GEOXYGEN: a global long-term dissolved oxygen dataset based on biogeochemistry-aware machine learning framework and multi-source observations

Zhenguo Wang[1], Weiwei Fu[1,2], Cunjin Xue[3,4], Guihua Wang[1]

[1]Department of Atmospheric and Oceanic Sciences, Fudan University, Shanghai, 200438, China

[2] Institute of Eco-Chongming (IEC), 1050 Baozhen, Lühua Town, Chongming District, Shanghai 202151, China

[3]International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China;

[4]Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

*Correspondence to: Weiwei Fu (*wwfu@fudan.edu.cn*)*

Abstract. Dissolved oxygen (DO) serves as an essential indicator of marine ecosystem health. However, sparse and uneven observations have limited our ability to characterize its full spatiotemporal variability, underscoring the continued need for long-term, high-resolution, and physically consistent global DO datasets. Here, we present GEOXYGEN, a global dataset of monthly DO fields at 0.5° × 0.5° resolution spanning 1960–2024 and depths from the surface to 5500 m (Wang et al., ~~2025~~2026, ~~https://doi.org/10.5281/zenodo.17615657).~~https://doi.org/10.12157/IOCAS.20260223.002), GEOXYGEN is generated with a hierarchical modeling framework that accounts for regional and vertical heterogeneity. By ~~integrating~~combining physical and biogeochemical predictors with an adaptive feature- selection strategy, GEOXYGEN ~~achieves~~demonstrated high predictive accuracy ~~across all depth layers on an~~ ($R^2 > 0.9$) in independent ~~out-of-time test ($R^2 > 0.92$).~~temporal tests. The reconstructed spatial patterns align closely with the World Ocean Atlas 2023 climatology, and in subsurface ~~and deep~~ waters, GEOXYGEN demonstrates superior generalization relative to existing data-driven products. ~~A sensitivity~~Uncertainty analysis ~~further reveals~~shows that ~~including~~ the uncertainty in coastal data ~~in model training increases basin-wide uncertainty by~~ is approximately ~~7.5%, underscoring~~twice that ~~current observing systems~~ of the open ocean, while long-term deoxygenation trends remain ~~insufficient to reliably resolve nearshore DO dynamics.~~stable even without satellite-era sea-surface predictors. Additionally, a ship-only analysis of the Southern Ocean indicates that early reconstructions are robust, unaffected by the inclusion of Argo observations. GEOXYGEN ~~provides~~offers a consistent, physically informed baseline for ~~analyzing~~investigating global and regional DO variability ~~of DO. It also offers a valuable benchmark~~, providing an important tool for evaluating ~~and improving~~ the representation of DO in climate and Earth system models ~~and can support future studies on long-term deoxygenation trends and regional hotspots~~.

Keywords: Dissolved oxygen, Machine learning, Hierarchical modeling, Long-term dataset, Uncertainty analysis ~~Coastal uncertainty~~

样式定义: First Paragraph: (中文) 简体中文(中国大陆)

设置了格式: 字体: 倾斜

设置了格式: 字体: 倾斜

设置了格式: 字体: 10 磅

设置了格式: 期刊-正文 字符

设置了格式: 期刊-正文 字符

设置了格式: 期刊-正文 字符

设置了格式: 期刊-正文 字符

1 Introduction

Ocean dissolved oxygen (DO) concentration serves as an essential indicator of marine ecosystem health and biogeochemical status (Robinson, 2019; Grégoire et al., 2023). Beyond its ecological significance, DO plays a critical role in modulating climate-relevant biogeochemical feedbacks in the global carbon cycle (Gregoire et al., 2021; Oschlies, 2021; Yamaguchi et al., 2024). Observations over recent decades reveal marked spatiotemporal variability in DO, accompanied by a clear trend toward deoxygenation (Ito et al., 2017), particularly within tropical oxygen minimum zones (OMZs) and in subsurface waters at mid to high latitudes (Bopp et al., 2013; Li et al., 2020). This loss of oxygen is projected to persist under continued global warming (Gong et al., 2021; Zhou et al., 2022), with growing consequences for marine habitats, fisheries, and ecosystem services (Breitburg et al., 2018; Kim et al., 2023; Chen et al., 2024; Humphries et al., 2024).

Sparse and heterogeneous observational coverage ~~hampers an~~hinders accurate ~~estimate~~estimation of the global oxygen inventory~~ and the quantification of its long-term changes~~. A seminal study by Schmidtko et al. (2017) estimates a 2% decline (4.8 ± 2.1 Pmol) in the global ocean oxygen inventory from 1960 to 2009. Yet, the accuracy of such an estimate depends heavily on the observations used. Historically, DO measurements have been sourced from ship-based campaigns compiled in global databases such as the World Ocean Database (WOD) and the Global Ocean Data Analysis Project (GLODAP), which exhibit strong spatial and temporal sampling biases (Garcia et al., 1998). The resulting unevenness in data availability across time, space, and quality standards, especially in coastal waters, complicates robust quantification of deoxygenation rates, particularly in dynamic and vulnerable systems such as coastal shelves~~ and polar oceans.~~. These limitations underscore the pressing need for a spatially continuous, long-term, and accurate global DO reconstruction.

Multiple approaches have been developed to address these observational gaps. Earth system models (ESMs) simulate four-dimensional DO fields continuously but often suffer from systematic biases and incomplete representation of multi-scale processes (Cocco et al., 2013; Oschlies et al., 2018). Limited observational constraints further compound uncertainties in model evaluation and in projections. Traditional statistical interpolation methods can reproduce mean climatologies but frequently underestimate trends in data-sparse regions and fail to capture seasonal to interannual variability (Ito et al., 2024b; Cheng and Gouretski~~ et al., 2024b~~, 2024). In recent years, data-driven machine learning (ML) has emerged as a promising alternative, leveraging relationships between DO and physical or biogeochemical covariates to reconstruct continuous four-dimensional fields from sparse in situ measurements (Sharp et al., 2023; Garabaghi et al., 2023; Huang et al., 2023; Wang et al., 2024; Lu et al., 2024). In principle, ML can recover local variability and identify deoxygenation risk without relying on computationally expensive coupled simulations.

Despite this potential, several methodological challenges remain. First, many existing ML reconstructions ~~employ~~rely on a single model trained on global~~-~~ ocean data~~, which struggles to represent regional variations in~~ . However, the ~~dominant

physical– and biogeochemical controls on DO as well as the spatial heterogeneity of processesvary markedly across biogeochemical provincesregions, and the relationships between DO and its environmental drivers are strongly nonstationary in space and time (Garabaghi et al., 2023). ThisAs a result, a unified global approach tends to blur contrasts between water-mass regimes and degrade skillmapping can be unduly shaped by data-rich areas, so that inferred relationships in dynamically distinct regions such as oxygen minimum zones and boundary currentsdata-sparse regions are effectively constrained by remote analogs, yielding unstable long-term trends. Second, a common workflow is to first reconstruct DO at scattered profile locations and then interpolate these point estimates onto a regular grid, often using a limited set of predictors such as temperature and salinity (Sharp et al., 2023; Wang et al., 2024; Liu et al., 2025). In data-sparse regions, this two-step procedure encourages extrapolation, propagates local errors, and can generate spurious fine-scale structure that is not supported by the underlying observations, particularly near sharp DO gradients and in historically undersampled basins. Third, training models directly on raw profiles amplifies sampling biases: autonomous platforms such as Argo repeatedly sample specific regions and depth ranges, whereas historical ship-based surveys are concentrated along cruise tracks (Huang et al., 2023; Lu et al., 2024). Without explicit rebalancing or weighting, ML models place disproportionate emphasis on well-observed areas and generalize poorly elsewhere, leading to reconstructions that systematically underrepresent variability and trends in data-poor regions.

At the dataset level, available ML-based global DO products, such as GOBAI-O_2 (Sharp et al., 2023), G4D-DOC (Xue et al., 2024), and ML4O2ML4O2 (Ito et al., 2024a), represent important advances, providing monthly gridded DO fields at 1° resolution over multi-year to multi-decadal periods and resolving much of the upper and intermediate ocean. However, theythese products are generally typically either limited either to the Argo erain temporal coverage or to depths shallower than aboutdo not extend below approximately 1000–2000 m. To our knowledge, there is currently no single observation-based product that combines pre-Argo coverage from the 1960s, full-depth global fields, and sub-degree horizontal resolution.

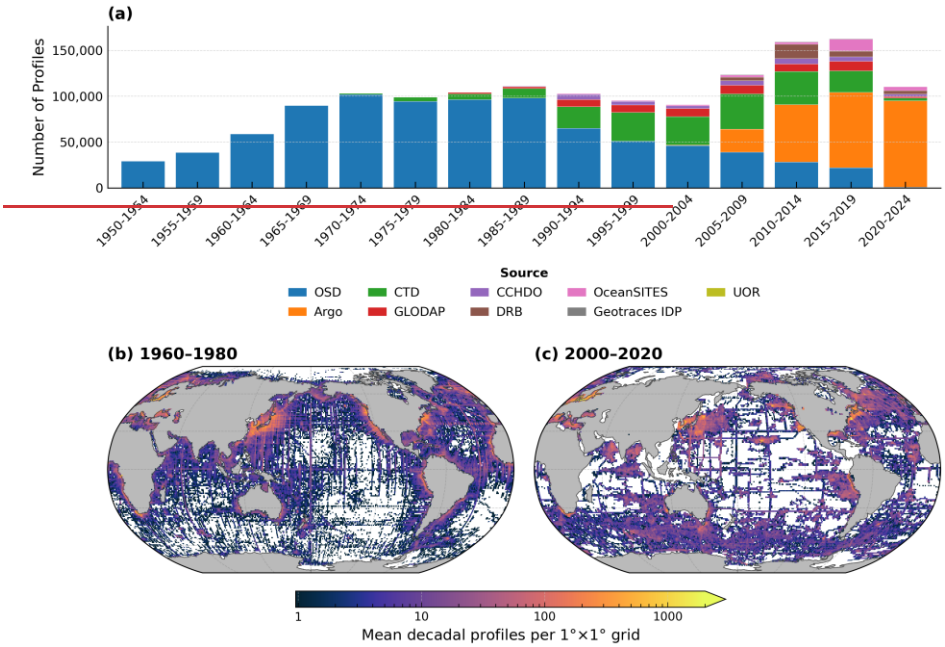
To address these methodological and dataset-level limitations, we generated GEOXYGEN, a monthly global DO dataset at 0.5° × 0.5° resolution on 187176 depth levels from 1960 to 2024 (Wang et al., 20252026, https://doi.org/10.5281/zenodo.17615657https://doi.org/10.12157/IOCAS.20260223.002). The dataset is generated by combining a global compilation of in situ DO profiles with objectively analyzed temperature–salinity fields and related sea-surface environmental variables, and by learning their relationships with DO through a regionally structured, depth-aware, and adaptively constrained machine-learning framework. Our approach explicitly accounts for strong spatiotemporalSpecifically, we implement heterogeneity by-based partitioning the ocean into ecogeographical macro-regions and by training separate submodels for differenteach depth layerslayer within each region and incorporating monthly climatological environmental state fields as prior information. Within each region–depth unit, we adaptively select predictive features from a suite of variables including temperature, salinity, oxygen saturation, physical indicators, carbonate-system

parameters, and bio-optical properties, thereby ensuring physical interpretability while minimizing redundancy. To mitigate sampling bias and boundary discontinuities, we ~~implement~~apply inverse-density weighting, ~~year-grouped~~decadal-block cross-validation, and cross-boundary fusion ~~techniques~~. The resulting GEOXYGEN product provides a consistent, long-term, and spatially complete representation of global DO suitable for quantifying global and regional deoxygenation, diagnosing underlying drivers, and evaluating Earth system and biogeochemical models.

2 Data

2.1 ~~In situ oxygen observations~~Oxygen Data

To ~~support~~reconstruct a ~~data-driven reconstruction of dissolved oxygen (DO),~~long-term, global ocean DO data set, we compiled ~~a global collection of~~ several complementary in situ ~~measurements by integrating six major international~~ data ~~sources: the~~products. These include CLIVAR and the Carbon Hydrographic Data Office (CCHDO), ~~the Global Ocean Data Analysis Project (~~GLODAP~~),~~ the GEOTRACES Intermediate Data Product ~~2021~~ (IDP2021), ~~Biogeochemical Argo (BGC-Argo), the World Ocean Database 2023 (WOD23; Mishonov et al., 2024), and the~~ the OceanSITES ~~fixed-point observatory~~mooring network~~. This compilation spans the period 1950–2024 and includes millions of DO~~ , and the internally consistent OSD/CTD and Argo DO data set of Gouretski et al. (2024). The OSD/CTD and Argo profiles ~~collected via bottle casts (OSD), Conductivity–Temperature–Depth (CTD) profilers~~are merged under a single automated QC framework, and Argo ~~floats~~DO biases are evaluated and corrected using contemporaneous reference measurements. This makes the quality more consistent across platforms and helps reduce platform-related systematic differences that can affect later modeling.



A rigorous dual-stage QC protocol ensured observational reliability. The primary stage involved standardizing metadata formats and units across disparate sources, retaining only observations flagged as "good" or "probably good". Spurious terrestrial signals were omitted via land-masking, and duplicate profiles—defined by coincidence criteria of ≤ 1 km spatial distance and ≤ 24 h temporal difference—were identified across archives. In cases of redundancy, we prioritized profiles with the highest vertical sampling density.

Standardization was further achieved by mapping observations to 176 vertical levels, adopting an expanding grid resolution (10 m intervals above 800 m, 20 m to 2000 m, and 100 m thereafter down to 5500 m). This configuration aligns with the vertical frameworks of CORA and ISAS17 (Szekely et al., 2019; Kolodziejczyk et al., 2023). For the original profiles, DO, temperature, and salinity were mapped using a shape-preserving piecewise cubic Hermite interpolator (PCHIP) strictly within the observed pressure range, precluding vertical extrapolation. Profiles from Gouretski et al. (2024) were utilized directly without further interpolation as they were pre-aligned to these standard levels.

To avoid adding values in large vertical gaps where observations are sparse while retaining only levels that are locally supported by sufficient observations, we applied an "observation constraint" filter to the interpolated standard-level data. For each standard level z, we first used a strict window $x(z)$: the level is kept if at least one observed pressure exists within

$\pm x(z)$. For levels that fail the strict window, we then applied a relaxed window $y(z)$, but required at least two observed pressures within $\pm y(z)$. The windows vary with depth y, as described in Eq. (1).

$$x(z) = \begin{cases} 2, & z \le 50 \\ 5, & 50 < z \le 800 \\ 10, & 800 < z \le 2000 \\ 20, & z > 2000 \end{cases} m, \quad y(z) = \begin{cases} 5, & z \le 50 \\ 15, & 50 < z \le 800 \\ 30, & 800 < z \le 2000 \\ 120, & z > 2000 \end{cases} m, \quad (1)$$

We also recorded an interpolation uncertainty term, σ_{interp}, to represent the local error scale introduced by vertical interpolation, as defined in Eq. (2). Using a first-order approximation, this uncertainty depends on the local vertical DO gradient and the distance to the nearest valid observed pressure. We estimated the local gradient $|\partial O_2 / \partial p(z)|$ from adjacent observed points along the profile, defined $\Delta p(z)$ as the pressure distance between the standard level and the nearest valid observed point.

$$\sigma_{interp}(z) = \left| \overline{\frac{\partial O_2}{\partial p}}(z) \right| \cdot \Delta p(z), \quad (2)$$

A second QC stage was applied to the standard levels. First, we excluded records where DO exceeded the 0–600 μmol kg^{-1} range. Next, to limit uncertainty from vertical interpolation, data with $\sigma_{\mathrm{interp}} > 3$ μmol kg^{-1} were removed, based on the 95th percentile of σ_{interp} across all standard-level samples. Finally, following TEOS-10, oxygen saturation percentage (Sat%) was computed using in-situ temperature and salinity data. For standard levels deeper than 200 m, records with O_2 saturation $\geq$ 120% were considered erroneous and removed.

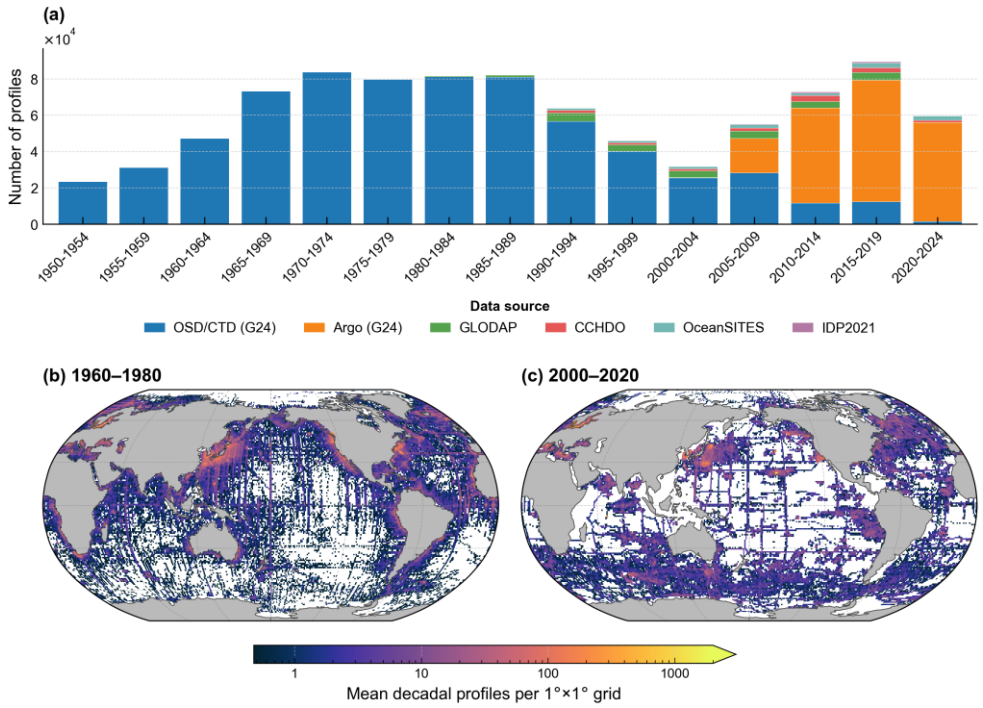


Figure 1: Global distribution and temporal coverage of DO profiles. (a) Changes in the number of profiles from major data sources during 1950–2024. (b–c) Spatial distribution of decadal-mean profile counts for 1960–1980 and 2000–2020, computed on a 1° × 1° grid~~, showing a transition from predominantly Northern Hemisphere sampling to much denser coverage in the Southern Hemisphere.~~. The color bar indicates the decadal-mean number of profiles per grid cell (log scale). G24 denotes the oxygen compilation of Gouretski et al. (2024).

