**IAP/CAS (Institute of Atmospheric Physics, Chinese Academy of Sciences) Ocean Dissolved Oxygen Profile Dataset**

**Format Description (monthly\_MAT)**

IAP/CAS (Institute of Atmospheric Physics, Chinese Academy of Sciences) provides quality-controlled and bias-corrected ocean dissolved oxygen profile data. This dataset includes in situ ocean oxygen profile data from 3 major instruments: OSD, CTD, and Argo. OSD instrumentation is represented by bottle casts with oxygen determined by the Winkler method. CTD profiles are obtained mainly through the electrochemical sensors, whereas Argo float profiles contain data mainly obtained by optodes. The total number of profiles from all three platforms exceeds 1.2 million from 1920 to 2022.

The quality-controlled (QC) was implemented by IAP-QC including 9 different QC checks to identify different kinds of erroneous data and/or outliers. We used Gourteski et al., (2024) oxygen correction scheme to correct the remaining systematic bias in the Argo data. Applying the new QC procedure and bias adjustment resulted in a new global ocean oxygen dataset from 1920 to 2023 with consistent data quality across bottle samples, CTD casts, and Argo floats.

This README provides the variable information for the monthly\_MAT (MATLAB) files (from January 1940 to August 2023) for dissolved oxygen *in-situ* observations. This is stored in monthly data. **However, it should be noted that due to the data gap, some monthly files are missed.**

The original data sources are (i) WOD (World Ocean Database) and (ii) GDAC (Global Data Assemble Center). **As the metadata format and definition varies between these two databases, in CODC-GOSD, we separated into two groups:**

1. **Group 1**: OSD and CTD profiles originated from WOD (DIRECTORY path: IAP\_Dissolved\_oxygen\_CTD\_OSD\_mat; <http://www.ocean.iap.ac.cn/ftp/cheng/IAP_oxygen_profile_dataset/IAP_Dissolved_oxygen_CTD_OSD_mat_202404/>)
2. **Group 2:** Argo (PFL) profiles originated from GDAC (DIRECTORY path: IAP\_Dissolved\_oxygen\_ARGO\_mat; <http://www.ocean.iap.ac.cn/ftp/cheng/IAP_oxygen_profile_dataset/IAP_Dissolved_oxygen_ARGO_mat_202404/>)

**VARIABLES in Group 1:**

**[1] DOprofile\_num\_all**: total number of profiles in the file.

**[2] DOXY\_ORIGIN**: original dissolved oxygen records without preforming any QC and adjustments (Dimension: 2000\* DOprofile\_num\_all)

**[3] Depth\_ORIGIN**: original depth records without preforming any QC and adjustments (Dimension: 2000\* DOprofile\_num\_all)

**[4] DOXY\_ADJUSTED\_IAPQC**: Quality Control flag of dissolved oxygen (adjusted data) by IAP-QC (Gouretski et al., 2024). These flags are corresponding to DOXY\_ADJUSTED. This is built upon < **DOXY\_origin\_QC>.** (Dimension: 2000\* DOprofile\_num\_all)

**[5]** **DOXY\_ADJUSTED\_IAPQC\_checks**: Quality Control flag for each QC module of dissolved oxygen (adjusted data) by IAP-QC (Gouretski et al., 2024). The checks order is: Geographical Location Check, Crude range check, Maximum oxygen solubility check, Stucked value check, Spike check, Multiple extrema check, Oxygen Vertical Gradient check, Local Climatological range check, Excessive flagged level percentage check. (Dimension: 2000\* DOprofile\_num\_all\*QC module numbers)

**[6] Depth\_QCed\_Adjusted\_IAP**: Interpolated depth (10m interval) adjusted after IAP-QC and bias adjustement (Gouretski et al., 2024). (Dimension: 2000\* DOprofile\_num\_all)

**[7] DOXY\_QCed\_ADJUSTED\_IAP**: in-situ dissolved oxygen measurements after 10m interpolation, IAP-quality control and IAP bias corrections (Gouretski et al., 2024). This is used with < **Depth\_QCed\_Adjusted\_IAP** >. (Dimension: 2000\* DOprofile\_num\_all)

**[8]** **DO\_profile\_info\_record\_all**: Metadata information of each profile (float/double variables). There are 14 types of metadata in total. The definition of each line is obtained by < **DOprofile\_info\_record\_name**>. (Dimension: 15\* DOprofile\_num\_all)

**[9] DO\_profile\_info\_str\_all**: Metadata information of each profile (string variables). There are 14 types of metadata in total. The definition of each line is obtained by < **DOprofile\_info\_str\_name**>. (Dimension: 14\* DOprofile\_num\_all)

**VARIABLES in Group 2:**

**[1] DOprofile\_num\_all**: total number of profiles in the file.

**[2] DOXY\_origin**: original dissolved oxygen records without preforming any QC and adjustments (Dimension: 2000\* DOprofile\_num\_all)

**[3] PRES\_origin**: original depth records without preforming any QC and adjustments (Dimension: 2000\* DOprofile\_num\_all)

**[4] DOXY\_origin\_QC**: quality control flag of original dissolved oxygen records provided by GDAC (Dimension: 2000\* DOprofile\_num\_all)

