ANNUAL REVIEWS

Annual Review of Earth and Planetary Sciences

Are Simulated Ocean Deoxygenation Rates Consistent with the Observational Reconstructions?

Takamitsu Ito,[1] Yohei Takano,[2] Yassir A. Eddebbar,[3] Jerry F. Tiputra,[4] Zhankun Wang,[5] Shoshiro Minobe,[6] Lijing Cheng,[7] Juan Du,[8] and Yumi Abe[1]

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA; email: taka.ito@eas.gatech.edu

[2]British Antarctic Survey, Cambridge, United Kingdom

[3]Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA

[4]NORCE Research AS, Bjerknes Centre for Climate Research, Bergen, Norway

[5]National Centers for Environmental Information and National Environmental Satellite, Data, and Information Service, National Oceanic and Atmospheric Administration, Silver Spring, Maryland, USA

[6]Department of Earth and Planetary Sciences, Hokkaido University, Sapporo, Japan

[7]State Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[8]International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Annu. Rev. Earth Planet. Sci. 2026. 54:239–67

First published as a Review in Advance on January 13, 2026

The *Annual Review of Earth and Planetary Sciences* is online at earth.annualreviews.org

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

Keywords

dissolved oxygen, ocean deoxygenation, Earth system modeling

Abstract

It is currently debated whether Earth system models (ESMs) can reproduce observation-based long-term changes in global and regional deoxygenation rates. Both models and observations include uncertainties, which must be considered when evaluating their consistency. Based on 14 ESMs and 6 observational datasets, the models' climatological annual mean oxygen matches observations well near the surface. However, significant biases remain in the tropics and in the thermocline. Based on the same set of models and three time-varying observation-based datasets, the models tend to underestimate deoxygenation trends from 1965 to 2014, except for the North Atlantic basin. However, the small number of observational datasets limits this conclusion. One dataset appears to significantly underestimate deoxygenation due to sparse data coverage. This review highlights the need for improvements in model process representations and the development of

more observation-based, quality-controlled datasets to better constrain and interpret oxygen changes in the ocean.

- Uncertainties of dissolved oxygen field in CMIP6 ESMs and observational reconstructions are quantified.
- The ESMs can skillfully reproduce long-term average oxygen near the surface, but challenges remain in the thermocline and tropics.
- The ESMs underestimate the deoxygenation trends except for the North Atlantic basin.

1. INTRODUCTION

The ocean is essential for sustaining biodiversity and human livelihood, health, and security. Key ecosystem services provided by the global ocean include climate regulation and biogeochemical cycling of carbon, nutrients, and dissolved gases such as oxygen. Living organisms in the oceans form a diverse range of ecosystems, spanning from microbes to marine mammals, in both coastal and open oceans. Climate and marine ecosystems are intimately linked, and historical observations from past decades have shown growing influences of anthropogenic perturbations on marine ecosystems through warming and ocean acidification due to the absorption of excess heat and anthropogenic carbon (Gruber et al. 2021, Pershing et al. 2015, Seidov et al. 2018, Wyatt et al. 2022). In addition to rising temperatures and acidity, marine ecosystems have experienced significant perturbations in recent decades through the loss of dissolved oxygen (O_2), known as ocean deoxygenation (Breitburg et al. 2018, Rabalais & Turner 2019, Stramma et al. 2020). The combination of these ecosystem stressors can lead to habitat compression and loss due to limited thermal tolerance (Schoepf et al. 2015, Sunday et al. 2010), metabolic stress (Deutsch et al. 2015, 2020), range shifts of species (Thomas 2010, Thuiller 2004), and changing food availability (Santora et al. 2020) that are potentially irreversible on centennial timescales (Bertini & Tjiputra 2022, Santana-Falcón et al. 2023).

The advection and convection of O_2-rich surface waters at high latitudes into the deep interior ocean is termed ocean ventilation, which is the primary source of preformed O_2 in the oceans. Variations in the subduction and mixing rates of the major ocean basins play essential roles in controlling the global O_2 inventory trends and variability, including the North Pacific Ocean (Abe & Minobe 2023, Kwon et al. 2016, Long et al. 2016, Mecking & Drushka 2024, Novi et al. 2024, Osafune & Yasuda 2013, Stramma et al. 2020, Toyama et al. 2015), the North Atlantic (Rhein et al. 2017, Tjiputra et al. 2018), the Southern Ocean (Abernathey & Ferreira 2015, Dove et al. 2021, Hollitzer et al. 2024, Kamenkovich et al. 2017, Marshall & Zanna 2014, Naveira Garabato et al. 2019), and the tropical Pacific (Busecke et al. 2019, Czeschel et al. 2012, Deutsch et al. 2011, Duteil et al. 2018, Ito & Deutsch 2013, Poupon et al. 2023). Global warming can lead to ocean deoxygenation, driven by the reduced oxygen solubility in seawater and reduced physical oxygen supply to the interior due to an increase in upper ocean stratification and lower ventilation rates (Katavouta et al. 2019, Seltzer et al. 2023, Trenberth et al. 2025, Yamaguchi & Suga 2019).

On the global scale, the observed oxygen decline is negatively correlated with ocean heat content (OHC) changes, where the ratio of oxygen to heat changes is greater than that expected from the temperature-solubility relationship (Ito et al. 2017, Oschlies et al. 2018, Takano et al. 2023), indicative of the reinforcing effects from nonsolubility processes (Keeling & Garcia 2002). The relationship is nonlinear, suggesting a delayed response in the O_2 inventory to the impacts of excess heat gain. The least-understood driver of O_2 change is the variability in biological O_2

consumption and production. There are potential linkages between subsurface respiration rates and climate variability, either through the temperature enhancement of microbial reaction rates (Brewer & Peltzer 2017) or in the amount of exported organic material from the surface euphotic layer (Deutsch et al. 2011, Lachkar et al. 2018). Brewer & Peltzer (2017) estimated the Q_{10} factor of 3.63 for the oxygen consumption rates in the upper water column of the Sargasso Sea, which translates to a 29% increase in ocean oxygen consumption rates under a 2°C warming. Oxygen minimum zones (OMZs) are regions where O_2 levels reach hypoxic conditions and are commonly found in the tropical thermocline and in the subsurface (typically 200–1,000 m) waters of the eastern boundary upwelling regions in all ocean basins (Busecke et al. 2019, Davila et al. 2023, Garçon et al. 2019, Stramma et al. 2010, Zhou et al. 2022). The volume and intensity of the OMZs are linked to climate variability through the wind-driven ocean circulation system. The low-O_2 condition of the eastern boundary upwelling region is dynamically maintained by an increased export of organic matter and subsequent respiratory O_2 loss, reinforcing the high-nutrient, low-O_2 condition induced by the upwelling itself (Buchanan & Tagliabue 2021, Deutsch et al. 2011). Monsoon wind variability plays a fundamental role in modulating the OMZ and seasonal hypoxia of the Arabian Sea (DiMarco et al. 2023, Lachkar et al. 2018). Studies show that biological O_2 consumption rates can covary with ocean circulation and modes of climate variability, and these processes may potentially modulate the spatial patterns and/or changes in O_2 inventory.

Earth system models (ESMs) are the primary tools for projecting future changes in the Earth system, including rates of ocean deoxygenation. They are based on our current understanding of the physical, chemical, and biological processes of the atmosphere, ocean, and land surfaces. These processes and their interactions are represented as coupled partial differential equations that are numerically integrated using high-performance computing systems. These models, however, are subject to biases due to the limited computational resources in resolving fine-scale processes, incomplete parameterizations of physical and biogeochemical processes, and incomplete understanding of the initial and boundary conditions (Cabré et al. 2015, Séférian et al. 2020). Thus, it is crucial to evaluate a model's skills by comparing model simulations against observations. A better grasp of a model's skills has both scientific and social values. Scientifically, understanding how and where a model fails provides a valuable insight into which aspects of the model need improvement (e.g., enhanced model resolution versus improving parameterizations). Well-crafted model experiments can, in turn, explain the underlying mechanism behind unexplained observations, inform the attribution of forced versus internally generated trends (Long et al. 2016), and guide the observing system strategy by identifying critical scales of variability. More broadly, reliable models of ocean biogeochemistry and physics support our fundamental understanding of how dissolved oxygen sustains marine ecosystems through improved near-term predictions and long-term projections of future ocean deoxygenation. Accurate simulations can inform current and future marine environmental change, such as the compound events of ocean warming, acidification, and deoxygenation, which are essential factors controlling the metabolic constraints of species and the transition of marine habitats (Deutsch et al. 2020, Gruber et al. 2021). Dissolved oxygen content is also an essential input for ecological models of higher trophic levels and fisheries [Fisheries and Marine Ecosystem Model Intercomparison Project (Fish-MIP)] (Tittensor et al. 2018). A dissolved oxygen dataset from simulations is also utilized to understand projected changes in the marine ecosystem and fisheries, as well as to assess the cumulative impacts of climate change in the following decades (Halpern et al. 2025). Oxygen dynamics are tightly intertwined with carbon and nutrient cycling, and thus interrogating model skill for oxygen has broad implications for improving models' simulation of the ocean carbon cycle and primary productivity, both of which showcase large uncertainties (Kwiatkowski et al. 2020). Therefore, the broader scientific community would benefit from understanding the reliability of model

simulations, with implications for fisheries and ocean resource management under future climate change.

It is currently debated how well ESMs can reproduce observed oxygen distribution and trends from the historical datasets. In an earlier assessment, ESMs were unable to reproduce observed patterns of thermocline oxygen changes and significantly underestimated the historical global deoxygenation rates (Oschlies et al. 2018). Kwiatkowski et al. (2020) argued that the range of observed global oxygen trends encompasses the multi-model mean oxygen trend of ESMs from Coupled Model Intercomparison Project phase 6 (CMIP6). Takano et al. (2023) also showed similar global deoxygenation trends between the fully coupled ESMs and the observed global deoxygenation rate from a single dataset [Ito 2022 (Ito22)]. Observed decadal deoxygenation and near-term future climate projections are exposed to uncertainties due to internal variability. This uncertainty could be better quantified by combining the observational dataset and ensembles of model simulations (e.g., Marotzke 2019). The underestimation of the internal variability of dissolved oxygen is also under debate. The factors driving internal variability are not fully understood, but it is crucial for the robust detection and attribution of ocean deoxygenation.

The focus here is to review our current understanding and quantification of the O_2 distributions and multi-decadal changes based on state-of-the-art ESMs and the synthesis of the latest observation-based datasets. Bindoff et al. (2019) summarized results from earlier observational studies of global ocean deoxygenation (Helm et al. 2011, Ito et al. 2017, Schmidtko et al. 2017) with varying estimates for deoxygenation rates ranging from 0.5% to 3.3% loss for 0–1,000-m depths for 1970 to 2010. This synthesis exhibited a large uncertainty in the deoxygenation trends. Many factors can affect estimates of deoxygenation trends; each study can use different land-ocean masks, vertical and horizontal grids, primary data sources, measurement platforms, data quality control (QC) methods, and interpolation mapping methods. Variations in any of these factors can lead to different estimates, resulting in large uncertainties in deoxygenation assessment. It is timely to revisit the model-observation comparison because several new O_2 datasets [Gouretski et al. 2024, Ito et al. 2024a (Ito24), Roach & Bindoff 2023 (RB23), Sharp et al. 2023] and the new World Ocean Atlas 2023 (WOA23) O_2 climatology (Garcia et al. 2024, Reagan et al. 2024), including Argo-O_2 profiles as an additional primary data source, have emerged. These new datasets are available in the public domain, enabling the detailed comparison of oxygen distributions and multi-decadal trends as well as the assessment of their uncertainties. Because oxygen represents an integrated signal of physical and biogeochemical changes (such as ventilation dynamics and biogeochemical sources and sinks), revisiting the model-observation comparison with emerging new datasets can inform future efforts to assess simulated ocean deoxygenation, taking into account the uncertainties in both models and observations.

2. APPROACHES TO ASSESS OCEAN OXYGEN CHANGES

2.1. Observations

Several new observation-based global O_2 datasets have been published in peer-reviewed journals, and these datasets have been made publicly available since 2022. These data products are accompanied by descriptions of the source data, QC steps, and mapping interpolation methods listed in **Table 1**. Technical details of each dataset are available in the references, and they are not repeated here. Instead, we provide an overview of the procedures and challenges in estimating global O_2 inventory that are common to all these datasets, starting from the primary data sources.