~~To ensure dataset consistency, we implemented a multi-stage quality-control (QC) procedure. First, we standardized formats and units across sources and retained only measurements flagged as "good/probably good." Second, we removed duplicates from different sources on the same date and at the same location with vertical separation less than 1 m. Third, we applied a physiologically plausible DO threshold of 0–600 μmol kg^{-1} to exclude outliers. Although differences in QC protocols and metadata conventions across datasets may introduce cross-source biases (Gregoire et al., 2021), prior assessments have indicated no significant systematic offsets between Winkler-calibrated high-accuracy data and sensor measurements at the global scale (Schmidtko et al., 2017). Therefore, we did not apply cross-source corrections. Instead, we focused on local outlier detection at each standard depth level, whereby within each 0.5° × 0.5° grid cell and a ±10-day time window, observations deviating by more than 3σ from the local mean were excluded. Across depth levels, removal rates range from 0 to 0.105%, with most levels below 0.08%, indicating that only a very small fraction of the quality-controlled observations was~~

~~flagged as outliers.~~

~~Following QC, the final observational archive for 1950–2024 comprises approximately 1.4 million DO profiles (Fig. 1a). This full 1950–2024 record is used for model training and evaluation, whereas the reconstructed DO fields are produced over 1960–2024. The OSD, CTD, and Argo collectively account for ~91.15% of all profiles. Vertical sampling distribution has evolved with technological advances: OSD dominates in surface waters (~80% above 100 m) but declines to less than 10% by 500 m, whereas Argo represents over half of all profiles below 500 m and is the primary source in the mesopelagic layer. This shift reflects broader observational trends: discrete bottle sampling prevailed through the 1960s, CTD profiling expanded in the 1970s–1990s through programs such as GLODAP and CCHDO, and autonomous float observations increased by an order of magnitude after 2005. By 2010, Argo became the dominant platform in the open ocean.~~

~~Spatial coverage has shifted from a strongly Northern Hemisphere bias to a more Southern Hemisphere focus, but remains highly uneven (Fig. 1b–c). From 1960 to 1980, sampling was concentrated along ship tracks in the Northern Hemisphere, with limited data in the Southern Hemisphere and basin interiors. Quantitatively, analysis of our profile dataset shows that the number of Southern Hemisphere profiles increased from roughly one quarter of the Northern Hemisphere count in 1960–1980 to about 2.4 times the Northern Hemisphere count in 2000–2020, reflecting a pronounced shift of observing effort into the Southern Hemisphere. Since 2000, autonomous platforms have substantially improved coverage across the Southern Ocean and the open ocean, partly alleviating the historical Southern Hemisphere undersampling. Nevertheless, marginal seas and high-latitude ice-covered regions remain undersampled. In our reconstruction approach, we mitigate these biases through regionalized modeling, gridding to standard depth levels, and inverse-density weighting.~~

~~2.2 Physical–biogeochemical variables~~

~~Accurately predicting~~After two-stage QC, we assembled a high-quality archive of 930,252 DO profiles, of which 94% are derived from the OSD, CTD, and Argo platforms. Figure 1a captures the decisive transition in the historical data record: a long-standing reliance on ship-based OSD and CTD profiles was superseded post-2000 by the exponential growth of the Argo network, which now serves as the backbone of global oxygen monitoring. Platform-specific dominance is also depth-dependent, defining the vertical representativeness of the dataset. Above 800 m, the combined prevalence of OSD, CTD, and Argo remains robustly above 90%. However, the intermediate depths (800–1800 m) mark a transitional zone where Argo's increasing share compensates for the thinning OSD/CTD coverage. Beyond 2000 m, where Argo penetration is limited, the deep-ocean constraint returns to the OSD and CTD platforms, which maintain a prevalence of at least 80%. This tiered vertical structure ensures that secondary data products remain supplementary to the core observational framework.

The geometric evolution of sampling has shifted from ship-track-oriented linear clusters to a near-planar global distribution. This shift is most evident in the Southern Hemisphere, where coverage increased from extreme sparsity prior to 1980 to

approximately 36% of the global total in the 2000–2020 interval. This globalization of the data stream, driven by the Argo array, has mitigated historical hemispheric imbalances, although systematic gaps persist in marginal seas, the central Pacific, and ice-influenced Arctic regions.

Acknowledging that uneven observational density poses a risk of artificial signal dominance from high-coverage eras, our methodology incorporates specific safeguards to ensure trend stability. By adopting heterogeneity-based partitioning modeling and decadal-scale transferability assessments, we reduce the risk of data-sparse regions being dominated by modern sampling patterns. These efforts are further supported by an inverse-density weighting scheme and the use of grid-representative values, which together enhance the influence of sparse historical records and maintain the physical interpretability of multi-decadal oxygen trends.

2.2 Features Data

Accurate reconstruction of ocean deoxygenation ~~requires denser, longer~~needs both long, dense DO observations~~,~~ and drivers that ~~characterize~~describe physical transport and ~~cross-scale~~ biogeochemical processes across scales (Oschlies et al., 2018). ~~To meet this need, we complement objectively analyzed~~ Consequently, we leverage three-dimensional thermodynamic fields (temperature~~–~~ and salinity ~~fields with~~) alongside a curated suite of sea~~-~~ surface environmental variables (SSEVs~~), providing physically consistent and~~) to establish a process-~~informed predictors for DO reconstruction~~based predictor space (Table 1).

~~We derive three~~Three-dimensional temperature and salinity fields are taken from the Coriolis Ocean Dataset for Reanalysis ~~(~~, CORA~~)~~ (Szekely et al., 2025~~),~~). CORA is a CMEMS objective analysis that ~~compiles~~uses the ISAS objective mapping system to merge in situ temperature and salinity ~~using the ISAS objective mapping system, integrating~~ observations from ships, Argo floats, and other ~~in situ~~ platforms. ~~This dataset undergoes~~It also applies delayed~~-~~ mode quality control to ~~ensure~~support long-term stability and global ~~coherence. Using~~consistency. From these fields, oxygen saturation (O_2 *Sat*) is derived according to the TEOS-10 standard~~, we further calculate oxygen saturation (O_2 sat) from CORA temperature and salinity data~~ (IOC, 2010)~~; its deviation from~~). The resulting discrepancy between observed DO ~~reflects biological~~and O_2 *Sat* serves as a critical proxy for the integrated influence of microbial respiration and physical ~~mixing~~ventilation dynamics.

~~In parallel, we assemble~~Complementarily, we assembled a multidimensional suite of SSEVs~~ spanning~~, encompassing thermodynamic, dynamical, bio-optical, and carbon-chemistry ~~processes, supporting data-driven reconstruction~~features (Shao et al., 2024; Ma et al., 2025). Wind vectors (~~zonal and meridional components,~~ U and V) are taken from NASA's Cross-Calibrated Multi-Platform (CCMP) product (Mears et al., 2022). Mixed-layer depth (MLD) is obtained from the CMEMS Multi-Observation Global Ocean 3D product (Guinehut et al., 2012). Dynamical variables include sea surface height (SSH) and eddy kinetic energy (EKE), both derived from AVISO satellite altimetry (Hauser et al., 2020). Bio-optical variables comprise photosynthetically active radiation (PAR) and chlorophyll a (Chl-a) from NASA Level-3/Level-4 ocean-color

设置了格式: 字体: 10 磅

products (NASA Ocean Biology Processing Group, 2018). Carbon-chemistry variables include dissolved inorganic carbon (DIC), total alkalinity, pH, sea surface partial pressure of CO_2 (pCO_2), and CO_2 flux, all obtained from the CMEMS Surface Ocean Carbon Fields product (Chau et al., 2022; Chau et al., 2024). All feature variables ~~were,~~ last accessed in March 2025, were standardized onto a uniform monthly 0.5° grid to maintain spatial consistency across the reconstruction.

~~For consistency across sources, we regridded each variable to a uniform 0.5° × 0.5° grid, aggregated them to a monthly resolution, and aligned them with both the DO observation locations and the reconstruction grid. This standardization ensures coherent feature availability during model training and prediction while retaining large-scale physical and biogeochemical signals essential for reconstructing the deoxygenation trend. To preserve historical information and reduce imputation uncertainty, we retain all QC-passed DO data, even if some covariates (e.g., SSEVs) are missing. This sampling library provides standardized inputs for subsequent model training.~~

Table 1. Details of the ~~Physical–biogeochemical variables~~Features Data

Variable	Description	Spatial Resolution	Temporal Resolution	Temporal Coverage	Data Source
Temperature (°C)	Seawater temperature	0.~~25~~5° × 0.~~25~~5°; 187 standard depth levels (surface–5500 m)	Monthly	1960/01 – 2024/06	(Szekely et al., 2025)
Salinity	Seawater salinity	0.~~25~~5° × 0.~~25~~5°; 187 standard depth levels (surface–5500 m)	Monthly	1960/01 – 2024/06	
~~O_2_sat~~	~~Seawater oxygen saturation (calculated)~~	~~0.25° × 0.25°; 187 standard depth levels (surface–5500 m)~~	~~Monthly~~	~~1960/01 – 2024/06~~	
U (m s^{-1})	U-wind vector component at 10 meters	0.25° × 0.25°	Monthly	1993/01 – 2023/08	(Mears et al., 2022)
V (m s^{-1})	V-wind vector component at 10 meters				
MLD (m)	Ocean mixed layer depth	0.25° × 0.25°	Monthly	1993/01 – 2022/12	(Guinehut et al., 2012)
DIC (~~μmol~~μmol	Surface ocean			1985/01 –	(Chau et al.,

kg^{-1})	dissolved inorganic carbon	0.25° × 0.25°	Monthly	2023/12	2022; Chau et al., 2024)
pH	Surface pH on total scale				
pCO_2 (µatm)	Surface aqueous partial pressure of CO_2				
CO_2 flux (mol m^{-2} yr^{-1})	Surface downward flux of total CO_2				
Alkalinity (~~µmol~~µmol kg^{-1})	Total alkalinity in surface seawater				
PAR (mol m^{-2} d^{-1})	Photosynthetically available radiation	4 km / 9 km	Monthly	1997/10 – 2025/~~2~~02	(NASA Ocean Biology Processing Group, 2018)
Chl-a (mg m^{-3})	Mass concentration of chlorophyll in surface water				
SSH (m)	Sea surface height above geoid	0.25° × 0.25°	Monthly	1993/01 – 2023/08	(Hauser et al., 2020)
EKE (cm^2 s^{-2})	Surface averaged eddy kinetic energy				

~~**2.3 Depth-dependent relationships between oxygen and drivers**~~

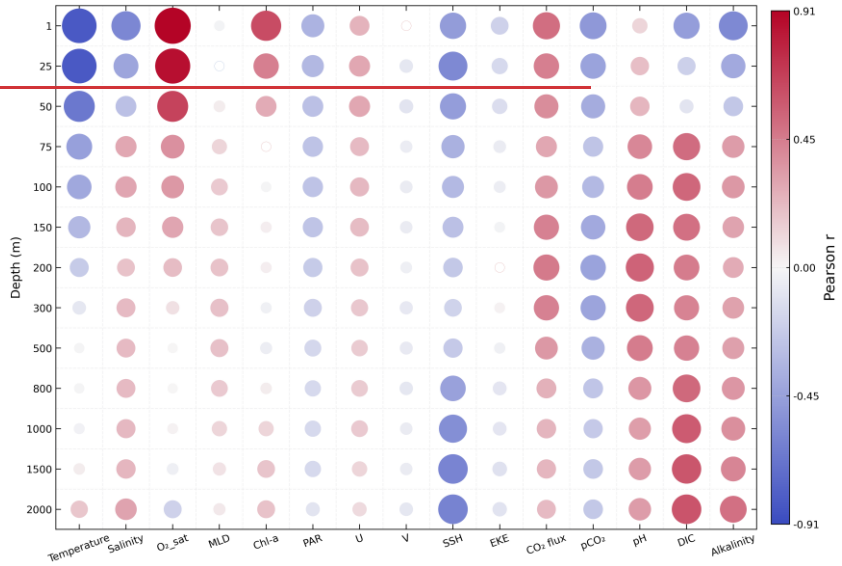
~~We characterized the depth-dependent relationships between dissolved oxygen (DO) and a suite of environmental variables (Fig. 2). Near the surface, DO exhibits a strong negative correlation with temperature, consistent with solubility control, while salinity contributes via the salting-out effect. With increasing depth, the direct influence of solubility declines, yet temperature and salinity remain valuable as proxies for water-mass identity and ventilation history. For reference, we also show correlations with TEOS-10 derived oxygen saturation (O_2_sat), which serves as a physicochemical baseline reflecting equilibrium conditions.~~

~~Sea-surface environmental variables (SSEVs) display distinct vertical patterns in their coupling with DO. Sea surface height (SSH) maintains a strong and stable correlation through much of the water column, while surface wind components (U, V) contribute dynamical context related to wind-driven advection and upwelling (Hollitzer et al., 2024). Carbonate system~~

~~variables retain explanatory power below the euphotic zone, indicative of remineralization signals and association with specific water masses. In contrast, Chl-a is mainly informative within the upper tens of metres, reflecting its biogeochemical role in the sunlit layer. Several SSEVs also display notable nonlinear relations with subsurface DO, which supports their value for reconstructing underwater DO (Ping et al., 2024; Cao et al., 2024).~~

设置了格式: 图案: 清除 (白色)

~~Overall, these correlation patterns illustrate how surface forcings and water-mass structure jointly shape subsurface oxygen variability. They provide a theoretical and empirical foundation for subsequent modeling of the often nonlinear relationships between these variables and DO.~~



~~**Figure 2: Vertical correlations between DO and physical–biogeochemical variables. Each bubble represents the Pearson correlation coefficient ((r)) between DO and an environmental variable at a standard depth level. Bubble color encodes the sign and magnitude (red = positive; blue = negative), and bubble area scales with (|r|). Filled bubbles denote correlations significant at (q<0.05) after Benjamini–Hochberg false-discovery-rate control; hollow bubbles indicate non-significant results.**~~

3 Method

~~Here, we develop a hierarchical modeling framework that resolves regional and vertical~~ The overall workflow for constructing the GEOXYGEN dataset is illustrated in Fig. 2, comprising data collection and preprocessing, heterogeneity ~~-by-~~based partitioning ~~the ocean into ecogeographic regions,~~ and model training ~~depth-stratified, region-specific CatBoost models~~ with ~~adaptive feature selection and year-grouped cross-validation.~~

~~3.1 Partitioning based on heterogeneity~~

~~Dissolved oxygen (DO) trends exhibit substantial variability across ocean basins, latitude bands, and depth layers, with~~ 带格式的: 正文文本

~~dominant controlling mechanisms shifting both regionally and vertically (Ma et al., 2025). To capture regional~~integrated physical~~,~~ and biogeochemical ~~coupling while mitigating sampling biases, we implement a modeling framework that is stratified in both~~predictors. By incorporating a spatial clustering descriptor that captures the ~~vertical and horizontal dimensions~~large-scale background structure, and depth-/region-adaptive feature selection, we develop a biogeochemistry-aware machine-learning framework for gridded DO reconstruction.

~~**Vertical stratification.** Given~~

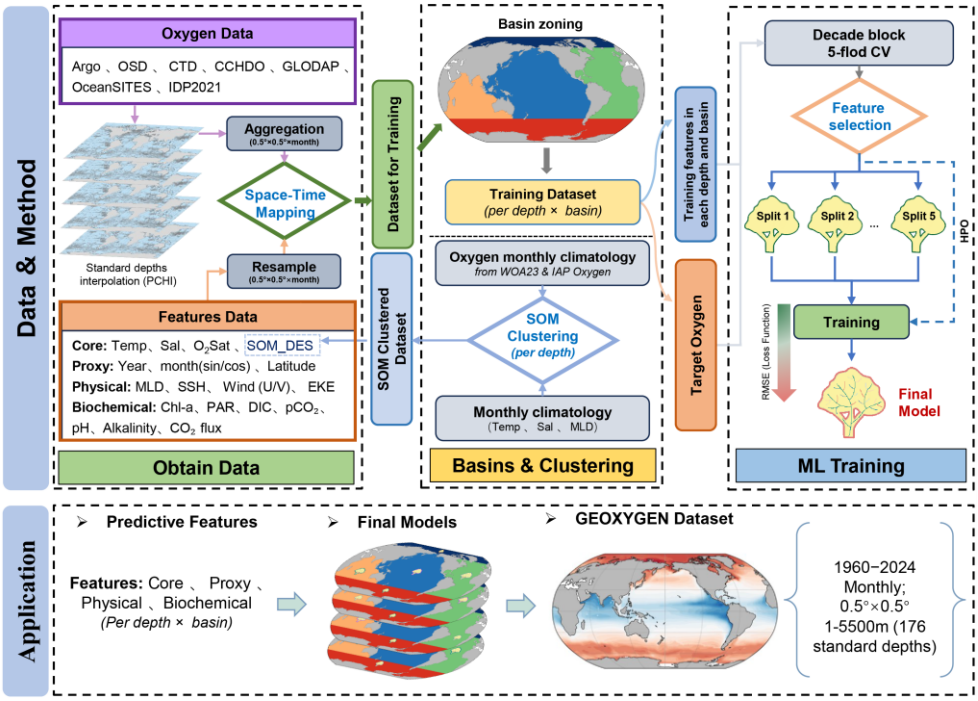


Figure 2: Overall workflow of the GEOXYGEN dataset construction.

