https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

Open Access Earth System Science Data Discussions

Reconstructing Global Monthly Ocean Dissolved Oxygen (1960–2023) to Nearly 6000 m Depth Using Bayesian Ensemble Machine Learning

Mingyu Han[1], Yuntao Zhou[1]*

1. School of Oceanography, Shanghai Jiao Tong University, Shanghai, China

Corresponding author

*Correspondence to: Yuntao Zhou, ytzhou@sjtu.edu.cn, ORCID: 0000-0001-9714-5385

Abstract

Oceanic oxygen levels, crucial for marine ecosystems and biogeochemical cycles, have declined significantly over the past few decades, driven by climate change and posing severe environmental risks. However, historical dissolved oxygen (DO) measurements, especially below 2000 m, remain sparse, limiting comprehensive annual and seasonal analyses. Here we introduce the BEM-DOR framework, a Bayesian-optimized ensemble of six machine-learning models (Random Forest, XGBoost, LightGBM, CatBoost, Extremely Randomized Trees and Histogram-based Gradient Boosting) fused via dynamic weighting, to reconstruct global monthly DO distributions at 1° ×1° resolution from the surface to 5902 m depth over 1960-2023. Validation against an independent dataset demonstrates that BEM-DOR outperforms existing products. Our dataset captures depth-dependent deoxygenation, with the most pronounced decline occurring between 150 and 200 m, and reveals dramatically accelerated oxygen loss in the Arctic Ocean and North Atlantic over the past decade. We quantify uncertainties from measurement errors, gridding processes, and model algorithms, providing the first long-term, high-resolution, uncertainty-quantified DO product from ocean surface to nearly 6000 m depth. The extension of DO data into the bathypelagic zone in this work is a significant contribution to deep ocean oxygen dynamics and global biogeochemical cycles.

Keywords

dissolved oxygen, machine learning, ensemble learning, Bayesian optimization, data reconstruction

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

Open Access Earth System Science Data Discussions

1 Introduction

Over the past few decades, dissolved oxygen (DO) levels in open oceans have been continuously decreasing (Breitburg et al., 2018; Keeling et al., 2010), primarily driven by climate change (Deutsch et al., 2011). This decline has severe impacts on marine organisms and biogeochemical processes, disrupting marine productivity, biodiversity, and biogeochemical cycles (Gruber, 2011; Stramma et al., 2012). Climate models predict that global warming will further accelerate this deoxygenation (Oschilies et al., 2018), potentially adversely affecting aerobic marine organisms within this century (Sampaio et al., 2021), and altering biogeochemical cycles (Gruber, 2004; Berman-Frank et al., 2008). Therefore, it is important to develop a comprehensive, high-resolution reconstruction of ocean DO across both space and depth to accurately quantify historical deoxygenation trends, identify regional hotspots, and inform future ecosystem and climate projections.

Despite significant progress in oceanographic data collection, severe gaps in historical DO data persist, hindering comprehensive analysis. For instance, the World Ocean Database (WOD) (Mishonov et al., 2024) compiles DO profiles from research cruises and floats, yet most ocean regions still lack any observations. This sparse spatial coverage severely This sparse spatial coverage severely limits the use of data imputation methods to reconstruct planar or three-dimensional DO fields. Furthermore, although many Earth System Models (ESMs) attempt to simulate global oceanic DO, these models lack adjustments based on DO observation data, leading to error propagation (Pathak et al., 2023). Thus, numerical models diverge significantly from in-situ observations and consistently underestimate the actual DO decline trends (Bopp et al., 2013; Cocco et al., 2013; Long et al., 2016; Kwiatkowski et al., 2020), restricting studies related to ocean deoxygenation, Oxygen Minimum Zones (OMZs), biogeochemical cycles, and so forth.

Classical geostatistical and interpolation methods have long been employed to map oceanic DO. Zhou et al. (2022) combined geostatistical regression with Monte Carlo methods to estimate changes in the area of Oxygen Minimum Zones (OMZs) globally and regionally from 1960 to 2019. Garcia et al. (2024) applied objective analysis in WOA23 to produce internally consistent annual and monthly DO fields from 1965 to 2022. Gouretski et al. (2024) developed an automated quality control procedure to detect outliers and correct biases in ocean oxygen profiles, producing a consistent global dataset from 1920 to 2023. Roach and Bindoff (2023) used Data Interpolating Variational Analysis (DIVA) to generate a global high-resolution oxygen atlas from 1955 to 2018. Recently, machine learning methods exhibit higher computational efficiency, capable of rapidly processing large-scale datasets. Giglio et al. (2018) utilized a Random Forest Regression Model to present an estimate of oxygen at 150 m in the Southern Ocean based on Argo data during 2008-2012. Sharp et al. (2022) reconstructed a global DO dataset called GOBAI-O2 using feedforward neural networks and Random Forest Regression, spanning the years 2004 – 2022 with a monthly resolution, and extending from the ocean surface to a depth of 2 km. Ito et al. (2024) developed a machine-learning ensemble of neural networks and random forests trained on historical shipboard and biogeochemical Argo O2 profiles to generate gridded monthly oxygen fields. While some of the DO data reconstruction studies focus on specific regions, some

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

span longer time spans, and some achieve higher temporal or spatial resolutions, it is challenging to simultaneously address all aspects.

Here we introduce the Bayesian Ensemble Machine-learning Dissolved Oxygen Reconstruction (BEM-DOR) framework , which integrates six tree-based learners, Random Forest, XGBoost, LightGBM, CatBoost, Extremely Randomized Trees and Histogram-based Gradient Boosting, each tuned via Bayesian hyperparameter optimization. Model outputs are fused with dynamic “soft” weights combining global cross-validation skill and local error performance. BEM-DOR produces a global 1° × 1° monthly DO dataset from 1960 to 2023 down to 5902 m, filling critical deep-ocean gaps. We validate against The Global Ocean Data Analysis Project version 2 (GLODAPv2) (Olsen et al., 2016) with eight-fold temporal cross-validation, compare spatially with Gridded Ocean Biogeochemistry from Artificial Intelligence - Oxygen (GOBAI) (Sharp et al., 2022), ITO’s product (Ito et al., 2024) and World Ocean Atlas 2023 (WOA23) (Garcia et al., 2024), and quantify measurement, gridding and algorithm uncertainties. Finally, we analyze global, basin and depth-resolved DO distribution and deoxygenation trends. We divided the global ocean into ten basins to capture regional differences in oxygen storage and trends: North Pacific (NP), Equatorial Pacific (EP), South Pacific (SP), North Atlantic (NA), Equatorial Atlantic (EA), South Atlantic (SA), Indian Ocean north of the equator (NI), Indian Ocean south of the equator (SI), Southern Ocean (SO) and Arctic Ocean (AO). Basin boundaries follow Schmidtko et al. (2017).

2 Data and methods

2.1 Data

2.1.1 In-situ data of dissolved oxygen

We assembled our observational DO database by merging quality‐controlled profiles from the Array for Real-Time Geostrophic Oceanography dataset (Argo, https://argo.ucsd.edu) (Wong et al., 2020) with CTD and OSD measurements archived in the World Ocean Database (WOD, https://www.ncei.noaa.gov/products/world-ocean-database) (Mishonov et al., 2024). Each profile consists of oxygen concentrations sampled at multiple depths at a given date and location. We retained only those records flagged as good, then de‐duplicated overlapping casts by keeping the version with finer vertical sampling. We discarded any profile showing unrealistically high or low values and excluded casts in which oxygen fell below 10 $\mu mol\ kg^{-1}$ at any depth, which likely indicate incorrect unit descriptions. Following Schmidtko et al. (2017), we treated the combined dataset as free of systematic errors.

