



Supplement of

Global monthly ocean dissolved oxygen (1960–2023) reconstructed to 5902 m with BLENDR, a Bayesian-optimized ensemble learning framework

Mingyu Han et al.

Correspondence to: Yuntao Zhou (ytzhou@sjtu.edu.cn)

The copyright of individual parts of the supplement might differ from the article licence.

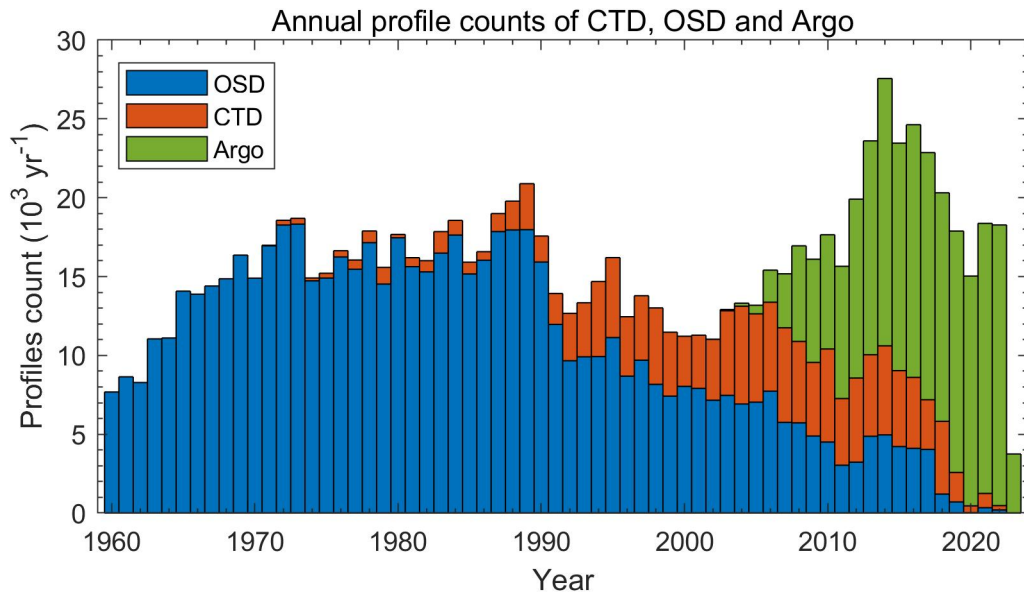


Figure S1. Annual profile counts of OSD (blue), CTD (orange), and Argo (green) observations used in this study. Bars show the number of dissolved oxygen profiles (in $10^3 \text{ profiles yr}^{-1}$) from each dataset, highlighting the dominance of ship-based OSD/CTD measurements before the 2000s and the rapid increase in Argo float profiles in recent years.