A unique challenge in estimating deoxygenation trends is that sampling density for O_2 is significantly lower in time and space than that of temperature and salinity. This difference results in substantial data gaps that are difficult to fill when generating time-varying datasets. **Figure 1**

Table 1 List of observational datasets used in this study

Dataset	Mapping method	Primary data	Time period	Frequency	Reference
WOA23	Objective analysis	Bottle, CTD, Argo	Climatology	Monthly	Garcia et al. 2024
IAP	Ensemble optimum interpolation	Bottle, CTD, Argo	Climatology	Monthly	Cheng et al. 2024
Ito22	Optimum interpolation	Bottle	1965–2020	Annual	Ito 2022
RB23	DIVA	Bottle, CTD	1955–2018	Annual	Roach & Bindoff 2023
GOBAI	ML	Bottle, CTD, Argo	2004–2022	Monthly	Sharp et al. 2023
Ito24	ML	Bottle, CTD, Argo	1965–2020	Monthly	Ito et al. 2024a

(*blue*) The top two rows are independently quality-controlled and mapped climatological mean datasets. (*orange*) The bottom four rows are time-varying datasets. All time-varying datasets, except for Ito22, provide climatological means; Ito22 includes anomalies only. Abbreviations: CTD, conductivity, temperature, and depth; DIVA, Data Interpolating Variational Analysis; GOBAI, Gridded Ocean Biogeochemistry from Artificial Intelligence; IAP, Institute of Atmospheric Physics; Ito22, Ito 2022; Ito24, Ito et al. 2024a; ML, machine learning; RB23, Roach & Bindoff 2023; WOA23, World Ocean Atlas 2023.

shows the distribution of quality-controlled, shipboard [bottle and conductivity, temperature, and depth (CTD)] and Argo-O_2 measurements based on World Ocean Database (WOD) version 2023 (Mishonov et al. 2024). WOD is an international collaboration among national data centers, oceanographic research institutions, and investigators to provide a comprehensive dataset of quality-controlled oceanographic variables. The number of profiles taken each year or month fluctuates significantly. Bottle samples are measured using Winkler titrations with an accuracy of O(0.1%) or ~0.3 μmol/kg (Carpenter 1965). Before 1990, most O_2 profiles were taken by

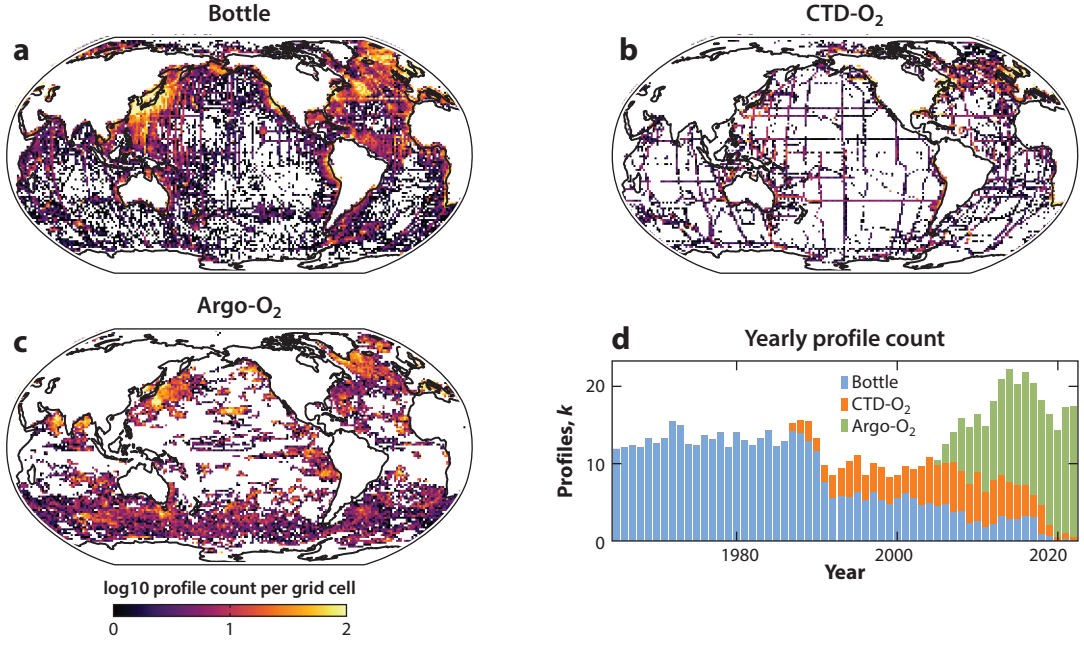

Figure 1

Sampling density of O_2 data in the World Ocean Database 2023 from 1965 to 2022. (*a–c*) The base-10 logarithm of the profile count within the 1 × 1 longitude-latitude grid cell for bottles (Winkler titration), CTD-O_2, and Argo-O_2 sensors. (*d*) Annual profile count for the three measurement methods: bottle (*blue*), CTD-O_2 (*orange*), and Argo-O_2 (*green*) sensors. Abbreviation: CTD, conductivity, temperature, and depth.

ship-based bottle measurements. Following the 1990s, the number of CTD-O_2 sensors increased, making them a major source of O_2 data. Most CTD-O_2 profiles are calibrated against a smaller number of bottle measurements from the same cruise.

Since the mid-2000s, the number of Argo-O_2 profiles has steadily increased. Argo-O_2 data are distributed by the Argo Global Data Assembly Center in real-time and delayed mode. Argo-O_2 sensors often return data that are out of calibration, and the delayed-mode adjusted O_2 profiles provide adjustments to improve the data accuracy. The Argo-O_2 dataset and its calibration methods are still evolving (Maurer et al. 2021), and the current uncertainty estimates of delayed-mode adjusted oxygen are O(1%) or ~3 μmol/kg. The finite response time of optode sensors can further influence calibration bias for Argo-O_2, potentially causing systemic biases in the oxycline regions (Bittig & Körtzinger 2017, Bittig et al. 2014).

Historical measurements were collected for different scientific purposes, and the measurement uncertainty is not expected to be uniform over time. Data QC protocols can play a crucial role in developing unified datasets from different measurement platforms with distinct accuracy. Primary QC typically includes checking the data format and internal consistency for corrupted data and making sure that the data are in common units. Questionable data points and outliers are statistically identified and flagged, and the QC flags are recorded with the raw data. Secondary QC steps can then be applied, which may include recalibrations and adjustments of ocean community data by subject matter experts. While WOD provides O_2 profiles with the QC flags, different groups may apply additional secondary QC, which could lead to varying estimates of oxygen distribution and its changes.

Traditionally, bottle data are considered to have the highest accuracy and have been the preferred primary data source for gridded datasets such as *World Ocean Atlas 2018* (Garcia et al. 2018) and Ito (2022). However, the amount of bottle data has been declining since the 1990s (**Figure 1*d***), limiting its data coverage, especially for recent decades. Thus, newer gridded oxygen datasets include bottle, shipborne CTD-O_2, and Argo-O_2 sensors (Garcia et al. 2024, Ito et al. 2024a, Sharp et al. 2023), while others may include bottle and CTD-O_2 data. Differences in measurement platforms for the primary data can lead to variations in estimates of O_2 inventory. Discrepancies between different measurement platforms can be evaluated by comparing profiles that are measured within close proximity over space and time. For example, Wang et al. (2025) and Bushinsky et al. (2025) have identified systematic differences between shipboard measurements and Argo-O_2 data in the subsurface waters where Argo-O_2 tends to underestimate the shipboard data. While this offset is still within the Argo-O_2 stated accuracy of O(1%), systemic bias of this magnitude can significantly impact the estimates of global deoxygenation rates (Ito et al. 2024a).

To estimate the ocean's O_2 inventory, it is necessary to fill data gaps in space and time when observations are not available. All datasets listed in **Table 1** used different gap-fill approaches. These mapping methods create spatially and temporally continuous fields from discrete measurements, and we re-remap these datasets onto a common 1° × 1° grid system for the comparison. World Ocean Atlas mapping is based on an objective analysis (Garcia et al. 2024). The Institute of Atmospheric Physics (IAP) dataset is based on ensemble optimal interpolation (Cheng et al. 2024), whereas Ito (2022) used a simpler implementation of optimal interpolation using a Gaussian covariance function. Roach & Bindoff (2023) employed the Data Interpolating Variational Analysis method (Barth et al. 2014, Troupin et al. 2012), an optimization-based interpolation technique that minimizes a cost function based on data misfit and field smoothness. Machine learning–based methods are new, data-driven approaches that train algorithms to predict oxygen concentrations from related predictor data, which can represent complex and nonlinear patterns (Ito et al. 2024a, Sharp et al. 2023).

In our analysis, we utilize six published global O_2 data products (**Table 1**), each of which incorporates different data sources, QC methods, and mapping approaches. These products can exhibit notable differences in spatial patterns and temporal trends, reflecting the complexity and variability inherent in the reconstruction of ocean biogeochemical states. The discrepancies among the datasets arise from a combination of factors, including the choice of primary observational data (bottle, CTD-O_2, and autonomous floats), the coverage of measurement platforms, the QC procedures applied to the raw data, and the specific mapping and interpolation techniques used to generate gap-filled gridded fields. Because these components are deeply interlinked and often applied through the workflow, it is challenging to isolate and attribute the overall differences in the final products to any single factor. Some datasets prioritize long-term consistency and stability by selecting shipboard data only (e.g., Ito22, RB23). The others focus on maximizing data coverage by including Argo-O_2 data [e.g., IAP, Gridded Ocean Biogeochemistry from Artificial Intelligence (GOBAI), Ito24)], which can further contribute to the differences. With these complexities in mind, our approach does not attempt to disentangle the influence of each factor. Instead, it treats the ensemble spread across the six products as a pragmatic estimate of observational uncertainty. By quantifying the differences between the products, in terms of both mean state and temporal evolution, we aim to characterize the bounds of uncertainty in O_2 observations, which is essential for validating models, detecting long-term trends, and assessing regional variability in the global ocean.

2.2. Earth System Models

Many models participating in CMIP6 include biogeochemical model components that simulate the cycling of carbon, nutrients, and oxygen in the ocean, as well as terrestrial carbon fluxes. These coupled ESMs undergo an initial spin-up phase under preindustrial (circa 1850) conditions, allowing the physical and biogeochemical components to reach a dynamic quasi-equilibrium. However, there is no consistent and strict criterion for this procedure, and each model adopts their independent methods and lengths of spin-ups (Séférian et al. 2016). Following spin-up, the models are integrated forward using the historical forcing protocol described by Eyring et al. (2016), which spans the period from 1850 to 2014. These historical simulations are driven by time-evolving atmospheric forcings, including observed greenhouse gas concentrations (e.g., CO_2, CH_4, N_2O), anthropogenic and volcanic aerosols (both stratospheric and tropospheric), solar variability, and land-use changes such as deforestation and agricultural expansion. The realism and fidelity of these models in simulating the climatological distribution of ocean biogeochemical states have been comprehensively documented by Séférian et al. (2020), while the transient changes in oxygen, OHC, and CO_2 fluxes have been evaluated by Takano et al. (2023) and Liddicoat et al. (2021). For this review, model output from the historical simulations is analyzed over the period 1965–2014, a time frame that overlaps with the modern observational era and is summarized in **Table 2**.

Differences in the simulated oxygen arise primarily from two sources: model structural differences and internal variability. Model structure refers to the design and configuration of each ESM. This includes, e.g., the representation of physical climate and biogeochemical processes, such as precipitation, clouds, cryospheric processes, ocean mixing, primary productivity, the export of organic particles from the surface ocean, remineralization rate, and land-atmosphere interactions, as well as choices in spatial resolution and coupling strategies between system components. Even though CMIP6 models use standardized external forcings, differences in model design can lead to varying climate states, contributing to what is often referred to as structural uncertainty. In addition, the simulated deoxygenation rates may be influenced by model drift. The inconsistency in

Table 2 The list of Coupled Model Intercomparison Project phase 6 Earth system models whose historical simulations are used in this study

Model	Reference(s)
Australian Community Climate and Earth System Simulator Earth System Model version 1.5 (ACCESS-ESM1-5)	Ziehn et al. 2020
Euro-Mediterranean Centre on Climate Change Earth System Model version 2 (CMCC-ESM2)	Lovato et al. 2022
Centre National de Recherches Météorologiques Earth System Model version 2-1 (CNRM-ESM2-1)	Séférian et al. 2019
Canadian Earth System Model version 5 (CanESM5)	Christian et al. 2022, Swart et al. 2019
Canadian Earth System Model version 5 Canadian Ocean Ecosystem model (CanESM5-CanOE)	Christian et al. 2022, Swart et al. 2019
Energy Exascale Earth System Model version 1-1 (E3SM-1-1)	Burrows et al. 2020, Moore et al. 2004
Geophysical Fluid Dynamics Laboratory Climate Model version 4 (GFDL-CM4)	Dunne et al. 2020a, Held et al. 2019
Geophysical Fluid Dynamics Laboratory Earth System Model version 4 (GFDL-ESM4)	Dunne et al. 2020b, Stock et al. 2020
Institut Pierre-Simon Laplace Climate Model version 6A Low Resolution (IPSL-CM6A-LR)	Boucher et al. 2020
Model for Interdisciplinary Research on Climate, Earth System version 2 for Long-term simulations (MIROC-ES2L)	Hajima et al. 2020
Max Planck Institute for Meteorology Earth System Model version 1.2 Low Resolution (MPI-ESM1-2-LR)	Ilyina et al. 2013, Mauritsen et al. 2019, Paulsen et al. 2017
Meteorological Research Institute Earth System Model version 2.0 (MRI-ESM2-0)	Yukimoto et al. 2019
Norwegian Earth System Model version 2 (NorESM2-LM)	Tjiputra et al. 2020
UK Earth System Model version 1.0 (UKESM1-0-LL)	Sellar et al. 2019, Yool et al. 2020

When there are multiple ensemble members, only the first ensemble member is used.

the preindustrial spin-up procedure mentioned above leads to different initial states at the start of the historical period for each ESM and may introduce various degrees of model drift contributing to the divergence in the projected long-term oxygen evolution, especially for deep waters.

Internal variability reflects the ESM-dependent spontaneous and chaotic fluctuations within the climate system that occur without changes in external forcings (e.g., Deser et al. 2020). These fluctuations can result from processes such as ocean-atmosphere interactions [e.g., El Niño–Southern Oscillation and associated teleconnections (Eddebbar et al. 2017)], random weather patterns, or turbulent mixing and can delay the emergence of anthropogenic deoxygenation until later in the twenty-first century over large parts of the thermocline (Long et al. 2016). Abe & Minobe (2023) quantified the impact of internal variability on the North Pacific deoxygenation rates using the ESMs with multiple ensemble members, showing the diverging deoxygenation rates of the Subarctic North Pacific due to internal dynamics alone.

Disentangling the relative contributions of model structure and internal variability to the total spread in deoxygenation projections is essential for robust uncertainty attribution and model evaluation. However, such an analysis typically requires large ensembles from individual models, which are beyond the scope of this review. Instead, we treat the multi-model ensemble spread as a combined measure of total model uncertainty, recognizing that both sources contribute to varying degrees depending on the timescale and spatial scale of interest.

2.3. Uncertainty Quantification Approaches

Our overarching question is whether the current ESMs can reproduce the observed O_2 inventories and their long-term trends. Evaluation of consistency between observation-based fields and models must consider the uncertainties in both. Exact matches are obviously not expected, but their differences can be compared to the uncertainties to determine whether the differences can be attributed to a systematic bias or to random fluctuations.

