from models requires high-resolution model output in both time and space to extract data points closely approximating the locations of data contained in the original database. Therefore, the gap-filling method is not directly applied to the model output due to the limitation in the spatiotemporal resolution of the model output. Instead, we begin our discussion by analyzing gridded products and focusing on the broader consequences of limited observational data coverage. Therefore, we do not address bias due to gap-filling methods in this study.

The primary diagnostic we use in this study is $\Delta O2\_300$ (ΔO_2 at 300m for the gridded observational dataset and at 310m for the model output). Weighted-area mean $\Delta O2\_300$ time series are calculated using two different approaches from MPI-GE. One uses sub-sampled model output that is masked according to the observational data coverage information from 1958-2015, as introduced in the previous section (Section 2.1 and Fig. 1b and 1d). The other approach utilizes all of the model output in time and space (i.e., full sampling, with no sampling bias) (see Fig. 1a and 1c for the global mean time series based on two different approaches). Feeding in the spatial and temporal sparsity of the observational coverage allows us to compare the observational and model estimates of $\Delta O2\_300$ variability and trend. The advantage of this sub-sampling approach is that the model then also includes a potential bias due to uneven sampling in both time and space. This could eventually influence the estimate of $\Delta O2\_300$ trends, which is an essential indicator of ocean deoxygenation.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

2.3 Ensemble Statistics

Within the ensemble members, we calculate the $\Delta O2_{\_300}$ time series for both full-sampled (e.g., Fig. 1a) and sub-sampled (e.g., Fig. 1c) cases. We quantify the ICV based on the spread of ensemble members calculated as the ensemble standard deviation. The uncertainties due to ICV have been quantified in other large ensemble modeling studies, and this metric is commonly used to estimate the range of ICV based on large ensemble simulations (Deser et al., 2020; Lehner et al., 2020; Maher et al., 2019). To quantify the impact of limited sample size on ICV of $\Delta O2_{\_300}$, we calculate the ensemble standard deviation (SD) for each year for both full- and sub-sampled $\Delta O2_{\_300}$ time series and then take the ratio of the ensemble SD of the two cases as written as follows:

$$Ens\ SD_{ratio} = \frac{SD_{sub-sampled}}{SD_{full-sampled}} = \frac{\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(x_{m,sub-sampled} - \overline{x_{sub-sampled}})^2}}{\sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(x_{m,full-sampled} - \overline{x_{full-sampled}})^2}}$$

The subscripts *sub* and *full* denote sub-sampled and full-sampled members, *m* is the ensemble member, and the overbar denotes the ensemble mean. *M* equals 100, as that is the number of ensemble members for MPI-GE. Note that the Ens SD ratio (and sub-sampled and full-sampled SDs are all time-dependent. The presentation of the Ens SD ratio (e.g., Fig. 1d, red line) is multiplied by 100 and expressed as a percentage.

To estimate multi-decadal ocean deoxygenation, we calculated linear trends using the least squares method, applied to both the time series and the gridded datasets that combine observations with ensemble simulations. The results are summarized in box-and-whisker plots (Fig. 3), patterns (Fig. 4), and histograms (Fig. 6), which are based on linear trends derived from time series of $\Delta O2_{\_300}$. Model estimates include both fully sampled cases (without sampling bias) and observationally sub-sampled cases. The spread of linear trends in the box-and-whisker plots represents the range of trends attributable to ICV, effectively illustrating the ensemble probability density function (PDF) of linear trends from ensemble members. The ensemble mean of these trends corresponds to the component driven by the forced response.

2.4 Oxygen Saturation and Apparent Oxygen Utilization (AOU) Calculations

To quantify the effects of oxygen saturation-driven changes and circulation-biology-driven changes in oceanic dissolved oxygen, we utilized temperature and salinity data from the ocean reanalysis and model output. The ocean reanalysis is from the ECMWF ORA-S4 (Balmaseda et al., 2013), following the observational study by Ito et al., 2017. The oxygen saturation ($O_{2,sat}$) calculation is based on Weiss's formula (Weiss et al., 1970), the same formula used in the model (Ilyina et al., 2013). The circulation-biology driven change in dissolved oxygen is quantified using Apparent Oxygen Utilization (AOU), defined as follows,

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

$$AOU = O_{2,sat}(T,S) - O_2$$

O_2 is the dissolved oxygen concentration from observations or model output, $O_{2,sat}$ is calculated using temperature (T) and salinity (S), and AOU is the residual between the two. In this study, we focus on analyzing $O_{2,sat}$, and AOU anomalies at around 300m (denoted as $\Delta O_{2,sat\_300}$ and $\Delta AOU_{\_300}$) following ΔO_{2_300}.

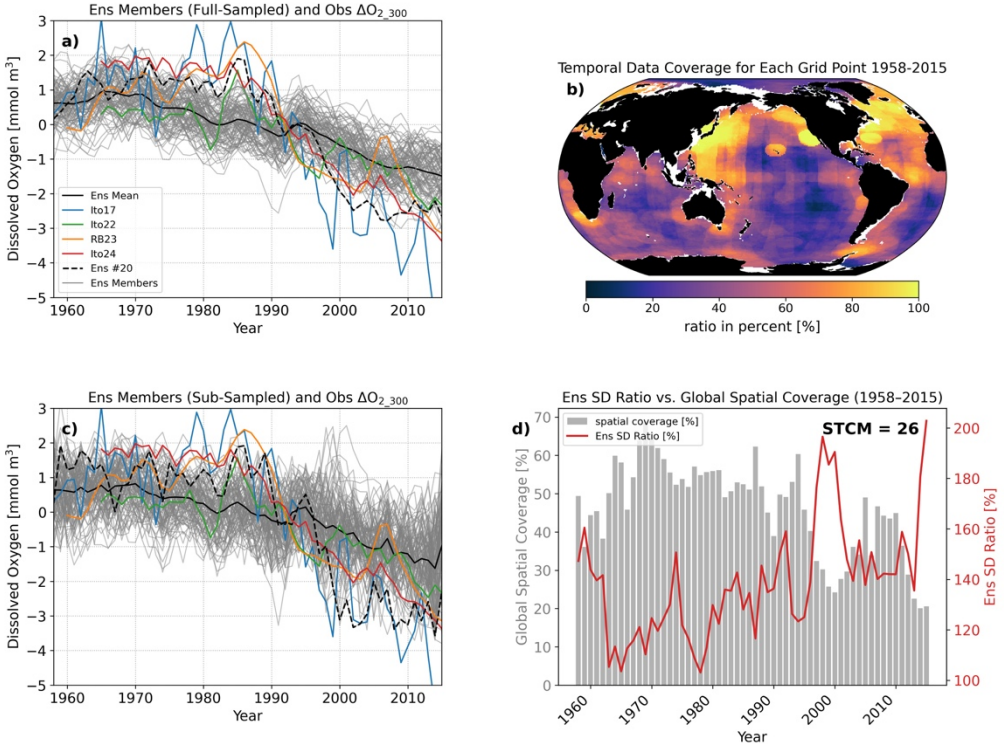


Figure 1: Temporal evolution of global mean oxygen anomalies at approximately 300 m depth, derived from four observational datasets (at 300 m) and large ensemble simulations (at 310 m, MPI-GE). Panels a) and c) show simulated ΔO_{2_300} time series based on a) fully sampled model output and c) observationally sub-sampled model output, respectively, for the period 1958–2015. In panels a) and c), gray lines represent ΔO_{2_300} time series from 100 ensemble members (historical simulations extended with RCP8.5 to 2015). The black solid line shows the ensemble mean (forced response), and the black dashed line indicates the time series from ensemble member #20. Member #20 shows the largest negative global ΔO_{2_300} trend and is highlighted as an example of pronounced ocean deoxygenation. Colored solid lines correspond to observational datasets: Ito et al. (2017) (Ito17), Ito (2022) (Ito22), Roach and Bindoff (2023) (RB23), and Ito et al. (2024b) (Ito24). Panel b) shows the percentage of annual data coverage at each grid point from 1958–2015, based on the gridded O_2

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

data from Ito17 at 300 m. Panel d) presents the global spatial coverage as a percentage of 1-degree ocean grid cells containing data for each year (100% indicates full coverage). The red line in panel d) shows the ratio of ensemble standard deviation (Ens SD ratio) of the global mean ΔO_{2_300} time series (see Section 2.3 for details). The STCM is the *Spatiotemporal Coverage Metric* (STCM) (see Section 2.1 for details).

3 Quantifying Observational Sampling Uncertainties

3.1 Global and Regional Variability of Dissolved Oxygen near the Thermocline

We first examine the ΔO_{2_300} time series from the ensemble simulations and assess observational sampling biases in both global and regional annual ΔO_{2_300} anomalies, based on the large ensemble simulations (see Section 2.3 for details). The regional partitions follow Schmidtko et al. (2017) (Fig. 2a). We compare the simulated ΔO_{2_300} with four gridded observational datasets derived from various sources, including bottle samples, CTD casts, and Argo floats. These datasets show substantial differences in their ΔO_{2_300} time series (Figs. 1ac and 2), primarily due to differences in data sources, coverage, and gap-filling methods. The Ito17 dataset exhibits significant temporal and spatial gaps (Figs. 1b, 1c), which may lead to higher apparent interannual and decadal variability due to the limited gap-filling and smoothing capabilities. In contrast, the Ito22 dataset applies global interpolation and a five-year running mean to address data gaps, resulting in a smoother temporal and spatial evolution of ΔO_{2_300} (Figs. 1ac, 2; green lines). We regard Ito22 as a conservative estimate of oxygen variability and trends. The more recent datasets, Ito24 and RB23, fall between these two in terms of variability and trend. They show similar decadal trends in ΔO_{2_300} but differ in interannual variability (Figs. 1ac, 2; orange and red lines), reflecting differences in data sources (e.g., the inclusion of BGC-Argo in Ito24) and gap-filling approaches. Keeping these observational differences and spatiotemporal gaps in mind, we further analyze the simulated ΔO_{2_300} variability and trend in comparison with these four datasets.

