https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Open Access Earth System Science Data Discussions

A consistent ocean oxygen profile dataset with new quality control and bias assessment

Viktor Gourteski[1], Lijing Cheng[1], Juan Du[1], Xiaogang Xing[2], Fei Chai[3,4]
[1] Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[2] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China
[3] State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China
[4] College of Ocean and Earth Sciences, Xiamen University, Xiamen, China

Correspondence to: Viktor Gouretski (viktor.gouretski@posteo.de);
Lijing Cheng (chenglij@mail.iap.ac.cn)

Abstract. The global ocean oxygen levels have declined in the past decades, posing threats to marine life and human society. High-quality and bias-free observations are crucial to understanding the ocean oxygen changes and assessing their impact. Here, we propose a new automated quality control procedure for ocean profile oxygen data. This procedure consists of a suite of nine quality checks, with outlier rejection thresholds being defined based on underlying statistics of the data. The procedure is applied to three main instrumentation types: bottle casts, CTD (Conductivity-Temperature-Depth) casts, and Argo profiling floats. Application of the quality control procedure to several manually quality-controlled datasets of good quality suggests the ability of the scheme to successfully identify outliers in the data. Collocated quality-controlled oxygen profiles obtained utilizing the Winkler titration method are used as unbiased references to estimate possible residual biases in the oxygen sensor data. The residual bias is negligible for electrochemical sensors typically used on CTD casts. We explain this as the consequence of adjusting to the concurrent sample Winkler data. However, our analysis finds a prevailing negative residual bias for the delayed-mode quality-controlled adjusted Argo profiling floats varying from -4 to -1 µmol kg^{-1} among the data adjusted by different Argo data assembly centers (DACs). The respective overall DAC-specific corrections are suggested. Applying the new QC procedure and bias adjustment resulted in a new global ocean oxygen dataset from 1920 to 2022 with consistent data quality across bottle samples, CTD casts, and Argo floats. The adjusted Argo profile data is available at the Marine Science Data Center of the Chinese Academy of Sciences (Gouretski et al., 2023, http://dx.doi.org/10.12157/IOCAS.20231208.001)

1 Introduction

Progressive warming caused by the human-induced increase of the greenhouse gases in the Earth's atmosphere leads to the decline of the dissolved oxygen concentration in the global ocean because of the reduction in oxygen solubility, the increase in stratification which hampers the exchange between the surface layer and the ocean interior, and the accompanying change of ocean circulation (Keeling et al., 2010; Gregoire et al., 2021). Another factor related to human activities is the increasing input of nutrients from agriculture and wastewater. Nutrients facilitate a dense growth of phytoplankton and microbes subsequently decrease of oxygen after the phytoplankton dies. Recognizing the crucial role of dissolved oxygen for marine aerobic organisms, oceanographers started to measure oxygen in the late 19th century

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Open Access Earth System Science Data Discussions

using the chemical method developed by Winkler (1888). Since then, Winkler titration has been a standard method used on oceanographic ships and in laboratories (Langdon, 2010), and the technique has an accuracy estimated to be 0.1% or ±0.3 μmol kg^{-1} (Carpenter, 1965).

With the rapid technological progress during the 1960-70s and the development of the electronic CTD (Conductivity-Temperature-Depth) profilers, the first electrochemical sensors appeared, providing the possibility for continuous oxygen profiling, which is not possible with the Winkler method restricted by water samples from several depth levels. Electrochemical sensors are based on a Clark polarographic membrane (Clark et al., 1953). Oxygen concentration outside the membrane and oxygen diffusion through the membrane determine the sensor response. Electrochemical Clark-type sensors possess a very fast time response (<1 s), with an initial accuracy of 2% of oxygen saturation and precision of about 1 μmol kg^{-1} (Coppola et al., 2013). Sensor drift due to membrane fouling and changes in electrolyte over time requires periodic calibration. The first type of sensors applied on Biogeochemical Argo profiling floats (BGC floats) were Clark-type electrodes (Riser and Johnson, 2008).