As discussed in Section 2.1, there are three types of observational uncertainties: measurement uncertainty, sampling uncertainty, and analysis uncertainty (**Supplemental Figure 2**). Each of them can include systematic biases and random errors. Measurement error arises from the precision of the instruments used and the accuracy of the calibration, which are specific to the monitoring platforms. Sampling uncertainty depends on the measurement patterns and frequency as well as the spatio-temporal variability of O_2 distribution. Sparse sampling patterns may not fully capture the real O_2 variability and may potentially cause biases in representing seasonal, interannual, and decadal variability. Analysis uncertainty depends on many factors, including the data QC methods and criteria selection, bias correction, and vertical and horizontal data interpolation errors. Our approach is to use the spread among multiple observation-based products as a measure of bulk observational uncertainties. Specifically, five datasets are used for the long-term mean climatologies (**Table 1**, as Ito22 includes anomalies only), and three datasets are used for multi-decadal trends since 1965. There are two caveats. First, there are significant overlaps in the primary data; thus, all datasets include similar measurement and sampling errors. Thus, our estimates of the observational uncertainties primarily reflect analysis errors. It is then possible to compare the estimated analysis uncertainties against the typical measurement error of O(0.1%) for bottle and CTD-O_2 data and O(1%) for Argo-O_2 data. Second, there are only three time-varying datasets covering the period since 1965, which are limited in characterizing the uncertainty. Therefore, the uncertainties in long-term mean climatologies are more robust than those of multi-decadal trends; however, trend analysis helps check the uncertainties in the deoxygenation rates. The standard deviation among the five observational products is used as the observational uncertainty for climatological means, and the range (maximum–minimum) among the three observational products is used for the multi-decadal trends.

Supplemental Material >

Model uncertainties comprise two primary sources: model structure differences and internal variability, as discussed in Section 2.2. Multi-model standard deviation across the 14 models can provide a measure of the combined model uncertainties. When averaged over wide areas or large volumes, it is often assumed that structural uncertainties dominate at global scales and over long timescales. Local O_2 anomalies caused by internal variability may cancel out when integrated across vast regions, whereas the forced signal (e.g., anthropogenic warming and associated ocean stratification) is widespread and tends not to average out over wide regions. This conceptual separation may be useful and will be checked as we evaluate the differences between modeled and observed oxygen trends. Similar arguments can be made for observational uncertainties as integrating over large volumes can often average out random noise in the local concentrations, potentially leading to relatively small uncertainties in the volume-integrated inventories.

3. CLIMATOLOGICAL O_2 DISTRIBUTIONS AND INVENTORIES

The climatological O_2 distribution refers to the time-mean spatial pattern, computed over a multi-year period, typically several decades. Here, the period is standardized to 1965–2014, except for the GOBAI dataset, which covers 2004–2022 (Argo era). The long-term averaging smooths out short-term variability, providing a stable baseline condition. Climatological means thus provide a reasonable first step to gauge the observational and model uncertainties and how well ESMs

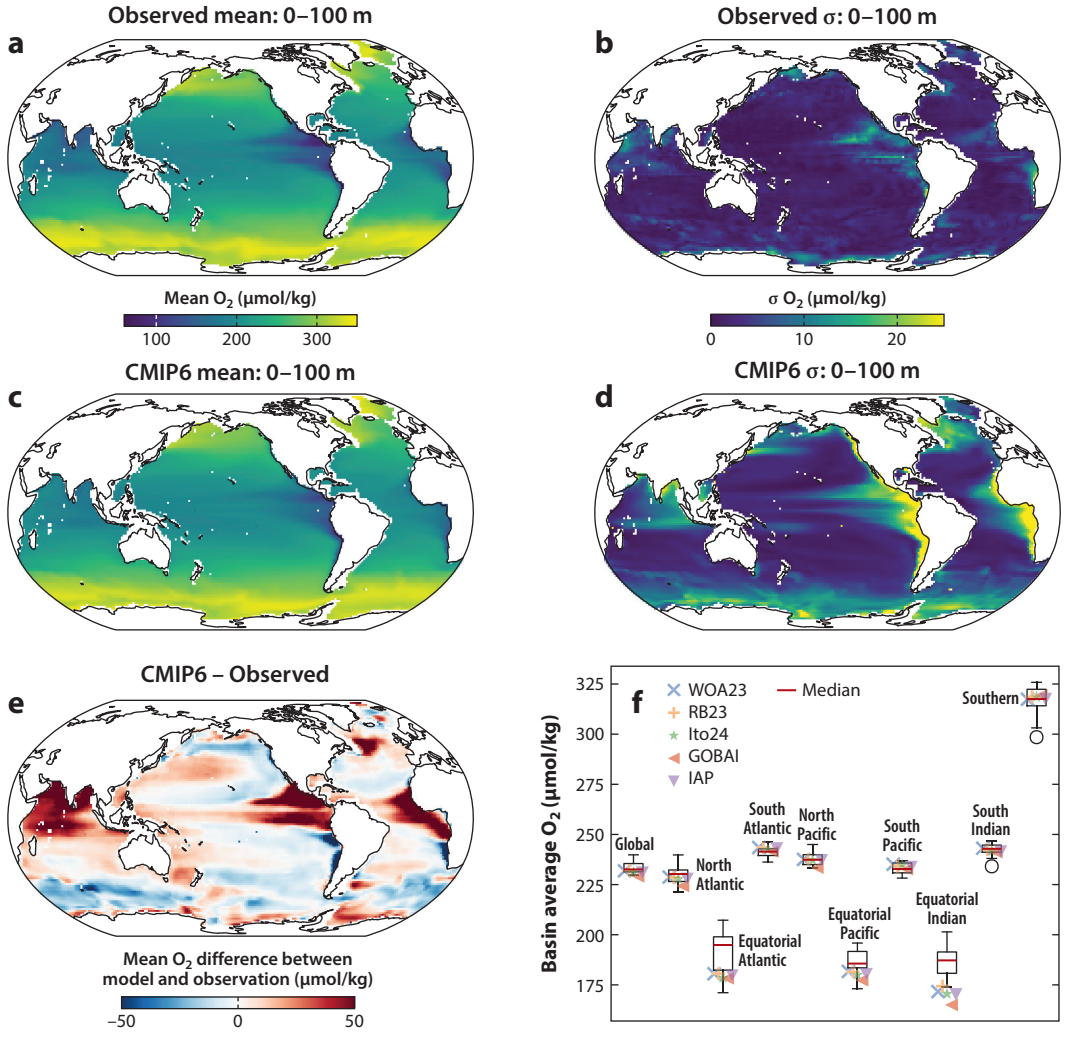

Figure 2

Climatological mean O_2 patterns averaged over 0–100-m depth. (*a*,*b*) Mean and standard deviation among the multiple observational products listed in **Table 1** (except for Ito22). (*c*,*d*) The same for the CMIP6 ESMs listed in **Table 2**. The standard deviation represents interobservation (*b*) and intermodel (*d*) spread. (*e*) Difference between CMIP6 multi-model mean and observational mean. (*f*) Basin-scale breakdown of the model statistics (*box-whisker*) and observational products (*colored markers*). Definitions of the basins are provided in **Supplemental Figure 1**. The box extends from the first quartile to the third quartile of the data, with a line at the median. The whiskers extend from the box to the farthest data point lying within 1.5 times the interquartile range from the box, and the open circles represent the outliers from the ESMs. Abbreviations: CMIP6, Coupled Model Intercomparison Project phase 6; ESM, Earth system model; GOBAI, Gridded Ocean Biogeochemistry from Artificial Intelligence; IAP, Institute of Atmospheric Physics; Ito22, Ito 2022; Ito24, Ito et al. 2024a; RB23, Roach & Bindoff 2023; WOA23, World Ocean Atlas 2023.

Supplemental Material >

can reproduce observed O_2 distributions. Before calculating the climatological means, all observational and model datasets are interpolated to the standard grid with 1° × 1° longitude-latitude resolution and 47 depth levels, focusing on the upper 1,000 m. Climatological annual mean O_2 is averaged vertically for near-surface waters (0–100 m) in **Figure 2** and thermocline (100–600 m) in **Figure 3**, including its observational and model ensemble means and their standard deviation.

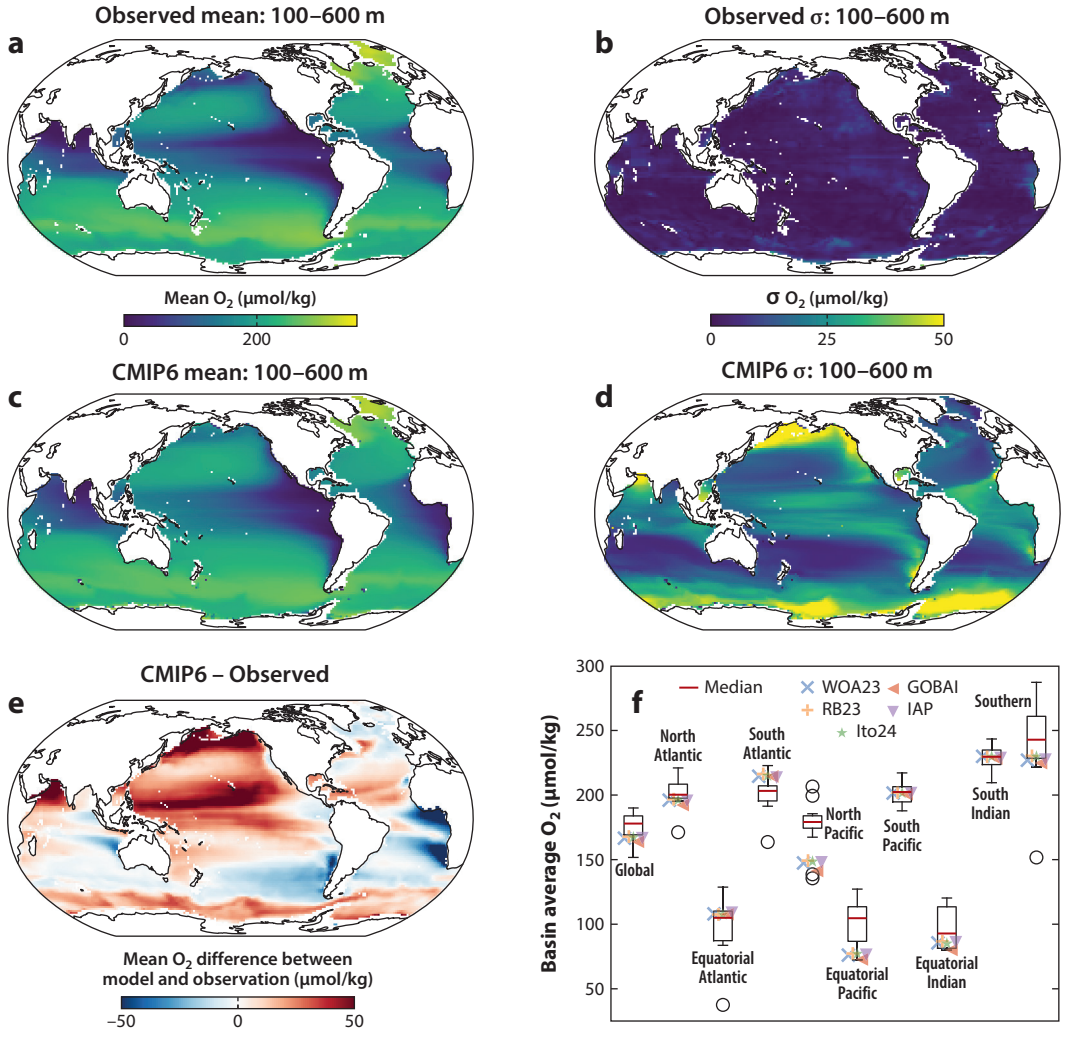

Figure 3

Climatological mean O_2 patterns averaged over 100–600-m depth. (*a*,*b*) Mean and standard deviation from the observational products listed in **Table 1** (except for Ito22). (*c*,*d*) The same for the CMIP6 ESMs listed in **Table 2**. The standard deviation represents interobservation (*b*) and intermodel (*d*) spread. (*e*) Difference between CMIP6 multi-model mean and observational mean. (*f*) Basin-scale breakdown of the model statistics (*box-whisker*) and observational products (*colored markers*). The open circles represent the outliers from the ESMs. Abbreviations: CMIP6, Coupled Model Intercomparison Project phase 6; ESM, Earth system model; GOBAI, Gridded Ocean Biogeochemistry from Artificial Intelligence; IAP, Institute of Atmospheric Physics; Ito22, Ito 2022; Ito24, Ito et al. 2024a; RB23, Roach & Bindoff 2023; WOA23, World Ocean Atlas 2023.

The plotted O_2 maps do not include the Arctic Ocean and marginal basins (e.g., the Mediterranean Sea) because some of the observational datasets cover only major ocean basins, namely the Atlantic, Pacific, Indian, and Southern Oceans.

The near-surface O_2 distribution from observations and models (**Figure 2*a*,*c***) shares many similar structures with increased O_2 levels at high latitudes, and low-O_2 waters are concentrated in the tropics, especially near the eastern boundary of the basins. When we compare the CMIP6 models and observational data products, the pattern correlation is 0.98 between the mean of five

observational products and the multi-model mean of the 14 CMIP6 ESMs. The models slightly overestimate with a global mean bias of 2.1 μmol/kg. The root mean square error (RMSE) is 9.5 μmol/kg, which measures the grid-scale differences. The positive biases are mainly focused on the tropical oceans of all basins, with the models overestimating by ∼20 μmol/kg (**Figure 2*e***). The models tend to underestimate near-surface O_2 in the eastern boundary upwelling regions, and the model biases in the Southern Ocean are patchy, including both positive and negative biases. The global ocean is divided into nine regions: the North Atlantic, Equatorial Atlantic, South Atlantic, North Pacific, Equatorial Pacific, South Pacific, Equatorial Indian, South Indian, and Southern Oceans. The northern/southern boundaries of the equatorial regions are set to 15°N/15°S. The northern boundary of the Southern Ocean is at 50°S. The basin-wise mean O_2 shows general agreement between CMIP6 models and the five observational datasets except for the tropical oceans, where the middle fiftieth percentile model is generally outside of the observed range (**Figure 2*f***). For the tropical Atlantic and Pacific basins, observations align with the bottom twenty-fifth percentile of the model, but in the tropical Indian Ocean, many observational datasets fall outside the model range.