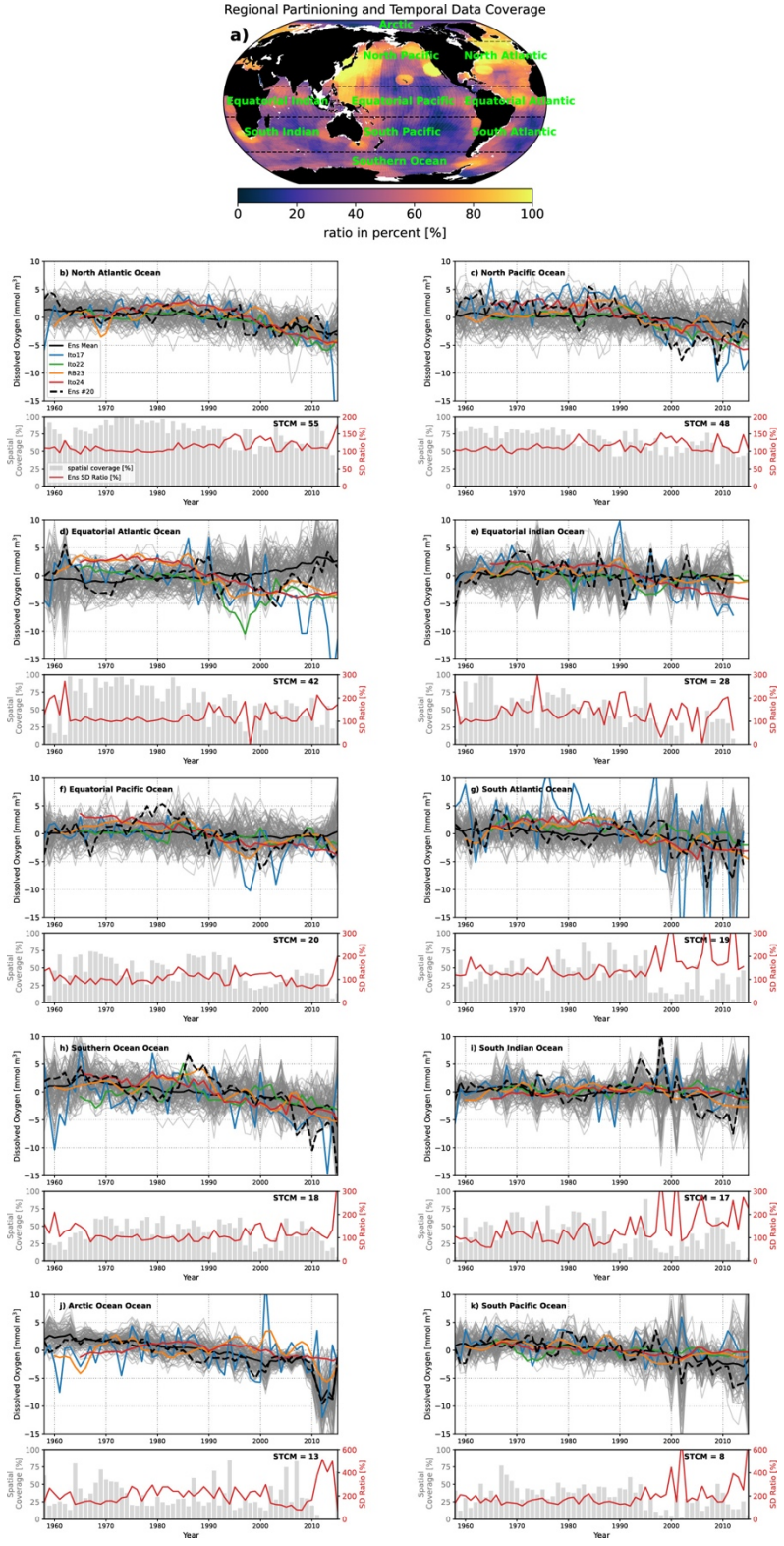
In general, the variability of both global and regional ΔO_{2_300} in observations is broadly consistent with the spread of ensemble model simulations (ICV), particularly when the effect of observational data gaps is considered (Figs. 1ac and 2). Although the ΔO_{2_300} variability from the Ito17 dataset tends to lie near the edge or even outside of the ensemble spread in some regions, the other three datasets generally fall within the model range. The global ΔO_{2_300} time series from ensemble member #20 closely resembles the observed ΔO_{2_300} time series, suggesting that ICV could explain multi-decadal deoxygenation trends similar to those observed. However, it is essential to consider how temporal and spatial variations in data coverage influence the observed time series. When the data coverage decreases in certain years, the global and regional ΔO_{2_300} time series tend to exhibit pronounced positive or negative peaks, reflecting signals from specific regions. This feature is apparent in both observational datasets, particularly highlighted in the Ito17 dataset, as well as in ensemble simulations. Moreover, unresolved physical and biogeochemical processes in the model may also affect regional ΔO_{2_300}

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

variability. Limited spatial coverage in observations can emphasize localized features, making the observed time series more sensitive to regional anomalies. Therefore, without accounting for observational sampling biases, direct comparisons between modelled and observed ΔO_{2_300} variability may be misleading. Sub-sampling the model output based on the spatiotemporal coverage of observational data increases the ensemble spread, indicating that sampling bias has a significant impact on ICV estimation.

It should also be noted that observational datasets are subject to uncertainties arising from factors such as quality control procedures, high-frequency variability, and the methods used for gap filling and spatiotemporal interpolation (Ito et al., 2022, 2024a, 2024b; Roach and Bindoff, 2023; Schmidtko et al., 2017). The magnitude of these uncertainties could range from -10 to 10 mmol/m³ (Schmidtko et al., 2017) and is generally of the same order as the ICV spread derived from ensemble simulations. This highlights the need for caution when directly comparing observational data with model outputs. Furthermore, the ΔO_{2_300} variance in the fully sampled case tends to align more closely with the Ito22 dataset, likely because the spatial and temporal gaps have been filled using interpolation and smoothing techniques, which suppress variability. Similarly, ensemble simulations subjected to full spatiotemporal sampling also exhibit smoothed and attenuated variability (Fig. 1a).

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025



https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Figure 2: Temporal evolution of regional mean oxygen anomalies at approximately 300 m depth, derived from four observational datasets (at 300 m) and large ensemble simulations (at 310 m, MPI-GE). Panel a) displays a map with regional partitioning overlaid on the percentage of annual data coverage at each grid point from 1958 to 2015 (the same as in Fig. 1b). Panels b) to k) show simulated $\Delta O2_{\_300}$ time series based on observationally sub-sampled model output for the period 1958–2015. In the upper rows in panels b) to k), gray lines represent $\Delta O2_{\_300}$ time series from 100 ensemble members (historical simulations extended with RCP8.5 to 2015). The black solid line shows the ensemble mean (forced response), and the black dashed line indicates the time series from ensemble member #20. Member #20 shows the largest negative global $\Delta O2_{\_300}$ trend and is highlighted as an example of pronounced ocean deoxygenation. Colored solid lines correspond to observational datasets: Ito et al. (2017) (Ito17), Ito (2022) (Ito22), Roach and Bindoff (2023) (RB23), and Ito et al. (2024b) (Ito24). In the lower rows in panels b) to k), the gray bar charts present the regional spatial coverage as a percentage of 1-degree ocean grid cells containing data for each year (100% indicates full coverage). The red lines in panels b) to k) show the ratio of ensemble standard deviation (Ens SD ratio) of the regional mean $\Delta O2_{\_300}$ time series (see Section 2.3 for details). The STCM is the *Spatiotemporal Coverage Metric* (STCM) (see Section 2.1 for details).

To quantify the uncertainty in ICV arising from observational sampling bias, we calculate the ratio of ensemble standard deviations (ens SD ratio) between the observationally sub-sampled and fully sampled cases, using 100 ensemble members. The ensemble SD represents the magnitude of ICV variability captured within the large ensemble simulation. By comparing the SDs of the two sampling approaches, we assess the extent to which spatiotemporal data gaps introduce uncertainty into ICV estimates (see Section 2.3). When the ens SD ratio (Figs. 1ac and 2) exceeds 100%, the ICV of $\Delta O2_{\_300}$ at that time is overestimated relative to the fully sampled case; conversely, ratios below 100% indicate underestimation. The ens SD ratio varies over time and generally remains above 100%, but can exceed 500% in certain regions, particularly when observational coverage is sparse. This result highlights that spatiotemporal gaps in observations introduce a time-varying bias into the $\Delta O2_{\_300}$ time series, most often leading to overestimation of ICV. Although rare, there are cases where the ens SD ratio drops below 100%, indicating a potential underestimation of ICV. For example, this occurs in Fig. 2d for the Equatorial Atlantic around 1998 and in Fig. 2f for the Equatorial Pacific around 2010, possibly due to disproportionate sampling of regions with inherently low ICV. Importantly, these over- and underestimations of ICV are not apparent from observational datasets alone, due to their limited spatial and temporal coverage and difficulties in disentangling ICV from observational datasets. Thus, these sampling-related biases represent a form of **hidden uncertainty**, uncertainties that can only be revealed when combining observational data with ensemble-based modeling. Therefore, caution is warranted when interpreting ICV estimates based solely on observational datasets.

Another notable feature is the decrease in spatial data coverage observed in many regions since the 1980s (Figs. 1d and 2). This reduction is associated with an overall increase in the ensemble SD ratio, suggesting that time-varying **hidden**

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

uncertainty may affect estimates of the long-term trend in ΔO_{2_300}. Moreover, extremely sparse data coverage at certain times can lead to artificial signals in the ΔO_{2_300} time series, due to large fluctuations in the ICV range, as observed in the ensemble-based time series analyses. For example, the spike detected in the South Pacific Ocean in 2002 (Fig. 2k) may be a result of sampling bias due to the limited number of observations. The ensemble spread notably widens in this year, highlighting the potential influence of sampling variability on ICV estimates. These data gaps could thus introduce biases into the ΔO_{2_300} time series, complicating the detection and interpretation of long-term oxygen trends. Consistently across regions, the ens SD ratio of ΔO_{2_300} tends to increase as spatial data coverage decreases (Figs. 1d and 2). This relationship highlights the importance of considering observational sampling effects, particularly when evaluating linear trends in ΔO_{2_300} to assess ocean deoxygenation.

3.2 Effects of Observational Sampling Bias on Dissolved Oxygen Trend

To assess the impact of observational sampling bias on the estimated trends of ΔO_{2_300}, we analyze both observational data and ensemble simulations, applying the same spatiotemporal data gaps (Figs. 1b and 1d), over two distinct periods: 1981–2000 (representing a 20-year, shorter-term trend) and 1961–2010 (representing a 50-year, longer-term trend). The 20-year period was selected because the observed global ΔO_{2_300} concentration shows a marked decline after the 1980s (Figs. 1a and 1c). Additionally, on this timescale, the influence of ICV on the linear trend is relatively large, as the ΔO_{2_300} spectrum exhibits enhanced variance at decadal scales (Ito and Deutsch, 2010). Short-term trends are therefore more sensitive to ICV-related uncertainties than longer-term trends. Including the 50-year period allows us not only to provide a longer-term comparison but also to place our findings in the context of previous studies on the detection and quantification of anthropogenic deoxygenation, based on both observational and modeling approaches (e.g., Ito et al., 2017; Oschlies et al., 2017).