3.1 Gridding and Aggregation

To match the spatiotemporal scale of the supervised learning labels with the target reconstruction grid and to reduce sample-weight bias caused by uneven sampling in space and time, all observations are aggregated to a monthly 0.5° by 0.5° grid. Within each grid cell, observations are summarized into a single representative value, defined as the median to limit the influence of outliers and extreme events on the labels. The within-unit dispersion is also computed using the median absolute deviation (MAD), which serves as an empirical proxy for uncertainty.

To quantify representativeness error introduced by finite sampling and sub-grid variability within each cell-month-depth ~~dependence of processes governing DO~~ unit, within-unit dispersion is used to estimate σ_{rep}. Let a unit contain n_{obs} observations $\{x_i\}_{i=1}^{n_{\mathrm{obs}}}$. The representative value is defined as the robust central statistic, as shown in Eq. (3). We then compute the dispersion and define it as the representativeness error metric, σ_{rep}, as described in Eq. (4), where the factor 1.4826 makes σ_{rep}comparable to the standard deviation under an approximate normal distribution, facilitating comparison across units.

$$\tilde{x} = \mathrm{median}(x_i), \quad (3)$$

$$MAD = \mathrm{median}(|x_i - \tilde{x}|), \quad \sigma_{rep} = 1.4826 \times \mathrm{MAD}, \quad (4)$$

At the same spatiotemporal scale, all environmental variables and DO data are mapped in space and time. To preserve historical information, all QC-passed DO data are retained even when some covariates such as SSEVs are missing. This sample archive provides standardized inputs for later model training.

3.2 Heterogeneity-Based Partitioning

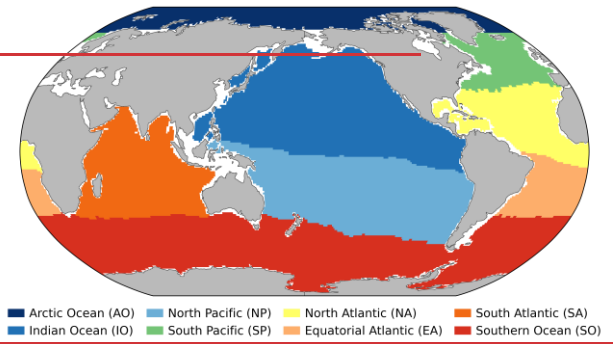
Our reconstruction framework utilizes a spatiotemporally stratified approach to address the shifting controlling mechanisms of ocean deoxygenation across basins and depths (Ma et al., 2025; Ito et al., 2024a). Following the basin definitions in the World Ocean Atlas 2023 (WOA23; Garcia et al., 2024), we ~~divide the 0–5500 m water column into 187 standard depth levels, following the~~additionally treat the Southern Ocean as a dedicated domain. Accordingly, the horizontal grid is divided into five primary modeling domains (Atlantic, Pacific, Indian, Southern, and Arctic), which are held constant across vertical ~~grid of the~~ layers to maintain training coherence (Fig. 3). By avoiding highly intricate province boundaries, this design reduces sensitivity of variability and trend estimates to boundary effects and makes cross-boundary continuity easier to maintain. We further mask the Mediterranean, Red Sea, and other semi-enclosed marginal seas to focus the reconstruction on open-ocean dynamics dominated by large-scale circulation and transport.

The vertical architecture utilizes 176 standard levels, as discussed in Sect. 2.1, similar to the CORA ~~objective/ analysis and~~ ISAS17 ~~climatology~~frameworks (Szekely et al., 2019; Kolodziejczyk et al., 2023). ~~An independent model is trained at~~At each standard depth~~, allowing~~ below 3000 m, the framework ~~to resolve depth-varying controls while preserving realistic vertical structure.~~

~~**Horizontal regionalization.** To account for spatial heterogeneity in~~ converges into a single global model, a necessary adaptation to accommodate the physical~~ biogeochemical controls on DO, we follow the global ocean biogeochemical province classification of Fay and Mckinley (2014) and refine it using basin boundaries and the statistical robustness of sample distributions, allowing region-specific submodels to better represent distinct DO–environment relationships. Gaps in the original map are filled by nearest-neighbor interpolation constrained by ocean connectivity, combined with a KD-tree projection. To avoid local complexities, we exclude enclosed or exchange-limited seas (e.g., the Mediterranean Sea and the~~

设置了格式: 字体颜色: 蓝色

~~Red Sea) and mask coastal waters shallower than the 200 m isobath. A sensitivity analysis of this coastal threshold is presented in Sect. 4.3~~ constraints of the seafloor and data limitations. For each basin and depth layer, separate models are trained based on the partitioned regions.



Within each depth, the SOM_DES predictor accounts for localized variations in water-mass properties and frontal dynamics by encoding the large-scale, climatological background state. Specifically, for each month and grid cell, we form a four-dimensional vector [SST, SSS, MLD, DOclim] from the corresponding climatological monthly fields and assign the grid cell to a discrete SOM class, yielding a categorical environmental-state label. The global ocean is clustered into 25 biogeochemical regimes. The clustering is based on the joint patterns of multiple environmental fields, and, at this step, the O_2 climatology is implicitly assigned a higher weight to obtain a seasonally varying, dynamic partitioning. The SOM is trained directly on these monthly climatological fields, such that the DO climatology implicitly exerts a stronger influence on the resulting state partition and ensures that the partitioning evolves coherently with the seasonal DO cycle. This feature-engineering step leverages the SOM's ability to map multivariate climatological structure onto a discrete set of regimes while preserving topological dependencies, thereby providing clear seasonal and spatial context for month-scale DO reconstruction. The background climatological fields are anchored in WOA23 (upper 1500 m) and the IAP Global Ocean Oxygen gridded product (IAP Oxygen; Cheng and Gouretski, 2024), ensuring vertically and horizontally comprehensive baseline states. Overall, SOM_DES captures the large-scale climatological structure of DO and supplies background-state information that supports refined month-scale DO modeling, improving diagnostic consistency across heterogeneous regimes and reinforcing the physical interpretability of the global DO product.

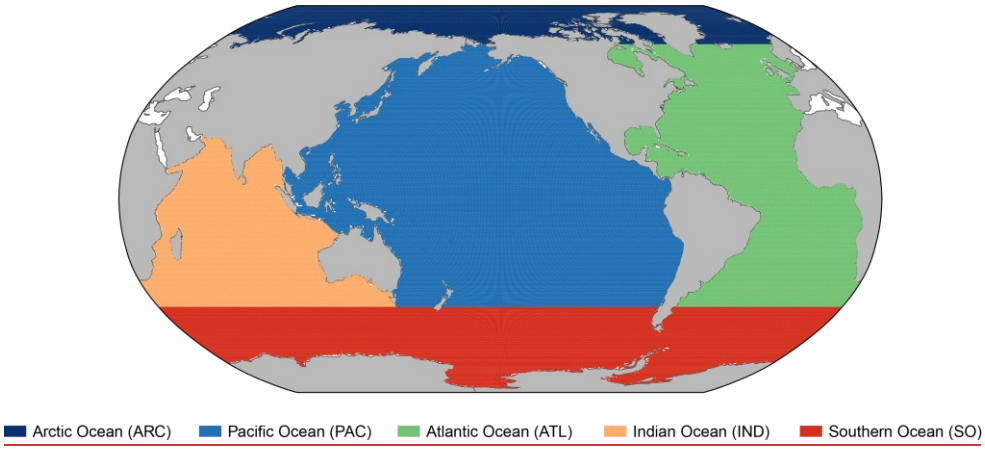


Figure 3: Partitioning of the global open ocean into ~~eight macro biogeochemical provinces~~five basins.

~~The open ocean is partitioned into eight macro biogeochemical provinces (Fig. 3): the Arctic Ocean (AO), North Pacific (NP), South Pacific (SP), North Atlantic (NA), Equatorial Atlantic (EA), South Atlantic (SA), Indian Ocean (IO), and Southern Ocean (SO). Each province serves as an independent modeling unit, enabling regionally tailored representation of key processes, reducing spurious extrapolation across biogeochemical domains, and introducing structured constraints that enhance the physical consistency of the reconstructed DO fields.~~

~~3.2~~3.3 Adaptive ~~modeling~~Regional Modeling

~~Within each modeling unit, we employ~~We train independent sub-models for each ocean basin partition across 176 standard depth levels, employing the CatBoost gradient- boosting framework~~ an interpretable decision tree method to learn the functional~~ to resolve the non-linear mapping between sparse ~~dissolved oxygen (~~DO~~)~~ observations and ~~their~~diverse environmental predictors. ~~This method builds ensembles~~CatBoost's implementation of oblivious decision trees ~~with~~and ordered boosting~~, a training scheme that reduces~~ is strategically utilized to mitigate variance ~~and limits~~while precluding target leakage~~. The algorithm offers several advantages for our application: it directly accommodates missing covariates~~ a critical requirement for stable multi-decadal reconstruction. This framework is algorithmically predisposed to this task due to its innate capacity to process missing predictors without imputation, ~~incorporates~~its support for cost-sensitive sample weighting~~ to address spatial biases~~, and ~~employs multiple~~its robust regularization ~~strategies including~~ suite (L2 leaf regularization ~~on leaf values, subsampling,~~ and early stopping~~ to prevent~~) that safeguards against overfitting.~~ in data-sparse regimes.~~

~~For historical periods (e.g., 1960–1997) where~~ During the period before the satellite ~~derived predictors are unavailable, the model treats missing covariate values as a distinct input state, allowing it to rely primarily on temperature, salinity, oxygen saturation,~~ launch (for example, from the 1960s to the 1980s), this model would automatically identify and properly handle the missing SSEVs. This non-imputational strategy allows the reconstruction to hinge predominantly upon thermodynamic variables (*Temperature*, *Salinity*, *O_2 Sat*) and spatiotemporal coordinates~~. This strategy avoids~~ without introducing the systematic biases ~~that can arise from statistical imputation and maintains consistency across the reconstruction period.~~inherent in statistical filling. We tested longitude as a candidate predictor; however, it induced spurious banded (stripe-like) artifacts in data-sparse regions and was therefore excluded from the final predictor set. The CatBoost model uses ~~19~~20 predictor variables as inputs~~. They include: temperature, salinity, O_2_sat, year_norm~~: *Year*, *month_sin*, *month_cos*, *Latitude*, *Temperature*, *Salinity*, *O_2 Sat*, *SOM_DES*, *U*, *V*, *SSH*, *EKE*, *MLD*, *PAR*, *Chl-a*, *DIC*, *pCO_2*, *pH*, *Alkalinity*, and *CO_2 flux*. The ~~temporal~~*month_sin* and *month_cos* encodings are defined in Eq. ~~(1)–(3), where days_since_start denotes the number of days elapsed since 1950-01-01 00:00:00 UTC.~~(5) – (6).

~~year_norm = $\frac{\text{days\_since\_start}}{365.25}$,~~ ~~(1)~~

~~month_sin~~*month_sin* = $\sin\left(\frac{2\pi\,(m-1)}{12}\right)$,

~~(2~~ (5)

~~month_cos~~*month_cos* = $\cos\left(\frac{2\pi\,(m-1)}{12}\right), m \in \{1,2,\ldots,12\}$,

~~(3~~ (6)

To ~~enhance model~~ improve interpretability and generalization, ~~we implemented~~ a two-stage feature selection procedure ~~for~~ is applied to each region–depth submodel~~, using only~~. Only training data are used, and ~~a year-grouped~~ cross-validation ~~scheme consistent with our~~follows the same decadal-scale design used for model evaluation ~~framework~~. Tree-based models ~~tend to perform better with~~often benefit from compact~~, informative~~ feature sets~~, as~~ because redundant predictors can ~~dilute~~reduce predictive ~~accuracy~~skill (Garabaghi et al., 2023). First, we estimated permutation importance under five-fold cross-validation grouped by ~~year~~decade and retained an initial subset of features using an adaptive rule $K = \max(10, 2\sqrt{p})$, where $p =$ ~~19~~20 is the number of candidate features, discarding predictors with negligible contribution. Second, we perform recursive feature elimination with cross-validation (RFECV~~) using the same year-grouped folds,~~), iteratively removing the least important feature and selecting the combination that minimizes validation Root Mean Square Error (RMSE~~. The five independent test years are excluded from both stages.~~). This ~~design allows~~process facilitates a physical evolution of the feature ~~sets to adapt to regional and vertical regimes—~~space: surface models ~~emphasise rapidly varying surface and~~prioritize high-frequency biogeochemical ~~terms~~forcing, mid-depth models ~~highlight meridional gradients and~~emphasize water-mass transition

设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜
设置了格式: 字体: 倾斜

~~metrics~~markers, and deep-ocean models ~~rely primarily~~converge on ~~temperature–salinity structure, while avoiding temporal information leakage and reducing overfitting~~conservative thermodynamic tracers.

To mitigate biases arising from heterogeneous spatiotemporal sampling (Fig. 1), we applied inverse-density weighting within a fixed binning scheme. The sample domain is partitioned on a 5° × 5° latitude–longitude grid and non-overlapping 10-year time windows in each ~~partition~~cell. Sample weights were computed as the inverse of the observational density within each spatiotemporal bin and standardized at each depth level, thereby reducing the influence of over-sampled regions and periods during model training.

Let $b(i)$ denote the spatiotemporal bin containing sample i, and let $n_{b(i)}$ be the number of samples in that bin. The initial per-sample weight is the inverse square root of this count ~~(~~, as defined in Eq. ~~(4)~~):7).

$$\tilde{w}_i = \frac{1}{\sqrt{n_{b(i)}}}, \qquad (4 \text{ ~~(7)~~})$$

Specifically, each sample weight is set proportional to the inverse square root of the local sample density. To preserve the aggregate information content, we then normalize the weights to unit mean ~~(~~, as defined in Eq. ~~(5)~~):8).

$$w_i = \frac{\tilde{w}_i}{\frac{1}{N}\sum_{k=1}^{N} \tilde{w}_k}, \qquad (~~5~~ (8))$$

This strategy increases the influence of observations from sparse regions and earlier periods without altering the aggregate sample distribution.

3.~~3~~4 Hyperparameter optimization and validation

~~To efficiently explore the~~Independent hyperparameter ~~space and avoid ad hoc manual tuning, hyperparameters for each regional modeling unit were optimized independently~~optimization for each basin-depth unit is performed using Bayesian ~~optimization~~inference (Optuna~~; Table 2) with~~), targeting the objective of ~~minimizing~~ validation RMSE~~. We employed a minimization (Table 2). This automated search is integrated with a decadal-block~~ five-fold cross-validation scheme ~~grouped by calendar year, wherein the observational record was divided into five~~ to address the challenges of non-~~overlapping temporal blocks. Each block serves as a validation fold (~20%), ensuring that data from the same year never appears in both training and validation splits. This design prevents temporal information leakage and avoids inflated performance estimates that can arise from autocorrelated~~stationary ocean signals. By grouping observations ~~within narrow time windows~~into multi-year blocks, we decouple validation results from the short-term temporal dependencies that often inflate predictive skill in traditional random-split CV (Salazar et al., 2022). ~~During~~This structural separation ensures that the model learns large-scale climatic drivers rather than localized temporal artifacts. In each cross-validation fold, ~~an~~ early stopping ~~halts training~~is

设置了格式: 字体: 非倾斜

设置了格式: 字体: 非倾斜

activated if validation RMSE fails to improve for 50 consecutive iterations ~~and determines the optimal boosting rounds. The final model used the median of~~ , after which the optimal iteration ~~counts~~count is recorded. To secure a robust final model, we set the terminal iteration count to the median of these recorded values across all folds. ~~Early~~This iteration-locked retraining on the complete calibration set—with early stopping ~~was then~~ disabled~~, and the model was retrained on the full training set prior to reconstruction~~—prevents overfitting and ensures a stable convergence state.

~~To provide a final,~~ For an unbiased final assessment ~~of model performance, we construct an~~, a strictly independent ~~out-of-time~~global test set ~~by randomly selecting five calendar years—1964, 1972, 1985, 2012, and 2019 (without replacement)—and withholding them~~is constructed via decade-stratified random sampling (1960–2024), where one full calendar year per decade is entirely excluded from ~~training~~feature selection and hyperparameter ~~optimization. Generalization capability was evaluated on this holdout~~tuning. The resultant test years (1961, 1970, 1984, 1993, 2003, 2012, and 2020) provide a temporally representative benchmark. This withheld-year test set is reserved exclusively for final performance assessment and inter-product benchmarking; all model selection and calibration are conducted via decade-block cross-validation within the remaining training data. Integrated predictions from regional sub-models are then evaluated against this set using RMSE, mean absolute error (MAE), and the coefficient of determination ~~(R^2), providing robust and interpretable metrics of predictive accuracy and strengthening confidence in the reconstruction. These five withheld years are used in Sect. 4.1, 4.2 and 4.5 for independent evaluation and product intercomparison.~~(R^2).

Table 2. CatBoost hyperparameters and ~~their~~ Optuna ~~prior~~ search ~~spaces~~priors.

Hyperparameter	Explanation	Search range
iterations	Maximum boosting rounds	~~100–2000~~200–2500
learning_rate	Learning rate (shrinkage)	0.02–0.~~10~~08
depth	Tree depth	~~4–8~~6–10
l2_leaf_reg	L2 regularization on leaf values	~~2–10~~5–20
bagging_temperature	Temperature for Bayesian bootstrap	0.~~05–~~1.–0.6

4 Results ~~and Discussion~~

4.1 Variable Associations and Feature ~~Importance~~Contributions

The relationship between DO and its environmental drivers undergoes a transformation from the surface to the deep ocean (Fig. 4). In the near-surface layers, thermodynamic solubility constraints dominate, manifesting as a robust negative correlation with temperature. As depth increases, this direct solubility control attenuates, allowing salinity to emerge as the primary

integrative proxy for water-mass properties and ventilation history, particularly tracking the signatures of deep-water formation and transport. These depth-dependent patterns provide the empirical foundation for resolving complex non-linear interactions within the water column (Ping et al., 2024; Cao et al., 2024).