2.1.2 Reanalysis data of environmental factors

In this study, we investigated the contribution of physical, chemical, and biological factors to ocean DO during the period of 1960 –2023. We obtained monthly ocean temperature (T, ℃), salinity (S, ‰), meridional velocity and zonal velocity (MV and ZV, m/s) from the Ocean Reanalysis System 5 (ORAS5) gridded ocean dataset with a spatial resolution of 0.25°×0.25° and 75 vertical levels (Table S2), ranging from ocean surface to nearly 6000 m in depth (https://cds.climate.copernicus.eu/datasets/reanalysis-oras5).

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

115 **2.2 BEM-DOR framework**

116 We developed the BEM-DOR (Bayesian Ensemble Machine-learning Dissolved Oxygen
117 Reconstruction) framework to reconstruct a global, monthly DO product from 1960 through 2023.
118 The process begins with assembling and preprocessing all available in situ DO profiles alongside
119 key environmental factors, such as temperature, salinity, and currents, onto a monthly grid
120 featuring 1°×1° horizontal resolution and 75 vertical levels. Next, each of six tree-based learners
121 (Random Forest, XGBoost, LightGBM, CatBoost, Extremely Randomized Trees and
122 Histogram-based Gradient Boosting) undergoes Bayesian hyperparameter tuning via Optuna's
123 TPE sampler, ensuring that each model's configuration minimizes cross-validation RMSE (Akiba et
124 al., 2019). Once optimized, the models train on the full gridded dataset and predict DO at every
125 valid grid cell, producing six complete five-dimensional DO fields. Those outputs are then merged
126 through a dynamic weighting scheme: global "prior" weights reflect each model's time-CV skill,
127 while local "dynamic" weights adjust according to the magnitude of agreement with nearby
128 observations, yielding a soft-weighted ensemble that adapts in space and time (Dietterich, 2000).
129 Finally, we validate and quantify uncertainty by performing eight-fold temporal cross-validation,
130 independent evaluation against GLODAPv2 (Olsen et al., 2016) and comparisons with GOBAI
131 (Sharp et al., 2022), ITO (Ito et al., 2024) and WOA23 (Garcia et al., 2024). We then analyze depth
132 and basin-resolved deoxygenation trends to reveal the full vertical and regional patterns of
133 oxygen change.

134 **2.2.1 Data processing**

135 In this study, all ocean DO observation data include temporal and spatial information, including
136 year, month, day, longitude, latitude, and measurement depth. Longitude and month are both
137 periodic features. For instance, longitude ranges from 0° to 360°, with 360° overlapping
138 with 0°, and months cycle annually. To address this issue, we followed the approach of Gade
139 (2010) and Tang et al. (2019) by converting the longitude and month attributes to polar
140 coordinate systems, using sine and cosine functions to simulate these features, thus preserving
141 their cyclical nature in the model.

142 $$coordinates = \begin{pmatrix} \sin(latitude \cdot \frac{\pi}{180}) \\ \sin(longitude \cdot \frac{\pi}{180}) \cdot \cos(latitude \cdot \frac{\pi}{180}) \\ -\cos(longitude \cdot \frac{\pi}{180}) \cdot \cos(latitude \cdot \frac{\pi}{180}) \end{pmatrix} \quad (1)$$

143 $$time = \begin{pmatrix} \cos(month \cdot \frac{2\pi}{12}) \\ \sin(month \cdot \frac{2\pi}{12}) \end{pmatrix} \quad (2)$$

144

145 We use the ORAS5 reanalysis grid (1° × 1° horizontal, 75 depth levels) as the target for
146 gridding all variables. DO observations are binned to each grid cell by averaging all points that fall
147 within the cell the same month and depth level. We upscaled the other environmental factors
148 with finer resolution, using inverse distance weighting of surrounding pixel values to match these
149 resolutions. To address potential multicollinearity, which can lead to instability in subsequent
150 modeling and increase the risk of overfitting, we analyzed correlations between the 11 factors.

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

No correlation coefficient exceeded 0.4, so variable selection was not necessary in this case. A complete list of predictors, with abbreviations and data sources are shown in Table 1.

Table 1. Predictors, abbreviation and products/reference of the 19 environmental factors.

Predictor	Abbreviation	Product/Reference
sin(latitude·π/180)	coord_1	WOD (CTD+OSD) & Argo
sin(longitude·π/180) ·cos(latitude·π/180)	coord_2	
-cos(longitude·π/180) ·cos(latitude·π/180)	coord_3	
Year	Year	
cos(month·2π/12)	time_cos	
sin(month·2π/12)	time_sin	
Depth	Depth	
Temperature	T	ORA-S5
Salinity	S	
Zonal Velocity	ZV	
Meridional Velocity	MV	

Note: The observational data come from WOD and Argo. The data from ORA-S5 are 0.25°x0.25° monthly mean profile data.

2.2.2 Machine learning models

We used six tree- based algorithms to reconstruct dissolved oxygen. Each model offers a different balance of bias, variance and speed. We chose them for their strong performance in regression tasks and their ability to handle nonlinear relationships. All models were trained on the same input features and tuned via Bayesian optimization (Sect. 2.2.3). Below we describe each model. Random Forest (RF) builds many decision trees on bootstrap samples and averages their outputs (Breiman, 2001). It selects a random subset of features at each split. This randomness reduces overfitting. RF handles large datasets well and is robust to outliers. CatBoost is a gradient-boosting library designed for categorical features (Prokhorenkova et al., 2018). It uses ordered target statistics to avoid target leakage. It grows symmetric trees and applies efficient leaf pruning. CatBoost often converges faster and needs less tuning of learning rate. XGBoost implements gradient boosting with second-order optimization (Chen & Guestrin, 2016). It adds regularization to control tree complexity. It uses approximate split finding to speed up training on large data. XGBoost balances accuracy and runtime efficiency. LightGBM uses histogram-based binning and leaf-wise tree growth (Ke et al., 2017). It buckets continuous features into bins, reducing memory. Trees grow by selecting splits that yield the largest loss reduction. LightGBM is highly efficient for large feature sets and large datasets. Histogram-based Gradient Boosting (Hist_GBT) follows Friedman's original gradient boosting framework (Guryanov, 2019; Friedman, 2001). It fits a sequence of weak learners to the negative gradient of the loss. It also uses histogram binning for faster split evaluation. Hist_GBT offers good accuracy in high-dimensional settings. Extremely Randomized Trees (ERT) introduces extra randomness compared to RF (Geurts et al., 2006). It picks split thresholds at random rather than searching for the best cut. It uses the full dataset, rather than bootstrapping. This strong randomization further lowers variance at modest cost in bias.

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

2.2.3 Bayesian parameter optimization

To optimize hyperparameters across different machine learning models in a systematic and efficient manner, we employed Bayesian optimization using the Optuna framework (Akiba et al., 2019). This approach selects hyperparameter configurations based on the history of performance evaluations, aiming to minimize the prediction error of each model.

Bayesian optimization constructs a probabilistic surrogate model of the objective function f(x), where x is a vector of hyperparameters. The optimization seeks to identify the optimal x* that minimizes f:

$$\mathrm{x}^{*} = \arg\min_{x \in \chi} f(x) \tag{3}$$

Here, χ denotes the hyperparameter space. Optuna models the objective function using a Tree-structured Parzen Estimator (TPE), which fits two probability densities: one for good parameter configurations and another for all others. The next sampling point is chosen to maximize the Expected Improvement (EI):

$$EI(X) = \int_{-\infty}^{y^*} (y^* - y) \cdot p(y \mid x) dy \tag{4}$$

where y* is the current best objective value. The sampling focuses on regions with high EI.

To reduce temporal overfitting and preserve model generalizability across decades, hyperparameter optimization was conducted using a subset of data from eight years (1960, 1968, 1976, 1984, 1992, 2000, 2008, 2016). The objective function minimized the Mean Squared Error (MSE) on an independent validation set derived from other eight test years (1967, 1975, 1983, 1991, 1999, 2007, 2015, 2023). The objective function was defined as:

$$MSE = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{5}$$

Each model was optimized over its own hyperparameter space, with the best-performing configuration recorded for final training and subsequent prediction on independent test data. This consistent, data-driven approach ensured fair comparability across all six learners and minimized bias from manual tuning. Below we summarize the search space and optimal parameters in Table 2.