A comparison of ΔO_{2_300} trends between the observationally sub-sampled and fully sampled ensemble members shows a tendency for trend overestimation in data-sparse regions, both for the 20-year and 50-year periods (Figs. 3a and 3b). For global ΔO_{2_300} trends, a standard metric used in previous studies (e.g., Schmidtko et al., 2017), the ranges of the 20-year and 50-year trends are 0.32 and 0.12 mmol m^{-3} yr^{-1} for the sub-sampled cases, and 0.27 and 0.09 mmol m^{-3} yr^{-1} for the fully sampled cases, respectively (Figs. 3a and 3b; values indicated above box plots). Although the range of trends slightly increases in the sub-sampled cases, this increase is much smaller than the amplification seen in the ens SD ratio shown in Fig. 1d and 2. While the ens SD ratio, reflecting variability in the ΔO_{2_300} time series due to sampling bias, can reach up to 500%, the corresponding increase in trend range is limited to approximately 150%. Notably, both increases and decreases in trend ranges are also observed in some data-dense regions, such as the North Atlantic and Pacific Oceans (e.g., Fig. 3a, Natl and NPaci), suggesting that uncertainties in the ΔO_{2_300} time series do not directly scale to the 20-year trend estimates. This may be because trend estimates are sensitive to the choice of start and end years (Fay et al., 2014; McKinley et al., 2011) and

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

because multi-decadal averaging reduces the influence of high-frequency variability that contributes to ensemble spread (Schlunegger et al., 2019, 2020).

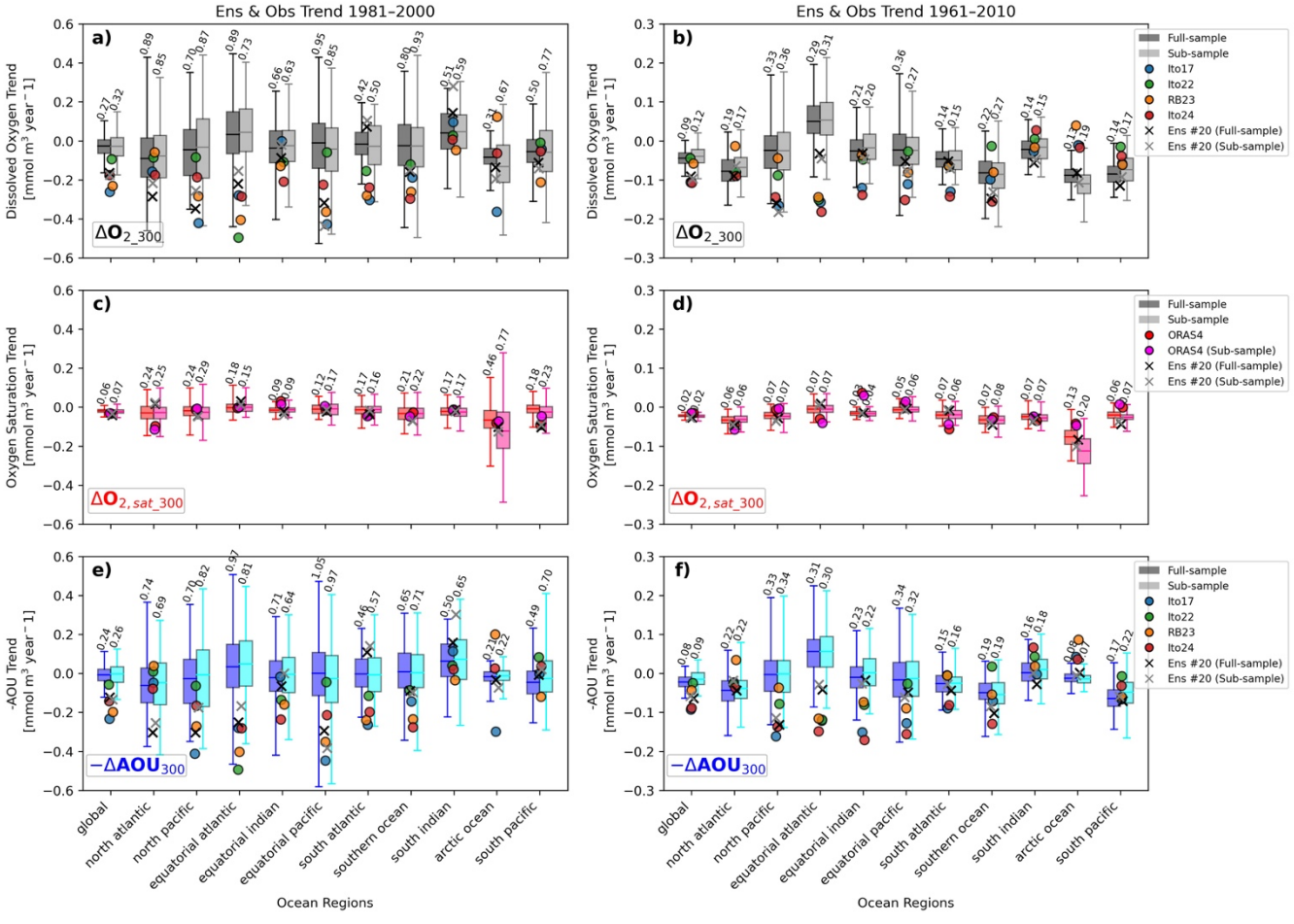


Figure 3: Box-whisker plots of linear trends in observed and simulated global and regional ΔO_{2_300}, oxygen saturation ($\Delta O_{2\_sat\_300}$), and apparent oxygen utilization ($-\Delta AOU_{300}$) over two periods: 1981–2000 and 1961–2010. Panels a)–f) show box-whisker plots of global and regional trends based on both fully sampled and sub-sampled time series. The regions are ordered from most data-rich (e.g., the North Atlantic Ocean) to most data-sparse (e.g., the South Pacific Ocean), according to the Spatiotemporal Coverage Metric (STCM; see Section 2.1). The plots display the ensemble mean and range of ΔO_{2_300}, $\Delta O_{2\_sat\_300}$, and $-\Delta AOU_{300}$ trends across all ensemble members for fully sampled cases (dark-colored boxes) and sub-sampled cases (light-colored boxes). Black and gray crosses indicate trends from ensemble member #20, based on fully sampled and sub-sampled time series, respectively. Member #20 shows the largest negative global ΔO_{2_300} trend and is

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

355 highlighted as an example of pronounced ocean deoxygenation. Colored circles represent observed trends for $\Delta O2_{\_300}$ and -$\Delta AOU_{\_300}$ from four datasets: Ito et al. (2017) (Ito17), Ito (2022) (Ito22), Roach and Bindoff (2023) (RB23), and Ito et al. (2024b) (Ito24). The $\Delta O2_{\_sat\_300}$ values are calculated from temperature and salinity data from ECMWF ORA-S4 and are used in all -$\Delta AOU_{\_300}$ calculations. Note that $\Delta AOU_{\_300}$ values are multiplied by −1 to align the sign convention with $\Delta O2_{\_300}$ changes.

360

To further evaluate the robustness of the observational sampling strategy, we analyze the contributions of thermal and biogeochemical mechanisms to the $\Delta O2_{\_300}$ trends (Figs. 3c-f). Specifically, we decompose $\Delta O2_{\_300}$ trends into a saturation-driven component ($\Delta O2_{,sat\_300}$), primarily determined by the ocean's thermal state, and a component driven by circulation and biological processes, represented by apparent oxygen utilization (-$\Delta AOU_{\_300}$). Our results suggest that, outside the Arctic
365 Ocean, $\Delta O2_{,sat\_300}$ is not the dominant driver of multi-decadal deoxygenation. Instead, the large ensemble simulations indicate that the pronounced deoxygenation observed during the 1980s–1990s is primarily attributed to changes in circulation and biology, as reflected in the magnitude of the -$\Delta AOU_{\_300}$ component, a finding consistent with previous studies (Ito et al., 2017; Takano et al., 2023). Overall, the sub-sampled ensemble $\Delta O2_{\_300}$ trends closely match the fully sampled trends, with sub-sampling introducing a bias of approximately 5%–25% in most regions (Fig. 3). This relationship
370 holds not only for total $\Delta O2_{\_300}$ trends but also for the individual components, $\Delta O2_{,sat\_300}$ and -$\Delta AOU_{\_300}$, except for the $\Delta O2_{,sat\_300}$ trend in the Arctic, where deviations are larger. The equatorial Atlantic Ocean is the only region where the 50-year ensemble trends consistently show positive values (i.e., box plots in the positive range in Fig. 3b). Although tropical oxygenation is a common feature in future Earth system model (ESM) projections (Bopp et al., 2013; Cabré et al., 2015; Kwiatkowski et al., 2020; Takano et al., 2018), it has yet to be confirmed in observational records, possibly due to masking
375 by both sub-sampling bias and ICV. Alternatively, this feature may reflect a systematic bias in model projections.

The observed global $\Delta O2_{\_300}$ decreased by approximately 6 mmol m^{-3} over the 20 years after the 1980s, following a period of relative stability during the 1960s and 1970s (black lines in Figs. 1a and 1c; see also Ito et al., 2017). Our analysis suggests that this decline cannot be entirely attributed to the forced response alone, implying a substantial contribution from
380 ICV. Supporting this, the ens SD ratio increased after the 1980s, reaching around 200% by the year 2000 (Fig. 1d). This increase coincides with the period of pronounced observed decline in global $\Delta O2_{\_300}$. The 20-year observed global $\Delta O2_{\_300}$ trends during the 1980s partly fall outside the range of trends simulated by the ensemble, which accounts for ICV (Fig. 3a). Specifically, two observational estimates (Ito17 and RB23) exceed the simulated range, while the other two (Ito22 and Ito24) remain within it, underscoring the uncertainties inherent in the observed trends. Regional observations indicate that the
385 global decline is driven by consistent negative trends across multiple regions. In most regions, the observed 20-year $\Delta O2_{\_300}$ trends fall below the lower quartile (25th percentile) of the ensemble distribution (see colored circles in Figs. 3a and 3b), though this varies by dataset. Ensemble simulations highlight the North Atlantic and North Pacific as key contributors to the