设置了格式: 图案: 清除 (白色)

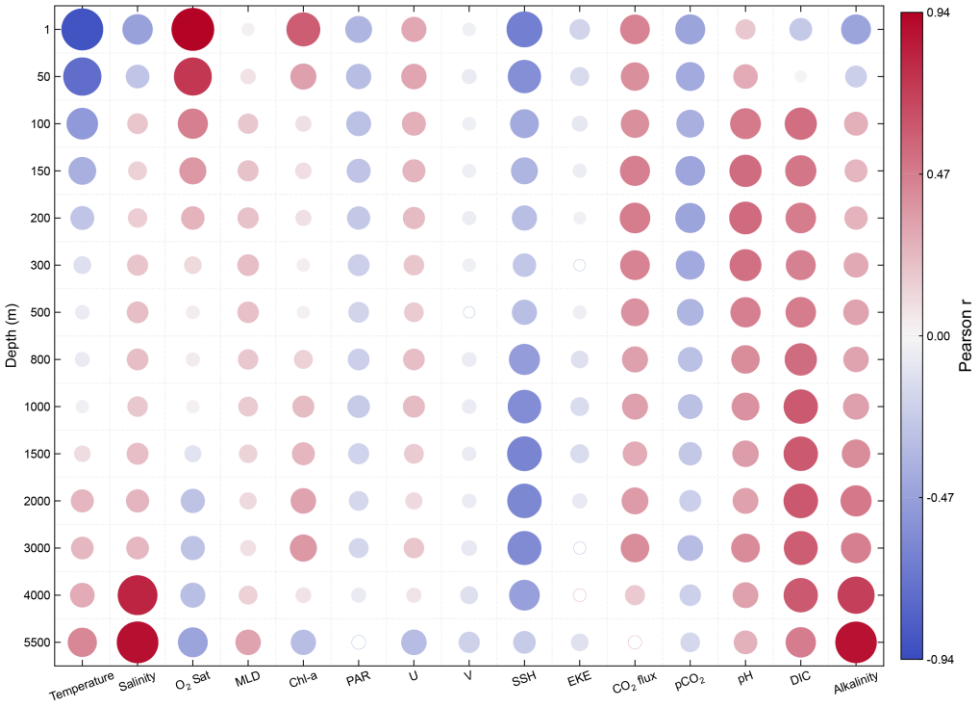
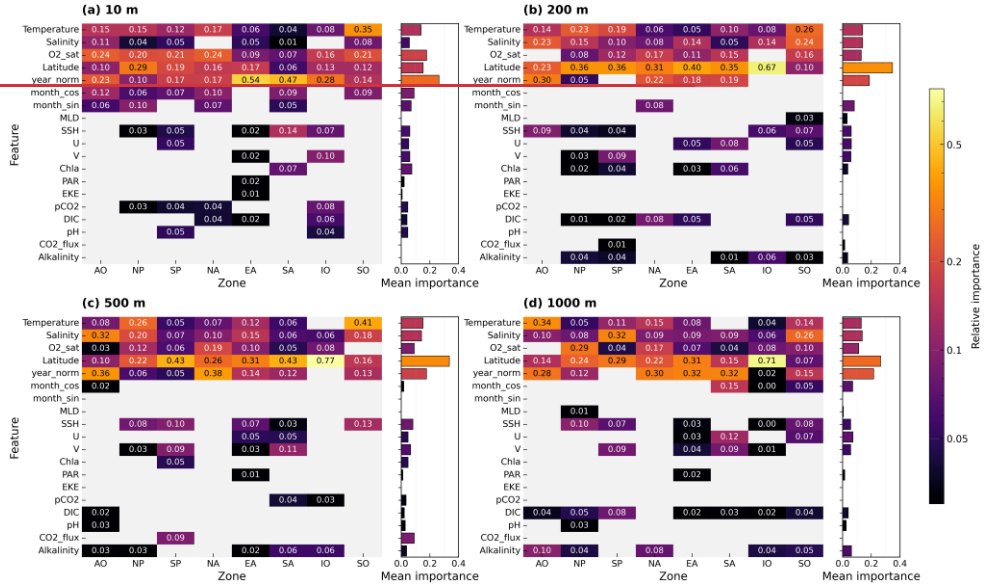


Figure 4: Vertical correlations between DO and physical–biogeochemical variables. Bubble color encodes the sign and magnitude (red = positive; blue = negative), and bubble area scales with (|r|). ~~Under the adaptive feature-selection framework, each biogeographic region develops depth-specific optimal predictor sets (Fig. 4). Across all region–depth submodels, the final feature sets typically contain 5–10 predictors. At 10 m, year_norm is the dominant control. Among sea-surface environmental variables (SSEVs), sea surface height (SSH) and partial pressure of CO_2 (pCO_2) consistently enhance model performance in multiple regions. In the Indian Ocean, for instance, SSH and meridional wind velocity (V) account for substantial DO variance—a pattern consistent with wind-driven transport and upwelling processes that modulate surface oxygen concentrations. Similarly, surface pCO_2 patterns have been shown to closely reflect the upwelling of carbon-rich, oxygen-poor waters (Franco et al., 2014), underscoring its utility as a predictor of DO variability in upwelling and OMZ regions.~~

~~At 200 and 1000 m, latitude replaces year_norm as the most influential variable, particularly in the Indian Ocean. This result~~

~~suggests that latitude serves as an effective spatial proxy, capturing broad-scale oxygen gradients and the influence of unobserved physical or biogeochemical drivers. Under this dominant latitudinal effect, the marginal contributions of most SSEVs are generally subdued, though not negligible (Milà et al., 2024). In summary, although SSEVs contribute little on average across depth levels, they remain important predictors in specific regions at certain depths. These results demonstrate that region and depth jointly shape the effective input space. Adaptive regional modeling preserves key controls and removes redundancy, improving both accuracy and physical interpretability. Feature subsets are not universal across regions, supporting the regionalized approach's advantage and smaller bias on the independent test. Our feature selection provides mechanistic evidence for these findings.~~



Filled bubbles denote correlations significant at (q < 0.05) after Benjamini–Hochberg false-discovery-rate control; hollow bubbles indicate non-significant results.

Correlations between sea surface environmental variables (SSEVs) and subsurface DO represent teleconnected pathways rather than local drivers. Surface dynamical forcing and upper-ocean mixing modulate water-mass formation and reventilation, thereby constraining the DO budget at depth. Specifically, sea surface height (SSH) exhibits persistent vertical coherence across much of the water column, while wind stress (U, V) provides a dynamical context for advection and upwelling (Hollitzer et al., 2024). In contrast, the Chl-a signal remains confined to the epipelagic zone, reflecting its role as a productivity indicator. Beyond the euphotic zone, carbonate system variables, notably alkalinity, maintain high correlations with DO, delineating the remineralization background linked to deep-water aging.

Feature-importance diagnostics are reported to describe model dependence and do not imply causality. The adaptive feature-selection results show that the model's reliance on predictors varies strongly with depth and region. In the upper 10 m, temperature and O_2 saturation are consistently among the most informative predictors, whereas at intermediate depths (1000–2000 m) salinity tends to contribute more strongly in certain high-latitude regimes, including the Arctic and Southern Oceans (Fig. 5). This depth-dependent pattern motivates the use of depth-specific predictor sets to better represent distinct hydrographic contexts. In addition, several regionally relevant covariates (e.g., SSH in the Indian Ocean and DIC/alkalinity in low-oxygen environments) are retained more frequently by the selection procedure, indicating that they provide useful contextual information for prediction under specific regimes (Franco et al., 2014).

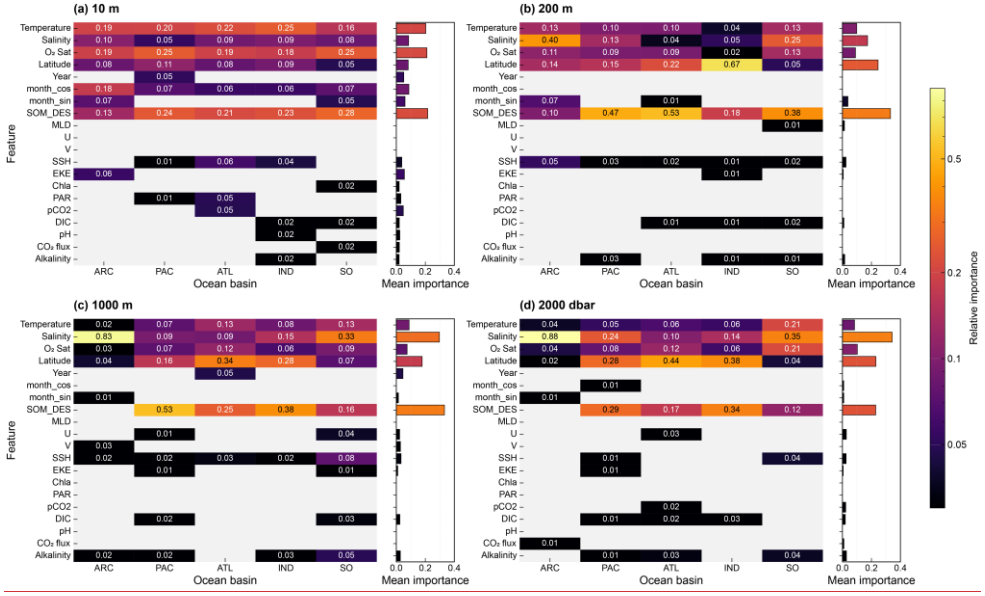


Figure 45: Heatmap of relative feature importance across depths and ~~provinces~~basins. Colors are on a logarithmic scale. The bar chart on the right shows each feature's mean importance computed over the basin provinces in which that feature is available.

Although latitude serves as a dominant proxy for broad-scale gradients (Milà et al., 2024), specialized SSEVs are vital for refining local accuracy. Our regionalized architecture prioritizes these idiosyncratic dynamics, utilizing variables like *SOM_DES* to resolve high-frequency seasonal variations in the Southern Ocean. This adaptive approach mitigates the biases inherent in spatially stationary parameterizations, ensuring the reconstruction respects the intrinsic heterogeneity of the global oxygen cycle.

4.2 Model evaluation

Vertical stratification of error profiles reveals a coherent link between RMSE and the intensity of the oxycline (Fig. 6). In most

basins, uncertainty is concentrated in the upper 600 m, where biological consumption and physical mixing generate high spatiotemporal variance. The Indian Ocean (IND) presents the most significant upper-ocean RMSE peak, likely due to its unique biogeochemical configuration, whereas the Arctic Ocean (ARC) shows fluctuations in RMSE at intermediate depths, but its R^2 remains relatively high. This vertical decoupling of error—where water-mass stability increases with depth before decreasing—represents a common feature across the Atlantic, Pacific, IND, and Southern Oceans, further confirming the appropriateness of the regional partitioning approach.

Except for the sparsely sampled Arctic, all ocean basins maintain high R^2 values (typically > 0.8) across the sampled water column. While absolute errors are inherently higher in the upper ocean where DO variability is maximized, the high R^2 values indicate that the underlying spatial and temporal patterns are accurately recovered. The analysis demonstrates that uncertainty is largely a function of the vertical DO structure, with the deep-ocean regime providing a stable and highly predictable anchor for long-term deoxygenation trends.

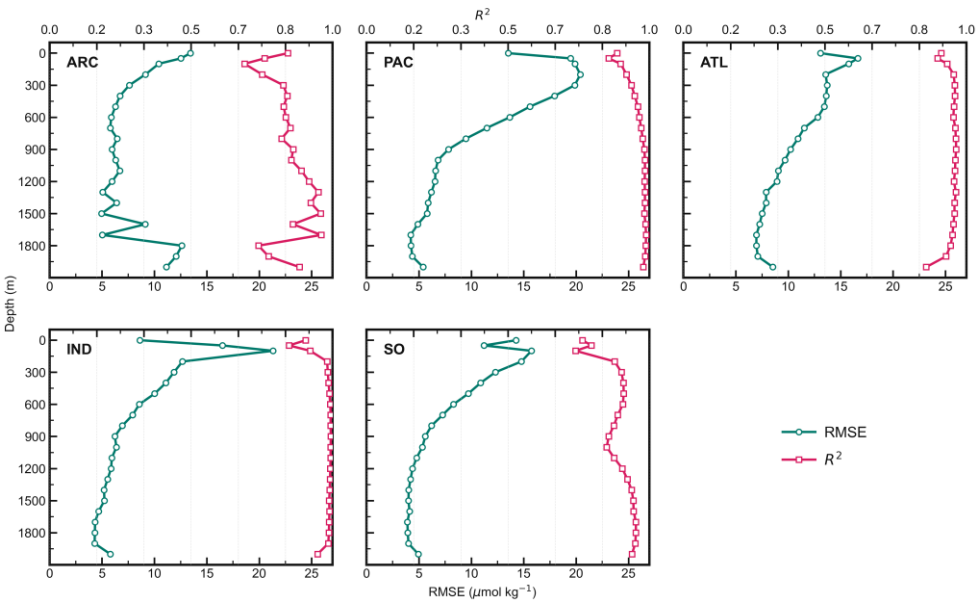
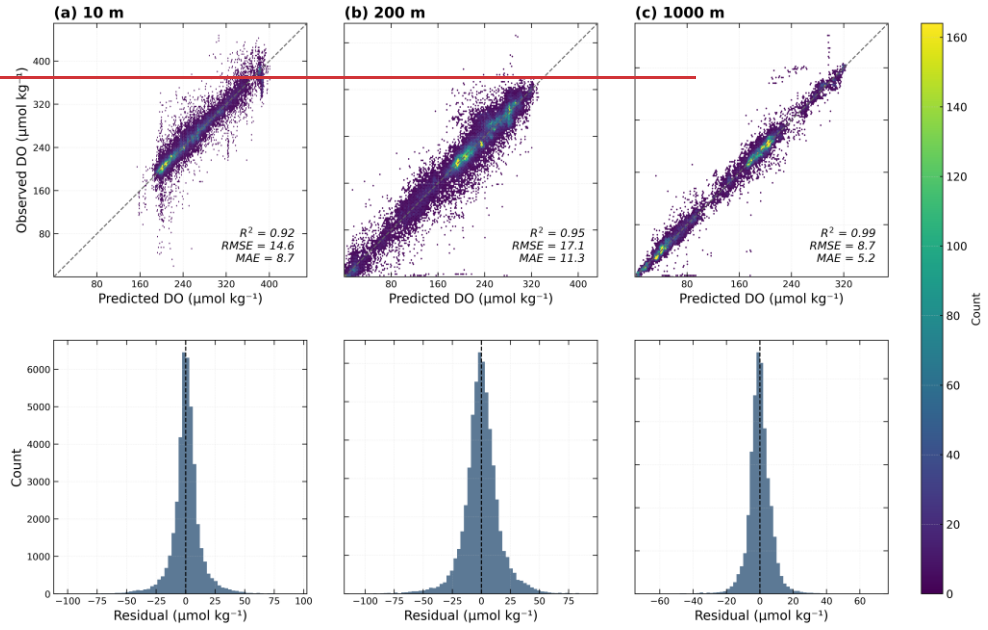


Figure 6: Depth profiles of model performance across basin provinces.

The reconstructed ~~dissolved oxygen (DO)~~DO fields demonstrate high accuracy ~~across depth layers and biogeographic regions. Overall model performance was evaluated by aggregating predictions from all regions and comparing them with observed values (Fig. 5). The model demonstrates high accuracy in estimating DO concentrations, with the~~ throughout the global ocean, with scatter ~~plot closely following~~plots confirming that estimates follow the 1:1 line. across the entire concentration range (Fig. 7). At representative depths, the global RMSE/MAE on the independent test set is on the order of ~~14.6/8.7 μmol~~12.8/8.0 μmol

kg⁻¹ at 10 m, ~~17.1~~16.8/11.~~3 μmol~~4 μmol kg⁻¹ at 200 m, and ~~8.~~7.5/5.2 ~~μmol~~μmol kg⁻¹ at 1000 m. Most R^2 values are above 0.~~92~~9, and the R^2 for deep-layer DO reconstruction reaches 0.99. A slight ~~positive bias is noted under severely hypoxic conditions, where extremely~~overestimation in regions with relatively low ~~DO values are marginally overestimated—a known behavior~~values reflects the inherent smoothing characteristic of regularized ~~tree ensembles that~~ decision trees, which tend to ~~smooth outliers~~pull extreme values toward local means. ~~These cases are rare, confined to strongly hypoxic conditions, and have negligible impact on~~However, the symmetric distribution of residuals around zero indicates that these effects are localized and do not introduce a systematic large-scale ~~statistics. It is also worth noting that some of the lowest recorded DO values may reflect measurement uncertainty. Residuals are generally symmetric and centered near zero, indicating that the model produces unbiased estimates with only occasional negative outliers~~bias. This performance confirms that the modeling architecture is well-suited for resolving the non-linearities inherent in ocean oxygen dynamics.