In the thermocline, the pattern correlation remains high at 0.96 between the mean observational products and the multi-model CMIP6 ESMs mean. The models overestimated by a mean bias of 9.0 μmol/kg, O(3%). The RMSE is 22 μmol/kg (**Figure 3*a*,*c***). The model biases are approximately a factor of four greater in the thermocline than in the near-surface waters. The comparison of the multi-model mean and observations shows a strong overestimation in the tropical and Subarctic North Pacific (**Figure 3*e***), which is partially offset by the underestimation in the South Atlantic and South Pacific Oceans. The multi-model mean also overestimates the polar Southern Ocean. The intermodel differences are greater in the thermocline (**Figure 3*f***), including several outliers in the North Pacific with a large spread among the models in the Subarctic North Pacific (**Figure 3*d***). In the North Pacific basin, observational datasets fall outside the model range, indicating widespread model biases. The South Pacific and South Indian appear to be consistent with the observation on the basin average sense, but the spatial pattern shows compensating positive and negative biases within the basin. There are relatively large model uncertainties in the tropical oceans of all basins, but even so, the observational datasets only overlap with the bottom twenty-fifth percentile of the model in the Equatorial Pacific.

The model-observation differences can be evaluated with respect to the uncertainties as measured by the spread of observational and modeled O_2 fields, respectively. For the 0–100 m climatologies, the spread among the observational datasets is much smaller than that of CMIP6 models. Horizontally averaged standard deviation (calculated as the square root of the horizontally averaged variance) is 2.2 μmol/kg or approximately O(1%) for the observational datasets and is 8.6 μmol/kg or approximately O(3%) for the CMIP6 ESMs. As visualized in **Figure 2*b*,*d***, the model uncertainties are approximately a factor of three greater than the observations at 0–100 m. Some CMIP6 models' climatologies can lie outside of the observational range, even though the multi-model means are very close (within 1%) to the observations.

In the thermocline (100–600 m), the average standard deviations across the observational products are 3.5 μmol/kg, which are slightly higher than the near-surface values. However, the mean standard deviation across the CMIP6 ESMs is 23 μmol/kg, which is a factor of three greater than the near-surface waters and a factor of seven greater than the observational uncertainty in the thermocline. The high mean bias and RMSE indicate that the model skills are weaker in the thermocline, and the models have greater uncertainties. This highlights the importance of considering model structural uncertainties and biases, as climatological means tend to average out variability. As a result, many CMIP6 models' climatologies can lie outside of the observational range in the thermocline. Regionally, tropical oceans in all basins show an overestimation of O_2 in the

Table 3 Climatological O_2 inventories from observational datasets in units of $PmolO_2$ ($=10^{15}$ $molO_2$)

	0–100 m	100–300 m	300–600 m	600–1,000 m	0–1,000 m
WOA23	7.05	11.40	14.36	17.12	49.92
RB23	7.06	11.47	14.49	17.27	50.28
Ito24	7.05	11.44	14.41	17.20	50.11
IAP	7.03	11.39	14.36	17.13	49.91
GOBAI	6.98	11.19	14.12	16.91	49.19
Ensemble mean	7.03	11.38	14.35	17.13	49.88
Ensemble standard deviation	0.03	0.10	0.12	0.12	0.37
Ensemble standard deviation (%)	0.42	0.87	0.86	0.71	0.75

The quasi-global domain, excluding the Arctic Ocean and marginal basins, is used for global integration. Abbreviations: GOBAI, Gridded Ocean Biogeochemistry from Artificial Intelligence; IAP, Institute of Atmospheric Physics; Ito24, Ito et al. 2024a; RB23, Roach & Bindoff 2023; WOA23, World Ocean Atlas 2023.

near-surface (0–100 m) waters despite the larger model uncertainties. In the thermocline, tropical oceans, as well as the North Pacific and Southern Oceans, show greater model uncertainties.

Next, we compare the volume-integrated O_2 inventory in the upper 1,000 m based on the annual mean climatology. A quasi-global land-ocean mask is developed based on ocean grid cells, incorporating all datasets that contain O_2 values, including observations and models. Applying the standard land-ocean mask is crucial for the intercomparison of the O_2 inventory. This quasi-global domain has somewhat less ocean volume compared to the realistic global ocean bathymetry because it excludes the Arctic Ocean and all marginal basins. The common units for O_2 concentrations are moles per unit mass, and seawater density needs to be considered before integrating over the ocean volume. For simplicity, we use a uniform constant density of 1,025 kg/m^3 such that differences in O_2 inventories arise solely from the climatologies.

Observation-based O_2 inventory is approximately 50 $PmolO_2$ for the 0–1,000-m depth range in the quasi-global domain (**Table 3**). Ensemble standard deviation for different depth levels ranges from 0.03 to 0.37 $PmolO_2$, or equivalently, 0.42% to 0.87%, smaller than the grid-level uncertainty of ∼1%. Comparing individual datasets, the GOBAI O_2 inventory is the smallest across all depth ranges, likely reflecting that all datasets have the same temporal coverage (1965–2014), except for the WOA23 (1965–2022) and GOBAI (2004–2022). Thus, GOBAI covers only recent periods, and considering the potential impact of the multi-decadal loss of O_2, we anticipate better agreement when the GOBAI O_2 inventory is excluded. Indeed, the ensemble standard deviation ranges from 0.01 to 0.15 $PmolO_2$, or equivalently, 0.13% to 0.35% when recalculated without the GOBAI dataset. In this case, the observation-based O_2 inventories are closely consistent with one another, with less than O(0.1%) uncertainty with the standardized analysis period.

The CMIP6 ESMs' O_2 inventory exhibits much larger uncertainties. The temporal coverage of the models is standardized from 1965 to 2014, matching the observations (except for GOBAI). Across all depth ranges, the model-based inventory is slightly higher than the observations (**Table 4**). The multi-model mean of 51.7 $PmolO_2$ for the 0–1,000-m-depth range is about 3.6% higher than the observational ensemble mean. The multi-model standard deviation ranges from 0.1 to 4.0 $PmolO_2$, or equivalently, 1.6% to 13.1%, which is about an order of magnitude greater than the observational uncertainty. The most significant model uncertainty, at 13.1%, is found in the intermediate water of the 600–1,000-m depth range, followed by the lower thermocline, with 8.9% uncertainty in the 300–600-m depth range. Model uncertainties are relatively smaller in the shallower waters. Climatological means tend to average out the internal variability, so the

Table 4 Climatological O_2 inventories from the Coupled Model Intercomparison Project phase 6 historical runs in units of PmolO_2 (=10^{15} molO_2)

	0–100 m	100–300 m	300–600 m	600–1,000 m	0–1,000 m
Australian Community Climate and Earth System Simulator Earth System Model version 1.5 (ACCESS-ESM1-5)	7.05	12.34	15.36	16.96	51.71
Canadian Earth System Model version 5 (CanESM5)	7.31	12.92	16.34	20.51	57.08
Canadian Earth System Model version 5 Canadian Ocean Ecosystem model (CanESM5-CanOE)	7.07	12.77	16.42	19.25	55.51
Euro-Mediterranean Centre on Climate Change Earth System Model version 2 (CMCC-ESM2)	7.01	11.53	13.64	16.25	48.42
Centre National de Recherches Météorologiques Earth System Model version 2-1 (CNRM-ESM2-1)	7.05	11.30	13.98	14.80	47.13
Energy Exascale Earth System Model version 1-1 (E3SM-1-1)	7.32	11.24	11.25	11.20	41.01
Geophysical Fluid Dynamics Laboratory Climate Model version 4 (GFDL-CM4)	7.19	12.51	15.59	18.79	54.07
Geophysical Fluid Dynamics Laboratory Earth System Model version 4 (GFDL-ESM4)	7.14	12.29	15.72	17.56	52.71
Institut Pierre-Simon Laplace Climate Model version 6A Low Resolution (IPSL-CM6A-LR)	7.17	12.32	16.06	17.66	53.22
Model for Interdisciplinary Research on Climate, Earth System version 2 for Long-term simulations (MIROC-ES2L)	7.09	12.04	14.89	17.83	51.86
Max Planck Institute for Meteorology Earth System Model version 1.2 Low Resolution (MPI-ESM1-2-LR)	6.97	11.81	15.10	16.93	50.81
Meteorological Research Institute Earth System Model version 2.0 (MRI-ESM2-0)	7.27	12.39	16.35	20.01	56.02
Norwegian Earth System Model version 2 (NorESM2-LM)	7.00	11.85	15.05	17.46	51.36
UK Earth System Model version 1.0 (UKESM1-0-LL	7.07	11.33	14.47	19.47	52.33
Ensemble mean	7.12	12.05	15.02	17.48	51.66
Ensemble standard deviation	0.11	0.53	1.34	2.29	3.95
Ensemble standard deviation (%)	1.57	4.41	8.91	13.08	7.65

The quasi-global domain is used for the volume integration, consistent with the one used for the observation-based dataset.

relatively large intermodel differences indicate the importance of model structural uncertainties in the model representation of ocean ventilation and respiratory O_2 loss in the lower thermocline and intermediate waters. Despite a relatively large spread across individual models of O(10%), the multi-model mean O_2 inventory is much closer to the observations within a few percent, indicating that the CMIP6 models, when averaged together, are better compared to the observed climatological mean inventories.

4. MULTI-DECADAL O_2 CHANGES

Multi-decadal changes in O_2 inventories are calculated from four observation-based datasets (Ito22, RB23, GOBAI, Ito24) and the 14 CMIP6 historical runs using the same land-ocean masks. The observational ensembles comprise datasets based on bottle data only (Ito22), bottle and CTD-O_2 (RB23), bottle and Argo-O_2 (GOBAI), and bottle, CTD-O_2, and Argo-O_2 data sources (Ito24), utilizing different types of mapping interpolation methods, ranging from statistical to machine learning–based approaches. Directly comparing observation-based and model-based O_2 inventories is challenging due to the large intermodel spread of O(10%). Thus, the comparison is performed on the O_2 inventory anomalies after removing the mean seasonal cycle and subtracting the decadal means from 1965 to 1975 from each dataset (**Figure 4*a*,*b***). The comparison focuses on the depth ranges of 0–1,000 m and 100–600 m for the period 1965–2014, except for GOBAI, which spans 2004–2022. To calculate the O_2 inventory anomalies for the GOBAI dataset, the 1965–1974 decadal mean from the Ito24 dataset is subtracted, forming a nearly continuous time series with a consistent offset for the baseline (1965–2014) condition.

Observed multi-decadal trends exhibit a wide range of variation. The multi-decadal trends from the three datasets (Ito22, RB23, and Ito24) are −107, −225, and −309 Tmol/decade (1 Tmol = 10^{12} mol) (equivalent of 1.1%, 2.6%, and 3.1% O_2 loss) for 0–1,000 m and −69, −129 and −174 Tmol/decade (equivalent of 1.3%, 2.5%, and 3.4% O_2 loss) for 100–600 m over the 50-year period from 1965 to 2014. The GOBAI dataset is limited to recent periods (2004–2022). Due to the difference in time period, the GOBAI dataset is not included for further comparison of multi-decadal trends, but it appears to follow close trajectories with the Ito24 for the overlapping period (**Figure 4*a*,*b***). This leaves only three datasets, Ito22, RB23, and Ito24, that fully cover the 50-year period from 1965 to 2014. It is not adequate to fully characterize the observational uncertainties of O_2 inventory trends with only three datasets, which makes our assessment preliminary.

The weakest trend from the observations is found in the Ito22 dataset, with 107 and 69 Tmol/decade for depths of 0–1,000 m and 100–600 m, respectively. These trends from Ito22 appear to be much smaller than those in the other two datasets, and their interpretation requires caution. The Ito22 dataset is based solely on the bottle data, and data gaps are filled using the optimal interpolation method. The optimal interpolation method tends to fill data gaps with diminishing anomalies when there are no nearby observational data points. In the case of Ito22, a Gaussian covariance is assumed with a 750-km e-folding length scale. The amount of bottle O_2 data has been declining steadily since the 1990s, resulting in wide data gaps after the 1990s (**Figure 1**). When the optimal interpolation is applied to such sparse data, many grid cells contain no bottle samples within a few multiples of the e-folding scale, and it significantly underestimates the O_2 anomalies (Ito et al. 2024b). The RB23 dataset includes bottle and CTD-O_2 data sources, whereas the Ito24 dataset includes bottle, CTD-O_2, and Argo-O_2 data sources offering more extensive data coverage, especially after the 2010s.