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

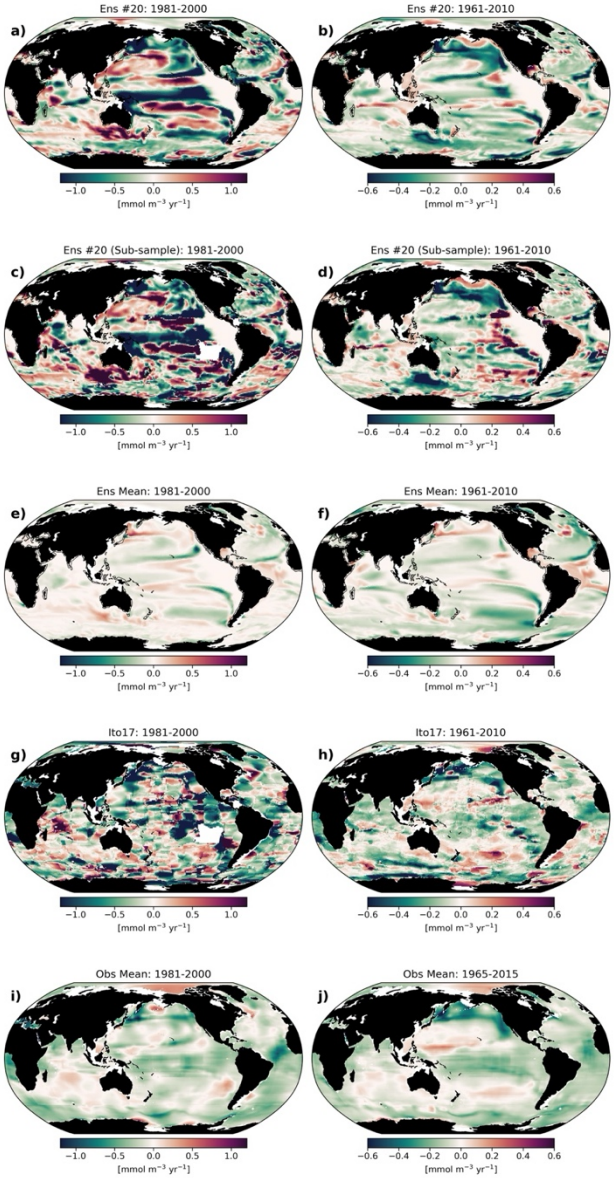
global decline, partly due to higher data coverage in these regions. The 20-year trends in these basins frequently reach values below -0.4 mmol m^3 yr^{-1}. While most observed regional trends fall within the ensemble range, the models generally fail to reproduce the magnitude of the global observed $\Delta O2_{\_300}$ trend reported by three datasets (Ito17, RB23, and Ito24). This discrepancy likely arises because strong negative trends in different regions do not always occur simultaneously in individual ensemble members. This point is exemplified by ensemble member #20, which exhibits the largest negative global $\Delta O2_{\_300}$ trend (see, for instance, the South Atlantic trends for member #20, marked by black and gray crosses in Fig. 3). The 50-year global $\Delta O2_{\_300}$ trend exhibits similar behavior, although regional trends differ slightly from the 20-year results (Figs. 3a and 3b). As expected, the magnitudes of 50-year trends are smaller, yet the influence of ICV remains significant in many regions, as indicated by the comparison between the forced trend and the ensemble spread over the 1961–2010 period.

3.3 Effects of Observational Sampling Uncertainties on Spatial Patterns of Dissolved Oxygen Trend

The large ensemble simulations and observed $\Delta O2_{\_300}$ trends exhibit diverse spatial patterns at both the 20-year and 50-year timescales, mainly driven by internal climate variability (ICV) (Fig. 4). Although trend magnitudes are generally larger over the shorter term, ICV strongly influences both timescales, as shown by comparisons between single ensemble members and the forced response (Figs. 4ab, 4ef). For comparison, we selected ensemble member #20, which displays the largest negative global $\Delta O2_{\_300}$ trend among the 100 ensemble members (see global ocean in Figs. 3ab). To recall, this member's global $\Delta O2_{\_300}$ timeseries most closely resembles the observed $\Delta O2_{\_300}$ timeseries. The ensemble member shows diverse patterns, with both strong negative and positive regional trends. The forced response contributes locally, such as near oxygen minimum zones (OMZs) and parts of the Southern Ocean, but ICV largely determines the overall trend patterns. The observational trend based on Ito17 is notably spatially heterogeneous compared to the model results (Figs. 4g-h), likely reflecting sparse and uneven observational coverage. This heterogeneity is also evident in comparisons between fully sampled and observationally sub-sampled ensemble cases (Figs. 4a-d). In contrast, the mean trend from the three recent observational datasets (Ito22, RB23, and Ito24; Figs. 4ij) appears more spatially smooth. This may result from the ensemble averaging of datasets, which tends to smooth regional differences. Nonetheless, negative trends in the North Pacific, tropical Pacific, and Southern Ocean appear consistent across datasets. Our results also indicate that sampling bias can lead to sign reversals in some regions. For example, in the central equatorial and southeastern Pacific, the 50-year trend differs between sampling cases (Figs. 5a–d). Another factor contributing to spatial heterogeneity may be model limitations in resolving mesoscale processes, such as eddies, which coarse-resolution models do not capture. Furthermore, the model tends to underestimate both the variability and trends of $\Delta O2_{\_300}$ within OMZs when compared to observations (Figs. 4a-d, 4g-j). Similar biases have been reported in other CMIP5 and CMIP6 models (Cabré et al., 2015; Ilyina et al., 2013; Séférian et al., 2020). In MPI-ESM 1.1, OMZ extent is generally overestimated (see also Fig. A1). In these regions, near-zero climatological oxygen levels suppress oxygen consumption, which may reduce the variability and trend amplitude of $O2_{\_300}$. Finally, the number of observations used in our sub-sampled trend analysis varies across grid points. In some areas, limited observational coverage may result in local $\Delta O2_{\_300}$ trends appearing opposite to the true underlying trend, a limitation not

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

apparent when examining observational data alone. We also note that our assessment of sampling bias is based on model ensemble simulations, which may overestimate the confidence in the ability of sampling bias correction to accurately capture regional trend directions.



425

Figure 4: Spatial patterns of ΔO_{2_300} trends over two periods, 1981-2000 (20-year) and 1961-2010 (50-year), from a single ensemble member (#20), the forced trend (ensemble mean), and observational datasets. Panels a) and b) show the 20-year and 50-year trends from ensemble member #20, based on fully sampled data. Panels c) and d) show the corresponding trends

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

from ensemble member #20, based on observationally sub-sampled data. Panels e) and f) display the fully sampled trends based on the ensemble mean (forced trend). Sub-sampling is applied to the model output at each grid point and time step, and linear trends are calculated from the sub-sampled time series. Ensemble member #20, selected for having the largest negative global $\Delta O2_{\_300}$ trend among the 100 ensemble members, is highlighted as an example of pronounced ocean deoxygenation. Panels g) and h) show $\Delta O2_{\_300}$ trends derived from the Ito17 observational dataset. Panels i) and j) present trends based on the mean of three observational datasets: Ito22, RB23, and Ito24.

4 Implications of Near-Future Observational Sampling Bias for Interpreting Global Deoxygenation

4.1 Extending Sub-Sampling to Near-Future Projections

The Biogeochemical Argo (BGC-Argo) program (https://biogeochemical-argo.org/) is an international initiative that deploys free-drifting robotic floats to collect high-quality biogeochemical data from the global ocean. The program aims to establish and maintain a global array of approximately 1,000 autonomous profiling floats for long-term ocean monitoring (Biogeochemical-Argo Planning Group, 2016). With current and planned BGC-Argo deployments in mind, we designed a set of hypothetical sub-sampling experiments using model outputs from near-future projections (2020–2039). The overview of these experimental designs is presented in Fig. 5. The experiments explore the potential impacts of limited observational coverage on estimating near-future decadal $\Delta O2_{\_300}$ trends. We conducted three types of sub-sampling experiments - **sustain, plan, and cut** - based on random sampling from the native model output (MPI-OM), with an approximate horizontal resolution of 1.5 degrees (~150 km). These scenarios reflect varying assumptions about the future evolution of BGC-Argo float coverage after 2020 (Stoer et al., 2023). We applied a Python random number generator to select model grid points annually, with sample locations varying over time.

- In **sustain**, the sample size remains fixed at $N = 460$ (the approximate number of floats in 2020, Stoer et al., 2023) throughout the 20-year period (2020–2039).
- In **plan**, the sample size gradually increases from $N = 460$ to $N = 1{,}059$ by 2028, reaching the global coverage target of ~1,000 floats, and remains stable thereafter.
- In **cut**, we simulate a hypothetical scenario where the sample size drops by half in 2026 (from $N = 834$ in 2025 to $N = 417$), representing an unforeseen reduction in measurements.

These sub-sampling scenarios are designed to test: (1) if the sampling effort remains unchanged, (2) if it increases as planned until 2028 and stabilizes afterward, and (3) if it unexpectedly declines from 2026. We assess how these scenarios affect the estimation of near-term ocean deoxygenation by analyzing their impact on the 20-year (2020-2039) linear trend of $\Delta O2_{\_300}$.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Our experiments assume that, given the model's horizontal resolution (approximately 1.5 degrees), each grid cell represents variability over at least that scale. By comparison, the original Argo design was based on a horizontal decorrelation scale of about 300 km (about 3 degrees), treating individual profiles as statistically independent at that spacing. Thus, the model resolution nominally resolves finer-scale variability than the Argo sampling design. However, actual ocean decorrelation scales vary widely (25–400 km), depending on region, depth, and dynamics (Purkey and Johnson, 2010). Hence, a 1.5-degree scale may overlook finer features or anisotropic patterns. Finally, we assume that BGC-Argo floats remain stationary and sample from fixed model grid points, following the approach used in Majkut et al. (2014) for surface CO_2 studies. While simplifying the experimental design, this assumption may introduce bias into the reconstructed data fields (Kamenkovich et al., 2011).