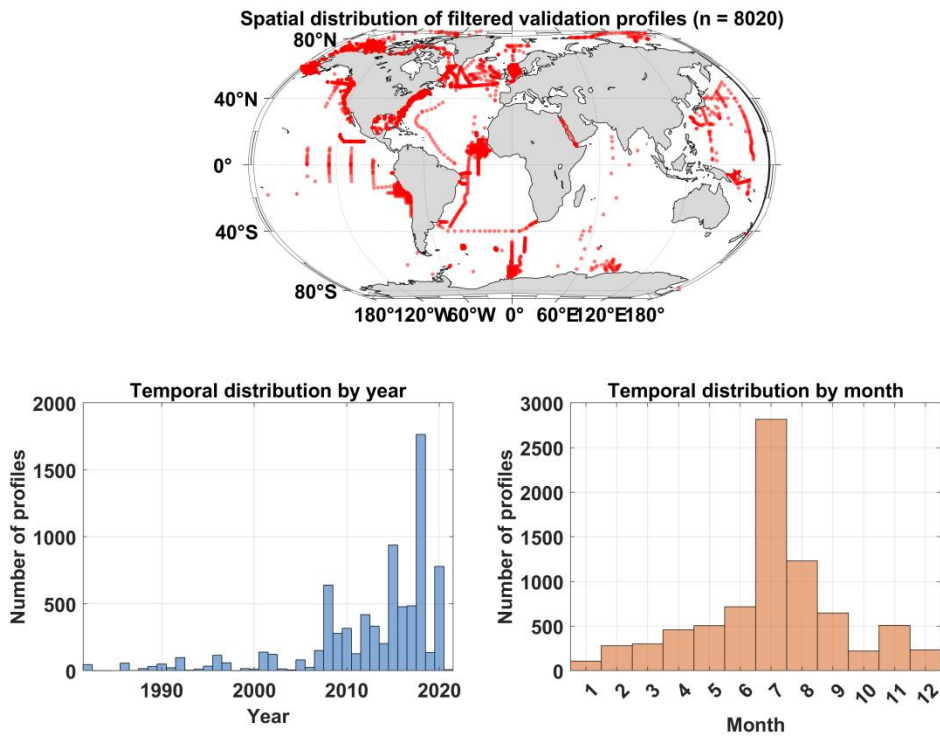


Figure S2. Spatial and temporal distribution of the filtered GLODAPv2 validation set. The upper panel shows the locations of the 8,020 unique validation profiles after excluding profiles matched with the CTD and OSD records used in model training. The lower panels show the temporal distribution of these profiles by year and by month.

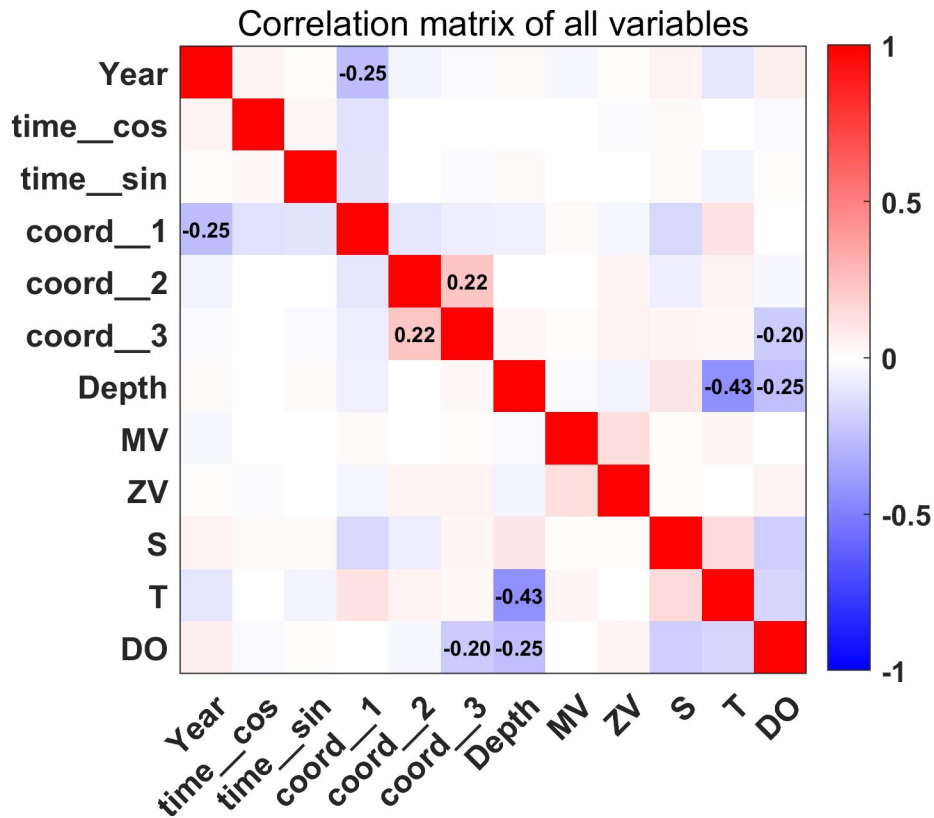


Figure S3. Correlation matrix of all predictor variables and dissolved oxygen (DO). Colors indicate Pearson correlation coefficients (red: positive, blue: negative, white: near zero), with the diagonal showing self-correlations. Numerical values are overlaid only for coefficients with absolute values greater than 0.2.