Optical oxygen sensors called "optodes" are based on the principle of fluorescence quenching of a fluorescent indicator embedded in a sensing foil (Körtzinger et al. 2005, Tengberg et al., 2006). The optode sensors appeared soon after the first iplementation of the Clark-type sensors on Argo floats. Compared to electrochemical sensors, optodes are characterized by long-term stability and high precision with the disadvantage of a slower response time. During the initial period of several years both Clarke-type and optode sensors were used on Argo floats (Claustre et al., 2020). However drift and initial calibration issues with electrochemical sensors has lead to the increased implementation of optodes on Argo floats (Claustre et al., 2020) , for which calibration using simultaneous water samples is not possible.

Different techniques have been applied in the past to collect ocean oxygen data, and the total number of oxygen profile data reached a total of more than 1.2 million until 2023. However, there are a lot of data quality issues in the historical oxygen database due to many reasons, including instrumental errors, data collection failure, data processing errors, improper sample storage, unit conversion and others. These quality issues impede the various applications of oxygen data, for instance, investigating how much oxygen the ocean has lost in the past decades (Gregoire et al., 2021). Furthermore, as different instruments have different data quality, merging different instrumentation types into an integrated database requires a proof of data consistency.

To provide a quality-consistent database for oxygen, this study presents an automated quality control procedure for ocean oxygen profiles and analyzes the quality of oxygen data obtained by different instrumentation types. We further assess remaining oxygen biases for CTD and Argo oxygen profiles, comparing them with the reference bottle sample data obtained through Winkler method. The rest of the paper is organized as follows. The data and methods employed in the study are presented in Section 2. The data quality-control procedure is introduced in Section 3, with the data quality assessment presented in Section 4. Results of the benchmarking of the automated quality control procedure using manually controlled datasets is shown in Section 5. Assessment of the residual bias for Argo and CTD profiles is conducted in Section 6. Data availability is described in Section 7. The results of the study are summarized and discussed in Section 8.

2 Global archive of dissolved oxygen profiles

For the current study we used data from two large depositories: 1) World Ocean Database (as of January 2023); 2) Oxygen profiles from the Argo Global Assembly Center (GDAC) (ARGO, 2000). The WOD Argo profiles represent a blend of not-adjusted (real-time) and adjusted (delayed-mode) profiles, whereas for the data from GDACs it is possible

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

to discriminated between adjusted and unadjusted profiles. For this reason, the analysis of biases for Argo data is based on the data from the Argo GDAC.

2.1 Oxygen profiles from the World Ocean Database and DACs

World Ocean Database (Boyer et al., 2018) represents the largest depositary of the dissolved oxygen profile data, which is composed of three main instrumentation types: 1) Ocean Station Data (OSD), 2) high-resolution CTD profiles, and 3) Argo profiling float (PFL) data. OSD instrumentation group 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, so a manual quality control of the global oxygen dataset is nearly impossible.

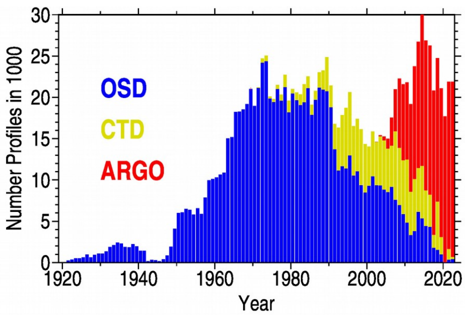

Figure 1. Yearly number of oxygen profiles from the World Ocean Database (OSD and CTD profiles) and from national DACs.