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

Table 2. Hyperparameter Search Spaces and Optimal Values

Model	Hyperparameter	Search Range	Best Value
ERT	n_estimators	50 – 500	482
	max_depth	3 – 20	20
	min_samples_split	2 – 20	6
	min_samples_leaf	1 – 10	2
	max_features	0.1 – 1.0	0.856
	bootstrap	{True, False}	False
CatBoost	iterations	50 – 1000	644
	depth	3 – 12	7
	learning_rate	0.005 – 0.3	0.246
	l2_leaf_reg	10^{-5} – 10	7.18
	random_strength	10^{-5} – 10	2.96
	bagging_temperature	0 – 1	0.071
	border_count	32 – 255	168
Hist_GBT	learning_rate	0.005 – 0.3	0.219
	max_iter	50 – 1000	748
	max_depth	3 – 12	9
	min_samples_leaf	5 – 50	8
	l2_regularization	10^{-5} – 10	1.24×10^{-5}
	max_bins	32 – 255	187
LightGBM	n_estimators	50 – 1000	928
	max_depth	3 – 12	9
	learning_rate	0.005 – 0.3	0.132
	num_leaves	10 – 300	118
	min_child_samples	5 – 50	7
	subsample	0.5 – 1.0	0.778
	colsample_bytree	0.5 – 1.0	0.965
	reg_alpha	10^{-8} – 10	3.24×10^{-8}
	reg_lambda	10^{-8} – 10	0.187
RF	num_trees	10 – 200	59
	min_leaf_size	10 – 100	10
XGBoost	n_estimators	50 – 1000	830
	max_depth	3 – 12	12
	learning_rate	0.005 – 0.3	0.265
	min_child_weight	1 – 10	1
	subsample	0.5 – 1.0	0.615
	colsample_bytree	0.5 – 1.0	0.908

2.2.4 Multi-model fusion and dynamic weighting strategy

We fuse six model predictions into one field. Our goal is to combine global model skill with local fit to observations. We assign each model a static "prior" weight. We then adjust those weights at each grid cell using the local agreement between prediction and observation.

We derive a prior weight w_i for model i from its time-cross-validation (Sect. 3.1) RMSE ε_i. We set a

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

decay parameter β. Then:

$$\omega_i = \frac{\exp(-\beta\varepsilon_i)}{\sum_{j=1}^{6}\exp(-\beta\varepsilon_j)}, \sum_{i=1}^{6}\omega_i = 1 \quad (6)$$

A smaller ε_i yields a larger w_i. We choose β=1 to balance influence among models.

At each grid cell x, we compute a dynamic weight $v_i(x)$. We use a tuning parameter α. For cells where an observation O(x) exists, we set:

$$v_i(x) = \exp(-\alpha \mid p_i(x) - O(x) \mid) \quad (7)$$

Here $p_i(x)$ is model i's prediction. A smaller local error makes $v_i(x)$ larger. We use α=1. For cells with no observation, we fall back on the static weight:

$$\mathrm{v}_i(x) = \omega_i \quad (8)$$

We compute the ensemble prediction E(x) by normalizing the dynamic weights:

$$E(x) = \frac{\sum_{i=1}^{6} v_i(x)\,p_i(x)}{\sum_{i=1}^{6} v_i(x)}, if \sum_i v_i(x) > 0 \quad (9)$$

and E(x)=NaN if all $p_i(x)$ are missing. This formula guarantees that models aligning well with local observations gain more influence, while the static weights keep poorly observed regions stable.

2.2.5 Data reconstruction

We produce a global, monthly dissolved oxygen (DO) dataset on a regular 1° × 1° grid and 75 depth levels (0–5902 m) spanning 1960–2023. First, we gather all predictor fields described in Table 1. Each field is remapped to the target grid and monthly time step following ORAS5. Next, we apply the six optimized machine-learning models (Sect. 2.2.2) at every valid grid cell and time. Each model ingests the full vector of predictors and returns a DO estimate only where all predictors are present. This yields six parallel prediction arrays of dimensions 360 × 180 × 75 × 12 × 64. We then merge these arrays using our dynamic weighting scheme (Sect. 2.2.4). Static "prior" weights reflect each model's cross-validation skill. Local weights adapt to agreement with any overlapping in situ observation. The weighted combination produces a single ensemble DO field at each grid cell and month. We packaged the ensemble field into a CF-compliant NetCDF file with coordinate variables, depth layers, time and global attributes documenting methods.

3 Model performance

3.1 Model Temporal Cross- Validation

We conducted eight-fold temporal cross-validation on each of the six models. In each fold f, data from eight test years $\{1960 + \mathrm{f} + 8k\}_{k=0}^{7}$ formed the test set, with remaining years for training.

We trained each model using its optimized hyperparameters (Sect. 2.2.3) on the training years, predicted the test years, and computed mean bias (ΔDO), mean absolute error (MAE), root-mean-square error (RMSE) and coefficient of determination (R^2) on the held-out data. These metrics collectively provide a comprehensive understanding of the model's predictive accuracy

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

and bias. The results appear in Tables 3-6.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (10)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (12)$$

$$\Delta DO = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \quad (13)$$

Table 2. Cross-validation ΔDO (μmol kg⁻¹)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	-0.160	0.253	-0.089	0.134	0.085	0.155	0.444	0.347
XGBoost	-0.403	0.313	-0.125	0.291	-0.007	0.158	0.325	0.211
LightGBM	-0.470	0.289	-0.173	0.250	0.048	0.193	0.391	0.255
Hist_GBT	-0.324	0.204	-0.170	0.174	0.052	0.186	0.378	0.285
ERT	-0.298	-0.120	-0.346	-0.069	-0.127	0.027	0.662	0.235
CatBoost	-0.287	0.128	-0.134	0.244	0.081	0.072	0.505	0.165

Table 3. Cross-validation MAE (μmol kg⁻¹)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	10.404	9.894	9.786	9.661	9.540	9.457	9.520	9.763
XGBoost	10.943	10.471	10.406	10.281	10.130	10.011	10.130	10.320
LightGBM	10.849	10.376	10.328	10.136	10.011	9.921	9.964	10.218
Hist_GBT	11.510	11.045	10.926	10.799	10.671	10.588	10.663	10.926
ERT	10.627	10.180	9.939	9.834	9.752	9.668	9.726	9.995
CatBoost	11.904	11.538	11.247	11.158	11.053	10.942	10.990	11.299

Table 4. Cross-validation RMSE (μmol kg⁻¹)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	17.294	16.610	16.317	16.337	15.999	16.028	16.344	16.302
XGBoost	17.620	16.968	16.814	16.829	16.479	16.391	16.756	16.752
LightGBM	17.405	16.789	16.634	16.506	16.260	16.273	16.526	16.503
Hist_GBT	18.048	17.477	17.130	17.154	16.892	16.859	17.210	17.179
ERT	17.325	16.801	16.238	16.433	16.016	16.094	16.321	16.349
CatBoost	18.501	18.068	17.464	17.607	17.312	17.324	17.568	17.582

Table 5. Cross-validation R^2

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	0.958	0.961	0.960	0.959	0.960	0.960	0.960	0.963
XGBoost	0.957	0.959	0.958	0.957	0.957	0.958	0.958	0.961

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

LightGBM	0.958	0.960	0.958	0.959	0.959	0.958	0.959	0.962
Hist_GBT	0.954	0.957	0.956	0.955	0.955	0.955	0.956	0.959
ERT	0.958	0.960	0.960	0.959	0.959	0.959	0.960	0.963
CatBoost	0.952	0.954	0.954	0.953	0.953	0.953	0.954	0.957

All six learners exhibit remarkably consistent skill across the eight temporal folds, with only minor spread in error metrics (Table 2-5). LightGBM's MAE varies by less than 0.9 µmol kg^{-1} (9.92–10.85) and its RMSE by under 0.9 µmol kg^{-1} (16.26–17.41), yielding an R^2 range of 0.958–0.962—small fluctuations that underscore stable performance year-to-year. RF delivers the lowest RMSE (15.99–17.29) and highest R^2 (0.958–0.963). In contrast, CatBoost and Hist_GBT register higher mean errors (MAE up to 11.90 and 11.51, RMSE up to 18.50 and 18.05) and slightly larger inter-fold variability, indicating they are more sensitive to the specific training/test split (Table 2–5). ERT and XGBoost fall between these extremes, with moderate error levels and consistent R^2. Crucially, no model ever produces an outlier fold with dramatically degraded skill—each maintains MAE < 12 µmol kg^{-1} and R^2 > 0.95. This uniformity across folds confirms strong temporal generalization and validates our choice of an ensemble approach (Bergmeir & Benítez, 2012; Roberts et al., 2017).