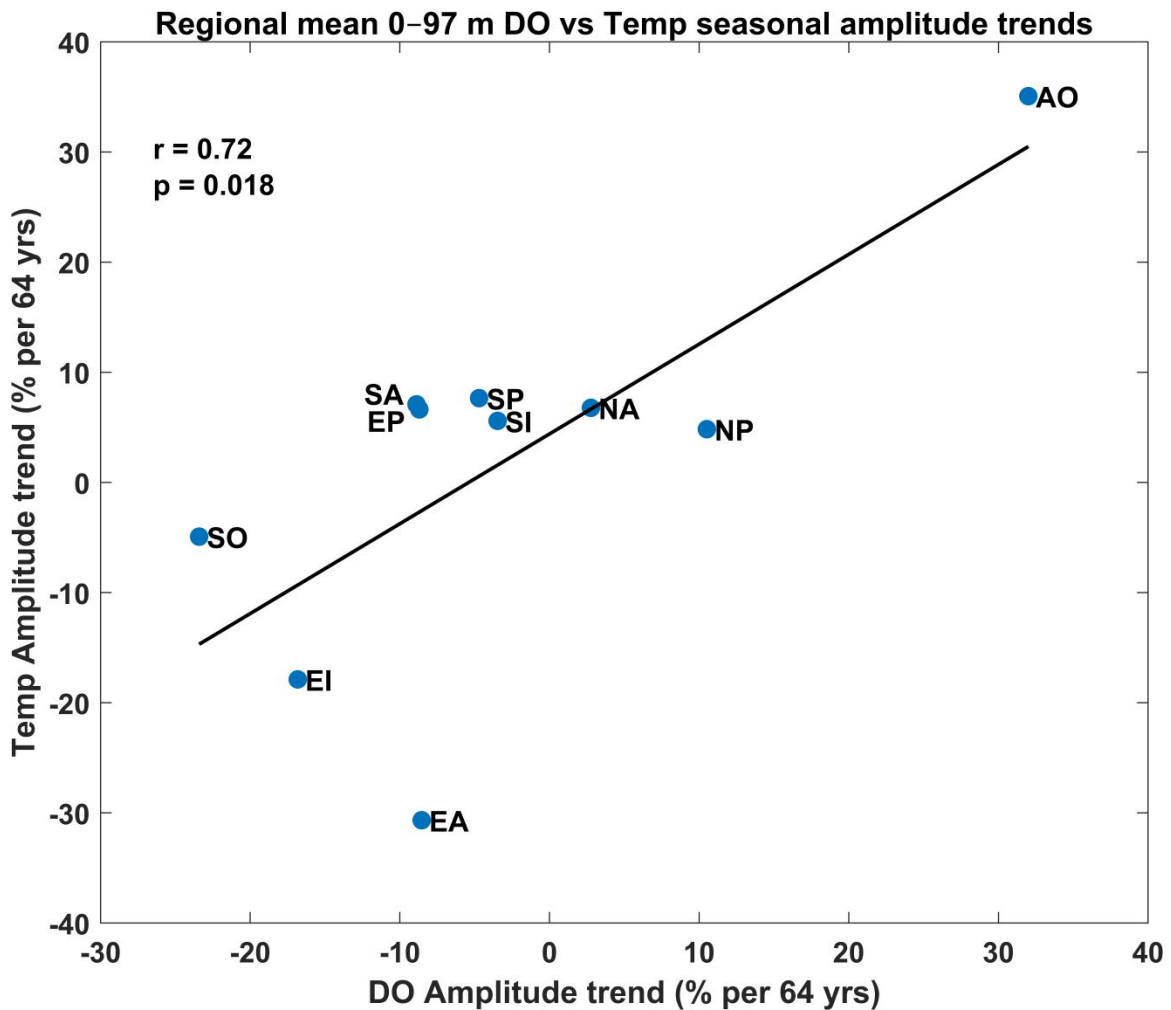


Figure S4. Relationship between regional trends in the seasonal amplitude of DO and temperature in the upper 0 – 97 m. Each point shows the basin - mean trend in DO seasonal amplitude (x axis) and temperature seasonal amplitude (y axis), expressed as percent change over 1960 – 2023, for the ten regions defined in Figure 6a of the main text (AO, NA, NP, EA, EI, EP, SA, SP, SI, SO). The solid line denotes the least-squares linear fit across basins, with correlation coefficient $r = 0.72$ and $p = 0.018$.