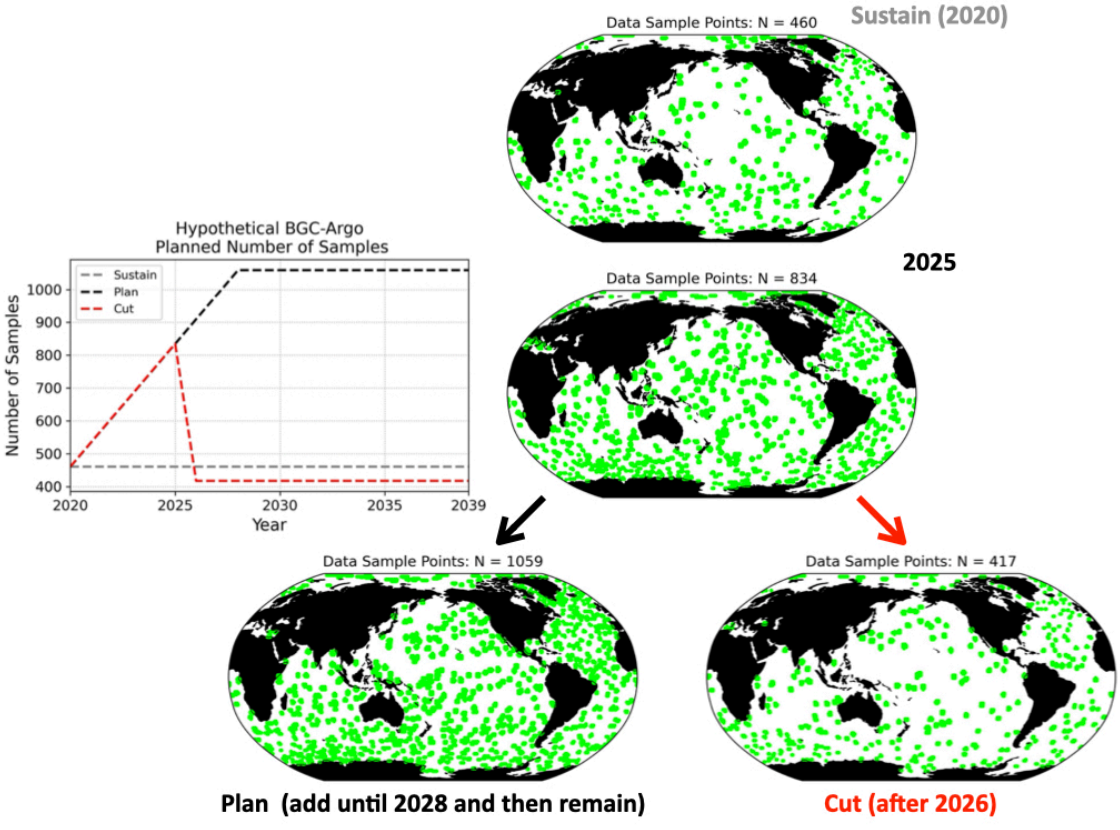


Figure 5: Schematic illustration of the three sub-sampling experiment designs based on model output. The time series plot illustrates the evolution of sample numbers under various scenarios (**sustain, plan, and cut**), inspired by current and planned BGC-Argo deployments (Biogeochemical-Argo Planning Group, 2016; Stoer et al., 2023). The maps provide examples of randomly selected sample locations corresponding to different total sample sizes.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

4.2 Implications of Future Observational Sampling Bias for Quantifying Near-Future Global Deoxygenation

What are the potential consequences of limited observational coverage on detecting near-future deoxygenation under global warming? Using three near-future sub-sampling experiments (see Section 4.1) based on large ensemble simulations, we assess how varying sampling sizes influence the detection of dissolved oxygen ($\Delta O2_{\_300}$) trends from 2020-2039. We begin with the fully sampled case (Fig. 6d), comparing the ensemble-mean global $\Delta O2_{\_300}$ trends across three emission scenarios, RCP2.6, RCP4.5, and RCP8.5. As expected, the RCP8.5 scenario shows the strongest negative $\Delta O2_{\_300}$ trend due to intensified warming. The magnitude of the ensemble mean global $\Delta O2_{\_300}$ trend decreases by 34% between RCP8.5 and RCP2.6 (from −0.064 mmol m^{-3} yr^{-1} in RCP8.5 to −0.022 mmol m^{-3} yr^{-1} in RCP2.6). Climate mitigation under RCP2.6 thus appears to significantly reduce the likelihood of detecting negative $\Delta O2_{\_300}$ trends: the detection probability drops from 91.0% in RCP8.5 to 68.0% in RCP2.6. Despite differences in mean trend values, the spread (standard deviation) of 20-year trends among ensemble members remains similar across scenarios, ranging from 0.039 to 0.046 mmol m^{-3} yr^{-1}. The ensemble probability density functions (PDFs) of the global $\Delta O2_{\_300}$ trends are approximately symmetric across scenarios, as indicated by the nearly equal probabilities of trends falling above or below the ensemble mean (μ) (e.g., probabilities > μ and < μ in Fig. 6).

Examining global $\Delta O2_{\_300}$ trends from the near-future sub-sampling experiments (Fig. 6), we find that the forced response is consistently overestimated across all scenarios in "**sustain**", ranging from 117% to 180%. This bias is most pronounced under the RCP2.6 scenario, suggesting that the assessment of mitigation impacts on global $\Delta O2_{\_300}$ trends could be significantly distorted if observational coverage remains at the 2020 level (Fig. 6a). The planned expansion of the BGC-Argo network appears effective in reducing this bias, bringing the forced response estimates closer to those obtained in the fully sampled (model truth) case (Fig. 6bc). Interestingly, even when the number of samples drops sharply after 2026 (as in the "**cut**" experiment), both the estimated forced trend and the probability of detecting negative global $\Delta O2_{\_300}$ trends remain nearly unchanged across scenarios (Fig. 6bc). This suggests that, at the global scale, a relatively uniform spatial sampling may sustain detection skill despite a reduction in the number of observations. However, this finding is based on global averages; regional analyses may reveal a different sensitivity to sampling changes. We also note that in both the "**plan**" and "**cut**" experiments, the probability of detecting negative global $\Delta O2_{\_300}$ trends decreases slightly (by 4%–8%) in the RCP4.5 and RCP8.5 scenarios compared to the fully sampled case. These results highlight that time-dependent variations in sampling effort can influence global trend estimates, particularly under scenarios of strong warming, and emphasize the need for careful data treatment and trend analysis. Finally, it is essential to acknowledge that this assessment concentrates solely on the global $\Delta O2_{\_300}$ metric. The detection and interpretation of regional $\Delta O2_{\_300}$ trends may exhibit stronger sensitivity to changes in sampling coverage.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

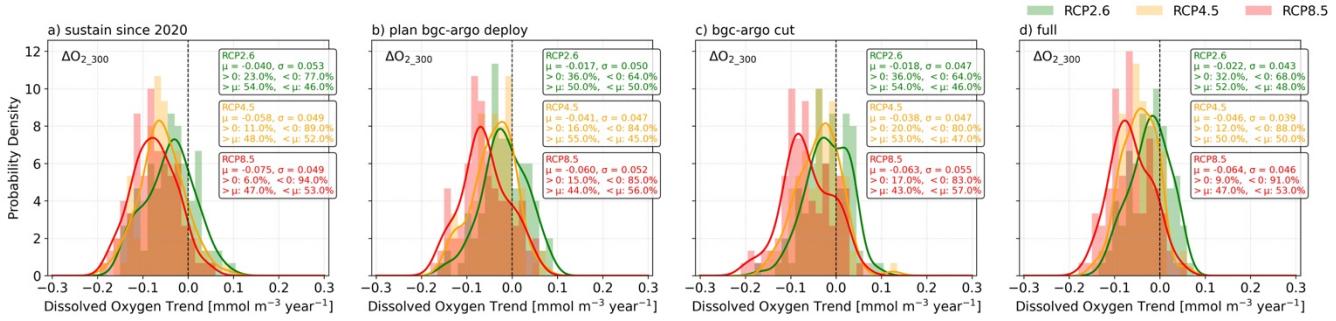


Figure 6: Histograms and probability density functions (PDFs) of 20-year linear trends (2020–2039) of simulated global ΔO_{2_300} from ensemble members under three future emission scenarios and three sub-sampling experiments (see Section 4.1 for details). Panels a) - c) show histograms and PDFs of global ΔO_{2_300} trends for the "**sustain**" (a), "**plan**" (b), and "**cut**" (c) sub-sampling experiments, each for three scenarios: RCP2.6 (green), RCP4.5 (orange), and RCP8.5 (red). Panel d) presents the same analysis as panels a)–c), but for the fully sampled case. Summary statistics of the 20-year trend distributions are shown in each panel, including the ensemble mean trend (μ), the ensemble spread (standard deviation, σ), the probabilities of positive and negative global ΔO_{2_300} trends (> 0 and < 0), and the probabilities of trends exceeding or falling below the ensemble mean (> μ and < μ).

5 Summary and Discussion

The expansion of the global ocean observing system is progressing, supported by the introduction of new autonomous platforms such as biogeochemical floats (e.g., BGC-Argo) and gliders equipped with ocean biogeochemical sensors (Biogeochemical-Argo Planning Group, 2016; Claustre et al., 2020; Stoer et al., 2023). The scientific community aims to deploy approximately 1,000 biogeochemical floats in the coming years to monitor ocean biogeochemistry and marine ecosystem stressors (Biogeochemical-Argo Planning Group, 2016; Germineaud et al., 2019). Our assessment of historical observational biases in dissolved oxygen, together with recent findings by Gloege et al. (2021) on air-sea CO_2 flux, suggests that spatiotemporal data gaps must be carefully considered when designing observing systems, interpreting observational data, and constructing ocean biogeochemical metrics such as area-weighted time series. Changes in the number and distribution of observations can introduce additional uncertainties, posing challenges for the interpretation of global and regional mean-based indicators.

In this study, we find that spatiotemporal gaps in observations can introduce significant bias into estimates of internal climate variability (ICV) in oceanic O_2. This bias can substantially inflate the apparent magnitude of ICV, increasing it by up to 150% at the global scale and by more than 500% in data-sparse regions when interpreting historical ΔO_{2_300} trends.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Considering observational sampling bias, our analysis reveals that the combined effect of ICV and sampling bias can introduce a deviation of 5%–25% in estimated multi-decadal ΔO_{2_300} trends. Furthermore, our results suggest that the observed decline in ΔO_{2_300} cannot be fully explained by the forced response alone, indicating a significant contribution from ICV, with sampling bias adding further uncertainty. Although uncertainties in the ΔO_{2_300} time series do not translate directly into uncertainties in multi-decadal trend estimates, the calculated trends may still depend on the selected analysis period (Fay et al., 2014; McKinley et al., 2011). Additionally, time-varying changes in observational coverage may introduce unexpected uncertainties in assessing long-term trends of deoxygenation.

Observed decadal variability and the decline in global ΔO_{2_300} levels may be biased by signals originating from data-dense regions, such as the North Atlantic and North Pacific Oceans, where ICV is substantial. A reduction in observational coverage after the 1980s could also introduce spurious signals, potentially affecting both global and regional ΔO_{2_300} trend estimates. Power spectrum analysis of the large ensemble simulations (figure not shown) suggests that ICV may also influence the temporal variability of the regional ΔO_{2_300} time series. To accurately assess the spectral characteristics of regional ΔO_{2_300} variability, a more uniform observational sampling is required, an objective that cannot be achieved without expanding both the number and spatial coverage of observations.