The OSD profiles are most abundant between 1960s to 2000s, CTD profiles between 1990s to 2010s, and Argo profiles dominate after 2010 (Fig. 1). The geographical distribution of oxygen profiles is inhomogeneous (Fig. 2). The OSD profiles exhibit considerably better sampling compared to CTD and Argo, with dense sampling in near-coastal areas and a sparser sampling in the central parts of the oceans (Fig. 2a). The CTD profiles are limited to transoceanic sections, leaving large data gaps especially in the central regions of Pacific, Indian, and Southern oceans (Fig. 2b). Oxygen profiles from ten national Argo DACs have been used for the current study, with the number of profiles given in Table 1. The most considerable contribution comes from two DACs: Atlantic Oceanographic and Meteorological Laboratory (AOML) and French CORIOLIS Center (Coriolis). Together these two DACs contribute with 71% of all oxygen profiles. The global sampling by Argo floats is characterized by big gaps in the tropical belt of the World Ocean (Fig. 2c) and in the marginal seas with shallow bottom depth.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

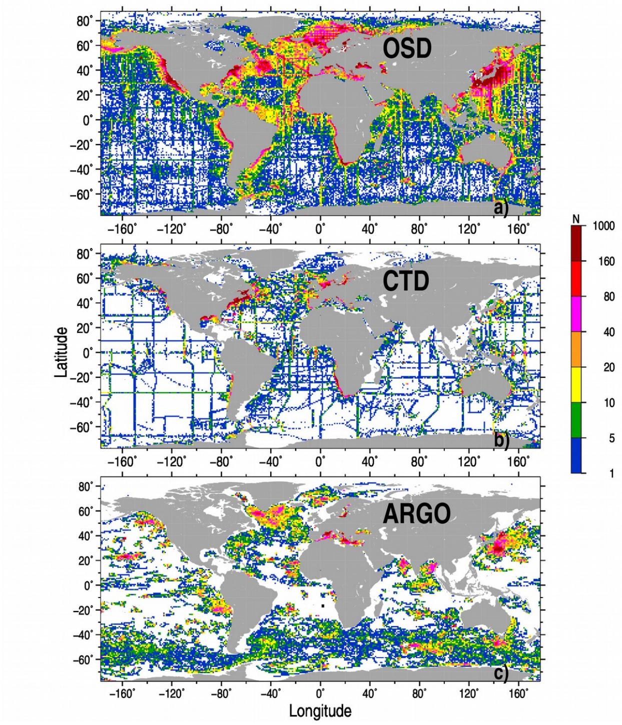

Figure 2 Number of profiles (N) in 1°×1° latitude/longitude squares for OSD (a), CTD (b), and Argo (c) data

The DACs report oxygen data along with quality flags set after the quality control procedure performed in each DAC. Spatial distribution of the profiles from each DAC is shown in Fig. 3. Only the AOML dataset is characterized by a more or less global coverage. The profiles from the second large Coriolis dataset are concentrated mostly in the Atlantic and Southern oceans. Other DACs are characterized by a regional scope: JMA data come from the Pacific Ocean east of Japan, CSIRO profiles cover the Southern Ocean, CSIO mainly provides profiles in the subtropical and tropical western Pacific Ocean, BODC profiles are located in the Atlantic Ocean. Profiles from KORDI and KMA, the smallest two datasets, are concentrated in the southern part of the Sea of Japan.

3 Data quality control

Quality evaluation of hydrographic data typically consists of two parts: data quality control for random errors and evaluation of systematic errors or biases. These two issues often are treated separately but both essentially represent the entire quality control procedure. A unified quality control procedure has yet to be suggested for the global archive of oxygen profile data, and oxygen-related studies often rely on WOD (Garcia et al., 2018) and Argo (Thierry et al., 2021) quality control procedures. The efforts undertaken under the IQuOD initiative (Cowley, 2021) resulted in a comprehensive study where different quality control procedures for temperature profiles were compared and evaluated (Good et al., 2022). As shown in the previous section, the characteristic feature of the global oxygen data archive is its heterogeneity. In the early years, a relatively small amount of data permitted expert quality control, but for the actual global archive automated quality control procedures (AutoQC) are required.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Table 1. Argo oxygen profiles from different national DACs