3.2 Evaluation on independent observations

We evaluated both the ensemble and each single model against an independent GLODAPv2 dissolved oxygen dataset, treated here as ground truth. GLODAPv2 profiles were averaged into the same 1°×1° grid and monthly time step as the reconstructions. We then identified grid cells where both the gridded GLODAPv2 values and model predictions were non-NaN. At those collocated points we computed four summary metrics: mean bias (ΔDO), mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2).

Table 6. Comparison of Ensemble and Single Models on GLODAPv2 Dataset

Model	MAE (µmol kg^{-1})	RMSE (µmol kg^{-1})	R^2	Δ DO (µmol kg^{-1})
Ensemble	**5.003**	**9.895**	**0.985**	0.311
Ensemble(static weight=1)	5.061	9.982	0.985	0.341
RF	5.802	10.979	0.982	**0.221**
XGBoost	6.198	10.943	0.982	0.330
ERT	6.757	11.960	0.979	0.405
LightGBM	7.637	12.825	0.975	0.543
Hist_GBT	8.632	13.875	0.971	0.589
CatBoost	9.179	14.474	0.969	0.693

Table 6 also includes an equal-weight ensemble (static weight = 1), which yields MAE = 5.061 µmol kg^{-1}, RMSE = 9.982 µmol kg^{-1} and R^2 = 0.985 (ΔDO = 0.341). Although this uniform blend already outperforms any single model, our RMSE-based prior weights push performance further, dropping MAE to 5.003 µmol kg^{-1}, RMSE to 9.895 µmol kg^{-1} and raising R^2 to 0.985, demonstrating that leveraging each learner's cross-validation skill yields a measurably better

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

ensemble than equal weighting. Among the individual algorithms, Random Forest (MAE 5.802, RMSE 10.979, R^2 0.982) and XGBoost (MAE 6.198, RMSE 10.943, R^2 0.982) follow most closely, while CatBoost (MAE 9.179, RMSE 14.474, R^2 0.969) and Hist_GBT (MAE 8.632, RMSE 13.875, R^2 0.971) sit at the lower end. All models keep bias under 0.7 μmol kg^{-1}. No single method ever exhibits a catastrophic fold, underscoring the robustness of our dynamic weighting in combining complementary strengths and minimizing weaknesses (Dietterich, 2000).

3.3 Uncertainty estimations

We quantify three independent sources of error in the reconstructed DO field. Estimating each component lets us understand their relative contributions and report a rigorous total uncertainty. Measurement uncertainty ($\triangle O_{meas}$) originates from the inherent precision limits of the in situ dissolved oxygen observations. In our study, we assume a constant uncertainty based on Ito et al., (2024): OSD and CTD data are assigned a measurement uncertainty of 1 μmol kg^{-1}, while Argo data are attributed a value of 3 μmol kg^{-1}. Thus, for any given observation, we represent the uncertainty as $\triangle O_{meas}$=1 μmol kg^{-1} for OSD or CTD data and $\triangle O_{meas}$=3 μmol kg^{-1} for Argo data. This constant assignment provides a pragmatic baseline for quantifying the observational error in the reconstructed dataset, acknowledging that, although regional variations might introduce additional variability, such effects are not considered in this baseline estimate.

Grid uncertainty ($\triangle O_{grid}$) quantifies the representation error associated with assigning a single dissolved oxygen value to a 1°×1° spatial cell across time. To estimate grid uncertainty, observations within each grid cell are collocated, and the standard deviation (σ) among these observations is computed. For a given grid cell, let the n observations be denoted by O_1, O_2, …, O_n ; then, the grid uncertainty is estimated by:

$$\Delta O_{\text{grid}} = \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{14}$$

where $\overline{O}$ is the mean value of the observations in that cell. These standard deviations are then averaged across the entire dataset, providing an overall estimate of the grid uncertainty. This method effectively captures the dispersion due to variable sampling density and spatial heterogeneity, reflecting the error introduced by mapping sparse in situ data onto a coarser grid.

Algorithm uncertainty ($\triangle O_{alg}$) reflects the error introduced by the machine learning reconstruction process. In this study, six ensemble models (RF, XGBoost, LightGBM, CatBoost, ERT, and Hist_GBT) were trained using a comprehensive set of dissolved oxygen observations and environmental factors. Each model was optimized via Bayesian hyperparameter tuning and validated using an eight-fold cross-validation procedure, yielding an MAE for each model, denoted by $error_1$ through $error_6$. We then compute the prior weight for the i-th model using an exponential decay function:

$$\omega_i = \frac{\exp(-error_i)}{\sum_{j=1}^{6}\exp(-error_j)} \tag{15}$$

The overall algorithm uncertainty is then estimated as the weighted average of the MAE values:

$$\Delta O_{\text{alg}} = \sum_{i=1}^{6}\omega_i error_i \tag{16}$$

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

This approach synthesizes the performance of the six different models into a single metric, representing the inherent uncertainty of the reconstruction algorithm across the entire dataset. The resulting component uncertainties are measurement uncertainty = 1.60 μmol kg^{-1}, grid uncertainty = 4.61 μmol kg^{-1}, and algorithm uncertainty = 10.17 μmol kg^{-1}. Finally, the total uncertainty in the reconstructed dissolved oxygen field is expressed as

$$\Delta O_{\text{total}} = \sqrt{\Delta O_{meas}^2 + \Delta O_{grid}^2 + \Delta_{a\lg}^2} \text{ =11.28 μmol kg}^{-1}.$$

This formulation integrates the observational, mapping, and model reconstruction uncertainties into a comprehensive framework for error quantification.

4 Data product comparison

4.1 Comparison with GLODAPv2 observations

To rigorously evaluate our DO reconstruction, we conducted a systematic comparison with two recent datasets, GOBAI (Sharp et al., 2023) and ITO (Ito et al., 2024), using the quality-controlled GLODAPv2 dataset (Olsen et al., 2016) an independent validation standard. Our reconstruction consistently achieves the lowest MAE and RMSE, and the highest R^2, with near-zero bias. Within the spatial, temporal, and depth ranges of GOBAI, our model also outperforms GOBAI by 15–20% in both MAE and RMSE, and by 0.01 in R^2. Within the spatial, temporal, and depth ranges of ITO, our model show a 30% reduction in MAE and RMSE compared to ITO, while the bias improved significantly, decreasing from 1.74 to 0.34 μmol kg^{-1}. These comparisons confirm that our ensemble delivers consistently better agreement with independent GLODAPv2 observations, both globally and within the individual domains of GOBAI and ITO.