Our study suggests that ICV and sampling bias play a significant role in explaining the magnitude of historical ocean deoxygenation. This has important implications for interpreting the apparent bias in Earth System Models (ESMs), which often simulate less deoxygenation than observed during the historical period. Specifically, if the historical ΔO_{2_300} trend contains a substantial contribution from ICV, and if this contribution is overstated due to sampling bias, then the underestimation of historical deoxygenation by ESMs may reflect insufficiently simulated variability (Oschlies et al., 2017, 2018), rather than deficiencies in representing the forced deoxygenation trend itself. In multi-member ESM ensemble mean, such as those in CMIP experiments, ICV averages out over decadal timescales, resulting in a small contribution to ensemble mean trends. Therefore, if observed ICV enhances negative ΔO_{2_300} trends, CMIP ensembles may systematically underestimate historical deoxygenation for reasons unrelated to model fidelity. This bias is further compounded when limited observational coverage skews trend estimates toward more negative values. The effect is exacerbated if the simulated magnitude of ICV is also underestimated.

The discrepancy between observed and simulated ΔO_{2_300} trends may also arise from limitations within ocean biogeochemical models. Common deficiencies in CMIP6-class models include the inability to represent key ecosystem processes, such as changes linked to ocean acidification (Doney et al., 2009; Hoegh-Guldberg et al., 2010), shifts in phytoplankton size structure (Barton et al., 2016), diel vertical migration of zooplankton (Bianchi et al., 2013), variable stoichiometric ratios (Galbraith et al., 2015; Kwiatkowski et al., 2018), heterogeneous reactivity of particulate organic carbon (Aumont et al., 2017), temperature-dependent remineralization (Laufkötter et al., 2017; Segschneider and Bendtsen,

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

2013), and particle aggregation and sinking dynamics (Maerz et al., 2020). These biological and biogeochemical processes influence O_2 variability and trends across a range of timescales. Additionally, models with coarse spatial resolution, such as those used in our study, struggle to resolve key physical processes, including mesoscale and submesoscale eddies and the equatorial current system (Getzlaff et al., 2013; Gnanadesikan et al., 2013). A more detailed overview of these limitations and their potential impacts is provided by Oschlies et al. (2017, 2018). Moving forward, analyses of CMIP6 multi-model ensembles, especially those including more advanced biogeochemical schemes (Séférian et al., 2020), may help clarify where and how such limitations affect our understanding of oxygen variability and trends.

Using idealized future sub-sampling experiments based on current and planned BGC-Argo float deployments (Biogeochemical-Argo Planning Group, 2016; Stoer et al., 2023), we show that the estimation of the forced $\Delta O2_{\_300}$ trend is not substantially biased by the limited and time-varying sample size of the present and near-future observing networks, at least over the next two decades under business-as-usual and intermediate scenarios. The recent expansion of the BGC-Argo network over the past five years appears to be effective in reducing sampling bias, as demonstrated by our sub-sampling experiments with large ensemble simulations (Fig. 6). This is particularly relevant for assessing decadal-scale ocean deoxygenation under climate mitigation efforts. However, accurately quantifying ICV with a limited number of observations remains both systematically challenging and essential for reliable estimation of the near-future forced O_2 trend. Our findings emphasize the importance of ongoing efforts to strengthen global ocean biogeochemical monitoring, including ship-based programs (e.g., GO-SHIP program; Sloyan et al., 2019), BGC-Argo, and glider observations, for improving both global and regional assessments of O_2 variability. Our results also highlight that globally uniform sampling, as currently planned, can help reduce uncertainty in detecting the forced trend. Moreover, large ensemble simulations of Earth System Models (ESMs) offer valuable methodological platforms for evaluating observational strategies under global warming. Strengthening collaboration between observational programs and ensemble modeling will be essential for guiding future ocean monitoring in a changing climate. Although climate mitigation is expected to slow the long-term pace of ocean deoxygenation (Hoegh-Guldberg et al., 2018), our analysis shows that near-future (2020–2039) decadal-scale assessments will likely remain uncertain due to ICV. Climate mitigation may reduce the magnitude of the forced negative O_2 trend, shifting its probability distribution towards less negative or even positive values. Nevertheless, the probability of observing a negative trend remains high, ranging from 68% to 91% depending on the scenario (Fig. 6). Accurate monitoring of such decadal-scale O_2 variability is critical not only for trend detection but also for understanding impacts on marine organisms and evaluating their capacity for adaptation and acclimation under long-term climate change.

Assessing decadal trends is crucial for monitoring and informed decision-making, particularly in interpreting the near-term effects of carbon emission reductions over the next few decades (IPCC, 2018). While the ocean may respond to climate change mitigation efforts, this response could be obscured by strong trends driven solely by ICV (Marotzke, 2019; Spring et al., 2020). To avoid misinterpreting the impacts of mitigation, ICV-driven trends must be carefully quantified. The outcomes

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

of emission reduction efforts over the coming decades may have significant implications for marine biodiversity, partly because marine ecosystems strongly depend on oceanic oxygen levels (Bindoff et al., 2020; Levin et al., 2015, 2018; Sperling et al., 2016). The physiological impacts on marine life (Pörtner et al., 2017), ecological shifts, and habitat compression (Sato et al., 2017) expected in the near-future will likely result from the combined effects of anthropogenic change and the influence of ICV on oxygen distributions and related environmental factors, such as ocean temperature (Deutsch et al., 2015, 2024).

Deoxygenation co-occurs with other human-driven changes in ocean biogeochemistry, such as pollution and ocean acidification, which interact with one another (Doney, 2010). Disentangling the magnitude of underlying drivers is a crucial step in understanding how their compounding effects impact marine organisms and ecosystems. For O_2, our results indicate that this requires a broadly distributed observing network to quantify the uncertainties imposed by ICV.

Data availability

The access information for MPI's large ensemble simulations, the Grand Ensemble Simulations (MPI-GE), can be found on the project website (https://mpimet.mpg.de/en/research/modeling/grand-ensemble). The observational dissolved oxygen dataset (in NetCDF format) from Ito et al., (2017) can be found through this link (https://o2.eas.gatech.edu/Itoetal_GRL_2017/o2_aan_mcl1950-2016_0147_QC3.nc). The access information for the three observational dissolved oxygen datasets can be found in the references: Ito (2022), Ito et al. (2024b), and Roach and Bindoff (2023).

Author contributions

Y.T led the observational and model data analysis and the manuscript writing. Both Y.T and T.I conceived the study and contributed to the writing.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgement

We thank Taka Ito, Nicola Maher, Sebastian Milinski, Dian Putrasahan, Friederike Fröb, and Fabrice Lacorix for fruitful discussions during this work. We extend our thanks to Peter Landschützer for the internal review. Special thanks to Sarah

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Schlunegger and Mathew Maltrud for constructive comments and proofreading of the manuscript. The model simulations were performed at the Swiss National Supercomputing Center (CSCS) and the German Climate Computing Center (DKRZ).

Financial support

This work is supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 641816 (CRESCENDO). YT is supported by a joint UK Natural Environment Research Council and US National Science Foundation Large grant NE/W009501/1 (C-Streams). TI was supported by the European Union's Horizon 2020 research and innovation programme (ESM2025 - ESMs for the Future - Grant 101003536), as well as the Deutsche Forschungsgemeinschaft (Germany's Excellence Strategy - EXC 2037 "CLICCS - Climate, Climatic Change, and Society" - project no. 390683824 of Universität Hamburg).

Appendix A: Model Evaluations

A1 Climatological Distributions of Dissolved Oxygen near the Thermocline

We conducted a climatological evaluation of simulated subsurface oxygen (O_2) using globally comprehensive observations to verify the realism of our model results. Overall, the large ensemble simulation from MPI-ESM1.1 (MPI-GE) reproduces the broad-scale climatological distribution of O_2 at around 300 m depth (O_{2_300}), near the thermocline. However, notable regional biases remain (Fig. A1). The spatial correlation coefficient is R = 0.91, and the root mean square error (RMSE) is 36.8 mmol m^{-3}. MPI-ESM1.1 tends to overestimate the O_{2_300} deficit in tropical Oxygen Minimum Zones (OMZs), a known bias among CMIP models (Cabré et al., 2015; Ilyina et al., 2013; Séférian et al., 2020). The model may also underestimate the internal climate variability (ICV) of O_{2_300} in these regions, likely because the simulated climatological mean state of O_{2_300} is close to anoxic conditions in tropical OMZs (Fig. A1). In the tropical Pacific, the model fails to simulate the westward extent of low O_{2_300} seen in observations, resulting in a positive bias in the eastern tropical Pacific (Fig. A1). This climatological bias could influence regional estimates of O_{2_300} variability, expressed as ensemble or temporal standard deviation (Section 3). Other biases include over-oxygenated water in the subpolar North Pacific and underestimated O_{2_300} levels in the mid-latitude Southern Ocean, where sub-Antarctic surface water forms. The cause of high oxygen concentrations in the subpolar North Pacific is unclear, but it may be related to weak upwelling in the model, which reduces the supply of oxygen-poor deep water. In the mid-latitude Southern Ocean, the unclear, but it may be related to weak upwelling in the model, which reduces the deficit, may result from weak surface and intermediate water mass formation, limiting oxygen supply from the surface to the ocean interior. These biases, related to upwelling, water mass formation, and the O_{2_300} mean state, may also affect O_{2_300} variability in these key regions.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

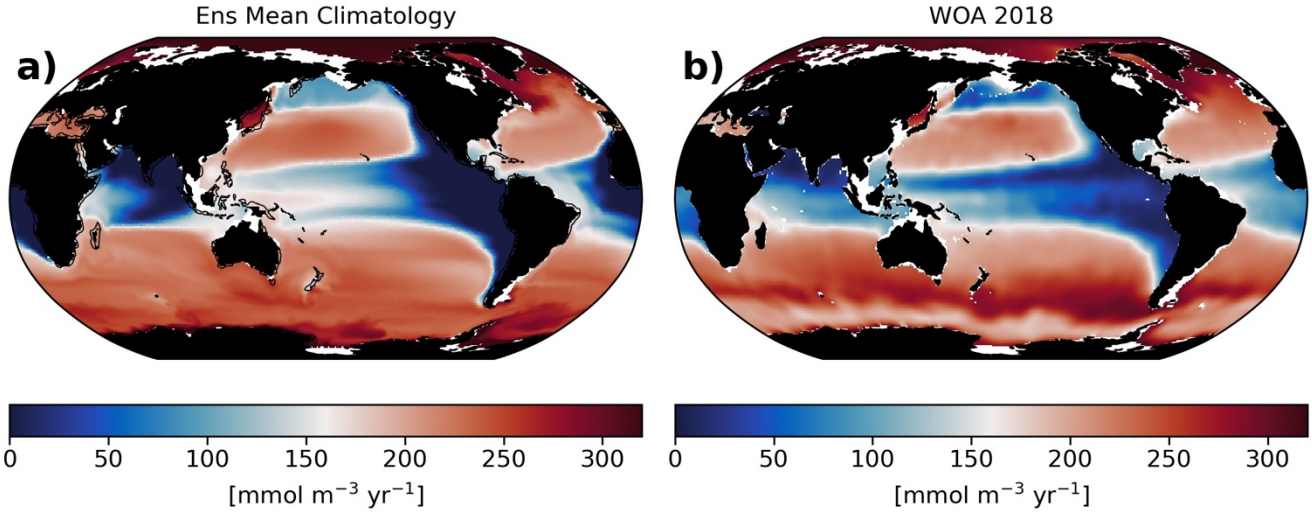


Figure A1: Climatological O_{2_300} distributions from the model (ensemble mean of MPI-GE) and observations, World Ocean Atlas 2018 (Garcia et al., 2018).