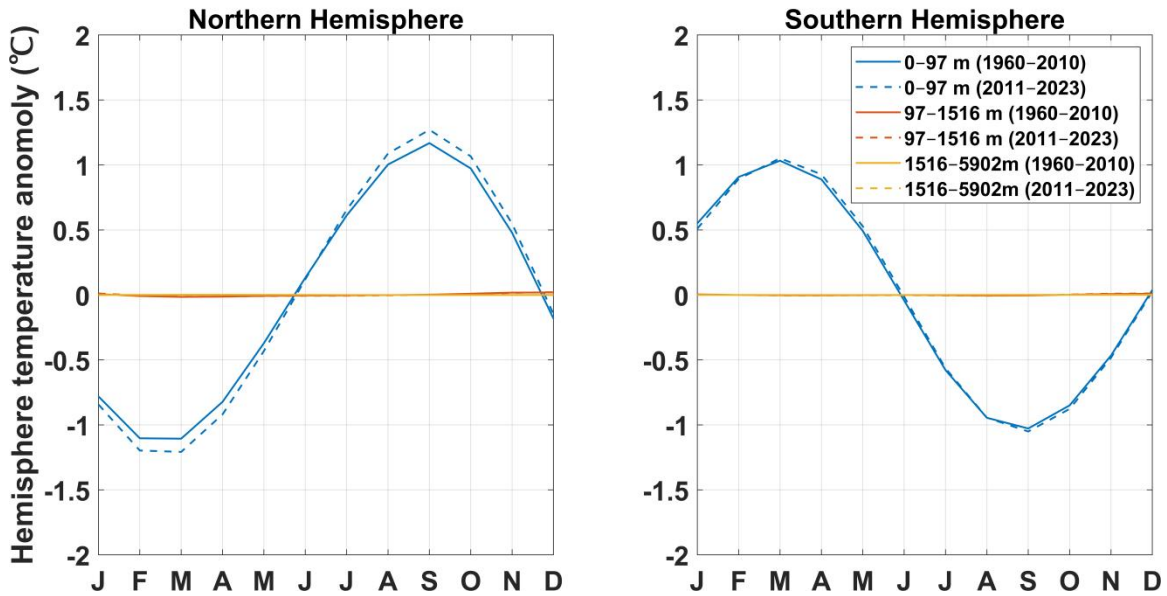


Figure S5. Climatological seasonal cycle of hemispheric mean temperature anomalies. The left panel shows the Northern Hemisphere and the right panel shows the Southern Hemisphere. The blue lines denote the surface layer (0 – 97 m), the red lines denote the thermocline and intermediate layer (97 – 1516 m), and the yellow lines denote the deep ocean (1516 – 5902 m). The solid lines represent anomalies from 1960 to 2010, and the dashed lines represent anomalies from 2011 to 2023.

Table S1. Number of available dissolved oxygen profiles below 2,000 m for each basin

Region	2000-3000m	3000-4000m	4000-5000m	5000-5902m
North Pacific	22867	10469	7230	3352
Equatorial Pacific	5988	3443	2307	809
South Pacific	5349	3357	2117	669
Arctic Ocean	8673	2555	58	3
North Atlantic	34629	12012	5678	1722
Equatorial Atlantic	4179	2953	2000	533
South Atlantic	4192	3380	2054	674
Southern Ocean	10098	6172	3262	545
North Indian Ocean	3664	2053	1351	340
South Indian Ocean	5148	2981	2043	648

Table S2. Vertical levels of the 75-layer depth used in this study.