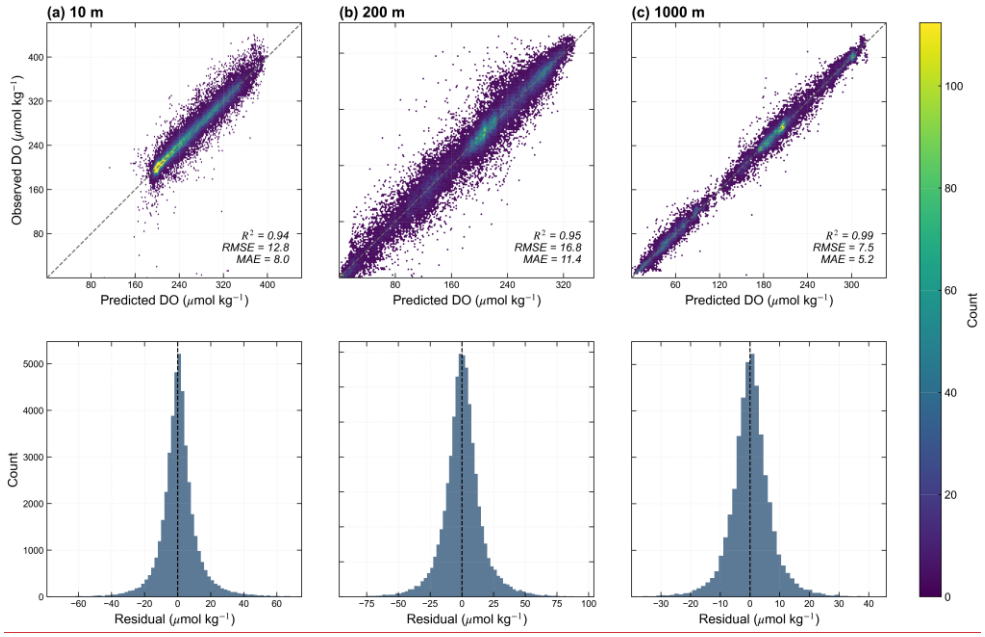


Figure ~~5~~7: Model performance for DO predictions across depth layers (independent test set). Top row: hexagon-binned scatterplots of predicted vs. observed values; the gray dashed line denotes the 1:1 reference. The color bar indicates sample counts per hexbin. ~~Each panel reports the R^2, RMSE, and MAE.~~ Bottom row: corresponding histograms of residuals (observed − predicted).

~~Vertical error structure exhibits consistent stratification across regions (Fig. 6): RMSE is highest in the surface and thermocline layers, decreasing with depth, consistent with a strong vertical gradient in the thermocline and the relative stability of intermediate to deep water masses. This pattern underscores the role of water mass stratification in shaping the estimation uncertainty of DO. Regional differences are primarily reflected in the position and amplitude of mid-layer peaks. Regarding error profile patterns, stable regions (e.g., SP_IO, NA_SA) show consistently high R^2, except at the surface, with RMSE monotonically or nearly monotonically decreasing with depth. In mid-layer-sensitive regions (e.g., NP), an RMSE peak is observed around 100–600 m, although R^2 remains relatively high. Lower surface R^2 values in the South Pacific, Equatorial Atlantic, and Indian Ocean arise primarily from limited DO variance in those layers rather than increased absolute error.~~

~~Deep layers, constrained by water masses, show a consistent pattern of low errors and high correlations. In the thermocline and OMZs, water properties and boundary positions are more sensitive to rapid seasonal and mesoscale–submesoscale fluctuations (Bettencourt et al., 2015). Together with limitations in historical coverage and resolution, this increases reconstruction uncertainty in these regions. Performance differences among regional models further confirm the pronounced spatial heterogeneity of DO. In summary, the regionalized modeling framework generalizes effectively, capturing the large-~~

scale DO distribution with high overall skill. Errors are primarily concentrated in regions and depths where oxygen gradients are strongest, aligning with expected patterns of physical and biogeochemical complexity.

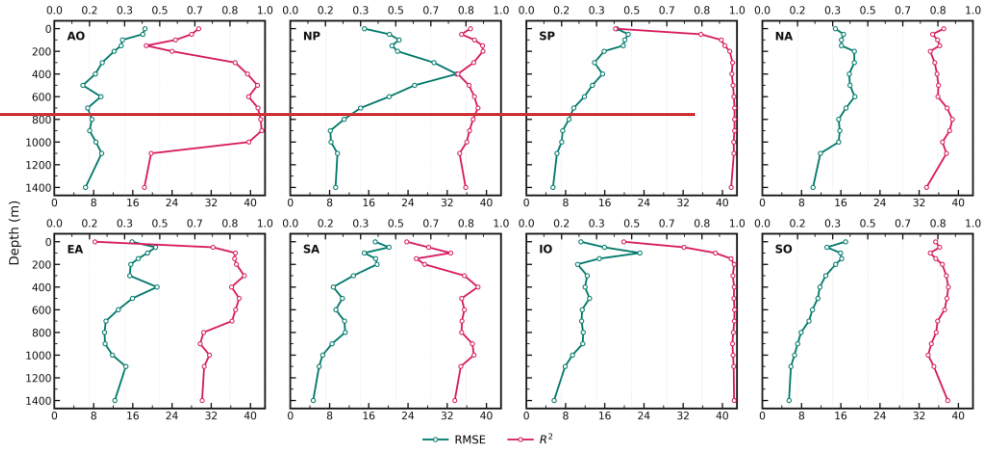


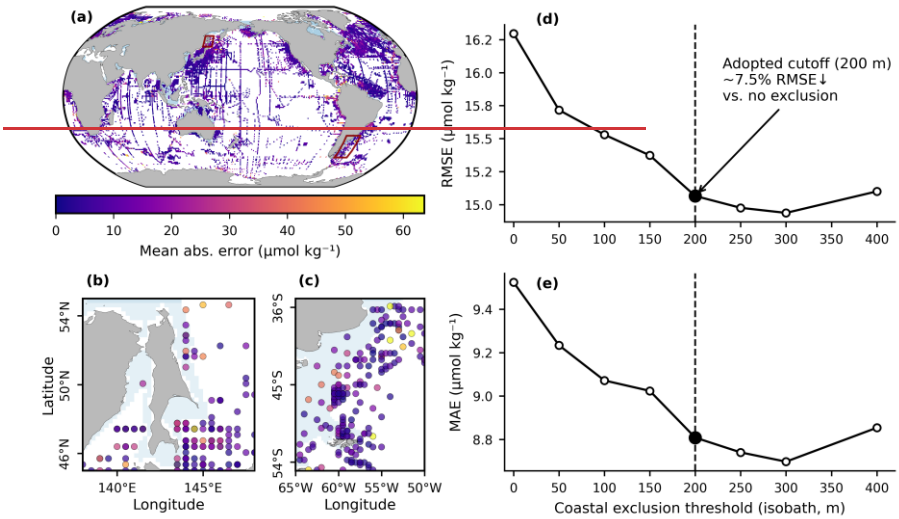
Figure 6: Depth profiles of model performance across provinces.

4.3 Coastal uncertainty

To assess the influence of nearshore regions on basin-wide reconstruction accuracy, we performed a depth-threshold sensitivity analysis. This involved progressively excluding coastal and marginal seas by increasing the bathymetric cutoff from 0 to 400 m, then evaluating model performance over the remaining open-ocean basin (Fig. 7). The results indicate that nearshore shallow waters are a major source of uncertainty. Increasing the threshold from 0 to 200 m yields a marked reduction in domain-mean error (**7.5%** relative to no exclusion; Fig. 7d–e). Further increasing the exclusion threshold to 250–300 m yields only marginal improvements, while a 400 m cutoff increases error—likely due to the loss of representative slope waters and a substantial decline in training sample size. In practice, the 200 m isobath offers a pragmatic coastal-exclusion standard when nearshore processes such as terrestrial inputs, eutrophication, fine-scale stratification, and benthic–pelagic coupling are not fully captured by our predictor set.

These findings are consistent with the physical and ecological characteristics of nearshore environments. Nearshore oxygen variability is driven by high-frequency, localized processes including phytoplankton bloom pulses, riverine discharge, tidal mixing, and anthropogenic effects, leading to strong diel and seasonal fluctuations that are poorly represented in existing open-ocean focused predictors (Gilbert et al., 2010; Regier et al., 2023; Giomi et al., 2023; Liu et al., 2024). Such nonlinear, rapidly varying processes produce error statistics that differ systematically from the open ocean. As a result, the model has limited ability to generalize features across regions (Valera et al., 2020). Statistically, shallow shelf systems are not representative of the large-scale oceanic DO background; their inclusion during model training and evaluation artificially elevates domain-

~~integrated error and obscures broad-scale oxygen patterns. Consequently, GEOXYGEN provides DO fields only for open-ocean grid cells deeper than 200 m, with shallower coastal grid cells flagged as missing values.~~



~~**Figure 7: Mean absolute error of the test set at the ocean surface (panel a). Blue shading indicates shallow coastal and marginal sea regions shallower than 200 m, and red boxes mark two representative high-error coastal areas. Panels (b) and (c) show zoomed views of the Sea of Okhotsk shelf and the Patagonian Shelf, respectively. Panels (d) and (e) show RMSE and MAE as functions of the coastal exclusion threshold, quantifying how model performance changes with the chosen depth cutoff.**~~

~~4.4~~ Long-term dataset

The trained models~~, forced by external physical~~ leverage CORA ocean analysis ~~fields,~~from CMEMS to generate spatially and temporally consistent monthly dissolved oxygen (DO) reconstructions. ~~We use the CORA ocean analysis from CMEMS to provide~~By utilizing temperature and salinity~~, together with the SSEVs, as large-scale physical and biogeochemical predictors on the target grid. Using these objectively analyzed gridded fields instead of interpolating from~~ grid product, the framework obviates interpolation from sparse in situ profiles ~~avoids an extra interpolation step and the~~and minimizes associated ~~secondary~~ uncertainties. ~~At each standard depth, the corresponding feature fields are evaluated by the matching regional model to generate monthly DO predictions, yielding 0.5° × 0.5° fields from the surface to 5500 m that are aligned with the CMEMS grid and calendar, but with coastal waters shallower than the 200 m isobath masked out. Because CORA analyses~~ These predictors undergo delayed-mode objective mapping and rigorous quality control~~, they provide~~ to ensure long-term consistency~~, global continuity~~ and full traceability. ~~Compared to methods that interpolate sparse in situ temperature and salinity observations, this approach reduces uncertainties associated with secondary statistical interpolation~~Monthly DO predictions are generated at

555 176 standard depth levels on a 0.5° × 0.5° grid extending to 5500 m.

To mitigate ~~potential discontinuities at the~~discontinuity artifacts at basin province boundaries ~~between biogeochemical provinces often termed a~~ described as the "step-effect" ~~we implement~~a boundary fusion protocol is implemented within ~~the~~ transition zones (Wagstaff and Bean, 2022). ~~Let $\mathbf{x}$ be the~~ For a prediction location, ~~$\hat{y}_i(\mathbf{x})$ $\mathbf{x}$,~~ the ~~prediction of~~fused estimate is derived from the ~~i~~ regional model ~~at $\mathbf{x}$,~~prediction $\hat{y}_i(\mathbf{x})$ and ~~d_i~~ the minimum great-circle distance ~~from $\mathbf{x}i$~~ to
560 the provincial boundary ~~of that region. Let the~~, as defined in Eq. (9). With a smoothing bandwidth ~~be defined as~~ of $S = 300$ km~~. The fused estimate is then defined as:~~, predictions from adjacent basins are smoothly blended in proportion to their distance from the edge. This approach preserves continuity across boundaries without excessive smoothing. Far from boundaries ($d > S$), only the local provincial model contributes to the estimate.

$$\hat{y}'(\boldsymbol{x}) = \frac{\sum_i w(d_i|S)\hat{y}_i(\boldsymbol{x})}{\sum_i w(d_i|S)}, \qquad w(d_i \mid S) = \begin{cases} \left(\frac{(S-d_i)}{S}\right)^2, & d_i \le S \\ 0, & d_i > S \end{cases}, \qquad \text{~~(6~~ (9)}$$

565 ~~Thus, far from a boundary ($d > S$), only the local provincial model contributes; within S of a boundary, predictions from adjacent provinces are blended smoothly in proportion to their distance from the edge. This preserves continuity across boundaries without excessive smoothing.~~

Finally, we obtained GEOXYGEN, a global DO dataset that provides monthly fields on a 0.5° × 0.5° latitude–longitude grid at ~~187~~176 standard depth levels from ~~the surface~~1 to 5500 m, spanning 1960–2024. ~~GEOXYGEN is distributed as~~The latitude
570 grid is not strictly uniform, with higher density in the high-latitude regions. Distributed via CF-compliant NetCDF files, ~~with one file per month~~each monthly volume (GEOXYGEN_DO_YYYYMM_0p5deg_v1.nc) ~~containing a~~includes the four-dimensional DO variable (time × depth × lat × lon) and ~~the corresponding coordinate variables (time, depth, latitude, longitude).~~associated coordinates. Time ~~is~~represents monthly means encoded as days since 1950-01-01 00:00:00 UTC ~~using a~~ (Gregorian calendar ~~and represents monthly means, and~~), with missing ~~values are flagged with~~ data identified by a large
575 sentinel value. ~~An accompanying biogeochemical province mask, which also serves as a valid-ocean mask (open-ocean grid cells deeper than 200 m), is provided in a~~A separate NetCDF file~~. All data files are openly~~ provides the basin provinces and valid-ocean masks. The complete dataset is available at ~~https://doi.org/10.5281/zenodo.17615657~~ https://doi.org/10.12157/IOCAS.20260223.002 (Wang et al., ~~2025~~2026).

4.~~5~~4 Comparison with other products

580 ~~We compare~~To evaluate the predictive fidelity of GEOXYGEN against four existing dissolved oxygen (DO) products (Table 3~~) with~~), we utilize the independent test dataset. ~~Performance is evaluated using spatiotemporally co-located samples from the five withheld years, analyzed separately for~~ stratified across discrete temporal windows: the early period (1960–1980) and

设置了格式: 字体颜色: 蓝色

设置了格式: 字体: 倾斜

设置了格式: 字体: 非倾斜

the recent epoch (2000–2020) periods (Table 4). Only ; Table 4). The intercomparison is strictly constrained to spatiotemporally co-located samples, ensuring that only grid-month-depth pointsintersections with concurrent observationsavailability across all products are included in the evaluation. This rigorous filtering preserves statistical equity and minimizes biases arising from disparate product masks.

Table 3. Summary of our product and other DO products

Product	Time coverage	Vertical levels	Temporal resolution	Horizontal resolution
Our product (GEOXYGEN))	1960-/01 – 2024-/06	01–5500 m (187176 levels)	Monthly	0.5° × 0.5°
ML4O2	1965-/01 – 2020-/12	6–1000 m (20 levels)	Monthly	1° × 1°
GOBAI-O_2	2004-/01 – 2023-/12	2.5–1975 m (58 levels)	Monthly	1° × 1°
G4D-DOC	2005-/01 – 2022-/12	10–1995 m (26 levels)	Monthly	1° × 1°
IAP Oxygen (Gouretski et al., 2024a)	1960-/01 – 2022-/12	01–6000 m (119 levels)	Monthly	1° × 1°

Table 4. Accuracy by depth for each product relative to observations

Product	Depth	1960-1980			2000-2020		
		RMSE	Bias	R²R²	RMSE	Bias	R²R²
Our product (GEOXYGEN)	10	**1512.2**	**0.13**	**0.9295**	**8.26.9**	**-0.54**	**0.9698**
IAP Oxygen		17.113.2	10.1	0.994	97.5	10.3	0.9497
ML4O2		17.313.4	0.71.5	0.8994	10.58.1	-1.7	0.9397
GOBAI-O_2		NA	NA	NA	107.5	1.1-0.4	0.9397
G4D-DOC		NA	NA	NA	7.43	0.2-1.3	0.9698
Our product (GEOXYGEN)	50	**2016.7**	**-0.31**	**0.993**	**13.512**	**-0.1**	**0.9495**
IAP Oxygen		2118.1	1.4	0.8992	13.817.2	2.50.9	0.994
ML4O2		2117.3	01.8	0.8993	16.113.4	0-1.6	0.9194
GOBAI-O_2		NA	NA	NA	1612.9	2.10	0.995
G4D-DOC		NA	NA	NA	13.6	0-1.2	0.9394
Our product (GEOXYGEN)	100	**20.317.4**	**-0.2**	**0.9395**	**15.314.4**	**0.1**	**0.9596**
IAP Oxygen		20.618.3	1.52	0.9394	19.417.2	1.80.9	0.9395
ML4O2		20.818.4	2.1	0.9394	17.218.6	-0.1	0.9395

设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标
格式化表格
设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标
设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 黑色
格式化表格
设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标
设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 黑色
设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标

GOBAI-O_2		NA	NA	NA	1815.6	0.71	0.9496
G4D-DOC		NA	NA	NA	1820.6	-0.7	0.9492
Our product (GEOXYGEN)		**1917.3**	**-0.14**	**0.9396**	**12.78**	**0.2**	**0.9798**
IAP Oxygen	200	20.217.6	-0.3	0.9296	16.914.8	10.4	0.9597
ML4O2		18.619.8	-0.1	0.9395	2016.2	-1.76	0.9496
GOBAI-O_2		NA	NA	NA	2115.5	-1.60.9	0.9397
G4D-DOC		NA	NA	NA	16.418.3	10.8	0.9695
Our product (GEOXYGEN)		**21.215.1**	**-0.3.6**	**0.9397**	**14.210.9**	**-0.3**	**0.9798**
IAP Oxygen	500	16.118.6	-2.90.6	0.9497	23.513	-2.20.1	0.9298
ML4O2		1916.6	-6.10.5	0.9497	2614.1	-41.7	0.9197
GOBAI-O_2		NA	NA	NA	28.413.9	-5.31.2	0.997
G4D-DOC		NA	NA	NA	16.112.7	0.13	0.9698
Our product (GEOXYGEN)		**13.17.4**	**-0.91**	**0.9799**	**6.5.2**	**0.2**	**0.991**
IAP Oxygen	1000	14.57.4	1.60	0.9699	8.26.3	30.4	0.99
ML4O2		14.68.7	-01.3	0.9699	8.17.6	0.5-2.2	0.99
GOBAI-O_2		NA	NA	NA	5.7.1	1.4-0.8	0.99
G4D-DOC		NA	NA	NA	711.7	-0.53.8	0.9998

The ~~comparison highlights~~comparative analysis reveals three ~~main features.~~dominant patterns. First, all products manifest elevated error magnitudes during the ~~data-sparse~~ early period, a trend predominantly attributable to the historical sampling architecture. Pre-Argo observations were characterized by a disproportionate concentration of coastal and shelf cruises, whereas the modern era is defined by the proliferation of autonomous Argo floats providing expansive open-ocean coverage. Because coastal environments are governed by pronounced small-scale heterogeneity and high-frequency nonstationarity, their reconstruction poses a greater challenge, inherently raising the error floor for the 1960–1980 interval.