**[5] DOXY\_ADJUSTED**: original depth records with adjustments provided by GDAC. (Dimension: 2000\* DOprofile\_num\_all)

**[6] DOXY\_ADJUSTED\_QC**: quality control flag of adjusted dissolved oxygen records provided by GDAC (Dimension: 2000\* DOprofile\_num\_all)

**[7] DOXY\_ADJUSTED\_ERROR**: Contains the error on the adjusted values as determined by the delayed mode QC process (Dimension: 2000\* DOprofile\_num\_all)

**[9] DOXY\_ADJUSTED\_IAPQC**: Quality Control flag of dissolved oxygen (adjusted data) by IAP-QC (Gouretski et al., 2024). These flags are corresponding to DOXY\_ADJUSTED. This is built upon < **DOXY\_origin\_QC>.** (Dimension: 2000\* DOprofile\_num\_all)

**[10]** **DOXY\_ADJUSTED\_IAPQC\_checks**: Quality Control flag for each QC module of dissolved oxygen (adjusted data) by IAP-QC (Gouretski et al., 2024). These flags are corresponding to <**DOXY\_ADJUSTED**>. The checks order is: Geographical Location Check, Crude range check, Maximum oxygen solubility check, Stucked value check, Spike check, Multiple extrema check, Oxygen Vertical Gradient check, Local Climatological range check, Excessive flagged level percentage check. (Dimension: 2000\* DOprofile\_num\_all\*QC module numbers)

**[11]** **PRES\_ADJUSTED**: Adjusted pressure measurements by GDAC (Global Data Assemble Center). (Dimension: 2000\* DOprofile\_num\_all)

**[12]** **PRES\_ADJUSTED\_QC**: Quality flag by GDAC (Global Data Assemble Center). (Dimension: 2000\* DOprofile\_num\_all)

**[13] PRES\_ADJUSTED\_ERROR**: Contains the error on the adjusted values as determined by the delayed mode QC process by GDAC. (Dimension: 2000\* DOprofile\_num\_all)

**[14] Depth\_QCed\_Adjusted\_IAP**: Interpolated depth (10m interval) adjusted after IAP-QC and bias adjustement (Gouretski et al., 2024). (Dimension: 2000\* DOprofile\_num\_all)

**[15] DOXY\_QCed\_ADJUSTED\_IAP**: in-situ dissolved oxygen measurements after 10m interpolation, IAP-quality control and IAP bias corrections (Gouretski et al., 2024). This is used with < **Depth\_QCed\_Adjusted\_IAP** >. (Dimension: 2000\* DOprofile\_num\_all)

**[16]** **DO\_profile\_info\_record\_all**: Metadata information of each profile (float/double variables). There are 14 types of metadata in total. The definition of each line is obtained by < **DOprofile\_info\_record\_name**>. (Dimension: 15\* DOprofile\_num\_all)

**[17] DO\_profile\_info\_str\_all**: Metadata information of each profile (string variables). There are 14 types of metadata in total. The definition of each line is obtained by < **DOprofile\_info\_str\_name**>. (Dimension: 14\* DOprofile\_num\_all)

**[18] PSAL\_origin, PSAL\_origin\_QC, PSAL\_ADJUSTED, PSAL\_ADJUSTED\_ERROR, PSAL\_ADJUSTED\_QC**: The same as DOXY etc., but for salinity measurements. (Dimension: 2000\* DOprofile\_num\_all)

**[19]** **TEMP\_origin, TEMP\_origin\_QC, TEMP\_ADJUSTED, TEMP\_ADJUSTED\_ERROR, TEMP\_ADJUSTED\_QC**: The same as DOXY etc., but for temperature measurements. (Dimension: 2000\* DOprofile\_num\_all)

**NOTES:**

**[1] We strongly recommended to use the variables < DOXY\_QCed\_ADJUSTED\_IAP > and < Depth\_QCed\_Adjusted\_IAP >**, which has been applied a new quality control method (IAP-QC) and the oxygen adjustment scheme developed by IAP (Gourteski et al., 2024) based on the <DOXY\_ADJUSTED> and <PRES\_ADJUSTED> provided by GDAC. Also, the data has been interpolated into 10m depth intervals.

**[2]** Each column represents the observations and its corresponding metadata information for each profile. For example, the method for extracting the 1001st profile is:

* Depth records: **Depth\_QCed\_Adjusted\_IAP** (1:1001);
* Dissoved oxygen records: **DOXY\_QCed\_ADJUSTED\_IAP** (:,1:1001);
* Metadata information:

DO\_profile\_info\_record\_all(:,1001);

DO\_profile\_info\_str\_all(:,1001);

**[3]** For a detailed description of each variable (such as unit, dimension, default value definition, long\_name, etc.), see the variable description in the netCDF file.

**DATA CITATIONS:**

[1] Viktor Gourteski, Lijing Cheng, Juan Du, Xiaogang Xing, Fei Chai, Zhetao Tan. 2024. A consistent ocean oxygen profile dataset with new quality control and bias assessment. *Earth System Science Data Discussions*, 2024, 2024: 1-27.

Contact:

Viktor Gouretski and Lijing Cheng

IAP/CAS

Viktor Gouretski (viktor.gouretski@posteo.de); chenglij@mail.iap.ac.cn

**14/04/2024**