Level	Depth (m)	Level	Depth (m)	Level	Depth (m)
1	0.51	26	120.00	51	1516.36
2	1.56	27	133.08	52	1652.57
3	2.67	28	147.41	53	1795.67
4	3.86	29	163.16	54	1945.30
5	5.14	30	180.55	55	2101.03
6	6.54	31	199.79	56	2262.42
7	8.09	32	221.14	57	2429.03
8	9.82	33	244.89	58	2600.38
9	11.77	34	271.36	59	2776.04
10	13.99	35	300.89	60	2955.57
11	16.53	36	333.86	61	3138.56
12	19.43	37	370.69	62	3324.64
13	22.76	38	411.79	63	3513.45
14	26.56	39	457.63	64	3704.66
15	30.87	40	508.64	65	3897.98
16	35.74	41	565.29	66	4093.16
17	41.18	42	628.03	67	4289.95
18	47.21	43	697.26	68	4488.15
19	53.85	44	773.37	69	4687.58
20	61.11	45	856.68	70	4888.07
21	69.02	46	947.45	71	5089.48
22	77.61	47	1045.85	72	5291.68
23	86.93	48	1151.99	73	5494.58
24	97.04	49	1265.86	74	5698.06
25	108.03	50	1387.38	75	5902.06

Table S3. Cross-validation Δ DO ($\mu\text{mol kg}^{-1}$)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	-0.191	0.222	-0.090	0.139	0.075	0.139	0.420	0.331
XGBoost	-0.605	0.177	-0.244	-0.001	-0.111	0.036	0.344	0.092
LightGBM	-0.525	0.353	-0.090	0.199	0.039	0.179	0.313	0.232
Hist_GBT	-0.262	0.203	-0.130	0.185	0.035	0.184	0.425	0.206
ERT	-0.331	-0.165	-0.355	-0.063	-0.118	0.002	0.634	0.199
CatBoost	-0.381	0.160	-0.122	0.101	0.059	0.153	0.379	0.265

Table S4. Cross-validation MAE ($\mu\text{mol kg}^{-1}$)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	10.381	9.882	9.781	9.607	9.513	9.434	9.378	9.753
XGBoost	11.750	11.313	11.256	11.089	10.787	10.787	10.863	11.158
LightGBM	11.035	10.543	10.482	10.300	10.170	10.107	10.161	10.371
Hist_GBT	11.422	11.014	10.875	10.686	10.626	10.544	10.532	10.827
ERT	11.123	10.686	10.422	10.317	10.199	10.143	10.143	10.448
CatBoost	11.036	10.586	10.452	10.288	10.208	10.097	10.086	10.383

Table S5. Cross-validation RMSE ($\mu\text{mol kg}^{-1}$)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8
RF	17.272	16.586	16.303	16.265	15.969	16.005	16.287	16.303
XGBoost	18.339	17.754	17.568	17.518	17.053	17.176	17.485	17.515
LightGBM	17.590	16.899	16.766	16.716	16.402	16.357	16.691	16.645
Hist_GBT	17.990	17.442	17.114	17.038	16.843	16.888	17.091	17.109
ERT	17.846	17.370	16.757	16.976	16.494	16.597	16.765	16.825
CatBoost	17.566	16.970	16.681	16.661	16.441	16.385	16.618	16.660

Table S6. 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	0.960	0.960	0.960	0.960	0.963
XGBoost	0.953	0.955	0.954	0.953	0.954	0.954	0.954	0.957
LightGBM	0.957	0.959	0.958	0.958	0.958	0.958	0.958	0.961
Hist_GBT	0.955	0.957	0.956	0.956	0.956	0.955	0.956	0.959
ERT	0.955	0.957	0.958	0.956	0.957	0.957	0.958	0.961
CatBoost	0.957	0.959	0.958	0.958	0.958	0.958	0.959	0.961

Table S7. Sensitivity of BLENDR performance on the filtered GLODAPv2 validation set to different spatiotemporal exclusion criteria.

	Remaining profiles	MAE	RMSE	R ²	Δ DO
$\pm 1^\circ$, same month	8020	10.204	18.139	0.968	-0.334
$\pm 5^\circ$, same year	2982	9.719	18.545	0.968	-0.788