N	National Data Assembly Center	Code Name	Number of Argo profiles	Number of Argo profiles collocated with Winkler profiles	Percent of Argo profiles having collocations with Winkler profiles
1	Atlantic Oceanographic and Meteorological Laboratory, US	AOML	89059	32396	41.08
2	CORIOLIS data Center, France	Coriolis	63220	33233	65.09
3	Commonwealth Scientific and Industrial Research Organisation, Australia	CSIRO	19183	3302	23.75
4	Japan Meteorological Agency, Japan	JMA	15981	11233	82.90
5	Indian National Centre for Ocean Information Services, India	INCOIS	9901	2069	33.09
6	Second Institute of Oceanography, Ministry of Natural Resources, China	CSIO	6455	3921	68.98
7	Marine Environmental Data Service, Canada	MEDS	4605	14.04	50.50
8	British Oceanographic Data Center, UK	BODC	3533	1905	61.57
9	Korea Ocean Research and Development Institute, Korea	KORDI	2239	0	0
10	Korea Meteorological Administration, Korea	KMA	93	0	0

The AutoQC procedure aims to identify and flag outliers, which represent observations significantly deviating from the majority of other data in the population. Monhor and Takemoto (2005) noted that there is no rigid mathematical definition of an outlier. The outliers do not necessarily represent erroneous data and can occur due to the natural variability of the measured variable. A quality control procedure defines outliers using a set of thresholds, which are based on physical laws (for instance, the maximum solubility of gases in the water) or are defined based on the statistical properties of the data population.

To increase the reliability in detecting erroneous data a set of quality-checks is applied to each profile. The larger the number of failed distinct quality checks, the higher the probability that the flagged observation represents a data outlier. Based on the available quality control schemes for oceanographic data (most of them were developed for temperature and/or salinity profiles) quality checks can be subdivided into the following groups:

1. Check of location, date and bottom depth of the profile.
2. Check of profile attributes (maximum sampled depth, number of levels, variables measured) correspond to the attributes of the instrumentation type.
3. Range check, e.g. comparison of observations at each level against minimum/maximum value thresholds, which are set for the entire ocean, oceanic basin (global ranges) or for the particular location and depth (local ranges).
4. Check of the profile shape, which is characterized by the vertical gradient of the measured variable at observed levels, by the number of local extrema, and by the presence of spikes.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