Table 7. Performance comparison on GLODAPv2

Product	MAE	RMSE	R^2	Δ DO
Our reconstruction (full GLODAPv2)	5.003	9.895	0.985	0.311
GOBAI on GLODAPv2	8.107	14.201	0.969	1.167
Our reconstruction in GOBAI coverage	**6.399**	**12.353**	**0.980**	**1.019**
ITO on GLODAPv2	9.417	16.790	0.961	1.742
Our reconstruction in ITO coverage	**6.507**	**11.991**	**0.981**	**0.344**

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

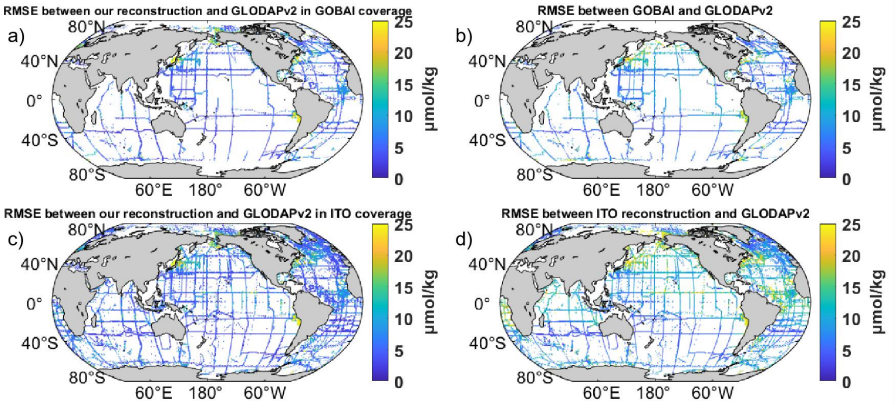


Figure 1. Global maps of RMSE between model reconstructions and GLODAPv2 observations over two domains. (a) RMSE of our reconstruction in GOBAI coverage. (b) RMSE of the GOBAI product. (c) RMSE of our reconstruction in ITO coverage. (d) RMSE of the ITO reconstruction. Color scale (µmol kg^{-1}) is the same in all panels; blue shading indicates low errors (< 10 µmol kg^{-1}), red indicates higher errors (up to 25 µmol kg^{-1}).

Across the domain of GOBAI (Fig.1a b), our ensemble consistently achieves RMSE values that are 3–8 µmol kg^{-1} lower than those of GOBAI, especially in the equatorial Pacific, Indian Ocean and Southern Ocean boundary currents. In contrast, both products show higher RMSE in high-latitude regions, reflecting sparse calibration data in these areas. Within the domain of ITO (Fig.1c d), our ensemble again shows lower RMSE across most subtropical and mid-latitude basins, with typical errors of 8–12 µmol kg^{-1} compared to 12–18 µmol kg^{-1} of ITO. The largest errors in ITO occur in the North Pacific. These spatial comparisons confirm that our ensemble reconstruction not only improves global summary statistics (Table 7) but also delivers consistently lower local errors than GOBAI and ITO. The reduction of RMSE in both open-ocean and dynamical boundary regions demonstrates the performance of our multi-model, Bayesian-weighted approach in capturing complex oxygen variability in the ocean.

4.2 Comparison of mean vertical profile

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

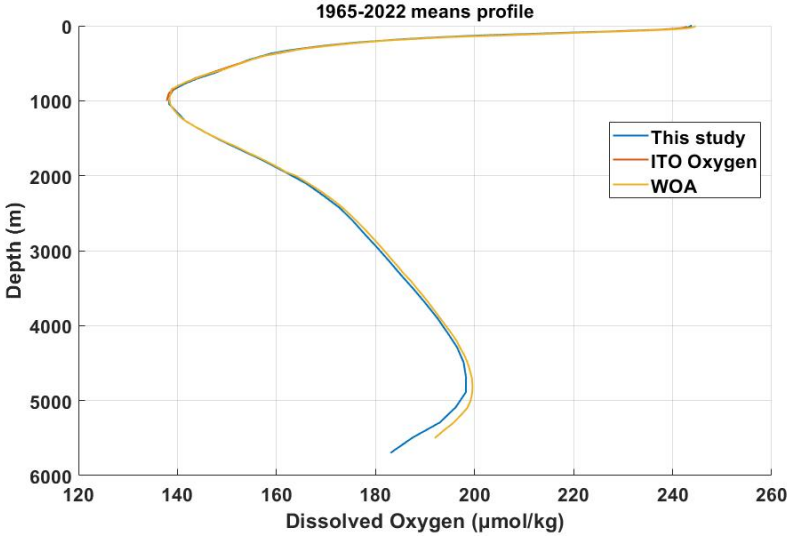


Figure 2. Global mean vertical profiles of different dissolved oxygen products (1965–2022). Solid lines show our reconstruction (blue), ITO's reconstruction (orange) and WOA23 climatology (yellow), plotted from the surface down to 5902 m.

Our DO reconstruction improves oxycline accuracy and extends coverage into the bathypelagic zone, filling gaps left by previous datasets. Figure 2 shows the mean profiles for our reconstruction (blue), ITO (orange; Ito et al., 2024), and WOA23 climatology (yellow; Garcia et al., 2024), averaged over the ITO/WOA23 domain for 1965–2022. At the surface, all three profiles align near 245–250 µmol kg^{-1}. In the main oxycline between 800 and 1,000 m, our reconstruction and WOA23 show almost identical profiles, while ITO falls 2–5 µmol kg^{-1} lower in this depth range. By blending multiple models and environmental factors, our ensemble better captures the true oxycline shape and matches the widely used WOA23 reference. Below 1000 m, ITO provides no data, therefore, the comparison focuses on our reconstruction with WOA23. The two profiles remain within 2–3 µmol kg^{-1} of each other down to 5902 m, demonstrating the reliability of our deep-ocean estimates. Our deep-water values fill a critical gap for studies of oxygen supply, interior circulation and biogeochemical cycling.

4.3 Spatial difference from climatological distribution

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

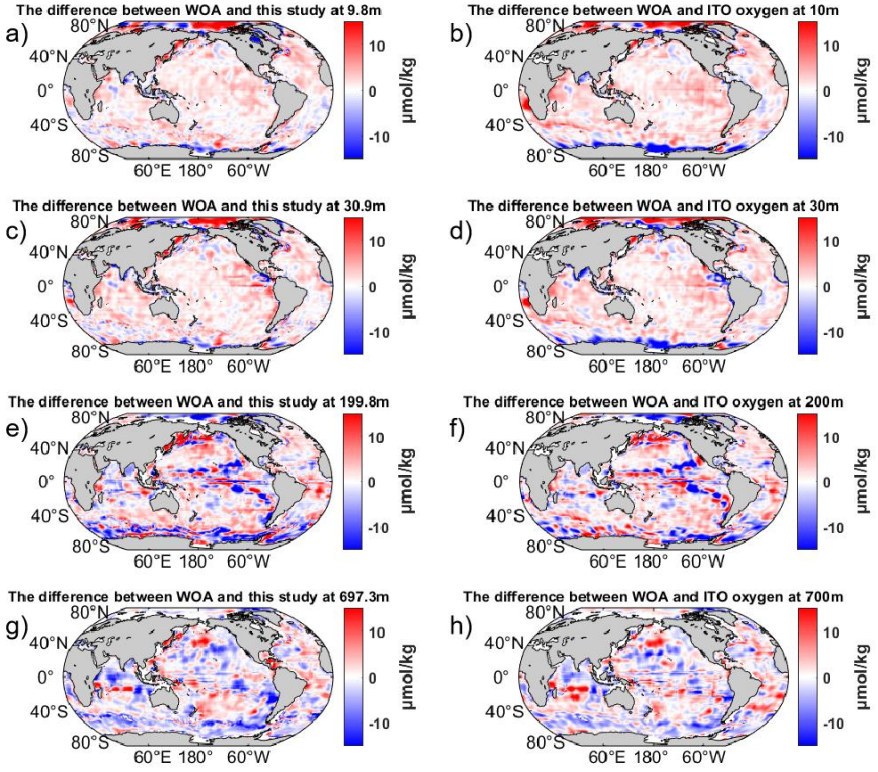


Figure 3. Maps of dissolved oxygen anomalies relative to WOA23 climatology at four depths (10 m, 30 m, 200 m and 700 m). (a,c,e,g) Left panels show the difference between WOA23 and our reconstruction; (b,d,f,h) right panels show the difference between WOA23 and ITO's reconstruction at the same depth and time period. White areas indicate near-zero bias, while reds and blues denote positive and negative offsets, respectively (±15 μmol kg^{-1}).