References

Aumont, O., van Hulten, M., Roy-Barman, M., Dutay, J.-C., Éthé, C., and Gehlen, M.: Variable reactivity of particulate organic matter in a global ocean biogeochemical model, Biogeosciences, 14, 2321–2341, https://doi.org/10.5194/bg-14-2321-2017., 2017.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver,: Evaluation of the ECMWF ocean reanalysis system ORAS4. Quarterly Journal of the Royal Meteorological Society, 139(674), 1132–1161, doi:10.1002/qj.2063, 2013.

Barton, A. D., Irwin, A. J., Finkel, Z. V., and Stock, C. A.: Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. Proceedings of the National Academy of Sciences, 113(11), 2964-2969., 2016.

Bianchi, D., Galbraith, E., Carozza, D. et. al.,: Intensification of open-ocean oxygen depletion by vertically migrating animals. Nature Geoscience, 6, 545–548, https://doi.org/10.1038/ngeo1837., 2013.

Bindoff, N. L., Cheung, W. W. L., Kairo, J. G., Arístegui, J., Guinder, V. A., Hallberg, R., Hilmi, N., Jiao, N., Karim, M. S., Levin, L., O'Donoghue, S., Purca Cuicapusa, S. R., Rinkevich, B., Suga, T., Tagliabue, A., and Williamson, P.: Changing Ocean, Marine Ecosystems, and Dependent Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by: Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N. M., in press., 2020.

Biogeochemical-Argo Planning Group (2016).: The scientific rationale, design and implementation plan for a Biogeochemical-Argo float array. https://doi.org/10.13155/46601., 2016.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

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.: 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, 2013.

Bopp, L., Resplandy, L., Untersee, A., Le Mezo, P., and Kageyama, M.: Ocean (de) oxygenation from the Last Glacial Maximum to the twenty-first century: insights from Earth System models. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2102), 20160323, 2017.

Boyer, T. P., Antonov, J. I., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., O'Brien, T. D., Paver, C. R., Reagan, J. R., Seidov, D., Smolyar, I. V., Zweng, M. M., et al.: World Ocean Database 2013. NOAA Atlas NESDIS 72, Silver Spring., 2013.

Breitburg, D, Levin L. A, Oschlies A, Grégoire M, Chavez F. P, Conley D. J, Garçon V, Gilbert D, Gutiérrez D, Isensee K, Jacinto G. S, Limburg K. E, Montes I, Naqvi S. W. A, Pitcher G. C, Rabalais N. N, Roman M. R, Rose K. A, Seibel B. A, Telszewski M, Yasuhara M, Zhang J.: Declining oxygen in the global ocean and coastal waters, Science, 359, 6371, https://doi.org/10.1126/science.aam7240, 2018.

Cabré, A., Marinov, I., Bernardello, R., and Bianchi, D.: Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends, Biogeosciences, 12, 5429–5454, https://doi.org/10.5194/bg-12-5429-2015., 2015.

Claustre, H., Johnson, K. S., Takeshita, Y.,: Observing the Global Ocean with Biogeochemical-Argo. Annual Review of Marine Science 12: 11.1-11.26 | DOI: 10.1146/annurev-marine-010419-010956., 2020.

Cocco, V., Joos, F., Steinacher, M., Frölicher, T. L., Bopp, L., Dunne, J., Gehlen, M., Heinze, C., Orr, J., Oschlies, A., Schneider, B., Segschneider, J., and Tjiputra, J.: Oxygen and indicators of stress for marine life in multi-model global warming projections, Biogeosciences, 10, 1849–1868, https://doi.org/10.5194/bg-10-1849-2013, 2013.

Deser, C., F. Lehner, K. B. Rodgers, T. Ault, T. L. Delworth, P. N. DiNezio, A. Fiore, C. Frankignoul, J. C. Fyfe, D. E. Horton, J. E. Kay, R. Knutti, N. S. Lovenduski, J. Marotzke, K. A. McKinnon, S. Minobe, J. Randerson, J. A. Screen, I. R. Simpson and M. Ting,: Insights from earth system model initial-condition large ensembles and future prospects. Nature Climate Change, doi: 10.1038/s41558-020-0731-2, 2020.

Deutsch, C., Ferrel, A., Seibel, B., Pörtner, H. O., and Huey, R. B.: Climate change tightens a metabolic constraint on marine habitats. Science, 348(6239), 1132-1135, doi: 10.1126/science.aaa1605., 2015.

Deutsch, C., Penn, J. L., & Lucey, N.: Climate, oxygen, and the future of marine biodiversity. Annual Review of Marine Science, 16(1), 217-245, 2024.

Doney, S. C.: The growing human footprint on coastal and open-ocean biogeochemistry. Science, 328(5985), 1512-1516, doi: 10.1126/science.1185198., 2010.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.,: Ocean Acidification: The Other CO_2 Problem. Annual Review of Marine Science, 1, 169-92., 2020.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Fay, A. R., McKinley, G. A., and Lovenduski, N. S.,: Southern Ocean carbon trends: Sensitivity to methods, Geophysical Research Letters., 41, 6833– 6840, doi:10.1002/2014GL061324., 2014.

Galbraith E. D., and Martiny A. C.,: A simple nutrient-dependence mechanism for predicting the stoichiometry of marine ecosystems. Proceedings of the National Academy of Sciences, 112, 8199–8204., 2015.

Garcia, H. E., K. Weathers, C. R. Paver, I. Smolyar, T. P. Boyer, R. A. Locarnini, M. M. Zweng, A. V. Mishonov, O. K. Baranova, D. Seidov, and J. R. Reagan,: World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 83, 38pp., 2018.

Germineaud, C., J. Brankart, and P. Brasseur,: An Ensemble-Based Probabilistic Score Approach to Compare Observation Scenarios: An Application to Biogeochemical-Argo Deployments. Journal of Atmospheric and Oceanic Technology., 36, 2307–2326, https://doi.org/10.1175/JTECH-D-19-0002.1., 2019.

Getzlaff, J., and Dietze, H.: Effects of increased isopycnal diffusivity mimicking the unresolved equatorial intermediate current system in an earth system climate model, Geophysical Research Letters., 40, 2166– 2170, doi:10.1002/grl.50419., 2013.

Giorgetta, M., Jungclaus,J., Reick, C., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.,: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. Journal of Advances in Modeling Earth Systems, 5(3), 572–597. https://doi.org/10.1002/jame.20038, 2013.

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., et al.,: Quantifying errors in observationally based estimates of ocean carbon sink variability. Global Biogeochemical Cycles, 35, e2020GB006788. https://doi.org/10.1029/2020GB006788., 2021.

Gnanadesikan, A., Bianchi, D., and Pradal, M.-A.,: Critical role for mesoscale eddy diffusion in supplying oxygen to hypoxic ocean waters, Geophysical Research Letters., 40, 5194– 5198, doi:10.1002/grl.50998., 2013.

Gruber, N.: Warming up, turning sour, losing breath: Ocean biogeochemistry under global change. Philosophical Transactions of the Royal Society A, 369, 1980–1996, https://doi.org/10.1098/rsta.2011.0003, 2011.

Gulev, S.K., P.W. Thorne, J. Ahn, F.J. Dentener, C.M. Domingues, S. Gerland, D. Gong, D.S. Kaufman, H.C. Nnamchi, J. Quaas, J.A. Rivera, S. Sathyendranath, S.L. Smith, B. Trewin, K. von Schuckmann, and R.S. Vose: Changing State of the Climate System. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 287–422, doi:10.1017/9781009157896.004., 2021.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Helm, K. P., Bindoff, N. L., and Church, J. A.: Observed decreases in oxygen content of the global ocean. Geophysical Research Letters, 38, L23602, doi:10.1029/2011GL049513, 2011.

Henson, S. A., C. Beaulieu, T. Ilyina, J. G. John, M. Long, R. Seferian, J. Tjiputra, and J. L. Sarmiento,: Rapid emergence of climate change in environmental drivers of marine ecosystems, Nature Communications, 8, 14682, doi:10.1038/ncomms14682., 2017.

Hoegh-Guldberg, O., and Bruno, J. F.: The impact of climate change on the world's marine ecosystems. Science, 328(5985), 1523-1528., 2010.

Hoegh-Guldberg, O., D. Jacob, M. Taylor, M. Bindi, S. Brown, I. Camilloni, A. Diedhiou, R. Djalante, K.L. Ebi, F. Engelbrecht, J. Guiot, Y. Hijioka, S. Mehrotra, A. Payne, S.I. Seneviratne, A. Thomas, R. Warren, and G. Zhou,: Impacts of 1.5ºC Global Warming on Natural and Human Systems. In: Global Warming of 1.5°C. An 14. IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press., 2018.

IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. In Press., 2018.

Ilyina, T., K. D. Six, J. Segschneider, E. Maier-Reimer, H. Li, and I. Núñez-Riboni
.,: The global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth System Model in different CMIP5 experimental realizations. Journal of Advances in Modeling Earth Systems, 5, 287–315, https://doi.org/10.1029/2012MS000178, 2013.

Ito, T., and C. Deutsch,: A conceptual model for the temporal spectrum of oceanic oxygen variability, Geophysical Research Letters., 37, L03601, doi:10.1029/2009GL041595., 2010.

Ito, T., Minobe, S., Long, M. C., and Deutsch, C.: Upper ocean O2 trends: 1958-2015. Geophysical Research Letters, 44, 4214–4223, https://doi.org/10.1002/2017GL073613, 2017.

Ito, T.: Optimal interpolation of global dissolved oxygen: 1965–2015. Geoscience Data Journal, 00, 1– 10, https://doi.org/10.1002/gdj3.130, 2022.

Ito, T., Garcia, H. E., Wang, Z., Minobe, S., Long, M. C., Cebrian, J., Reagan, J., Boyer, T., Paver, C., Bouchard, C., Takano, Y., Bushinsky, S., Cervania, A., and Deutsch, C. A.: Underestimation of multi-decadal global O_2 loss due to an optimal interpolation method, Biogeosciences, 21, 747–759, https://doi.org/10.5194/bg-21-747-2024, 2024a.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Ito, T., Cervania, A., Cross, K., Ainchwar, S., & Delawalla, S.: 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, 2024b.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and Storch, J. S.,: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. Journal of Advances in Modeling Earth Systems, 5, 422–446. https://doi.org/10.1002/jame.20023, 2013.