The CMIP6 ensemble's O_2 trends are 110 ± 46 and 74 ± 27 Tmol/decade from 1965 to 2014 in 0–1,000 m and 100–600 m depths, respectively (equivalent to $1.1 \pm 0.5\%$ and $1.4 \pm 0.5\%$ O_2 loss over the 50-year period). The model uncertainty is calculated as the ensemble standard deviation

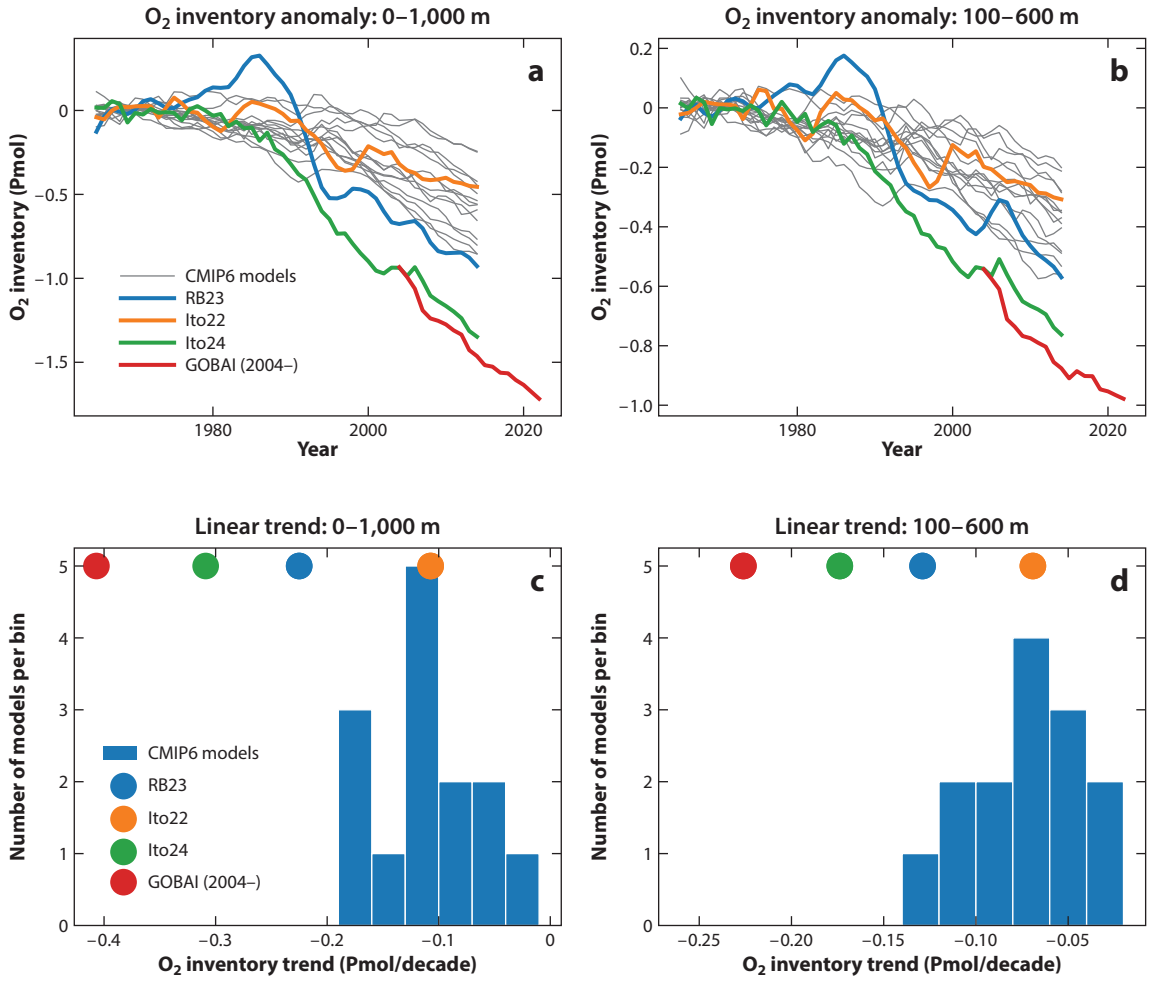

Figure 4

Observed and simulated O_2 inventory trends. (*a*,*b*) O_2 inventory anomaly time series relative to the (1965–1974) decadal means for the 0–1,000-m and 100–600-m depth ranges. (*c*,*d*) Distribution of CMIP6 models' O_2 inventory trends for the 0–1,000-m and 100–600-m depth ranges. The time period for the trend calculation is 1965–2014, except for the GOBAI dataset, which spans 2004–2022. Abbreviations: CMIP6, Coupled Model Intercomparison Project phase 6; GOBAI, Gridded Ocean Biogeochemistry from Artificial Intelligence; Ito22, Ito 2022; Ito24, Ito et al. 2024a; RB23, Roach & Bindoff 2023.

across the 14 models. Comparing the observed and modeled O_2 trends for both depth ranges, the CMIP6 models likely underestimated the multi-decadal O_2 trends, as the multi-model mean trends are separated from the two observational datasets (RB23 and Ito24) by more than two times the multi-model standard deviations. The CMIP6 ensemble mean overlaps with the weakest one of the three observational datasets (Ito22). However, the Ito22 dataset likely underestimated the O_2 inventory anomalies in recent years due to the very sparse primary data sources used.

Additional insights can be gained by inspecting the spatial patterns of the O_2 change (**Figure 5**). First, O_2 concentration is vertically integrated over the 0–1,000 m and 100–600 m layers separately, and then its linear trend is calculated for the period from 1965 to 2014, yielding two-dimensional (longitude-latitude) maps of multi-decadal O_2 trends. This is performed for the

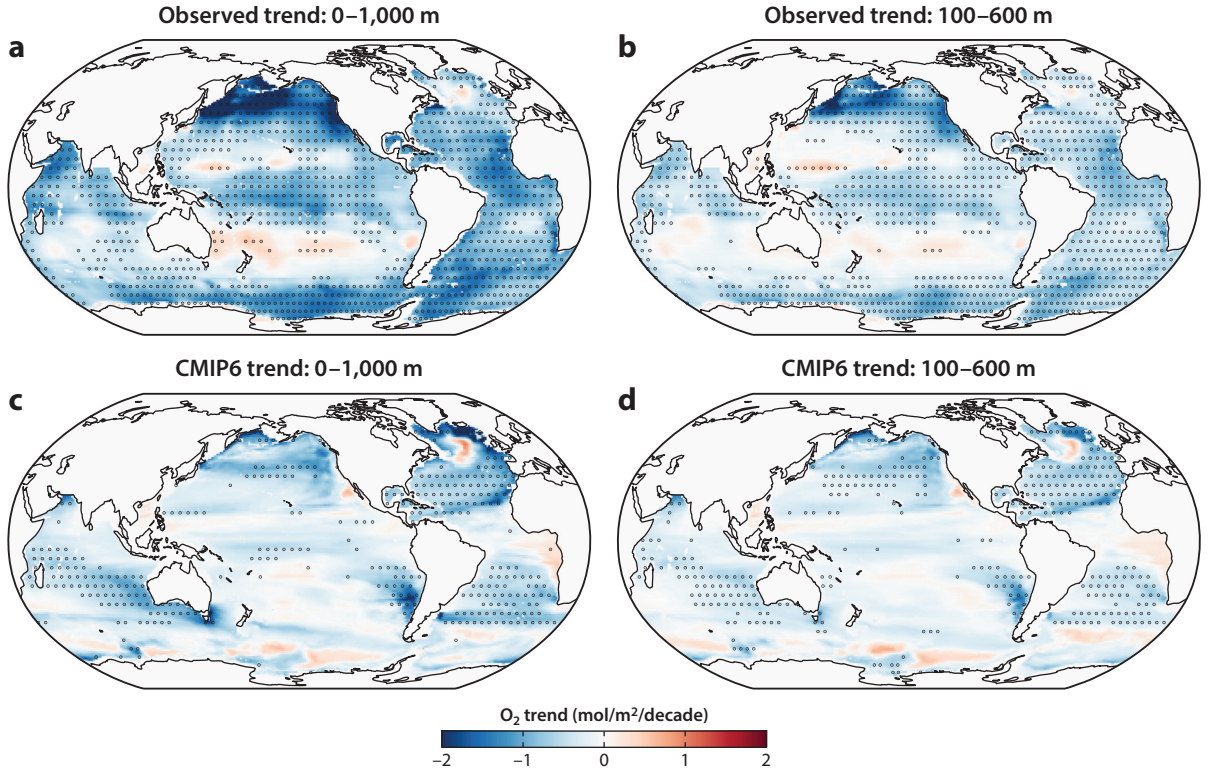

Figure 5

Observed and modeled patterns of multi-decadal O_2 trends from 1965 to 2014. (*a*,*b*) The observational trends are the average of the three datasets (Ito22, RB23, Ito24). The dots are placed where all three observational products agree on the sign of the trend. (*c*,*d*) The CMIP6 multi-model mean trends across the 14 models. The dots are placed where more than 80% of the models agree on the sign of the trend. Abbreviations: CMIP6, Coupled Model Intercomparison Project phase 6; Ito22, Ito 2022; Ito24, Ito et al. 2024a; RB23, Roach & Bindoff 2023.

Ito22, RB23, and Ito24 datasets, and their average values are plotted for the observation-based dataset (**Figure 5*a*,*b***). For the model-based trend patterns, the multi-model mean is plotted for the CMIP6 trend patterns (**Figure 5*c*,*d***). To indicate the robustness of the observed trend in the maps, a marker is placed on the grid cells where all datasets agree on the sign of the trend. For the models, a marker is placed on the grid cells where at least 80% of the models agree on the sign of the trend. Observation-based O_2 trend patterns are generally similar between the three datasets, indicating a robust spatial structure. Spatial patterns of the observation-based O_2 trends reveal a substantial decrease in O_2 in the subpolar North Pacific, tropics, and Southern Ocean. There are also regions of weak O_2 increase in the subtropical North Pacific, subtropical South Pacific, and subtropical South Indian Ocean. The Atlantic Ocean, overall, shows a clear moderate decrease in O_2, except for the subpolar warming hole in the north, which has been observed to exhibit multi-decadal variability (Rhein et al. 2017).

CMIP6 multi-model mean O_2 trends are weaker than the average of three observation-based datasets, as was the case for the inventory trends, but there are some regional differences. In the subtropical and subpolar North Atlantic, the CMIP6 O_2 trends are similar to the observations, and many models agree on the sign of the O_2 trend in this region, indicating robustness. In other regions, there is much weaker agreement among the models on the sign of O_2 trends. Models

disagree on the tropical O_2 trends, with minimal trends in the multi-model means in the tropical latitudes. Similarly, the subpolar North Pacific is a region of relatively weak multi-model mean trends, whereas the observational dataset shows a strong deoxygenation trend. Long et al. (2016) and Abe & Minobe (2023) demonstrated the strong influence of internal variability in the North Pacific, suggesting that the superposition of internal variability and anthropogenic trends may be behind the extreme O_2 trends observed, whereas models show diverging O_2 trends in this region. The tropical Pacific Ocean is influenced by the strong interannual variability associated with the El Niño–Southern Oscillation, and the combination of internal variability and anthropogenic trends is likely important there. The extratropical Southern Hemisphere oceans are also the regions where observations show a strong decline in oxygen, while the multi-model mean is close to neutral, with weak robustness across the models.

Both observational and model uncertainties must be considered when comparing O_2 inventories from different observations and CMIP6 models. The ranges of O_2 inventory trends can vary depending on the ocean basins. The distribution of the model-based O_2 inventory trends reveals a significant spread in the Pacific basin and the Southern Ocean (**Figure 6*a***). These regions also exhibit large observational uncertainty. The North Atlantic, South Pacific, and South Indian Oceans exhibit good agreement between observations and models. In these regions, observational and model uncertainties are of intermediate magnitude, and the observational estimates are within the range of the CMIP6 ensembles. The CMIP6 models agree on the O_2 trends in the Equatorial Atlantic and South Atlantic Oceans. However, there are significant differences between the model and observation-based O_2 trends in these regions. We note that the tropical Atlantic is a region with complex interannual variations (Koseki et al. 2023) with different physical dynamics affecting decadal oxygen variations at different depth levels (Hahn et al. 2017).

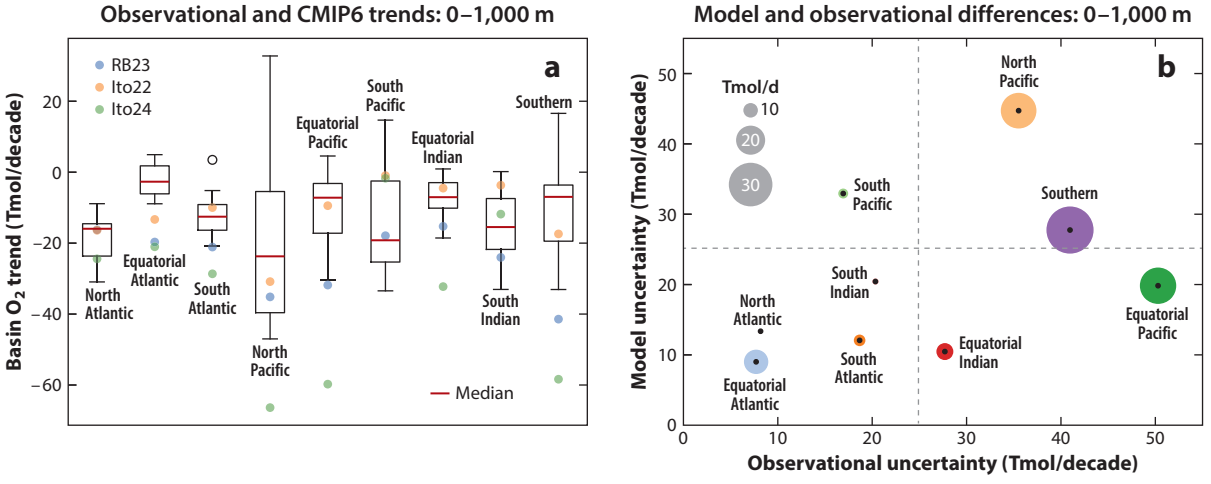

Figure 6

Regional O_2 inventory trends with observational and model uncertainties for the top 1,000 m. (*a*) A box-whisker diagram plotted for the nine regions based on the distribution of the CMIP6 multi-model statistics. Colored dots indicate the estimates from the three observation-based datasets. (*b*) Scatter plot of the nine ocean basins with observational uncertainty on the *x* axis and model uncertainty on the *y* axis. Observational uncertainties are set to the (maximum–minimum) range among the three datasets. Model uncertainties are set to two times the multi-model standard deviation. The size of the circle indicates the magnitude of the difference between observations and multi-model mean O_2 trends for each region. Dashed lines separate the four quadrants at the boundaries of 25 Tmol/decade. Abbreviations: CMIP6, Coupled Model Intercomparison Project phase 6; Ito22, Ito 2022; Ito24, Ito et al. 2024a; RB23, Roach & Bindoff 2023.

In general, the observational and model uncertainties appear to be linked (**Figure 6*b***), with a correlation coefficient of 0.7. To the leading order, it can be explained by the ocean volume. More voluminous basins contain more O_2 molecules, leading to greater differences in the inventory. The regions in the first quadrant of **Figure 6*b*** are large basins with significant internal variability and limited observational coverage. This domain encompasses the North Pacific and the Southern Ocean. CMIP6 multi-model means appear to underestimate the deoxygenation trends in these basins, but both observational and model uncertainties are high.

CMIP6 models do not agree on the sign of the trend in most of the North Pacific basin and the Southern Ocean (**Figure 5*c*,*d***), indicating substantial internal variability. The North Pacific O_2 trends are subject to the internal variability (Abe & Minobe 2023, Long et al. 2016, Mecking & Drushka 2024), which may be linked to the physical oceanographic and climate variability in this region (Andreev & Baturina 2006, Deutsch et al. 2005, Kwon et al. 2016, Osafune & Yasuda 2013, Stramma et al. 2020, Toyama et al. 2015).