Our reconstruction more closely aligns with the WOA23 climatology, particularly in the surface and upper-thermocline layers, demonstrating improved accuracy in low- and mid-latitude regions. At the surface layer around 10 m depth, our reconstruction shows light difference, with anomalies up to ±2 μmol kg^{-1} deviations in high-latitude regions and equatorial upwelling zones. ITO's reconstruction exhibits broader swaths of red (negative bias of 4-8 μmol kg^{-1}) in subtropical gyres and pronounced blue (positive bias of 6-10 μmol kg^{-1}) under the Antarctic Circumpolar Current. At 30 m our differences remain small in the gyres and mid-latitudes for our ensemble reconstruction, with slightly more variability in boundary currents. In contrast, ITO's reconstruction again shows larger negative offsets in oligotrophic subtropics and positive offsets in the southern high latitudes. Our reconstruction provides a clear advantage in matching WOA23 in surface ocean. This improvement likely arises from our use of diverse environmental factors and dynamic weighting of multiple learners, which together capture the characteristics of dissolved oxygen distribution in the surface and upper-thermocline more accurately.

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

At around 200 m, both our and ITO's reconstruction exhibit modest departures from the WOA23 reference, with anomalies mostly within ±10 µmol kg^{-1} in the tropical and subtropical regions. ITO's biases in the eastern boundary upwelling zones and the Arabian Sea also fall in a similar range. Deeper, around 700 m, our reconstruction and WOA23 remain within ±3 µmol kg^{-1} even in the deep basins of the Atlantic and Pacific, indicating comparable performance at mid-depths.

5 Global ocean DO distribution and trends

5.1 Spatial distributions at representative depths

We explore the spatial patterns of reconstructed DO at four depths: 0.5 m, 200 m, 857 m and 4093 m (Fig. 4). These levels represent the surface, the thermocline, the core of the global oxygen minimum zones (OMZs), and the deep bathypelagic, respectively. At 0.5 m, DO reaches relatively high values in high-latitude regions (> 300 µ mol kg^{-1}). At 200 m, the thermocline/upper oxycline produces strong horizontal gradients. Tropical OMZs off Peru and in the Arabian Sea appear as blue corridors (< 160 µ mol kg^{-1}), while subpolar zones remain green-yellow (> 200 µ mol kg^{-1}). At 857 m, OMZ cores deepen. The eastern Pacific and northern Indian Ocean drop below 150 µ mol kg^{-1}. The North Atlantic mid-depths maintain 180-200 µ mol kg^{-1}. At 4093 m, DO is more uniform (180-200 µ mol kg^{-1}) across basins, shaped by Antarctic Bottom Water and North Atlantic Deep Water signatures (Talley, 2013). These four depths capture the full vertical DO structure, from dynamic surface processes through OMZs to the slowly evolving deep ocean, and lay the groundwork for trend and basin-scale analyses.

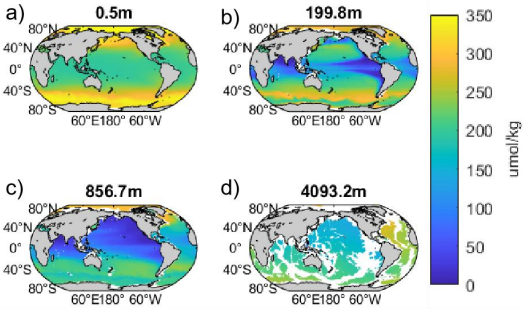


Figure 4. Global maps of reconstructed dissolved oxygen concentration at four representative depths: (a) 0.5 m (surface), (b) 200 m (thermocline), (c) 857 m (OMZ core) and (d) 4093 m (bathypelagic).

5.2 Global mean DO and deoxygenation profiles

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

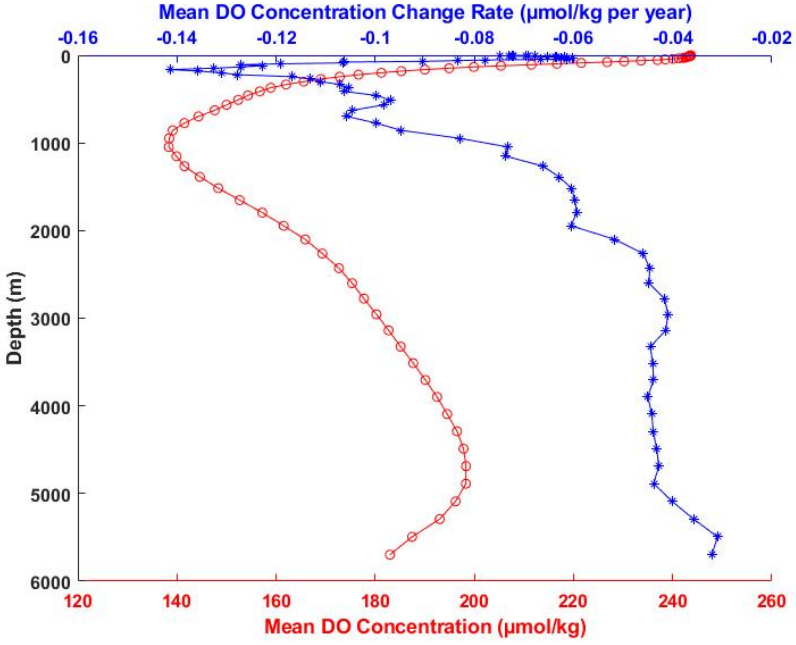


Figure 5. Global mean dissolved oxygen concentration (red circles, left axis) and its long-term change rate (blue stars, right axis) as a function of depth (1960–2023).

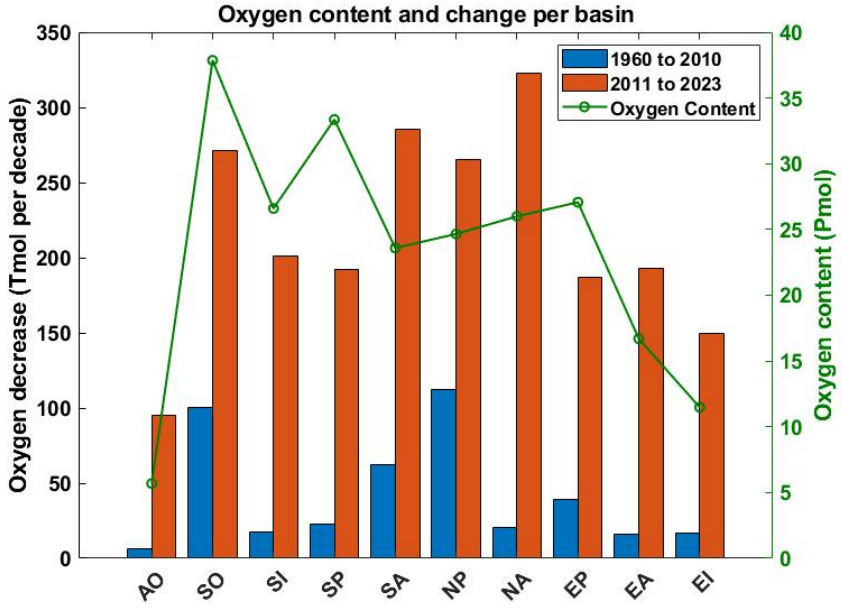
The vertical distribution of DO concentrations varies greatly with depth. Along the water column, DO concentration peaks at the surface around 250 μ mol kg^{-1} (Figure 5), where air-sea exchange and photosynthetic production maintain supersaturated conditions (Ryther, 1956; Kolber et al., 2000). Between approximately 70 m and 300 m, it falls rapidly to about 160 μ mol kg^{-1}, reflecting diminished gas exchange and ongoing microbial respiration (Keeling et al., 2010; Schmidtko et al., 2017). At around 1000 m, the profile exhibits a local minimum near the classical oxygen minimum zone (OMZ), then slowly rising toward 200 μ mol kg^{-1} at 2 000 m. Beyond 2 000 m DO gradually decreases further to 180 μ mol kg^{-1} near 6000 m.