Kamenkovich, I., W. Cheng, C. Schmid, and D. E. Harrison,: Effects of eddies on an ocean observing system with profiling floats: Idealized simulations of the Argo array, *J. Geophys. Res.*, 116, C06003, doi:10.1029/2010JC006910., 2011.

Keeling, R. F., A. Körtzinger, and Gruber, N.: Ocean deoxygenation in a warming world, Annual Review of Marine Science, 2, 199–229, doi:10.1146/annurev.marine.010908.163855, 2010.

Kwiatkowski, L., Aumont, O., Bopp, L., and Ciais, P.: The impact of variable phytoplankton stoichiometry on projections of primary production, food quality, and carbon uptake in the global ocean. Global Biogeochemical Cycles, 32, 516– 528, https://doi.org/10.1002/2017GB005799., 2018.

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.: 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, 2020.

Laffoley, D., and Baxter, J. M.: Ocean deoxygenation: Everyone's problem-causes, impacts, consequences and solutions. Gland, Switzerland: IUCN., 2019.

Laufkötter, C., J. G. John, C. A. Stock, and J. P. Dunne,: Temperature and oxygen dependence of the remineralization of organic matter, *Global Biogeochem. Cycles*, 31, 1038–1050, doi:10.1002/2017GB005643., 2017.

Levin, L. A, Liu K. K, Emeis K. C, Breitburg D. L, Cloern J, Deutsch C, Giani M, Goffart A, Hofmann E. E, Lachkar Z, Limburg K, Liu S. M, Montes E, Naqvi W, Ragueneau O, Rabouille C, Sarkar S. K, Swaney D. P, Wassman P, Wishner K. F.: Comparative biogeochemistry–ecosystem–human interactions on dynamic continental margins. Journal of Marine Systems, 141, 3-17., 2015.

Levin, L. A.: Manifestation, drivers, and emergence of open ocean deoxygenation. Annual Review of Marine Science, 10, 229-260., 2018.

Long, M. C., C. A. Deutsch, and T. Ito,: Finding forced trends in oceanic oxygen, Global Biogeochemical Cycles, 30, doi:10.1002/2015GB005310, 2016.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, N., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D.,

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Raddatz, T., Stevens, B. and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. Journal of Advances in Modeling Earth Systems, 11, 1-21. doi.org/10.1029/2019MS001639., 2019.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Marotzke, J.: Quantifying the irreducible uncertainty in near-term climate projections. WIREs Climate Change, 10:e563. https://doi.org/10.1002/wcc.563., 2019.

Majkut J. D., Carter, B. R., Frölicher. T. L., Dufour. C.O., Rodgers. K. B., Sarmiento. J. L.,: An observing system simulation for Southern Ocean carbon dioxide uptake. Phil.Trans. R. Soc. A 372:20130046. http://dx.doi.org/10.1098/rsta.2013.0046., 2014.

McKinley, G. A., A. R. Fay, T. Takahashi, and N. Metzl,: Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales, Nat. Geoscience., 4, 606– 610, doi:10.1038/Ngeo1193, 2011.

Meinshausen, M., and Coauthors,: The RCP greenhouse gas concentrations and their extension from 1765 to 2300. Climatic Change, 109, 213–241, doi:https://doi.org/10.1007/s10584-011-0156-z., 2011.

Maerz, J., Six, K. D., Stemmler, I., Ahmerkamp, S., and Ilyina, T.: Microstructure and composition of marine aggregates as co-determinants for vertical particulate organic carbon transfer in the global ocean, Biogeosciences, 17, 1765–1803, https://doi.org/10.5194/bg-17-1765-2020., 2020.

Moltmann, T., Turton, J., Zhang, H.M., Nolan, G., Gouldman, C., Griesbauer, L., Willis, Z., Piniella, A.M., Barrell, S., Andersson, E. and Gallage, C.,: A Global Ocean Observing System (GOOS), delivered through enhanced collaboration across regions, communities, and new technologies. Frontiers in Marine Science, 6, p.291., 2019.

Oschlies, A., Duteil, O., Getzlaff, J., Koeve, W., Landolfi, A. and Schmidtko, S.: Patterns of deoxygenation: sensitivity to natural and anthropogenic drivers. Philosophical Transactions of the Royal Society A, 375(2102), p.20160325, DOI 10.1098/rsta.2016.0325, 2017.

Oschlies, A., Brandt, P., Stramma, L., and Schmidtko, S.:, Drivers and mechanisms of ocean deoxygenation. Nature Geoscience, 11, 467–473, doi: 10.1038/s41561-018-0152-2, 2018.

Pörtner, H. O., Bock, C., and Mark, F. C.: Oxygen-and capacity-limited thermal tolerance: bridging ecology and physiology. Journal of Experimental Biology, 220(15), 2685-2696. doi:10.1016/j.jmarsys.2014.04.016., 2017.

Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler,: Representation of natural and anthropogenic land cover change in MPI-ESM. Journal of Advances in Modeling Earth Systems, 5, 459–482, doi: 10.1002/jame.20022, 2013.

Purkey, S. G., & Johnson, G. C.: Antarctic Bottom Water warming and freshening: Contributions to sea level rise, ocean freshwater budgets, and global heat gain. Journal of Climate, 26(16), 6105-6122., 2013.

Roach, C. J., & Bindoff, N. L.: 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, 2023.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Rodgers, K. B., Lin, J., and Frölicher, T. L.,: Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model, Biogeosciences, 12, 3301–3320, https://doi.org/10.5194/bg-12-3301-2015., 2015.

Sato, K. N., Levin, L. A., and Schiff, K.: Habitat compression and expansion of sea urchins in response to changing climate conditions on the California continental shelf and slope (1994–2013). Deep Sea Research Part II: Topical Studies in Oceanography, 137, 377-389, https://doi.org/10.1016/j.dsr2.2016.08.012., 2017.

Schmidtko, S., L. Stramma, and Visbeck, M.: Decline in global oceanic oxygen content during the past five decades, Nature, 542 (7641), 335–339, doi:10.1038/nature21399, 2017.

Schlunegger S., K. Rodgers, J.L. Sarmiento, J.P. Dunne, T.L. Frölicher, R. Slater, M. Ishii: Emergence of anthropogenic signals in the ocean carbon cycle. Nature Climate Change. doi:10.1038/s41558-019-0553-2., 2019.

Schlunegger, S., Rodgers, K.B., Sarmiento, J.L., Ilyina, T., Dunne, J.P., Takano, Y., Christian, J.R., Long, M.C., Frölicher, T.L., Slater, R. and Lehner, F.,: Time of Emergence and Large Ensemble Intercomparison for Ocean Biogeochemical Trends. Global Biogeochem. Cycles, 34: e2019GB006453. https://doi.org/10.1029/2019GB006453., 2020.

Séférian, R., Berthet, S., Yool, A. et al. Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. Current Climate Change Report, 6, 95–119.: https://doi.org/10.1007/s40641-020-00160-0
https://doi.org/10.5194/bg-17-3439-2020, 2020.

Segschneider, J., and J. Bendtsen,: Temperature-dependent remineralization in a warming ocean increases surface pCO_2 through changes in marine ecosystem composition, *Global Biogeochem. Cycles*, 27, 1214–1225, doi:10.1002/2013GB004684., 2013.

Sharp, J. D., Fassbender, A. J., Carter, B. R., Johnson, G. C., Schultz, C., and Dunne, J. P.: GOBAI-O_2: temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly 2 decades, Earth Syst. Sci. Data, 15, 4481–4518, https://doi.org/10.5194/essd-15-4481-2023, 2023.

Sloyan, B. M., Wanninkhof, R., Kramp, M., Johnson, G. C., Talley, L., Tanhua, T., ... and Katsumata, K.: The global ocean ship-base hydrographic investigations program (GO-SHIP): a platform for integrated multidisciplinary ocean science. Frontiers in Marine Science, 6, 445., 2019.

Sperling, E. A., Frieder, C. A., and Levin, L. A.: Biodiversity response to natural gradients of multiple stressors on continental margins. Proceedings of the Royal Society B: Biological Sciences, 283(1829), 20160637, doi:10.1098/rspb.2016.0637., 2016.

Spring, A., Ilyina, T., & Marotzke, J: Inherent uncertainty disguises attribution of reduced atmospheric CO2 growth to CO2 emission reductions for up to a decade. Environmental Research Letters, 15(11), 114058., 2020.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.,: Atmospheric component of the MPI-M Earth System Model: ECHAM6. Journal of Advances in Modeling Earth Systems, 5, 146–172. https://doi.org/10.1002/jame.20015, 2013.

https://doi.org/10.5194/egusphere-2025-3757
Preprint. Discussion started: 6 October 2025

Stoer AC, Takeshita Y, Maurer TL, Begouen Demeaux C, Bittig HC, Boss E, Claustre H, Dall'Olmo G, Gordon C, Greenan BJW, Johnson KS, Organelli E, Sauzède R, Schmechtig CM and Fennel K.,: A census of quality-controlled Biogeochemical-Argo float measurements. *Front. Mar. Sci.* 10:1233289. doi: 10.3389/fmars.2023.1233289., 2023.

Stramma, L., A. Oschlies, and S. Schmidtko,: Mismatch between observed and modeled trends in dissolved upper-ocean oxygen over the last 50 yr, Biogeosciences, 9, 4045–4057, doi:10.5194/bg-9-4045-2012, 2012.

Takano, Y., Ito, T., and Deutsch, C.,: Projected centennial oxygen trends and their attribution to distinct ocean climate forcings. Global Biogeochemical Cycles, 32, 1329–1349, https://doi.org/10.1029/2018GB005939., 2018.

Takano Y, Ilyina T, Tjiputra J, Eddebbar YA, Berthet S, Bopp L, Buitenhuis E, Butenschön M, Christian JR, Dunne JP, Gröger M, Hayashida H, Hieronymus J, Koenigk T, Krasting JP, Long MC, Lovato T, Nakano H, Palmieri J, Schwinger J, Séférian R, Suntharalingam P, Tatebe H, Tsujino H, Urakawa S, Watanabe M and Yool A.: Simulations of ocean deoxygenation in the historical era: insights from forced and coupled models. *Front. Mar. Sci.* 10:1139917. doi: 10.3389/fmars.2023.1139917, 2023.

Weiss, R. F.,: The solubility of nitrogen, oxygen and argon in water and seawater. Deep Sea Research and Oceanographic Abstracts, Vol. 17, No. 4, pp. 721-735, https://doi.org/10.1016/0011-7471(70)90037-9, 1970.