In the second quadrant, the South Pacific Ocean shows relatively large model uncertainty, but the observational uncertainty is small. There is a relatively small difference between the multi-model mean and the observed trend there. The Southern Ocean and the South Pacific Ocean exhibit similar model uncertainties; however, there are much higher observed uncertainties in the Southern Ocean. The uncertainties in the Southern Ocean can involve multiple factors, including the inherent complexity of the physical climate system and its interactions with the ocean biogeochemistry, the important roles of parameterized small-scale physics, and the sparse observations (prior to the Argo-O_2 era). The effects of increasing heat and freshwater fluxes on ocean ventilation may be partially compensated by the increasing westerly winds (Hollitzer et al. 2024, Lovenduski & Ito 2009), where the degree of compensation could affect the magnitude of the net change. Simulated ventilation characteristics of the Southern Ocean critically depend on the vertical profile mean state, response of ocean jets, mesoscale eddies, and small-scale meanders that are parameterized in the coarse-resolution ESMs, which can lead to model structure differences (Abernathey & Ferreira 2015, Bourgeois et al. 2022, Dove et al. 2021, Kamenkovich et al. 2017, Keppler et al. 2024, Marshall & Zanna 2014, Naveira Garabato et al. 2019).

The Atlantic basin and the South Indian Ocean are in the third quadrant, where both observational and model uncertainties are relatively small. In the North Atlantic and the South Indian Ocean, the multi-model means and observations show significant overlap. In the South Indian Ocean, however, the spatial patterns appear to differ between multi-model means and observations (**Figure 5**), and there are individual models that do not align with the observations. The CMIP6 models are consistent with the observed O_2 trends in the North Atlantic Ocean, and this is the only region that exhibits consistency between observations and models. Here, the robust broad deoxygenation signals have been shown to be related to the persistent warming in both observations and models (Takano et al. 2023). The Equatorial and South Atlantic basins show relatively small uncertainties; even so, the Equatorial Atlantic still exhibits significant differences between observation and model-based O_2 inventory trends as the model-based O_2 inventory trend is outside of the observational range.

Finally, the fourth quadrant includes the Equatorial Pacific and Equatorial Indian Oceans. These regions show exceptionally large observational uncertainty. Tropical air-sea interaction can drive strong internal variability in these regions. Furthermore, these tropical basins have been historically sparsely sampled by all three measurement platforms (bottle, CTD, and Argo) (see **Figure 1**). The combination of strong internal variability and low sampling density can pose significant challenges in filling data gaps and potentially aliasing the interannual variability in these regions (Busecke et al. 2019, Czeschel et al. 2012, Deutsch et al. 2011, Duteil et al. 2018, Ito & Deutsch 2013, Poupon et al. 2023).

5. DISCUSSION AND FUTURE DIRECTIONS

Both model outputs and observations are subject to uncertainty. To determine whether the models and observations are consistent, the model-observation differences need to be compared to the uncertainties. Considering the model uncertainties, it is often assumed that averaging over wide areas (or large volumes) tends to reduce noise from internal variability. Structural uncertainties dominate at global scales and over long timescales, where the forced signal (e.g., anthropogenic warming and associated ocean stratification) is strong, while local O_2 anomalies caused by internal variability may cancel out when integrated across vast regions. Conversely, internal variability tends to play a larger role at regional scales and over shorter timescales, where internal ocean-atmosphere dynamics and mesoscale processes may locally mask or amplify the forced trends. This separation provides a useful framework for interpreting differences between modeled and observed oxygen trends at different scales.

Observational uncertainties are combinations of multiple factors, including measurement errors, sampling errors, and analysis errors. Deconvolving the influence of each factor would require the coordinated intercomparison(s) of mapped O_2 concentration by changing one factor at a time, which is beyond the scope of this review. Here, the spread across the observational products was used as a pragmatic estimate of observational uncertainty, primarily reflecting the analysis errors because most data products rely on similar primary data sources but different data QC and mapping methods. The datasets analyzed in this study are remapped to a common $1° \times 1°$ resolution; thus, the primary focus is on global and basin-scale analysis (and not intended for coastal regions). Regarding sampling uncertainties, it is possible to revisit the model-observation comparisons using the sampling patterns of the historical observations and compare only the grid cells where and when observational profiles are available (Ito et al. 2024b) using model ensembles. Although it is beyond the scope of this review, further work is warranted to thoroughly evaluate sampling uncertainties.

For the long-term mean climatology, the CMIP6 multi-model mean is in close agreement with the observational climatology (pattern correlation > 0.9), with a slight overestimation of 2.1 μmol/kg in the near-surface layer (0–100 m). The model-observation differences mainly come from the tropical oceans at this depth range, where mesoscale processes, turbulent mixing, and equatorial circulation are poorly represented in ESMs (Busecke et al. 2019; Eddebbar et al. 2021, 2024; Calil 2023). In the thermocline (100–600 m), the pattern correlation remains high (>0.9). However, the mean bias is greater (9.0 μmol/kg), and there are widespread O(10%) intermodel differences, indicative of significant model structural differences for the physical and biogeochemical processes that affect the mid-depth O_2 representation (Cabré et al. 2015).

There are outstanding questions about whether the current ESMs can reproduce the long-term changes in global and regional deoxygenation rates within the bounds of model and observational uncertainties. There are two main conclusions. First, the CMIP6 models still exhibit significant mean state biases, especially in the tropics and in the thermocline. The spread in the volume-integrated O_2 inventories among the CMIP6 models was O(10%) and is significantly larger than the magnitude of the multi-decadal O_2 changes. Second, the CMIP6 models likely underestimated the ocean deoxygenation rates for the last several decades. Focusing on the linear trend of global O_2 inventories from 1965 to 2014, the CMIP6 models underestimated the observed quasi-global O_2 trend in the 0–1,000-m depth range (excluding the Arctic Ocean and marginal seas) by more than two times the multi-model standard deviation relative to the two recent datasets with better data coverage.

The caveat is that our conclusion is based on only three time-varying observational O_2 datasets, which are insufficient to robustly evaluate the uncertainty. Moreover, the observation

itself suffers from limited data points, which results in larger uncertainty. Comparison among the three observational datasets revealed that one of the datasets (Ito22) significantly underestimated the magnitude of the multi-decadal O_2 trends (in comparison to the other two datasets), likely due to the extremely sparse bottle data in recent decades. Ito22 used the bottle data only, and the lack of CTD-O_2 and Argo-O_2 data caused severe undersampling after the 1990s, leading to much weaker multi-decadal O_2 inventory trends compared to other data products that included CTD-O_2 and/or Argo-O_2 data. Additional observation-based data products are much needed to better characterize this uncertainty. While our assessment of the observational uncertainty is by no means complete, the regional-scale comparison between models and observations has provided some insights into the potential issues in detecting global and regional deoxygenation, as well as identifying critical areas for further investigation.

It is important to note that while Argo observations have significantly improved the spatial coverage of oceanic oxygen measurements, they cannot fully replace higher-quality ship-based observations. In particular, Argo data may introduce new biases (Bushinsky et al. 2025), which could pose challenges for accurately assessing long-term trends. In contrast, long-term repeat observation sites such as Hawaii Ocean Time-series (HOT), Ocean Station P (OSP), European Station for Time Series in the Ocean at the Canary Islands (ESTOC), and Bermuda Atlantic Time-series Study (BATS), as well as repeat sections like the Japanese 137°E line, Line P, and California Cooperative Oceanic Fisheries Investigations (CALCOFI), provide high-quality and reliable long-term datasets, despite their limited spatial coverage. Therefore, to ensure both adequate spatial coverage and the long-term stability of data quality, it is essential to integrate Argo and ship-based observations as complementary components of the global ocean observing system. Along with the increasing data stream from the autonomous floats, development of optimally extended observational data products will be crucial for better characterizing the uncertainties. We anticipate more observation-based O_2 datasets to emerge soon, utilizing machine learning approaches (Lu et al. 2024, Navarra et al. 2025, Ouala et al. 2026). Including those new datasets will be necessary for future intercomparison projects.

To further advance our understanding of ocean deoxygenation in ESMs, it is essential to analyze multiple ensemble members for each model, thereby separating model structural differences and internal variability. In this review, we did not separate these two sources of uncertainty, as we analyzed only a single ensemble member per model, as explained in Section 2.2. In particular, several ESMs in CMIP6 have large ensembles (e.g., Abe & Minobe 2023, Maher et al. 2021), which are useful for understanding the nature of internal variability in greater detail. This knowledge can provide valuable feedback in analyses of moderate ensemble sizes.

We also note that ESMs such as those in CMIP6 simulate model-dependent internal variability, which likely differs from the observations. As stated above, this internal variability may impose bias in the regional long-term trend in oxygen; hence, models may capture the observed long-term trend for the wrong reasons. Future studies should consider a better mechanistic understanding of the role of regional modes of climate variability in subsurface oxygen changes. Model studies on other biogeochemical cycles (such as carbon) have shown that incorrect interactions with physical processes from seasonal to interannual timescales lead to biases in regional long-term changes in the North Atlantic, North Pacific, Equatorial Pacific, tropical Atlantic, and Southern Ocean (Kessler & Tjiputra 2016, Rodgers et al. 2023, Vaittinada Ayar et al. 2022), all of which have been shown to experience significant deoxygenation in the recent decades. Alternatively, application of a large ensemble simulation could be adopted (Long et al. 2016). Concurrently, improving temporal coverage of observations in the above regions should be considered a priority in the context of establishing early warning indicators and detection of emerging interior oxygen changes in the coming decades (Tjiputra et al. 2023).

This review highlights the need for structural improvements in the physical and biochemical processes in the ESMs to better represent the mean states of O_2. The factors that influence ocean oxygen inventories are broadly categorized into four groups: (*a*) biases in model forcings and feedbacks, (*b*) biases in physical processes, (*c*) biases in biogeochemical processes, and (*d*) the role of internal climate variability. Combinations of all these factors can cause the differences between observation and models. First, carbon dioxide and other greenhouse gas emissions, as well as natural and anthropogenic aerosols, are the forcings of the climate system changes, together with effects of various feedbacks in the climate system. The Earth is currently out of the radiative balance—Earth energy imbalance (EEI) at the top of the atmosphere (Boland et al. 2023, Cheng et al. 2022). Most of the excess heat (positive EEI) goes into the ocean and increases OHC, which in turn causes changes in oxygen solubility, stratification, and circulation changes (Cheng et al. 2022, 2025). Biases in forcings and feedbacks may induce biases in the model simulations of climate change (Myhre et al. 2025). Another potential external forcing is the biogeochemical forcing of natural and anthropogenic aerosols. Aerosol deposition can be a significant source of macro- and micronutrients at the air-sea interface, which can substantially modulate surface ocean productivity and subsurface oxygen consumption on all timescales (Hamilton et al. 2020, 2022; Ito 2015). Second, physical oceanographic processes can influence the representation of oxygen in models. Such physical processes may include subduction and ventilation of the thermocline waters, which can depend on the model resolution and parameterization. Mesoscale ocean eddies are generally not resolved in coarse-resolution ocean models and are parameterized, which can have profound influences on the ventilation of oxygen-rich surface waters into the interior ocean along with other poorly resolved small-scale circulation features (e.g., equatorial zonal jets, turbulence, etc.) (Brandt et al. 2015, Calil 2023, Couespel et al. 2024, Eddebbar et al. 2024, Frenger et al. 2018, Levy et al. 2022, Margolskee et al. 2019). Third, biological oxygen consumption in the water column depends on the production of sinking organic matter, its sinking speed, rates of decomposition, and residence time of interior water masses, all of which depend on changing water temperatures (e.g., Couespel et al. 2025, Laufkötter et al. 2017, Maerz et al. 2020, Segschneider & Bendtsen 2013). Finally, simulations of the atmospheric and sea ice states in coupled ESMs and their coupling with the ocean (e.g., via air-sea interactions, cloud-aerosol interactions, sea and land ice melt, etc.) differ significantly between models and may influence the structure and variability of oxygen, irrespective of ocean models' physical or biogeochemical fidelity. Models can generate their own trajectories of low-frequency climate variability (e.g., Pacific decadal variability, Atlantic multidecadal oscillation) that can influence the multi-decadal O_2 trends. Because oxygen integrates many physical and biogeochemical factors, both external and internal to the oceans, there is an opportunity to revisit assumptions and fundamental knowledge of linkages among atmospheric, oceanic, and biogeochemical processes. A list of potential future community projects is provided here to make significant progress in terms of both understanding uncertainties and improving model representation:

- Develop a community-standard protocol and suite of key features (metrics) for observed ocean oxygen content that explicitly account for sampling gaps, mapping uncertainty, and natural variability.
- Evaluate the role of model spatial resolution and physical parameterization on oxygen biases and variability through an intercomparison of a hierarchy of models.
- Develop a comprehensive observation-observation, model-model, and observation-model comparison of gridded oxygen datasets, including ensembles of model simulations in the multi-model context.

- Design model experimental protocols for hindcast simulations that include a consistent spin-up strategy and model initialization to better constrain oxygen mean-state biases and long-term trends.
- Develop a ventilation-focused model intercomparison with standardized diagnostics (water age, transient tracers, transit time distribution) in parallel with the above hindcast simulations.
- Develop a biogeochemically focused model intercomparison with standardized physical forcings and diagnostics (e.g., oxygen production and remineralization rates, detailed oxygen budget terms for process attributions).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review. The views, opinions, and practices used to produce this dataset/software are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

ACKNOWLEDGMENTS

We are thankful for the community of oceanographers who collected the O_2 samples and contributed to the global dataset. This review would not be possible without their dedication and hard work. We acknowledge the World Climate Research Program, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We are thankful to Drs. Hernan E. Garcia, Christopher Roach, Jonathan D. Sharp, and an anonymous reviewer for providing helpful comments. T.I. is supported by the US National Science Foundation (OCE-2123546, 2446011). Y.T. is supported by a joint UK Natural Environment Research Council and US National Science Foundation Large grant (NE/W009501/1°C-Streams). Y.A.E. is supported by the National Oceanic and Atmospheric Administration Climate Program Office's Climate Variability & Predictability grant NA24OARX431C0018. J.F.T. acknowledges funding from the Research Council of Norway project Navigate (352142) and EU project OceanICU (101083922).