The deoxygenation rate also shows a distinct depth-dependent pattern. The deoxygenation rate (Figure 5) is small (-0.06 μ mol kg^{-1} yr^{-1}) in the surface. Below 60 m, the decline accelerates sharply. Between 150 and 200 m, it steepens sharply to about -0.14 μ mol kg^{-1} yr^{-1}. This points to an amplification of shallow subsurface oxygen loss, most likely driven by stronger stratification that inhibits ventilative exchange and by increased microbial respiration (Keeling et al., 2010; Schmidtko et al., 2017). Below 2000 m, the deoxygenation rate lessens, dropping to about -0.05 μ mol kg^{-1} yr^{-1} at 4000 m and further to -0.04 μ mol kg^{-1} yr^{-1} at abyssal depths. The weakening of deoxygenation below 2000 m reflects the long renewal times and strong isolation of deep ocean water masses, which buffer them against rapid surface driven changes. Observational syntheses likewise report smaller long-term oxygen declines at depth compared to the upper ocean, consistent with slower ventilation and lower remineralization rates in these remote layers (Schmidtko et al., 2017).

Our reconstruction captures the full vertical extent of historical deoxygenation from surface to

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

462 abyss, providing continuous monthly fields down to nearly 6 000 m depth and offering a unique,
463 high-resolution resource that enables explicit evaluation of deep-water oxygen changes and
464 supports future climate, biogeochemical and ecosystem research.
465
466 **5.3 Basin-scale oxygen content and deoxygenation rates**



467
468 **Figure 6. Basin-scale integrated oxygen content and deoxygenation rates.** Green circles (right axis, in Pmol) show
469 each basin's mean oxygen content (1960–2023) for North Pacific (NP), Equatorial Pacific (EP), South Pacific (SP),
470 North Atlantic (NA), Equatorial Atlantic (EA), South Atlantic (SA), Southern Ocean (SO), Arctic Ocean (AO), North
471 Indian (NI) and South Indian (SI). Blue bars (left axis, in Tmol decade^{-1}) indicate deoxygenation rates (i.e., linear
472 trend to each basin's time series of integrated oxygen) in 1960–2010; orange bars show rates in 2011–2023.
473
474 Basin-scale analysis reveals a clear intensification of basin-scale deoxygenation, with dramatic
475 increases in deoxygenation rates observed in the North Atlantic, Southern Ocean, and Arctic,
476 particularly since 2010. During the period from 1960 to 2010, deoxygenation rates are modest.
477 The North Pacific shows the largest loss (112 Tmol decade^{-1}), followed by the Southern Ocean
478 (100 Tmol decade^{-1}) and South Atlantic (63 Tmol decade^{-1}). Tropical basins (EI, EA) and the Arctic
479 exhibit minimal change (6-17 Tmol decade^{-1}). In contrast, the period after 2010 reveals
480 dramatic increases across all basins. The North Atlantic loss more than 15 times to 323 Tmol
481 decade^{-1}. The South Atlantic and Southern Ocean both exceed 270 Tmol decade^{-1}. Even the Arctic
482 accelerates from near zero to 95 Tmol decade^{-1}. The fastest deoxygenation occurs in the Arctic
483 Ocean and the North Atlantic, with both basins experiencing roughly 15 times increases in oxygen
484 loss. In the Arctic Ocean, rising temperatures have reduced oxygen solubility (Matear et al., 2003),
485 while accelerated sea-ice melt since the 2010s has freshened surface waters, strengthened
486 stratification, and suppressed deep-to-surface exchange (Solomon et al., 2021; Farmer et al.,

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

2021). Tropical areas also show a marked uptick in recent deoxygenation, likely driven by enhanced stratification, reduced mid-water ventilation and increased oxygen consumption for organic matter decomposition (Stramma et al., 2008; Schmidtko et al., 2017; Breitburg et al., 2018).

6 Conclusion and discussion

Our study generates the first global 1° × 1° monthly DO dataset from 1960 to 2023 that extends to nearly 6 000 m depth, achieved through the BEM-DOR (Bayesian Ensemble Machine-learning Dissolved Oxygen Reconstruction) framework. BEM-DOR integrates six tree based learners: Random Forest, XGBoost, LightGBM, CatBoost, ERT and Hist_GBT. These models are optimized using by Bayesian optimization and combined via dynamic weighting. Our ensemble moves well beyond simpler blends and applies soft weighting that combines each model's global cross-validation skill with local error performance, allowing the best-performing learner to dominate regionally and depth-wise while down-weighting those with larger local biases. As a result, our reconstruction achieves lower MAE and RMSE on GLODAPv2 and WOA comparisons, better resolves sharp oxycline features and deep ocean signals, and by integrating diverse model architectures with auxiliary environmental factors captures spatial and vertical DO variability more faithfully than fixed-weight or single-model approaches.

The vertical deoxygenation profile reveals accelerating oxygen loss between 150 and 200m at rates around -0.14 μ mol kg^{-1} yr^{-1}, while surface declines remain modest (-0.06 μ mol kg^{-1} yr^{-1}). Since 2010, basin-scale trends show dramatic acceleration, particularly in the North Atlantic and Arctic Ocean, consistent with observations of rising temperature, strengthening stratification (Matear et al., 2003; Solomon et al., 2021; Farmer et al., 2021).

We note several limitations and avenues for improvement. The 1° × 1° grid smooths small-scale features such as narrow boundary currents. Also, sensor biases in BGC-Argo still propagate into our training data, especially around steep oxyclines (Bittig et al., 2017; Bittig et al., 2018; Gouretski et al., 2024). Future work should incorporate more precisely calibrated Argo data and finer regional grids.

Overall, our dataset offers a unified, long-term view of ocean deoxygenation from surface to abyss and, by extending coverage into the bathypelagic realm, fills a critical observational void that enables studies of deep-ocean oxygen dynamics. Packaged in NetCDF with documented uncertainties, it provides a benchmark for Earth system models and a foundation for impact studies on marine habitats and biogeochemical cycles, and invites the community to explore trends, calibrate models and guide policy on ocean health under climate change.

Data Availability

The reconstructed global monthly dissolved oxygen dataset produced in this study is publicly available in CF-compliant NetCDF format via Zenodo at https://doi.org/10.5281/zenodo.15361819 (Han et al., 2025) under a Creative Commons Attribution 4.0 license. Source DO profile observations were obtained from the International Argo Program and the national programs that contribute to Argo (https://argo.ucsd.edu), the World

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

Ocean Database (WOD) maintained by the U.S. National Oceanic and Atmospheric Administration (NOAA; https://www.ncei.noaa.gov/products/world-ocean-database), and the Global Ocean Data Analysis Project version 2 (GLODAPv2). Environmental predictor fields were drawn from the ORAS5 ocean reanalysis provided by the European Centre for Medium-Range Weather Forecasts (ECMWF; https://cds.climate.copernicus.eu/datasets/reanalysis-oras5).

Acknowledgement

This research was supported by the National Natural Science Foundation of China (42341201, 42276201), the National Key Research and Development Program of China (2023YFF0805004), and the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (SL2022PT204). The authors also thank the International Argo Program and the national programs for providing Argo data (https://argo.ucsd.edu), National Oceanic and Atmospheric Administration (NOAA) for providing WOD data (https://www.ncei.noaa.gov/products/world-ocean-database), European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ORA-S5 data (https://cds.climate.copernicus.eu/datasets/reanalysis-oras5).

Author Contribution

M.H. and Y.Z. conceived and designed the study. M.H. performed the research and wrote the initial draft of this paper. M.H. and Y.Z. reviewed and edited the paper.

Competing interests

The authors declare no competing interests.