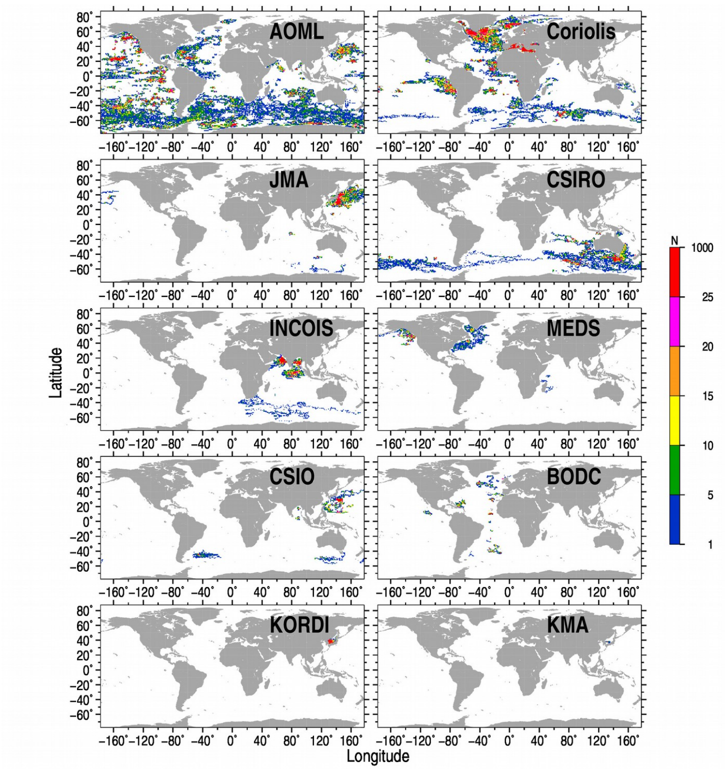

Figure 3. The number (N) of oxygen profiles in 1°×1° spatial boxes for different DACs. The name of each DAC is shown on Asia.

Quality control procedures often assume Gaussian distribution law, and outliers are defined in terms of multiple times the standard deviation from the mean value (Z-score method). However, distributions of oceanographic parameters are typically skewed, and the assumption of Gaussian distribution leads to false data rejection. Hubert and Vandervieren (2008) developed the adjusted Tukey's boxplot method for skewed distribution with a more accurate representation of possible outliers. Following this approach, Gouretski (2018) and Tan et al. (2023) applied quality control checks taking into account skewness of temperature distribution.

Developing the quality control procedure which consists of a suite of distinct checks we assume that oxygen data obtained by the reference Winkler method are superior in their quality compared to the sensor data. For this reason, we use OSD oxygen profiles in order to set the local accepted oxygen ranges at depth levels. In several other quality checks, instrument-dependent thresholds are used. There are a total of nine distinct quality checks which are introduced in the following sections.

3.1 Geographical Location Check

One possible indication of an erroneous geographical profile location can be the case when the depth of the deepest profile measurement exceeds the local ocean bottom depth. We use GEBCO 0.5-minute resolution digital bathymetry

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

map to define thresholds for this check. For each 0.1x0.1-degree grid node the difference between the maximum GEBCO depth within the 111km radius and the grid-node GEBCO depth is calculated (Fig. 4). If the difference between the deepest profile measurement depth and the local GEBCO depth exceeds the threshold the geographical location of the profile is considered to be in error and data at all levels are flagged.

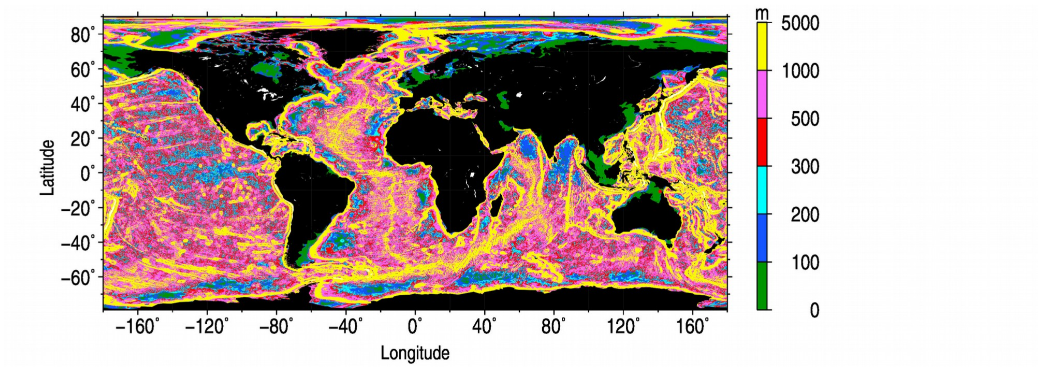

Figure 4. Threshold (in meters) for the difference between the profile's deepest sample depth and the local GEBCO bottom depth.

3.2 Crude range check

The test is applied to identify observations that are grossly in error (the so-called 'blunders') (Fig. 5). These data correspond to the cases of the total instrumentation fault or crude errors introduced during the data recording or formatting. The overall minimum/maximum oxygen ranges are defined based on the entire archive of the OSD profiles. These overall ranges are set both for depth levels and for temperature surfaces, because the maximum oxygen solubility depends on temperature. The relative frequencies serve as guidance to produce the overall oxygen minimum and maximum limits, which approximately correspond to the relative frequency of 0.05%.