LITERATURE CITED

Abe Y, Minobe S. 2023. Comparison of ocean deoxygenation between CMIP models and an observational dataset in the North Pacific from 1958 to 2005. *Front. Mar. Sci.* 10:2023

Abernathey R, Ferreira D. 2015. Southern Ocean isopycnal mixing and ventilation changes driven by winds. *Geophys. Res. Lett.* 42:10357–65

Andreev AG, Baturina VI. 2006. Impacts of tides and atmospheric forcing variability on dissolved oxygen in the subarctic North Pacific. *J. Geophys. Res.* 111(C7):C07S10

Barth A, Beckers JM, Troupin C, Alvera-Azcárate A, Vandenbulcke L. 2014. divand-1.0: *n*-dimensional variational data analysis for ocean observations. *Geosci. Model Dev.* 7:225–41

Bertini L, Tjiputra J. 2022. Biogeochemical timescales of climate change onset and recovery in the North Atlantic interior under rapid atmospheric CO_2 forcing. *J. Geophys. Res. Oceans* 127:e2021JC017929

Bindoff NL, Cheung WWL, Kairo JG, Arístegui J, Guinder VA, et al. 2019. Changing ocean, marine ecosystems, and dependent communities. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, ed. H-O Pörtner, DC Roberts, V Masson-Delmotte, P Zhai, M Tignor, et al. Cambridge University Press

Bittig HC, Fiedler B, Scholz R, Krahmann G, Körtzinger A. 2014. Time response of oxygen optodes on profiling platforms and its dependence on flow speed and temperature. *Limnol. Oceanogr.* 12:617–36

Bittig HC, Körtzinger A. 2017. Technical note: update on response times, in-air measurements, and in situ drift for oxygen optodes on profiling platforms. *Ocean Sci.* 13:1–11

Boland EJD, Dittus AJ, Jones DC, Josey SA, Sinha B. 2023. Ocean heat content responses to changing anthropogenic aerosol forcing strength: regional and multi-decadal variability. *J. Geophys. Res. Oceans* 128:e2022JC018725

Boucher O, Servonnat J, Albright AL, Aumont O, Balkanski Y, et al. 2020. Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst.* 12:e2019MS002010

Bourgeois T, Goris N, Schwinger J, Tjiputra JF. 2022. Stratification constrains future heat and carbon uptake in the Southern Ocean between 30°S and 55°S. *Nat. Commun.* 13:340

Brandt P, Bange HW, Banyte D, Dengler M, Didwischus S-H, et al. 2015. On the role of circulation and mixing in the ventilation of oxygen minimum zones with a focus on the eastern tropical North Atlantic. *Biogeosciences* 12:489–512

Breitburg D, Levin LA, Oschlies A, Grégoire M, Chavez FP, et al. 2018. Declining oxygen in the global ocean and coastal waters. *Science* 359:eaam7240

Brewer PG, Peltzer ET. 2017. Depth perception: the need to report ocean biogeochemical rates as functions of temperature, not depth. *Philos. Trans. R. Soc. A* 375:20160319

Buchanan PJ, Tagliabue A. 2021. The regional importance of oxygen demand and supply for historical ocean oxygen trends. *Geophys. Res. Lett.* 48:e2021GL094797

Burrows SM, Maltrud M, Yang X, Zhu Q, Jeffery N, et al. 2020. The DOE E3SM v1.1 biogeochemistry configuration: description and simulated ecosystem-climate responses to historical changes in forcing. *J. Adv. Model. Earth Syst.* 12:e2019MS001766

Busecke JJM, Resplandy L, Dunne JP. 2019. The equatorial undercurrent and the oxygen minimum zone in the Pacific. *Geophys. Res. Lett.* 46:6716–25

Bushinsky SM, Nachod Z, Fassbender AJ, Tamsitt VM, Takeshita Y, William N. 2025. Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO_2. *Glob. Biogeochem. Cycles* 39(5):e2024GB008185

Cabré A, Marinov I, Bernardello R, Bianchi D. 2015. Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends. *Biogeosciences* 12:5429–54

Calil PHR. 2023. High-resolution, basin-scale simulations reveal the impact of intermediate zonal jets on the Atlantic oxygen minimum zones. *J. Adv. Model. Earth Syst.* 15:e2022MS003158

Carpenter JH. 1965. The accuracy of the Winkler method for dissolved oxygen analysis. *Limnol. Oceanogr.* 10:135–40

Cheng L, Li G, Long SM, Li Y, von Schuckmann K, et al. 2025. Ocean stratification in a warming climate. *Nat. Rev. Earth Environ.* 6:637–55

Cheng L, Pan Y, Tan Z, Zheng H, Zhu Y, et al. 2024. IAPv4 ocean temperature and ocean heat content gridded dataset. *Earth Syst. Sci. Data* 16:3517–46

Cheng L, von Schuckmann K, Abraham JP, Trenberth KE, Mann ME, et al. 2022. Past and future ocean warming. *Nat. Rev. Earth Environ.* 3:776–94

Christian JR, Denman KL, Hayashida H, Holdsworth AM, Lee WG, et al. 2022. Ocean biogeochemistry in the Canadian Earth System Model version 5.0.3: CanESM5 and CanESM5-CanOE. *Geosci. Model. Dev.* 15:4393–424

Couespel D, Davila X, Goris N, Jeansson E, Lauvset SK, Tjiputra J. 2025. Earth system models overestimate the sensitivity of apparent oxygen utilisation to age change in the deep ocean. Preprint, EGUsphere. **https://doi.org/10.5194/egusphere-2025-2566**

Couespel D, Lévy M, Bopp L. 2024. Stronger oceanic CO_2 sink in eddy-resolving simulations of global warming. *Geophys. Res. Lett.* 51:e2023GL106172

Czeschel R, Stramma L, Johnson GC. 2012. Oxygen decreases and variability in the eastern equatorial Pacific. *J. Geophys. Res.* 117(C11):C11019

Davila X, Olsen A, Lauvset SK, McDonagh EL, Brakstad A, Gebbie G. 2023. On the origins of open ocean oxygen minimum zones. *J. Geophys. Res. Oceans* 128:e2023JC019677

Deser C, Lehner F, Rodgers KB, Ault T, Delworth TL, et al. 2020. Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* 10:277–86

Deutsch C, Brix H, Ito T, Frenzel H, Thompson L. 2011. Climate-forced variability of ocean hypoxia. *Science* 333:336–39

Deutsch C, Emerson S, Thompson L. 2005. Fingerprints of climate change in North Pacific oxygen. *Geophys. Res. Lett.* 32:L16604

Deutsch C, Ferrel A, Seibel B, Pörtner H-O, Huey RB. 2015. Climate change tightens a metabolic constraint on marine habitats. *Science* 348:1132–35

Deutsch C, Penn JL, Seibel B. 2020. Metabolic trait diversity shapes marine biogeography. *Nature* 585:557–62

DiMarco SF, Wang Z, Chapman P, al-Kharusi L, Belabbassi L, et al. 2023. Monsoon-driven seasonal hypoxia along the northern coast of Oman. *Front. Mar. Sci.* 10:2023

Dove LA, Thompson AF, Balwada D, Gray AR. 2021. Observational evidence of ventilation hotspots in the Southern Ocean. *J. Geophys. Res. Oceans* 126:e2021JC017178

Dunne JP, Bociu I, Bronselaer B, Guo H, John JG, et al. 2020a. Simple global ocean Biogeochemistry with Light, Iron, Nutrients and Gas Version 2 (BLINGv2): model description and simulation characteristics in GFDL's CM4.0. *J. Adv. Model. Earth Syst.* 12:e2019MS002008

Dunne JP, Horowitz LW, Adcroft AJ, Ginoux P, Held IM, et al. 2020b. The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): overall coupled model description and simulation characteristics. *J. Adv. Model. Earth Syst.* 12:e2019MS002015

Duteil O, Oschlies A, Böning CW. 2018. Pacific decadal oscillation and recent oxygen decline in the eastern tropical Pacific Ocean. *Biogeosciences* 15:7111–26

Eddebbar YA, Long MC, Resplandy L, Rödenbeck C, Rodgers KB, et al. 2017. Impacts of ENSO on air-sea oxygen exchange: observations and mechanisms. *Glob. Biogeochem. Cycles* 31:901–21

Eddebbar YA, Subramanian AC, Whitt DB, Long MC, Verdy A, et al. 2021. Seasonal modulation of dissolved oxygen in the equatorial Pacific by tropical instability vortices. *J. Geophys. Res. Oceans* 126:e2021JC017567

Eddebbar YA, Whitt DB, Verdy A, Mazloff MR, Subramanian AC, et al. 2024. Eddy-mediated turbulent mixing of oxygen in the equatorial Pacific. *J. Geophys. Res. Oceans* 129:e2023JC020588

Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, et al. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9:1937–58

Frenger I, Bianchi D, Sührenberg C, Oschlies A, Dunne J, et al. 2018. Biogeochemical role of subsurface coherent eddies in the ocean: tracer cannonballs, hypoxic storms, and microbial stewpots? *Glob. Biogeochem. Cycles* 32:226–49

Garcia HE, Wang Z, Bouchard C, Cross SL, Paver CR, et al. 2024. *World Ocean Atlas 2023*, Vol. 3: *Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*. NESDIS

Garcia HE, Weathers K, Paver CR, Smolyar I, Boyer TP, et al. 2018. *World Ocean Atlas 2018*, Vol. 3: *Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation*. NESDIS

Garçon V, Karstensen J, Palacz A, Telszewski M, Aparco Lara T, et al. 2019. Multidisciplinary observing in the world ocean's oxygen minimum zone regions: from climate to fish—the VOICE initiative. *Front. Mar. Sci.* 6:722

Gouretski V, Cheng L, Du J, Xing X, Chai F, Tan Z. 2024. A consistent ocean oxygen profile dataset with new quality control and bias assessment. *Earth Syst. Sci. Data* 16:5503–30

Gruber N, Boyd PW, Frölicher TL, Vogt M. 2021. Biogeochemical extremes and compound events in the ocean. *Nature* 600:395–407

Hahn J, Brandt P, Schmidtko S, Krahmann G. 2017. Decadal oxygen change in the eastern tropical North Atlantic. *Ocean Sci.* 13:551–76

Hajima T, Watanabe M, Yamamoto A, Tatebe H, Noguchi MA, et al. 2020. Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geosci. Model Dev.* 13:2197–244

Halpern BS, Frazier M, O'Hara CC, Vargas-Fonseca OA, Lombard AT. 2025. Cumulative impacts to global marine ecosystems projected to more than double by mid-century. *Science* 389:1216–19

Hamilton DS, Moore JK, Arneth A, Bond TC, Carslaw KS, et al. 2020. Impact of changes to the atmospheric soluble iron deposition flux on ocean biogeochemical cycles in the Anthropocene. *Glob. Biogeochem. Cycles* 34:e2019GB006448

Hamilton DS, Perron MM, Bond TC, Bowie AR, Buchholz RR, et al. 2022. Earth, wind, fire, and pollution: aerosol nutrient sources and impacts on ocean biogeochemistry. *Annu. Rev. Mar. Sci.* 14:303–30

Held IM, Guo H, Adcroft A, Dunne JP, Horowitz LW, et al. 2019. Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.* 11:3691–727

Helm KP, Bindoff NL, Church JA. 2011. Observed decreases in oxygen content of the global ocean. *Geophys. Res. Lett.* 38:L23602

Hollitzer HAL, Patara L, Terhaar J, Oschlies A. 2024. Competing effects of wind and buoyancy forcing on ocean oxygen trends in recent decades. *Nat. Commun.* 15:9264

Ilyina T, Six KD, Segschneider J, Maier-Reimer E, Li H, Núñez-Riboni I. 2013. Global ocean biogeochemistry model HAMOCC: model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *J. Adv. Model. Earth Syst.* 5:287–315

Ito A. 2015. Atmospheric processing of combustion aerosols as a source of bioavailable iron. *Environ. Sci. Technol. Lett.* 2:70–75

Ito T. 2022. Optimal interpolation of global dissolved oxygen: 1965–2015. *Geosci. Data J.* 9:167–76

Ito T, Cervania A, Cross K, Ainchwar S, Delawalla S. 2024a. Mapping dissolved oxygen concentrations by combining shipboard and Argo observations using machine learning algorithms. *J. Geophys. Res. Mach. Learn. Comput.* 1:e2024JH000272

Ito T, Deutsch C. 2013. Variability of the oxygen minimum zone in the tropical North Pacific during the late twentieth century. *Glob. Biogeochem. Cycles* 27:1119–28

Ito T, Garcia HE, Wang Z, Minobe S, Long MC, et al. 2024b. Underestimation of multi-decadal global O_2 loss due to an optimal interpolation method. *Biogeosciences* 21:747–59

Ito T, Minobe S, Long MC, Deutsch C. 2017. Upper ocean O_2 trends: 1958–2015. *Geophys. Res. Lett.* 44:4214–23

Kamenkovich I, Garraffo Z, Pennel R, Fine RA. 2017. Importance of mesoscale eddies and mean circulation in ventilation of the Southern Ocean. *J. Geophys. Res. Oceans* 122:2724–41

Katavouta A, Williams RG, Goodwin P. 2019. The effect of ocean ventilation on the transient climate response to emissions. *J. Clim.* 32:5085–105

Keeling RF, Garcia HE. 2002. The change in oceanic O_2 inventory associated with recent global warming. *PNAS* 99:7848–53

Keppler L, Eddebbar YA, Gille ST, Guisewhite N, Mazloff MR, et al. 2024. Effects of mesoscale eddies on Southern Ocean biogeochemistry. *AGU Adv.* 5:e2024AV001355

Kessler A, Tjiputra J. 2016. The Southern Ocean as a constraint to reduce uncertainty in future ocean carbon sinks. *Earth Syst. Dyn.* 7:295–312