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

Reference

Akiba T, Sano S, Yanase T, et al. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, **2019**: 2623-2631.

Bergmeir C, Benítez J M. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, **2012**, 191: 192-213.

Berman-Frank I, Chen Y B, Gao Y, et al. Feedbacks between the nitrogen, carbon and oxygen cycles. *Nitrogen in the marine environment. Amsterdam The Netherlands: Elsevier Inc*, **2008**, 1537-63.

Bittig H C, Körtzinger A, Neill C, et al. Oxygen optode sensors: principle, characterization, calibration, and application in the ocean. *Frontiers in Marine Science*, **2018**, 4: 429.

Bittig H C, Körtzinger A. Update on response times, in-air measurements, and in situ drift for oxygen optodes on profiling platforms. *Ocean Science*, **2017**, 13(1): 1-11.

Bopp L, Resplandy L, Orr J C, et al. Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models. *Biogeosciences*, **2013**, 10(10): 6225-6245.

Breiman L. Random forests. *Machine learning*, **2001**, 45: 5-32.

Breitburg D, Levin L A, Oschlies A, et al. Declining oxygen in the global ocean and coastal waters. *Science*, **2018**, 359(6371): eaam7240.

Chen T, Guestrin C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. **2016**: 785-794.

Cocco V, Joos F, Steinacher M, et al. Oxygen and indicators of stress for marine life in multi-model global warming projections. *Biogeosciences*, **2013**, 10(3): 1849-1868.

Deutsch C, Brix H, Ito T, et al. Climate-forced variability of ocean hypoxia. *Science*, **2011**, 333(6040): 336-339.

Dietterich T G. Ensemble methods in machine learning. *International workshop on multiple classifier systems. Berlin, Heidelberg: Springer Berlin Heidelberg*, **2000**: 1-15.

Farmer J R, Sigman D M, Granger J, et al. Arctic Ocean stratification set by sea level and freshwater inputs since the last ice age. *Nature Geoscience*, **2021**, 14(9): 684-689.

Friedman J H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, **2001**: 1189-1232.

Gade K. A non-singular horizontal position representation. *The journal of navigation*, **2010**, 63(3): 395-417.

Garcia, Hernan E., et al. *World Ocean Atlas 2023, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, Dissolved Oxygen Saturation and 30-year Climate Normal.* **2024**.

Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Machine learning*, **2006**, 63: 3-42.

Giglio D, Lyubchich V, Mazloff M R. Estimating oxygen in the Southern Ocean using Argo temperature and salinity. *Journal of Geophysical Research: Oceans*, **2018**, 123(6): 4280-4297.

Gouretski V, Cheng L, Du J, et al. A consistent ocean oxygen profile dataset with new quality control and bias assessment. *Earth System Science Data*, **2024**, 16(12): 5503-5530.

Gruber N. The dynamics of the marine nitrogen cycle and its influence on atmospheric CO 2 variations. *The ocean carbon cycle and climate*. Springer Netherlands, **2004**, 97-148.

Gruber N. Warming up, turning sour, losing breath: ocean biogeochemistry under global change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering*

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

Sciences, **2011**, 369(1943): 1980-1996.

Guryanov A. Histogram-based algorithm for building gradient boosting ensembles of piecewise linear decision trees. *Analysis of Images, Social Networks and Texts: 8th International Conference, AIST 2019, Kazan, Russia, July 17–19, 2019, Revised Selected Papers 8. Springer International Publishing*, **2019**: 39-50.

Han M, Zhou Y. Global Monthly Dissolved Oxygen Reconstruction via Bayesian Ensemble Machine Learning. *Zenodo*, **2025**. https://doi.org/10.5281/zenodo.15361819

Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, **2017**, 30.

Keeling R F, Körtzinger A, Gruber N. Ocean deoxygenation in a warming world. *Annual review of marine science*, **2010**, 2: 199-229.

Kolber Z S, Van Dover C L, Niederman R A, et al. Bacterial photosynthesis in surface waters of the open ocean. *Nature*, **2000**, 407(6801): 177-179.

Kwiatkowski L, Torres O, Bopp L, et al. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences*, **2020**, 17(13): 3439-3470.

Long M C, Deutsch C, Ito T. Finding forced trends in oceanic oxygen. *Global Biogeochemical Cycles*, **2016**, 30(2): 381-397.

Matear R J, Hirst A C. Long-term changes in dissolved oxygen concentrations in the ocean caused by protracted global warming. *Global Biogeochemical Cycles*, **2003**, 17(4).

Mishonov A.V., T. P. Boyer, O. K. Baranova, et al. World Ocean Database 2023. C. Bouchard, Technical Ed., *NOAA Atlas NESDIS 97, 206 pp*, **2024**.

Olsen A, Key R M, Van Heuven S, et al. The Global Ocean Data Analysis Project version 2 (GLODAPv2)–an internally consistent data product for the world ocean. *Earth System Science Data*, **2016**, 8(2): 297-323.

Oschlies A, Brandt P, Stramma L, et al. Drivers and mechanisms of ocean deoxygenation. *Nature Geoscience*, **2018**, 11(7): 467-473.

Pathak R, Dasari H P, Ashok K, et al. Effects of multi-observations uncertainty and models similarity on climate change projections. *npj Climate and Atmospheric Science*, **2023**, 6(1): 144.

Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, **2018**, 31.

Roach, C. J., Bindoff, N. L. Developing a new oxygen atlas of the world's oceans using data interpolating variational analysis. *Journal of Atmospheric and Oceanic Technology*, **2023**, 40(11): 1475-1491.

Roberts D R, Bahn V, Ciuti S, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, **2017**, 40(8): 913-929.

Ryther J H. Photosynthesis in the Ocean as a Function of Light Intensity 1. *Limnology and Oceanography*, **1956**, 1(1): 61-70.

Sampaio E, Santos C, Rosa I C, et al. Impacts of hypoxic events surpass those of future ocean warming and acidification. *Nature Ecology & Evolution*, **2021**, 5(3): 311-321.

Schmidtko, S., Stramma, L. & Visbeck, M. Decline in global oceanic oxygen content during the past five decades. *Nature*, **2017**, 542(7641): 335-339.

Sharp J D, Fassbender A J, Carter B R, et al. GOBAI-O 2: temporally and spatially resolved fields of

https://doi.org/10.5194/essd-2025-273
Preprint. Discussion started: 2 June 2025

ocean interior dissolved oxygen over nearly two decades. *Earth System Science Data Discussions*, **2022**, 2022: 1-46.

Solomon A, Heuzé C, Rabe B, et al. Freshwater in the arctic ocean 2010–2019. *Ocean Science*, **2021**, 17(4): 1081-1102.

Stramma L, Johnson G C, Sprintall J, et al. Expanding oxygen-minimum zones in the tropical oceans. *Science*, **2008**, 320(5876): 655-658.

Stramma L, Prince E D, Schmidtko S, et al. Expansion of oxygen minimum zones may reduce available habitat for tropical pelagic fishes. *Nature Climate Change*, **2012**, 2(1): 33-37.

Stramma L, Schmidtko S, Levin L A, et al. Ocean oxygen minima expansions and their biological impacts. *Deep Sea Research Part I: Oceanographic Research Papers*, **2010**, 57(4): 587-595.

Talley L D. Closure of the global overturning circulation through the Indian, Pacific, and Southern Oceans: Schematics and transports. *Oceanography*, **2013**, 26(1): 80-97.

Tang W, Li Z, Cassar N. Machine learning estimates of global marine nitrogen fixation. *Journal of Geophysical Research: Biogeosciences*, **2019**, 124(3): 717-730.

Wong A P S, Wijffels S E, Riser S C, et al. Argo data 1999–2019: Two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Frontiers in Marine Science*, **2020**, 7: 700.

Zhou Y, Gong H, Zhou F. Responses of horizontally expanding oceanic oxygen minimum zones to climate change based on observations. *Geophysical Research Letters*, **2022**, 49(6): e2022GL097724.