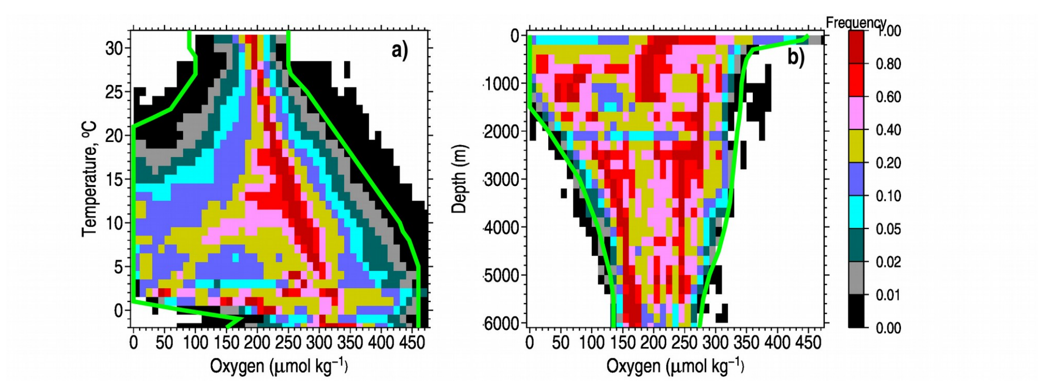

Figure 5. Normalized oxygen histograms used to define overall oxygen ranges versus temperature (a) and versus depth (b). Minimum and maximum overall oxygen limits are shown by solid green lines.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

3.3 Maximum oxygen solubility check

According to Henry's law, the quantity of an ideal gas that dissolves in a definite volume of liquid is directly proportional to the partial pressure of the gas. It is also known that gas solubility in the water typically decreases with increasing temperature. However, in the photic layer of the ocean, oxygen is produced by phytoplankton through photosynthesis so that oxygen supersaturation can evolve. Oxygen production due to photosynthesis leads to the formation of small bubbles (10-70 micron) with increasing oxygen supersaturation accompanied by a higher number of bubbles and their shift towards large sizes (Marks, 2008). Another factor leading to oversaturation is the inaccurate taking of oxygen samples accompanied by the building of air bubbles.

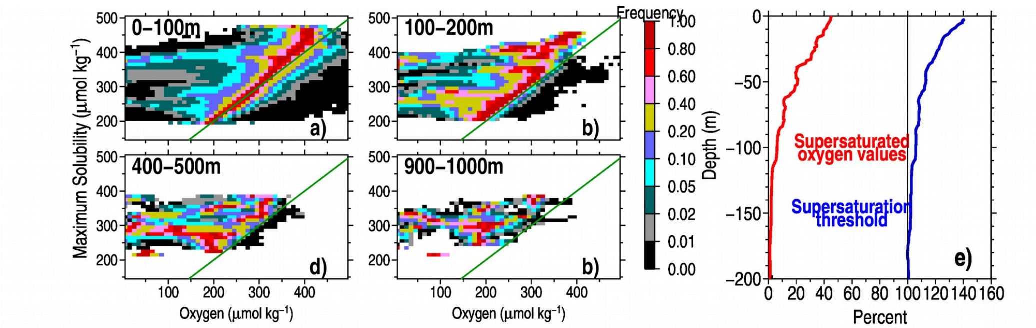

Figure 6. Supersaturation check: a-d) normalized frequency histograms between maximum solubility and the dissolved oxygen for different layers. Bin size is 10 μmol kg^{-1} ; e) percent of supersaturated oxygen values (red) and threshold percent for supersaturation applied in the quality control procedure.