Koseki S, Tjiputra J, Fransner F, Crespo LR, Keenlyside NS. 2023. Disentangling the impact of Atlantic Niño on sea-air CO_2 flux. *Nat. Commun.* 14:3649

Kwiatkowski L, Torres O, Bopp L, Aumont O, Chamberlain M, et al. 2020. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences* 17:3439–70

Kwon EY, Deutsch C, Xie S-P, Schmidtko S, Cho Y-K. 2016. The North Pacific oxygen uptake rates over the past half century. *J. Clim.* 29:61–76

Lachkar Z, Lévy M, Smith S. 2018. Intensification and deepening of the Arabian Sea oxygen minimum zone in response to increase in Indian monsoon wind intensity. *Biogeosciences* 15:159–86

Laufkötter C, John JG, Stock CA, Dunne JP. 2017. Temperature and oxygen dependence of the remineralization of organic matter. *Glob. Biogeochem. Cycles* 31:1038–50

Lévy M, Resplandy L, Palter JB, Couespel D, Lachkar Z. 2022. The crucial contribution of mixing to present and future ocean oxygen distribution. In *Ocean Mixing: Drivers, Mechanisms and Impacts*, ed. M Meredith, A Naveira Garabato. Elsevier

Liddicoat SK, Wiltshire AJ, Jones CD, Arora VK, Brovkin V, et al. 2021. Compatible fossil fuel CO_2 emissions in the CMIP6 Earth system models' historical and shared socioeconomic pathway experiments of the twenty-first century. *J. Clim.* 34:2853–75

Long MC, Deutsch C, Ito T. 2016. Finding forced trends in oceanic oxygen. *Glob. Biogeochem. Cycles* 30:381–97
Lovato T, Peano D, Butenschön M, Materia S, Iovino D, et al. 2022. CMIP6 simulations with the CMCC Earth System Model (CMCC-ESM2). *J. Adv. Model. Earth Syst.* 14:e2021MS002814
Lovenduski N, Ito T. 2009. The future evolution of the Southern Ocean CO_2 sink. *J. Mar. Res.* 67:597–617
Lu B, Zhao Z, Han L, Gan X, Zhou Y, et al. 2024. OxyGenerator: reconstructing global ocean deoxygenation over a century with deep learning. *Proceed. Mach. Learn. Res.* 235:33219–42
Maerz J, Six KD, Stemmler I, Ahmerkamp S, Ilyina T. 2020. Microstructure and composition of marine aggregates as co-determinants for vertical particulate organic carbon transfer in the global ocean. *Biogeosciences* 17:1765–803
Maher N, Milinski S, Ludwig R. 2021. Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth Syst. Dyn.* 12:401–18
Margolskee A, Frenzel H, Emerson S, Deutsch C. 2019. Ventilation pathways for the North Pacific oxygen deficient zone. *Glob. Biogeochem. Cycles* 33:875–90
Marotzke J. 2019. Quantifying the irreducible uncertainty in near-term climate projections. *WIREs Clim. Change* 10:e563
Marshall DP, Zanna L. 2014. A conceptual model of ocean heat uptake under climate change. *J. Clim.* 27:8444–65
Maurer TL, Plant JN, Johnson KS. 2021. Delayed-mode quality control of oxygen, nitrate, and pH data on SOCCOM biogeochemical profiling floats. *Front. Mar. Sci.* 8:683207
Mauritsen T, Bader J, Becker T, Behrens J, Bittner M, et al. 2019. Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *J. Adv. Model. Earth Syst.* 11:998–1038
Mecking S, Drushka K. 2024. Linking northeastern North Pacific oxygen changes to upstream surface outcrop variations. *Biogeosciences* 21:1117–33
Mishonov AV, Boyer TP, Baranova OK, Bouchard CN, Cross SL, et al. 2024. *World Ocean Database 2023*. NOAA National Environmental Satellite, Data, and Information Service
Moore JK, Doney SC, Lindsay K. 2004. Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Glob. Biogeochem. Cycles* 18:GB4028
Myhre G, Hodnebrog Ø, Loeb N, Forster PM. 2025. Observed trend in Earth energy imbalance may provide a constraint for low climate sensitivity models. *Science* 388:1210–13
Navarra G, Deutsch CA, Merchant C, Clare MCA, Sonnewald M. 2025. A Bayesian neural network approach to study dissolved oxygen in Southern Ocean water masses. Preprint, ESS Open Arch. **https://doi.org/10.22541/essoar.173655319.94315054/v2**
Naveira Garabato AC, Dotto TS, Hooley J, Bacon S, Tsamados M, et al. 2019. Phased response of the subpolar Southern Ocean to changes in circumpolar winds. *Geophys. Res. Lett.* 46:6024–33
Novi L, Bracco A, Ito T, Takano Y. 2024. Evolution of oxygen and stratification and their relationship in the North Pacific Ocean in CMIP6 Earth system models. *Biogeosciences* 21:3985–4005
Osafune S, Yasuda I. 2013. Remote impacts of the 18.6 year period modulation of localized tidal mixing in the North Pacific. *J. Geophys. Res. Oceans* 118:3128–37
Oschlies A, Brandt P, Stramma L, Schmidtko S. 2018. Drivers and mechanisms of ocean deoxygenation. *Nat. Geosci.* 11:467–73
Ouala S, Hidaoui O, Lachkar Z. 2026. A novel global gridded ocean oxygen product derived from a neural network emulator and in-situ observations. *Earth Syst. Data Sci.* 18:287–308
Paulsen H, Ilyina T, Six KD, Stemmler I. 2017. Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *J. Adv. Model. Earth Syst.* 9:438–64
Pershing AJ, Alexander MA, Hernandez CM, Kerr LA, Le Bris A, et al. 2015. Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science* 350:809–12
Poupon MA, Resplandy L, Lévy M, Bopp L. 2023. Pacific decadal oscillation influences tropical oxygen minimum zone extent and obscures anthropogenic changes. *Geophys. Res. Lett.* 50:e2022GL102123
Rabalais NN, Turner RE. 2019. Gulf of Mexico hypoxia: past, present, and future. *Limnol. Oceanogr. Bull.* 28:117–24
Reagan JR, Boyer TP, García HE, Locarnini RA, Baranova OK, et al. 2024. *World Ocean Atlas 2023*. NOAA National Centers for Environmental Information

Rhein M, Steinfeldt R, Kieke D, Stendardo I, Yashayaev I. 2017. Ventilation variability of Labrador Sea Water and its impact on oxygen and anthropogenic carbon: a review. *Philos. Trans. R. Soc. A* 375:20160321
Roach CJ, Bindoff NL. 2023. Developing a new oxygen atlas of the world's oceans using data interpolating variational analysis. *J. Atmos. Ocean. Technol.* 40:1475–91
Rodgers KB, Schwinger J, Fassbender AJ, Landschützer P, Yamaguchi R, et al. 2023. Seasonal variability of the surface ocean carbon cycle: a synthesis. *Glob. Biogeochem. Cycles* 37:e2023GB007798
Santana-Falcón Y, Yamamoto A, Lenton A, Jones CD, Burger FA, et al. 2023. Irreversible loss in marine ecosystem habitability after a temperature overshoot. *Commun. Earth Environ.* 4:343
Santora JA, Mantua NJ, Schroeder ID, Field JC, Hazen EL, et al. 2020. Habitat compression and ecosystem shifts as potential links between marine heatwave and record whale entanglements. *Nat. Commun.* 11:12
Schmidtko S, Stramma L, Visbeck M. 2017. Decline in global oceanic oxygen content during the past five decades. *Nature* 542:335–39
Schoepf V, Stat M, Falter JL, McCulloch MT. 2015. Limits to the thermal tolerance of corals adapted to a highly fluctuating, naturally extreme temperature environment. *Sci. Rep.* 5:17639
Séférian R, Berthet S, Yool A, Palmiéri J, Bopp L, et al. 2020. Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. *Curr. Clim. Change Rep.* 6:95–119
Séférian R, Gehlen M, Bopp L, Resplandy L, Orr JC, et al. 2016. Inconsistent strategies to spin up models in CMIP5: implications for ocean biogeochemical model performance assessment. *Geosci. Model Dev.* 9:1827–51
Séférian R, Nabat P, Michou M, Saint-Martin D, Voldoire A, et al. 2019. Evaluation of CNRM Earth system model, CNRM-ESM2-1: role of Earth system processes in present-day and future climate. *J. Adv. Model. Earth Syst.* 11:4182–227
Segschneider J, Bendtsen J. 2013. Temperature-dependent remineralization in a warming ocean increases surface pCO_2 through changes in marine ecosystem composition. *Glob. Biogeochem. Cycles* 27:1214–25
Seidov D, Mishonov A, Reagan J, Baranova O, Cross S, Parsons R. 2018. Regional climatology of the Northwest Atlantic Ocean: high-resolution mapping of ocean structure and change. *Bull. Am. Meteorol. Soc.* 99:2129–38
Sellar AA, Jones CG, Mulcahy JP, Tang Y, Yool A, et al. 2019. UKESM1: description and evaluation of the U.K. Earth System Model. *J. Adv. Model. Earth Syst.* 11:4513–58
Seltzer AM, Nicholson DP, Smethie WM, Tyne RL, Le Roy E, et al. 2023. Dissolved gases in the deep North Atlantic track ocean ventilation processes. *PNAS* 120:e2217946120
Sharp JD, Fassbender AJ, Carter BR, Johnson GC, Schultz C, Dunne JP. 2023. GOBAI-O_2: temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly 2 decades. *Earth Syst. Sci. Data* 15:4481–518
Stock CA, Dunne JP, Fan S, Ginoux P, John J, et al. 2020. Ocean biogeochemistry in GFDL's Earth System Model 4.1 and its response to increasing atmospheric CO_2. *J. Adv. Model. Earth Syst.* 12:e2019MS002043
Stramma L, Johnson GC, Firing E, Schmidtko S. 2010. Eastern Pacific oxygen minimum zones: supply paths and multidecadal changes. *J. Geophys. Res.* 115(C9):C09011
Stramma L, Schmidtko S, Bograd SJ, Ono T, Ross T, et al. 2020. Trends and decadal oscillations of oxygen and nutrients at 50 to 300 m depth in the equatorial and North Pacific. *Biogeosciences* 17:813–31
Sunday JM, Bates AE, Dulvy NK. 2010. Global analysis of thermal tolerance and latitude in ectotherms. *Proc. R. Soc. B* 278:1823–30
Swart NC, Cole JNS, Kharin VV, Lazare M, Scinocca JF, et al. 2019. The Canadian Earth System Model version 5 (CanESM5.0.3). *Geosci. Model Dev.* 12:4823–73
Takano Y, Ilyina T, Tjiputra J, Eddebbar Yassir A, Berthet S, et al. 2023. Simulations of ocean deoxygenation in the historical era: insights from forced and coupled models. *Front. Mar. Sci.* 10:1139917
Thomas CD. 2010. Climate, climate change and range boundaries. *Divers. Distrib.* 16:488–95
Thuiller W. 2004. Patterns and uncertainties of species' range shifts under climate change. *Glob. Change Biol.* 10:2020–27
Tittensor DP, Eddy TD, Lotze HK, Galbraith ED, Cheung W, et al. 2018. A protocol for the intercomparison of marine fishery and ecosystem models: Fish-MIP v1.0. *Geosci. Model Dev.* 11:1421–42
Tjiputra JF, Goris N, Lauvset SK, Heinze C, Olsen A, et al. 2018. Mechanisms and early detections of multidecadal oxygen changes in the interior subpolar North Atlantic. *Geophys. Res. Lett.* 45:4218–29

Tjiputra JF, Negrel J, Olsen A. 2023. Early detection of anthropogenic climate change signals in the ocean interior. *Sci. Rep.* 13:3006

Tjiputra JF, Schwinger J, Bentsen M, Morée AL, Gao S, et al. 2020. Ocean biogeochemistry in the Norwegian Earth System Model version 2 (NorESM2). *Geosci. Model Dev.* 13:2393–431

Toyama K, Iwasaki A, Suga T. 2015. Interannual variation of annual subduction rate in the North Pacific estimated from a gridded Argo product. *J. Phys. Oceanogr.* 45:2276–93

Trenberth KE, Cheng L, Pan Y, Fasullo J, Mayer M. 2025. Distinctive pattern of global warming in ocean heat content. *J. Clim.* 38:2155–68

Troupin C, Barth A, Sirjacobs D, Ouberdous M, Brankart JM, et al. 2012. Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA). *Ocean Model.* 52–53:90–101

Vaittinada Ayar P, Bopp L, Christian JR, Ilyina T, Krasting JP, et al. 2022. Contrasting projections of the ENSO-driven CO_2 flux variability in the equatorial Pacific under high-warming scenario. *Earth Syst. Dyn.* 13:1097–118

Wang Z, Garcia HE, Boyer TP, Reagan JR, Maurer T, et al. 2025. Bias evaluation for sensor-based dissolved oxygen from CTD and profiling floats in the World Ocean Database. *J. Atmos. Ocean. Technol.* 42:1263–80

Wyatt AM, Resplandy L, Marchetti A. 2022. Ecosystem impacts of marine heat waves in the northeast Pacific. *Biogeosciences* 19:5689–705

Yamaguchi R, Suga T. 2019. Trend and variability in global upper-ocean stratification since the 1960s. *J. Geophys. Res. Oceans* 124:8933–48

Yool A, Palmiéri J, Jones CG, Sellar AA, de Mora L, et al. 2020. Spin-up of UK Earth system model 1 (UKESM1) for CMIP6. *J. Adv. Model. Earth Syst.* 12:e2019MS001933

Yukimoto S, Kawai H, Koshiro T, Oshima N, Yoshida K, et al. 2019. The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component. *J. Meteorol. Soc. Jpn. Ser. II* 97:931–65

Zhou Y, Gong H, Zhou F. 2022. Responses of horizontally expanding oceanic oxygen minimum zones to climate change based on observations. *Geophys. Res. Lett.* 49:e2022GL097724

Ziehn T, Chamberlain MA, Law RM, Lenton A, Bodman RW, et al. 2020. The Australian Earth System Model: ACCESS-ESM1.5. *J. South. Hemisph. Earth Syst. Sci.* 70:193–214