Within this data-sparse window (1960–1980), GEOXYGEN ~~matches or exceeds the performance of IAP and ML4O$_2$ at~~ manifests superior predictive stability across the upper and intermediate layers (10~~–200 m depth, exhibiting~~ 500 m). It consistently achieves lower RMSE and ~~near-zero mean bias, and achieves the best performance at 1000 m. Although errors are marginally higher at 500 m, the high~~MAE than both IAP and ML4O2 across nearly all depth levels while maintaining robust R^2 ~~(>0.9) confirms~~values. These gains demonstrate that ~~large-scale oxygen structure is well reproduced~~the stratified, basin-specific modeling framework facilitates the recovery of mesoscale spatial heterogeneities and vertical gradients even under ~~limited~~sparse observational constraints. ~~Second, in the recent,~~ By leveraging high-resolution target fields, this architecture enhances the representation of regional variability, ensuring that GEOXYGEN maintains competitive skill during the pre-satellite and pre-Argo eras.

In the data-rich modern period (2000 – 2020), GEOXYGEN ~~systematically reduces RMSE and bias at 50–200 m~~exhibits even

设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 黑色
设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标
设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 黑色
设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标
设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 黑色
设置了格式: 非上标/ 下标
设置了格式: 非上标/ 下标
设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 黑色
设置了格式: 字体颜色: 自动设置

more pronounced performance gains. Within the critical mesopelagic range (100、200 and 500 m), the product shows a significant accuracy improvement relative to all reference ~~products~~datasets, while ~~maintaining $R^2 \geq 0.94$.~~ensuring that residual biases are smaller and more stable. In ~~this depth range~~contrast, several existing products exhibit persistent ~~positive or negative biases, whereas GEOXYGEN remains close to unbiased, consistent with the benefits of regionalized modeling, inverse-density weighting, and the inclusion of physically informed predictors. Third, at 1000 m, all products perform well, reflecting the stability of deep water masses, yet GEOXYGEN ranks among the top performers, with RMSE values that are slightly lower or comparable to the best alternatives. Overall, the out-~~systematic biases in this depth range, potentially compromising the representation of ~~time, co-located comparison indicates that GEOXYGEN provides reconstruction skill that exceeds existing products, particularly at mid-depths where deoxygenation signals and model spread are most pronounced~~ OMZs and steep vertical gradients. This suite of high-resolution, broad-coverage DO fields provides a reliable foundation for the integrated assessment of global deoxygenation trends and their underlying physical-biogeochemical drivers.

~~These outcomes align with our methodological strategy: the use of a long observational archive (1960–2024) enables training across diverse climate states; CORA temperature–salinity fields, process-relevant SSEVs, regionally adapted models, adaptive feature selection, inverse-density weighting, and boundary fusion collectively reduce bias, most notably between 50–500 m. GEOXYGEN agrees with existing products at the surface and in the deep ocean and achieves systematically lower or comparable RMSE and bias at key mid-depths. Taken together, it provides a high-resolution, broad-coverage, and accurate DO dataset that supports more comprehensive analyses of deoxygenation trends and their driving mechanisms.~~

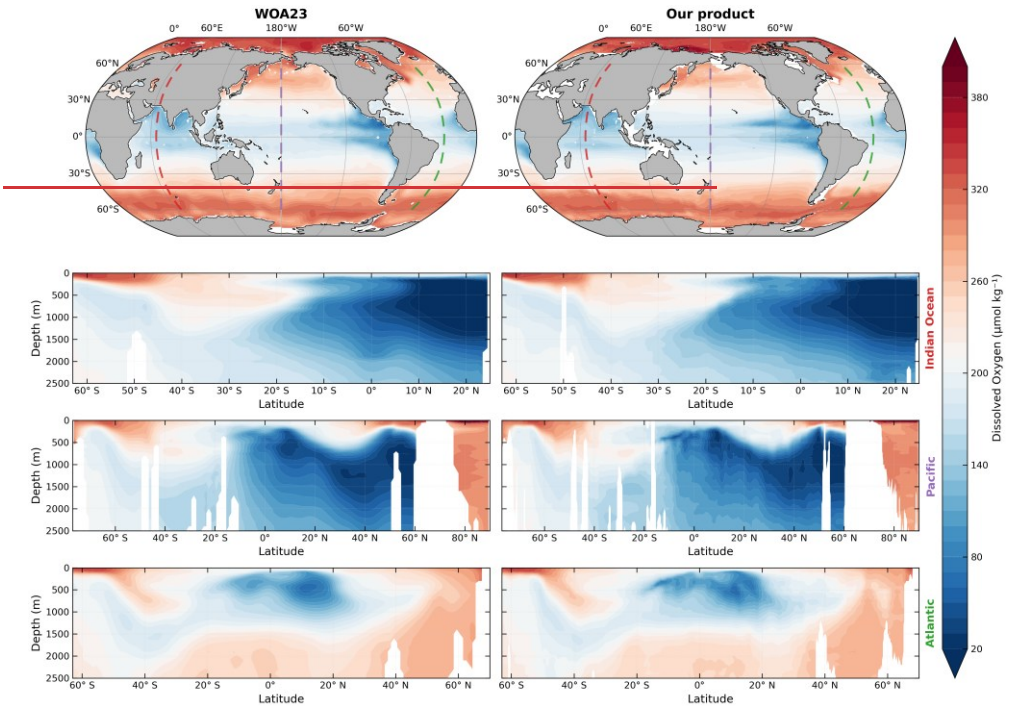
4.65 Comparison with WOA23

To evaluate the large-scale credibility and multi-year stability of our reconstructed dissolved oxygen (DO) fields, we compare our product's annual-mean DO climatology with ~~the World Ocean Atlas 2023 (WOA23; Garcia et al., 2024)~~WOA23 and examine vertical structure using representative profiles from three major basins (Fig. 8).

In the upper ocean (0–300 m, depth-averaged), both products capture consistent basin-scale spatial patterns. The subtropical gyres exhibit relatively high DO concentrations, while the equatorial region and eastern boundary upwelling systems form distinct oxygen-deficient belts. The spatial extent and location of these structures are in close agreement between the two climatologies. GEOXYGEN reproduces these banded structures continuously, with cross-frontal gradients in transition zones closely matching those in WOA23 ~~(Fig. 8, top row).~~

In the vertical, our product accurately depicts the largest hypoxic zone in climatology, which is mainly located in the intermediate depth, ranging from 300 to 1,300 m. Meridional sections ~~across the Pacific, Atlantic, and Indian Oceans show broad consistency between GEOXYGEN and WOA23 in the alignment and closure of key oxygen isopleths, as well as in the positioning of high-gradient transition layers. This structural agreement underscores the physical consistency of our product~~

~~across depth and region~~reveal that the model successfully reproduces the core depths and spatial extents of major hypoxic zones across the three primary ocean basins. The precise alignment of oxygen isopleths and the accurate representation of the transition layer steepness indicate that the reconstruction results are consistent with the climatological constraints of DO.



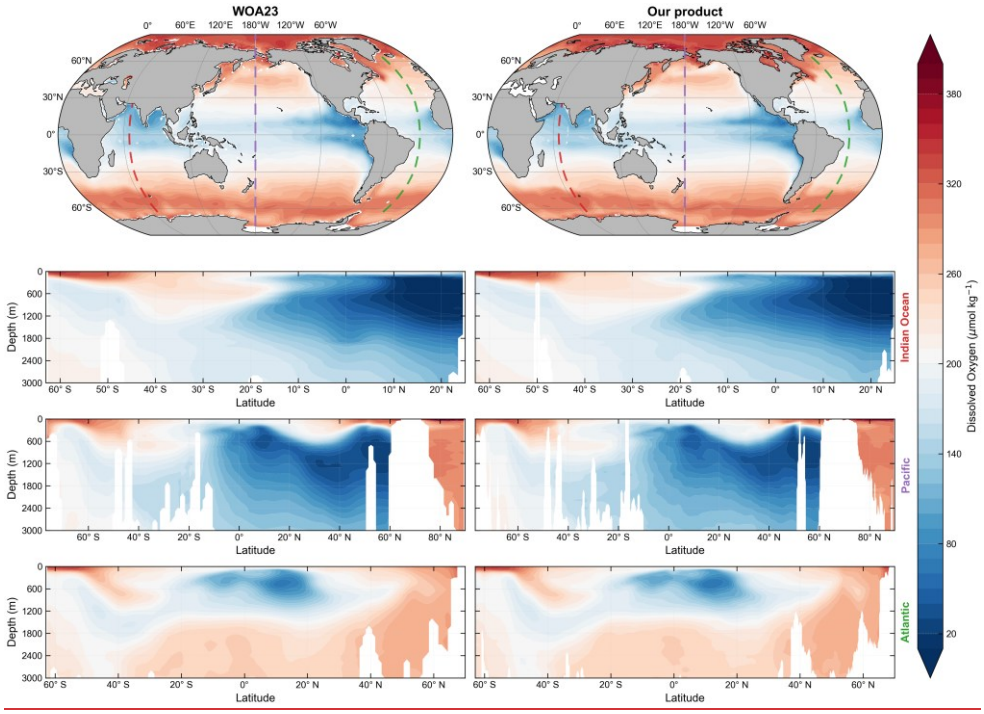


Figure 8: Climatological comparison between WOA23 and our product (GEOXYGEN). Row 1 shows the global-mean DO distribution averaged over 0–300 m. Colored dashed lines mark the locations of three sections: 65°E (red), 180° (purple), and 30°W (green). Rows 2–4 show DO cross-~~sectional DO~~ section distributions along these three ~~sections~~transects.

~~Most global DO products are distributed on a~~ Unlike the standard 1° grid~~.~~, used in ~~contrast~~most global products, GEOXYGEN ~~adopts~~employs a higher 0.5° horizontal ~~grid, chosen to match the~~ resolution ~~of the external analysis fields used as predictors. Within each ecogeographic region, the model learns~~to capture more detailed features. By training on localized relationships between DO and ~~objectively analyzed physical and biogeochemical fields and then applies these relationships across the grid. Consequently, the reconstruction is not merely a reflection of where DO observations are dense; it also~~ hydrographic analyses within discrete basin provinces, the model resolves ~~regional gradients and delineates the geometry~~the morphometry of OMZs ~~more clearly.~~with enhanced clarity. Although ~~performance is ultimately bounded by~~ the reconstruction fidelity remains conditioned upon the quality of ~~the underlying~~input observations~~ and external analysis products, comparisons with~~, benchmarking against 1° products ~~show~~reveals that the 0.5° configuration yields ~~consistently~~systematically lower~~ or comparable~~ errors ~~at key depths and can accurately resolve regions of strong DO gradients. We therefore regard 0.5° as~~ in high-gradient regimes. This selection constitutes an ~~appropriate~~optimized operating resolution that ~~exploits available~~maximizes information recovery from the underlying biogeochemical constraints without over-interpreting sparsely

sampled ~~areas~~intervals.

5 Discussion

5.1 Uncertainty Analysis

设置了格式: 英语(英国)

To quantify the credibility of the GEOXYGEN, we decompose total uncertainty into two components: observation-related uncertainty and mapping uncertainty. Observation-related uncertainty summarizes measurement error and the representativeness error introduced when observations are aggregated to the cell–month–depth scale. Mapping uncertainty describes prediction error from the machine-learning mapping between environmental predictors and DO. This decomposition separates label-side and model-side contributions. It also supports diagnosing elevated uncertainty in coastal regions and potential bias in earlier, data-sparse periods, and it improves traceability and clarity for product use.

Measurement error depends on the measurement technique and instrument. For Winkler titration, bottle measurements can include random errors on the order of 1 μmol kg^{-1} or smaller (Carpenter, 1965). Since observing platforms differ in sensor type, calibration strategy, and QC level, we assign a platform-specific measurement error scale σ_{meas} to each data source to represent differences in precision among platforms. Bottle-based sources including CCHDO Bottle, GLODAP, and GEOTRACES IDP are assigned 1 μmol kg^{-1}. Since Argo data are bias corrected, Argo, OSD and CTD, and CCHDO CTD are assigned 1.5 μmol kg^{-1}. OceanSITES is assigned 2.0 μmol kg^{-1}.

When forming supervised-learning labels at the cell–month–depth scale, observation-related uncertainty is defined as the combined contribution of measurement error, vertical mapping uncertainty, and representativeness error, as defined in Eq. (10).

$$U_{\text{obs}} = \sqrt{\sigma_{\text{meas}}^2 + \sigma_{\text{interp}}^2 + \sigma_{\text{rep}}^2} \quad (10)$$

Here, σ_{interp} is the vertical mapping uncertainty introduced when a profile is mapped to standard levels, and σ_{rep} is the representativeness error estimated from within unit dispersion. Their definitions and computation have been described earlier. Mapping uncertainty characterizes error generated during the mapping from environmental fields to DO. It reflects the combined effects of model structure, representativeness of training samples, and nonstationarity across regions and depth levels. On the independent test set, we define the residual following Eq. (11):

$$r = \hat{y} - y, \quad (11)$$

where $\hat{y}$ is the model prediction and y is the observed label for the corresponding test sample. Following Ito et al. (2024a), we estimate the mapping uncertainty at each grid cell from the test residuals by taking their second central moment, as defined

in Eq. (12).

$$\sigma_{\text{map}} = \sqrt{\overline{r^2} - (\overline{r})^2}, \quad (12)$$

Here, the overbar denotes the mean over all test samples within the same grid cell. This definition estimates the residual variance at the grid cell scale. It separates the systematic component $\bar{r}$ from the random error component captured by σ_{map}.

Based on the two components above, total uncertainty (σ_{tot}) is defined in Eq. (13).

$$U_{\text{total}} = \sqrt{U_{\text{obs}}^2 + \sigma_{\text{map}}^2}, \quad (13)$$

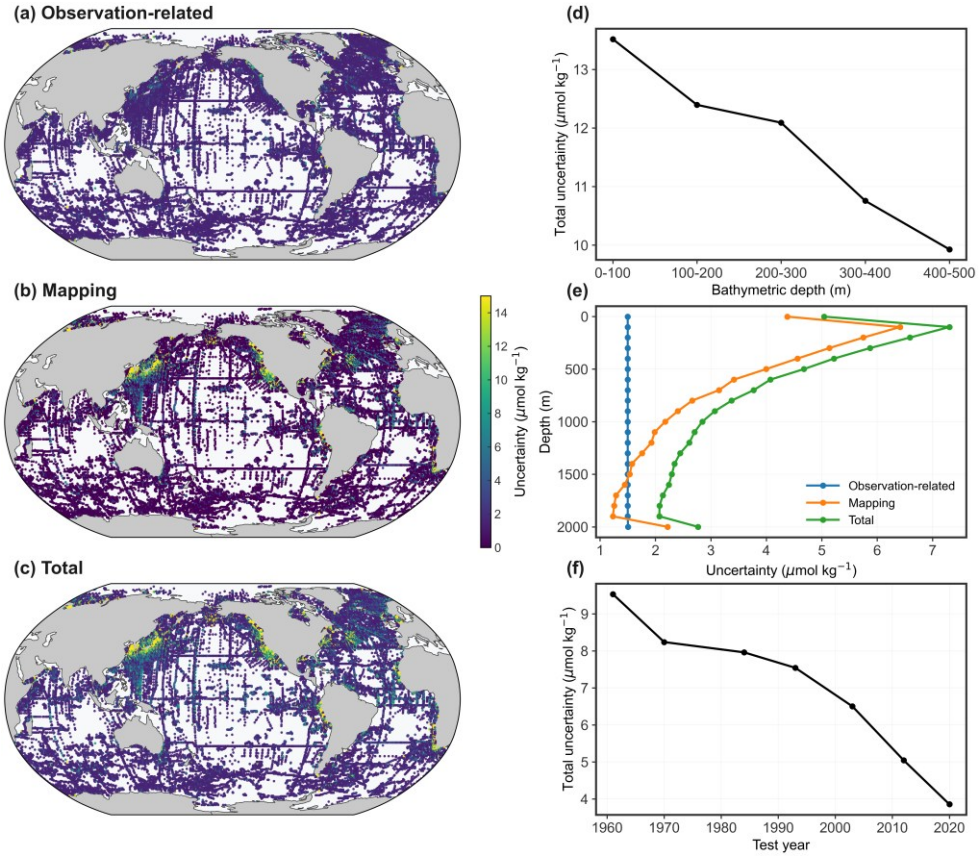


Figure 9: The left panels show the mean uncertainty distributions across multiple standard depth layers (0–2000 m), including observational uncertainty (a), mapping uncertainty (b), and total uncertainty (c). The right panels depict the nearshore spatial variability of total uncertainty (d), its vertical (depth-dependent) structure (e), and the decadal evolution of total uncertainty over the test years (f).