Histograms in Fig. 6a-c show frequency for different values of observed oxygen concentration (C_{obs}) and maximum oxygen solubility (C_{max}) for given temperature and salinity. For each maximum solubility bin, the frequencies are normalized by the number of the values in the most populated bin. The distribution mode for the layer 0-100m (Fig. 6a) follows the line C_{obs}=C_{max} progressively deviating to lower values above 300 μmol kg^{-1}. The diagram shows that a significant number of observed oxygen concentration values in this layer exceeds the maximum solubility. In the deeper layers (Fig. 6 b-d), the number of cases with supersaturation decreases. As illustrated by Fig. 6 e, the percentage of supersaturated values decreases from about 42 % in the near-surface layer to about 0.0% below the 200 m level. We used accumulated versions of the above histograms to put the supersaturation percent threshold (Fig. 6e) at each level above 200 m by 99-th quantile. Below 200 m all supersaturated oxygen values are flagged.

3.4 Stucked value check

Malfunctioning of sensors often results in stucked values when the same oxygen concentration is reported for all or most of the observed levels. To identify such profiles, we calculated oxygen standard deviations for each oxygen profile to build histograms (Fig. 7) for each instrumentation type. Only profiles with at least five oxygen levels are considered. Unlike the OSD and Argo data, for which the frequency of profiles drops for low standard deviation values, the CTD profiles are characterized by a distinct peak for the lowest standard deviation values. Guided by Fig. 7 we set the thresholds of 4 μmol kg^{-1} and 1 μmol kg^{-1} and for CTD and PFL profiles respectively, respectively. No thresholds are applied for OSD profiles, as stucked values are characteristics of the electronic sensors only.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

3.5 Spike check

Spikes are the data at levels that strongly deviate from those at neighboring levels. For each observed level z_k , the test value $s = s_1 - s_2$ is calculated, where $s_1 = |p_k - 0.5\ (p_{k-1} - p_{k+1})|$, $s_2 = |0.5\ (p_{k+1} - p_{k-1})|$ and p denotes the oxygen value. The observation is identified as outliers when s exceeds a threshold value. Due to the larger natural oxygen variability in the upper layers, we set depth-dependent spike thresholds using the respective accumulated spike frequency histograms. Shown in Fig. 8a-b are histograms for two layers, and the entire threshold profile is shown in the Fig. 8c. Spike threshold corresponds to the 99.8% frequency.

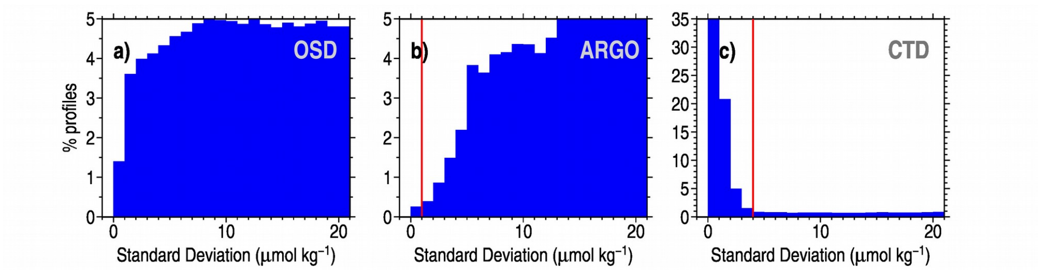

Figure 7. Oxygen profile standard deviation for OSD (a), Argo (b), and CTD (c) instrumentation types. Only profiles with at least five levels of oxygen data are considered. Red vertical lines show the respective threshold values for ARGO and CTD profiles.