As shown in Fig. 9, the spatial distribution of uncertainty in GEOXYGEN reveals a distinct bifurcation between observational noise and mapping residuals. Observational uncertainty manifests pronounced spatial stationarity across pelagic domains. Conversely, mapping uncertainty exhibits a structured regional geometry, with elevated error bands aligning with high-gradient kinetic regimes such as western boundary currents and upwelling systems. The highest uncertainty is observed in the Pacific (7.385 μmol kg^{-1}), followed by the Atlantic (6.148 μmol kg^{-1}), Arctic (4.439 μmol kg^{-1}), Indian (4.084 μmol kg^{-1}), and Southern Ocean (3.652 μmol kg^{-1}). The higher uncertainty in the Pacific and Atlantic Oceans is primarily due to the structural complexity and dynamic intensity of their oceanographic systems, as well as the coastal distribution of early observational data in these basins. The global-mean total uncertainty of 6.054 μmol kg^{-1} conceals a pronounced shallow-water divergence: in nearshore and shelf regions where bathymetric depth is shallower than ~200 m, total uncertainty rises to 12.917 μmol kg^{-1}—more than double the open-ocean baseline (5.970 μmol kg^{-1}). Consistent with the bathymetry-binned diagnostic, uncertainty increases monotonically with decreasing bathymetric depth, indicating progressively reduced predictability toward the shallow, dynamically heterogeneous coastal ocean (Fig. 9(d)). This intensification is predominantly driven by localized, high-frequency processes—including phytoplankton pulses, riverine influx, and tidal oscillation—which generate non-linear spatial gradients that challenge the transferability of pelagic-trained feature relationships (Gilbert et al., 2010; Regier et al., 2023; Giomi et al., 2023; Liu et al., 2024). These localized dynamics induce steep spatial gradients and temporal non-stationarity, which subsequently reduce the regional transferability of learned DO-environment associations (Valera et al., 2020). For nearshore and semi-enclosed bay environments, we recommend using GEOXYGEN with caution and interpreting results at larger spatial aggregation to reduce sensitivity to local high-frequency variability.

The vertical stratification of uncertainty reflects the underlying hydrographic stability and process coupling within the water column. As depicted in Fig. 9(e), the total uncertainty profile reaches its maximum in the epipelagic layer before undergoing monotonic attenuation toward the abyssal depths. This vertical variation is consistent with the change in model accuracy across depth layers. In contrast, the intermediate and deep layers provide stronger water-mass constraints and a more coherent oxygen field, facilitating a convergence of mapping uncertainty as the predictive relationship stabilizes.

Temporal trends in uncertainty serve as a diagnostic of the evolving global observing system. The progressive reduction in total uncertainty across the withheld test years (Fig. 9(f)) coincides with the expansion of the Argo float network, which transitioned ocean sensing from route-based ship surveys to a spatially distributed autonomous paradigm. This transition significantly improved the observational constraints in the Southern Hemisphere and remote ocean basins, effectively lowering the residual variance in model validation. While this decline highlights the structural evolution of sampling coverage, the resulting time series reflects the collective stability of the multi-decadal reconstruction rather than a year-specific local error estimate.

5.2 Impact of Removing Surface Predictors on Trends

A two-configuration sensitivity experiment was designed to test whether satellite-era sea-surface information alters the reconstructed dissolved oxygen (DO) signal. The full predictor configuration used the complete predictor suite, while the reduced predictor configuration excluded the sea-surface predictor group (U, V, SSH, EKE, MLD, PAR, Chl-a, DIC, pCO_2, pH, alkalinity, and CO_2 flux). All other parts of the reconstruction pipeline were kept the same, so differences between the two products isolate the effect of including this predictor group on the same grid and over the same period.

The comparison focuses on deseasonalized DO anomalies averaged over the 1–100 m depth range and further smoothed using a centered 12-month moving average. This metric targets the upper ocean, where sea-surface information is most likely to influence variability, while suppressing grid-scale noise that can obscure low-frequency signals. Predictor impact is quantified by the configuration-induced component, defined as Full − Reduced, which isolates the incremental effect of the excluded predictor group in anomaly space.

The full and reduced reconstructions exhibit strong agreement over 1960–2022 (Fig. 10), yielding a consistent depiction of upper-ocean (1–100 m) low-frequency variability: a sustained positive-anomaly regime through the 1970s–1980s followed by a marked decline after ~1990 and a transition into a persistently negative-anomaly regime by ~2000 (relative to the monthly climatology). The Full–Reduced difference remains close to zero prior to the mid-1980s and stays small compared with the total anomaly amplitude thereafter, indicating that the global-mean, decadal-scale signal is only weakly sensitive to the inclusion of satellite-era surface predictors. As an external benchmark, ML4O2 reproduces comparable variability during ~1965–2010 but shows more negative anomalies in the most recent decade, suggesting a stronger upper-ocean deoxygenation signal in anomaly space (relative to its monthly climatology) than GEOXYGEN over the same period; however, part of this divergence may arise from inter-product differences in baseline climatology, sampling, and reconstruction methodology. Overall, the close concordance between configurations supports the robustness of GEOXYGEN for decadal-scale assessments, with configuration-dependent deviations being minor relative to the dominant multi-decadal evolution.

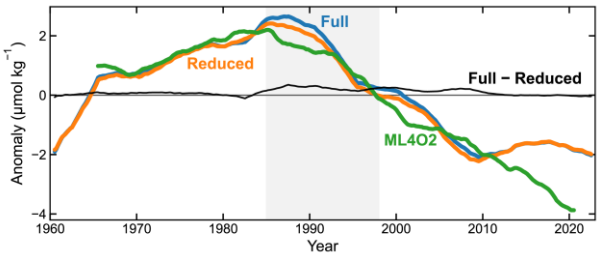


Figure 10: Depth-averaged (1–100 m) monthly dissolved-oxygen anomalies (1960–2022). The figure compares deseasonalized anomalies from the Full predictor reconstruction, the Reduced predictor reconstruction (excluding sea-surface predictors), and ML4O2, and overlays the difference between the Full and Reduced reconstructions (Full − Reduced). Gray shading indicates 1985–

1997, corresponding to the rapid expansion of satellite-derived sea-surface observations.

5.3 Ship-Only Analysis of Long-Term Trends

To quantify the dependency of the early reconstruction on modern autonomous sensing, a ship-only sensitivity experiment was conducted to assess potential retrospective signal contamination from the data-dense Argo era. Within the Southern Ocean (south of 45°S), we benchmarked an "Argo-excluded" configuration—utilizing non-Argo historical records—against an "Argo-included" configuration across the 1960–2000 period. This region serves as a critical diagnostic domain due to its radical transition from sparse historical sampling to comprehensive Argo coverage over the last two decades. By maintaining identical external physical analysis constraints, the experiment isolates the impact of Argo observations on the reconstructed climatological mean and vertical structure (Fig. 11).

The results manifest pronounced morphological invariance in the upper-ocean (1–100 m) DO climatology across both configurations. The annular high-oxygen band and its relative position within the circumpolar system remain effectively stationary, implying that Argo inclusion does not induce systematic structural rearrangement. Notably, the difference map reveals localized positive ΔDO hotspots concentrated around the Antarctic margin, with enhanced amplitudes in several coastal/shelf sectors, whereas departures in the open-ocean circumpolar interior are minimal and barely discernible. Consistent with this pattern, the area-weighted mean vertical profiles are nearly overlapping across most depths (panel d), with only a subtle upper-ocean positive offset in the Argo-included configuration that remains below ~0.5 μmol kg^{-1}, indicating modest amplitude refinement rather than depth-dependent reorganization. Collectively, these features suggest that the reconstructed vertical architecture is primarily constrained by consistent physical structure, while the denser modern sampling acts mainly to fine-tune regional magnitudes, thereby supporting the structural integrity of the historical reconstruction.

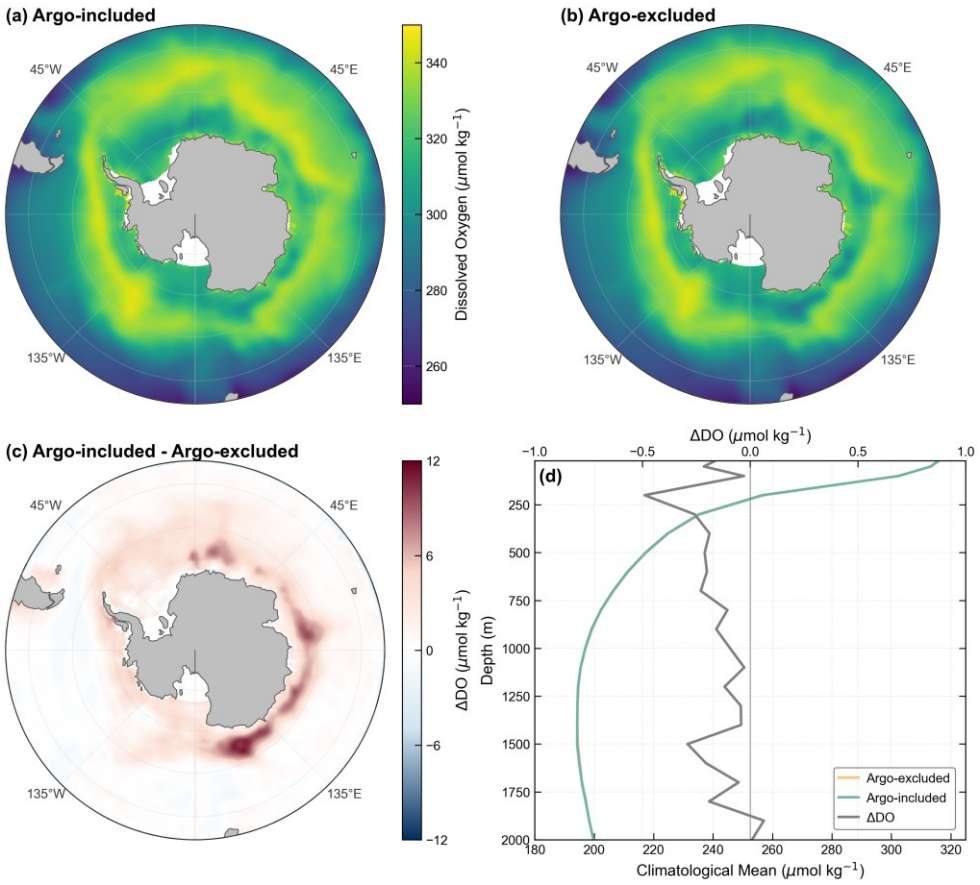


Figure 11: Southern Ocean dissolved oxygen (DO) climatology for 1960–2000. (a) Argo-included climatological mean DO averaged over 1–100 m. (b) Same as (a), but for the Argo-excluded configuration. (c) Difference map (Argo-included minus Argo-excluded) for the 1–100 m climatological mean DO. (d) Area-weighted mean vertical DO profiles south of 45°S from the Argo-included and Argo-excluded reconstructions; the upper x-axis shows the corresponding profile difference (Argo-included minus Argo-excluded).

5.4 Boundary Fusion Effects

While basin-based modeling represents regional environmental heterogeneity, it often induces "step-effect" discontinuities at basin boundaries, resulting in unrealistic shifts in long-term trends and monthly variability at adjacent grid points. To address these boundary artifacts, we implemented a fusion method to smooth inter-basin transitions. We validated this approach by constructing boundary diagnostic samples on a 0.5° by 0.5° grid, specifically selecting adjacent points across basin interfaces. The diagnostic experiment targeted the 100m depth layer, a region characterized by high spatial complexity and hydrographic sensitivity. Using a global non-partitioned model as a continuous baseline, we quantified discontinuity magnitudes via the

differential trend (Δtrend) and standard deviation (Δstd) at adjacent cells. Contrasting fused and unfused basin-specific outputs against this reference isolated the consistency gains.

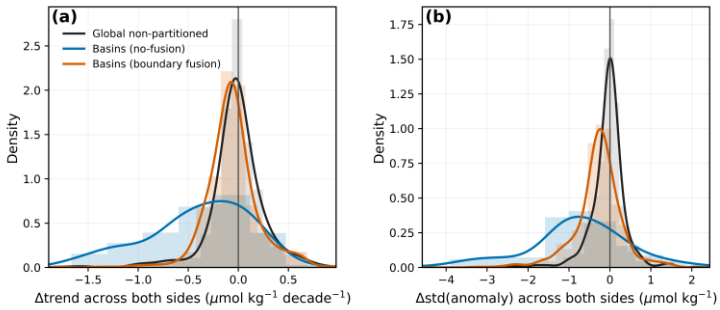


Figure 12: Spatial distribution of the "step-effect" before and after boundary fusion at the 100 m depth layer. (a) Compares the trend differences (Δtrend), and (b) compares the standard deviation differences (Δstd) across adjacent grid points at the basin boundaries.

Results reveal that the boundary fusion protocol mitigates abrupt statistical shifts, decreasing local biases in the partitioned map (Fig. 12). Improvements in Δtrend and Δstd demonstrate that the smoothing operator suppresses interface artifacts. Nevertheless, small discrepancies relative to the global non-partitioned baseline persist, indicating that independently trained regional submodels may retain slightly different statistical mappings that are only partially reduced by simple fusion.

Boundary fusion thus serves as an effective strategy for mitigating partition-induced "step-effect" artifacts, yielding a more coherent global spatial structure. While localized gradients remain in high-contrast hydrographic regions, their reduced magnitude indicates a meaningful alleviation of boundary inconsistencies. Future enhancements could involve adaptive fusion schemes or multi-task learning to explicitly share data across regional submodels while preserving specialization.

6 Conclusion

~~We introduce GEOXYGEN,~~ By integrating multi-source physical and biogeochemical predictors with an adaptive feature-selection strategy, we develop a biogeochemistry-aware hierarchical modeling framework. Building on this framework, we introduce GEOXYGEN—a monthly, four-dimensional global ocean dissolved oxygen (DO) product spanning 1960–2024 at 0.5° × 0.5° resolution~~, developed~~—designed to ~~address~~mitigate long-standing ~~challenges of data~~limitations arising from observation sparsity and strong spatiotemporal heterogeneity in historical DO ~~observations~~records. Evaluated on an independent out-of-time test composed of withheld years, the reconstruction demonstrates consistently high skill ~~across~~in all depth layers (~~typically~~ $R^2 > 0.9$~~2~~9), confirming its robustness under conservative validation protocols. The resulting high-resolution framework offers superior aggregate fidelity compared to existing 1° products, effectively capturing the fine-scale

spatial variance essential for understanding contemporary ocean change. Our approach is distinguished by three principal innovations:

~~In comparison with other existing global DO products, GEOXYGEN reproduces observed spatial variability and attains superior overall performance at mid-depths (50–200 m) and in the deep ocean while offering a higher spatial resolution (0.5° vs. 1°). Our approach is distinguished by three principal innovations:~~

- **Heterogeneity-~~aware hierarchical~~Based Partitioning modeling.** By combining vertical stratification with ~~biogeographic~~basin provincialization, we train CatBoost regressors within each ~~region~~basin–depth unit. This design directly addresses the limitations of single global ML models, which struggle to represent the spatially varying physical–biogeochemical controls on DO. In combination with adaptive feature selection, inverse-density weighting, year-grouped cross-validation, and cross-boundary fusion, the framework enhances robustness in undersampled regions, minimizes temporal leakage and boundary artefacts, and yields parsimonious, interpretable submodels. The same hierarchical strategy can be readily applied to other biogeochemical tracers or observing systems that exhibit strong regional and vertical heterogeneity.
- **Adaptive multi-source feature selection.** Starting from a rich set of physical, biogeochemical, and spatiotemporal predictors, we employ a two-stage feature-selection procedure within each ~~region~~basin–depth unit to retain only variables that add independent skill. This adaptive, region- and depth-aware integration of multi-source environmental predictors strengthens the representation of upper-ocean processes and water-mass transitions while suppressing noise and redundancy, providing physically interpretable feature sets.
- **A physically consistent, long-record, high-resolution product.** GEOXYGEN delivers global monthly DO fields from 1960 to 2024 on a consistent 0.5° × 0.5° horizontal grid, spanning depths from the surface to 5500 m. Its high ~~skill and climatological~~performance, climate consistency ~~support robust~~, and the robustness of the reconstructed trends enable reliable estimates of deoxygenation trends and decadal ~~variability~~variations, providing a ~~stringent~~rigorous benchmark for ~~assessing~~evaluating and constraining Earth system models.

Nonetheless, GEOXYGEN ~~is subject to certain~~has limitations ~~inherent in~~related to observational coverage and methodological assumptions. ~~It is also noted that uncertainty is elevated~~Uncertainty is generally higher in early decades, ice-covered high latitudes, data-sparse deep basins, and ~~unresolved~~ nearshore regions. ~~Extremely low DO values in severely hypoxic environments may be conservatively biased upward, and the product's sensitivity to uncertainties in external physical drivers merits further assessment.~~ Future ~~efforts~~work will focus on ~~cross-platform calibration, dynamically refined regionalization, and inclusion of additional subsurface constraints.~~refining the framework to further reduce uncertainty in these challenging regions and to improve its ability to represent time-varying biogeochemical regimes. Despite these limitations, ~~it~~GEOXYGEN

带格式的: 缩进: 左侧: 0.63 厘米, 悬挂缩进: 3.6 字符

provides a long-term, and internally consistent, and process informed foundation basis for detecting, attributing, and interpreting global assessing multi-decadal changes in ocean deoxygenation.dissolved oxygen. The GEOXYGEN, together with basic usage examples, is available at Zenodo and basins partition files can be obtained at the following website: https://doi.org/10.12157/IOCAS.20260223.002 (Wang et al., 2025, https://doi.org/10.5281/zenodo.17615657). Future updates will be versioned at the same DOI.2026)

Code and data availability

The dissolved oxygen (DO)GEOXYGEN dataset generatedproduced in this study, together with the biogeochemicalbasin province mask used for regionalization and coastal exclusion, is openlypublicly available at https://doi.org/10.5281/zenodo.17615657the Oceanographic Data Center, Chinese Academy of Sciences (CASODC) (Wang et al., 2025),2026, https://doi.org/10.12157/IOCAS.20260223.002), where details of the data files and metadata are documented. The codes used to train models and generate the data product in this paper are openly available at https://github.com/layne1202/GEOXYGEN-code.https://github.com/layne1202/GEOXYGEN_Code, which also documents the key training and evaluation settings (e.g., fixed random seeds) and provides the environment/dependency information needed to reproduce the experiments.

设置了格式: 字体: 非加粗

Author contribution

Z. W. and W. F. conceived the project. Z. W. collected and processed the data with contributions from C. X. Z. W. carried out the study and generated the GEOXYGEN product with contributions from W. F. Z. W. and W. F. wrote the manuscript with contributions from C. X. and G. W. All authors discussed the results and commented on the manuscript.

Acknowledgments

This research is sponsored by Natural Science Foundation of Shanghai under the grant number 24ZR1404500. The Argo data used here were collected and made freely available by the International Argo Program and by the national programs that contribute to it (https://doi.org/10.17882/42182). The OceanSITES data were collected and made freely available by the international OceanSITES project and the national programs that contribute to it.

Competing interests

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

References

~~Bettencourt, J. H., Lopez, C., Hernandez-Garcia, E., Montes, I., Sudre, J., Dewitte, B., Paulmier, A., and Garcon, V.: Boundaries of the Peruvian oxygen minimum zone shaped by coherent mesoscale dynamics, Nat. Geosci., 8, 937–940, https://doi.org/10.1038/ngeo2570, 2015.~~

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., and Seferian, R.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, Biogeosciences, 10, 6225–6245, https://doi.org/10.5194/bg-10-6225-2013, 2013.

Breitburg, D., Levin, L. A., Oschlies, A., Gregoire, M., Chavez, F. P., Conley, D. J., Garcon, V., Gilbert, D., Gutierrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., Telszewski, M., Yasuhara, M., and Zhang, J.: Declining oxygen in the global ocean and coastal waters, Science, 359, eaam7240, https://doi.org/10.1126/science.aam7240, 2018.

Cao, R., Wang, S., Bao, S., Li, X., Tan, J., and Shao, C.: SE-LeNet: A data reconstruction method for dissolved oxygen in tropical Pacific with deep learning, in: Proc. 2024 IEEE Int. Conf. Parallel Distrib. Process. Appl. (ISPA), https://doi.org/10.1109/ISPA63168.2024.00031, 2024.

Carpenter, J. H.: The accuracy of the winkler method for dissolved oxygen analysis 1, Limnol. Oceanogr., 10, 135-140, https://doi.org/10.4319/lo.1965.10.1.0135, 1965.

Chau, T.-T.-T., Gehlen, M., Metzl, N., and Chevallier, F.: CMEMS-LSCE: a global, 0.25∘, monthly reconstruction of the surface ocean carbonate system, Earth Syst. Sci. Data, 16, 121–160, https://doi.org/10.5194/essd-16-121-2024, 2024.

Chau, T. T. T., Gehlen, M., and Chevallier, F.: A seamless ensemble-based reconstruction of surface ocean pCO_2 and air–sea CO_2 fluxes over the global coastal and open oceans, Biogeosciences, 19, 1087–1109, https://doi.org/10.5194/bg-19-1087-2022, 2022.

Cheng, L. and Gouretski, V.: IAP Global Ocean Oxygen gridded product (1-degree) [data set], http://dx.doi.org/10.12157/IOCAS.20231214.006, 2024.

Chen, Z., Siedlecki, S., Long, M., Petrik, C. M., Stock, C. A., and Deutsch, C. A.: Skillful multiyear prediction of marine habitat shifts jointly constrained by ocean temperature and dissolved oxygen, Nat. Commun., 15, 900, https://doi.org/10.1038/s41467-024-45016-5, 2024.

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

Fay, A. R. and McKinley, G. A.: Global open-ocean biomes: mean and temporal variability, Earth Syst. Sci. Data, 6, 273–284, https://doi.org/10.5194/essd-6-273-2014, 2014.

Franco, A. C., Hernández-Ayón, J. M., Beier, E., Garçon, V., Maske, H., Paulmier, A., Färber-Lorda, J., Castro, R., and Sosa-Ávalos, R.: Air-sea CO_2 fluxes above the stratified oxygen minimum zone in the coastal region off Mexico, J. Geophys. Res.-Oceans, 119, 2923–2937, https://doi.org/10.1002/2013JC009337, 2014.

Garabaghi, F. H., Benzer, S., and Benzer, R.: Modeling dissolved oxygen concentration using machine learning techniques with dimensionality reduction approach, Environ. Monit. Assess., 195, 879, https://doi.org/10.1007/s10661-023-11492-3, 2023.

Garcia, H., Cruzado, A., Gordon, L., and Escanez, J.: Decadal-scale chemical variability in the subtropical North Atlantic deduced from nutrient and oxygen data, J. Geophys. Res.-Oceans, 103, 2817–2830, https://doi.org/10.1029/97JC03037, 1998.

Garcia, H. E., Wang, Z., Bouchard, C., Cross, S. L., Paver, C. R., Reagan, J. R., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Baranova, O., Seidov, D., and Dukhovskoy, D.: World Ocean Atlas 2023, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation, A. Mishonov (Ed.), NOAA Atlas NESDIS 91, 109 pp., https://doi.org/10.25923/rb67-ns53, 2024.

Gilbert, D., Rabalais, N. N., Díaz, R. J., and Zhang, J.: Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean, Biogeosciences, 7, 2283–2296, https://doi.org/10.5194/bg-7-2283-2010, 2010.

Giomi, F., Barausse, A., Steckbauer, A., Daffonchio, D., Duarte, C. M., and Fusi, M.: Oxygen dynamics in marine productive ecosystems at ecologically relevant scales, Nat. Geosci., 16, 560–566, https://doi.org/10.1038/s41561-023-01217-z, 2023.

Gong, H., Li, C., and Zhou, Y.: Emerging global ocean deoxygenation across the 21st century, Geophys. Res. Lett., 48, e2021GL095370, https://doi.org/10.1029/2021GL095370, 2021.

Gouretski, V., Cheng, L., Du, J., Xing, X., and Chai, F.: A quality-controlled and bias-adjusted global ocean oxygen profile dataset [data set], https://doi.org/10.12157/IOCAS, 2024a.

Gouretski, V., Cheng, L., Du, J., Xing, X., Chai, F., and Tan, Z.: A consistent ocean oxygen profile dataset with new quality control and bias assessment, Earth Syst. Sci. Data, 16, 5503–5530, https://doi.org/10.5194/essd-16-5503-2024, 2024b2024.

Gregoire, M., Garcon, V., Garcia, H., Breitburg, D., Isensee, K., Oschlies, A., Telszewski, M., Barth, A., Bittig, H. C., Carstensen, J., Carval, T., Chai, F., Chavez, F., Conley, D., Coppola, L., Crowe, S., Currie, K., Dai, M., Deflandre, B., Dewitte, B., Diaz, R., Garcia-Robledo, E., Gilbert, D., Giorgetti, A., Glud, R., Gutierrez, D., Hosoda, S., Ishii, M., Jacinto, G., Langdon, C., Lauvset, S. K., Levin, L. A., Limburg, K. E., Mehrtens, H., Montes, I., Naqvi, W., Paulmier, A., Pfeil, B., Pitcher, G., Pouliquen, S., Rabalais, N., Rabouille, C., Recape, V., Roman, M., Rose, K., Rudnick, D., Rummer, J., Schmechtig, C., Schmidtko, S., Seibel, B., Slomp, C., Sumalia, U. R., Tanhua, T., Thierry, V., Uchida, H., Wanninkhof, R., and Yasuhara, M.:

A global ocean oxygen database and atlas for assessing and predicting deoxygenation and ocean health in the open and coastal ocean, Front. Mar. Sci., 8, https://doi.org/10.3389/fmars.2021.724913, 2021.

Grégoire, M., Oschlies, A., Canfield, D., Castro, C., Ciglenecki, I., Croot, P., Salin, K., Schneider, B., Serret, P., and Slomp, C.: Ocean Oxygen: the role of the Ocean in the oxygen we breathe and the threat of deoxygenation, European Marine Board, Ostend, Belgium, https://doi.org/10.5281/zenodo.7941157, 2023.

Guinehut, S., Dhomps, A.-L., Larnicol, G., and Le Traon, P.-Y.: High resolution 3-D temperature and salinity fields derived from in situ and satellite observations, Ocean Sci., 8, 845–857, https://doi.org/10.5194/os-8-845-2012, 2012.

Hauser, D., Tourain, C., Hermozo, L., Alraddawi, D., Aouf, L., Chapron, B., Dalphinet, A., Delaye, L., Dalila, M., and Dormy, E.: New observations from the SWIM radar on-board CFOSAT: Instrument validation and ocean wave measurement assessment, IEEE Trans. Geosci. Remote Sens., 59, 5–26, https://doi.org/10.1109/TGRS.2020.2994372, 2020.

Hollitzer, H. A. L., Patara, L., Terhaar, J., and Oschlies, A.: Competing effects of wind and buoyancy forcing on ocean oxygen trends in recent decades, Nat. Commun., 15, https://doi.org/10.1038/s41467-024-53557-y, 2024.

Huang, S., Shao, J., Chen, Y., Qi, J., Wu, S., Zhang, F., He, X., and Du, Z.: Reconstruction of dissolved oxygen in the Indian Ocean from 1980 to 2019 based on machine learning techniques, Front. Mar. Sci., 10, 1291232, https://doi.org/10.3389/fmars.2023.1291232, 2023.

Humphries, N. E., Fuller, D. W., Schaefer, K. M., and Sims, D. W.: Highly active fish in low oxygen environments: Vertical movements and behavioural responses of bigeye and yellowfin tunas to oxygen minimum zones in the eastern Pacific Ocean, Mar. Biol., 171, 55, https://doi.org/10.1007/s00227-023-04366-2, 2024.

IOC, SCOR, and IAPSO: The international thermodynamic equation of seawater – 2010: calculation and use of thermodynamic properties, Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO, 196 pp., 2010.

Ito, T., Minobe, S., Long, M. C., and Deutsch, C.: Upper ocean O_2 trends: 1958–2015, Geophys. Res. Lett., 44, 4214–4223, https://doi.org/10.1002/2017GL073613, 2017.

Ito, T., Cervania, A., Cross, K., Ainchwar, S., and Delawalla, S.: Mapping dissolved oxygen concentrations by combining shipboard and argo observations using machine learning algorithms, J. Geophys. Res.: Mach. Learn. Comput., 1, e2024JH000272, https://doi.org/10.1029/2024JH000272, 2024a.

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

Kim, H., Franco, A. C., and Sumaila, U. R.: A selected review of impacts of ocean deoxygenation on fish and fisheries, Fishes, 8, https://doi.org/10.3390/fishes8060316, 2023.

Kolodziejczyk, N., Prigent-Mazella, A., and Gaillard, F.: ISAS temperature, salinity, dissolved oxygen gridded fields, SEANOE [data set], https://doi.org/10.17882/52367, 2023.

Li, C., Huang, J., Ding, L., Liu, X., Yu, H., and Huang, J.: Increasing escape of oxygen from oceans under climate change, Geophys. Res. Lett., 47, e2019GL086345, https://doi.org/10.1029/2019GL086345, 2020.

Liu, G., Yu, X., Zhang, J., Wang, X., Xu, N., and Ali, S.: Reconstruction of the three-dimensional dissolved oxygen and its spatio-temporal variations in the Mediterranean Sea using machine learning, J. Environ. Sci., 157, 710–728, https://doi.org/10.1016/j.jes.2025.01.010, 2025.

Liu, Q., Liu, C., Meng, Q., Su, B., Ye, H., Chen, B., Li, W., Cao, X., Nie, W., and Ma, N.: Machine learning reveals biological activities as the dominant factor in controlling deoxygenation in the South Yellow Sea, Cont. Shelf Res., 283, 105348, https://doi.org/10.1016/j.csr.2024.105348, 2024.

Lu, B., Zhao, Z., Han, L., Gan, X., Zhou, Y., Zhou, L., Fu, L., Wang, X., Zhou, C., and Zhang, J.: ~~OxyGenerator: reconstructing~~Oxygenerator: Reconstructing global ocean deoxygenation over a century with deep learning, ~~in: Proc. 41st Int. Conf. Mach. Learn. (ICML 2024), Vienna, Austria,~~arXiv preprint arXiv:.07233, https://doi.org/10.48550/arXiv.2405.0723, 2024.

Ma, D., Zhao, F., Zhu, L., Li, X., Wei, J., Chen, X., Hou, L., Li, Y., and Liu, M.: Deep learning reveals hotspots of global oceanic oxygen changes from 2003 to 2020, Int. J. Appl. Earth Obs. Geoinf., 136, 104363, https://doi.org/10.1016/j.jag.2025.104363, 2025.

Mears, C., Lee, T., Ricciardulli, L., Wang, X., and Wentz, F.: Improving the Accuracy of the Cross-Calibrated Multi-Platform (CCMP) Ocean Vector Winds, Remote Sens., 14, 4230, https://doi.org/10.3390/rs14174230, 2022.

Milà, C., Ludwig, M., Pebesma, E., Tonne, C., and Meyer, H.: Random forests with spatial proxies for environmental modelling: opportunities and pitfalls, Geosci. Model Dev., 17, 6007–6033, https://doi.org/10.5194/gmd-17-6007-2024, 2024.

NASA Ocean Biology Processing Group: Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Level-2 Ocean Color Data, version R2018.8, NASA Ocean Biology Distributed Active Archive Center [data set], https://doi.org/10.5067/ORBVIEW-2/SEAWIFS/L2/OC/2018, 2018.

~~Mishonov, A. V., Boyer, T. P., Baranova, O. K., Bouchard, C. N., Cross, S. L., Garcia, H. E., Locarnini, R. A., Paver, C. R., Reagan, J. R., Wang, Z., Seidov, D., Grodsky, A. I., and Beauchamp, J. G.: World Ocean Database 2023, NOAA Atlas NESDIS, 97, NOAA National Environmental Satellite, Data, and Information Service and NOAA National Centers for Environmental Information, Silver Spring, MD, USA, https://doi.org/10.25923/z885-h264, 2024.~~

Oschlies, A.: A committed fourfold increase in ocean oxygen loss, Nat. Commun., 12, ~~https://doi.org/10.1038/s41467-021-22584-4,~~https://doi.org/10.1038/s41467-021-22584-4, 2021.

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

Ping, B., Meng, Y., Su, F., Xue, C., and Li, Z.: Retrieval of subsurface dissolved oxygen from surface oceanic parameters based on machine learning, Mar. Environ. Res., 199, 106578, https://doi.org/10.1016/j.marenvres.2024.106578, 2024.

Regier, P. J., Ward, N. D., Myers-Pigg, A. N., Grate, J., Freeman, M. J., and Ghosh, R. N.: Seasonal drivers of dissolved oxygen across a tidal creek – marsh interface revealed by machine learning, Limnol. Oceanogr., 68, 2359–2374, https://doi.org/10.1002/lno.12426, 2023.

Robinson, C.: Microbial respiration, the engine of ocean deoxygenation, Front. Mar. Sci., 5, https://doi.org/10.3389/fmars.2018.00533, 2019.

Salazar, J. J., Garland, L., Ochoa, J., and Pyrcz, M. J.: Fair train-test split in machine learning: Mitigating spatial autocorrelation for improved prediction accuracy, J. Pet. Sci. Eng., 209, 109885, https://doi.org/10.1016/j.petrol.2021.109885, 2022.

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

Shao, J., Huang, S., Chen, Y., Qi, J., Wang, Y., Wu, S., Liu, R., and Du, Z.: Satellite-based global sea surface oxygen mapping and interpretation with spatiotemporal machine learning, Environ. Sci. Technol., 58, 498–509, https://doi.org/10.1021/acs.est.3c08833, 2024.

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

Szekely, T., Gourrion, J., Pouliquen, S., and Reverdin, G.: The CORA 5.2 dataset for global in situ temperature and salinity measurements: data description and validation, Ocean Sci., 15, 1601–1614, https://doi.org/10.5194/os-15-1601-2019, 2019.

Szekely, T., Gourrion, J., Pouliquen, S., Reverdin, G., and Merceur, F.: CORA, Coriolis Ocean Dataset for Reanalysis, SEANOE [data set], ~~https://doi.org/10.17882/46219.~~https://doi.org/10.17882/46219, 2025.

Valera, M., Walter, R. K., Bailey, B. A., and Castillo, J. E.: Machine learning based predictions of dissolved oxygen in a small coastal embayment, J. Mar. Sci. Eng., 8, 1007, https://doi.org/10.3390/jmse8121007, 2020.

Wagstaff, J. and Bean, B.: remap: Regionalized models with spatially smooth predictions, R J., 14, 160–178, https://doi.org/10.32614/RJ-2023-004, 2022.

Wang, Z., Fu, W., Xue, C., and Wang, G.: GEOXYGEN: ~~a~~A global ~~long-term~~monthly gridded dissolved oxygen ~~dataset (v1.~~product (0.5° × 0.5°) [data set], ~~https://doi.org/10.5281/zenodo.17615657, 2025~~https://doi.org/10.12157/IOCAS.20260223.002, 2026.

Wang, Z., Xue, C., and Ping, B.: A reconstructing model based on time–space–depth partitioning for global ocean dissolved oxygen concentration, Remote Sens., 16, 228, https://doi.org/10.3390/rs16020228, 2024.

Xue, C., Wang, Z., Yue, L., and Niu, C.: A global four-dimensional gridded dataset of ocean dissolved oxygen concentration retrieval from Argo profiles, Geosci. Data J., 11, 775–789, https://doi.org/10.1002/gdj3.251, 2024.

Yamaguchi, R., Kouketsu, S., Kosugi, N., and Ishii, M.: Global upper ocean dissolved oxygen budget for constraining the biological carbon pump, Commun. Earth Environ., 5, https://doi.org/10.1038/s43247-024-01886-7, 2024.

Zhou, Y., Gong, H., and Zhou, F.: Responses of Horizontally Expanding Oceanic Oxygen Minimum Zones to Climate Change Based on Observations, Geophys. Res. Lett., 49, https://doi.org/10.1029/2022gl097724, 2022.