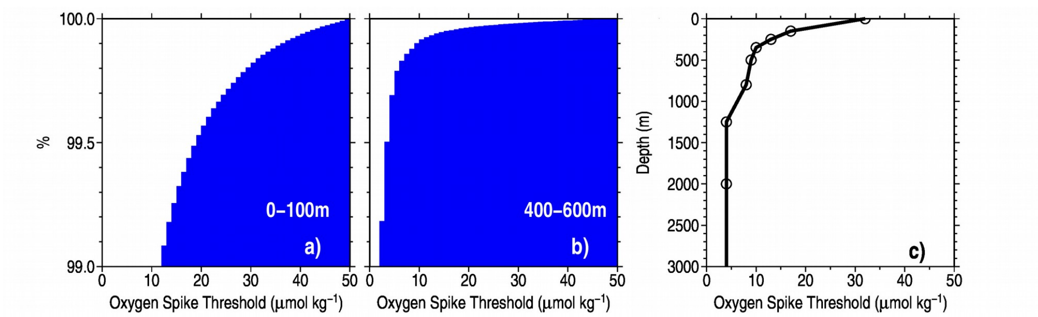

Figure 8. Spike magnitude histograms for the layer 0-100m (a), 400-600m (b) and spike threshold values versus depth (c).

3.6 Multiple extrema check

Multiple extrema check aims to identify profiles whose shape significantly deviates from majority of profiles. For each profile with at least 5 observed levels (see schematics in Fig.9), the number of local extrema (black dots) and their magnitudes (denoted as M_n in Fig.9) are calculated. Then frequency histograms of oxygen profiles for different combinations of the number of oxygen extrema and of the extremum magnitude are calculated (**Fig. 10**). As can be seen from the histograms the larger the extremum magnitude, the less frequent are the profiles with multiple extrema of this magnitude. Grey areas of histograms correspond to the outlier profiles, e.g. profiles exhibiting too many local extrema for a given extremum magnitude threshold value.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Open Access Earth System Science Data Discussions

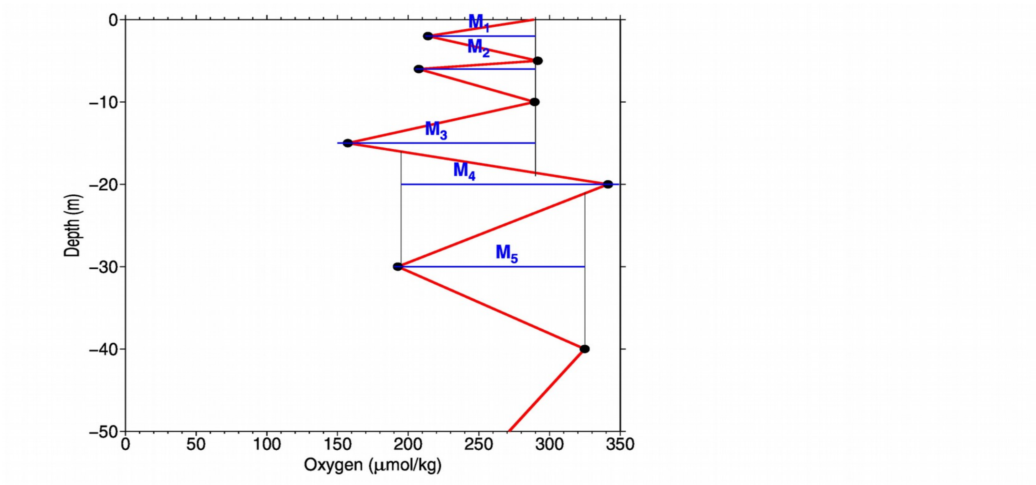

Figure 9. Schematics for the local extrema check. Black dots represent the local extrema (M), whereas extremum magnitudes are shown with blue lines.

Normalized histograms showing the frequency of profiles as the function of the number of extrema and the extremum magnitude value are given in Fig.10. Grey areas correspond to the outlier profiles, e.g. profiles exhibiting too many local extrema for a given extremum magnitude threshold value. The histograms were produced for three instrumentation types. The histogram for Argo profiles differs from those for OSD and CTD because it is based on profiles that the respective DACs have already validated.

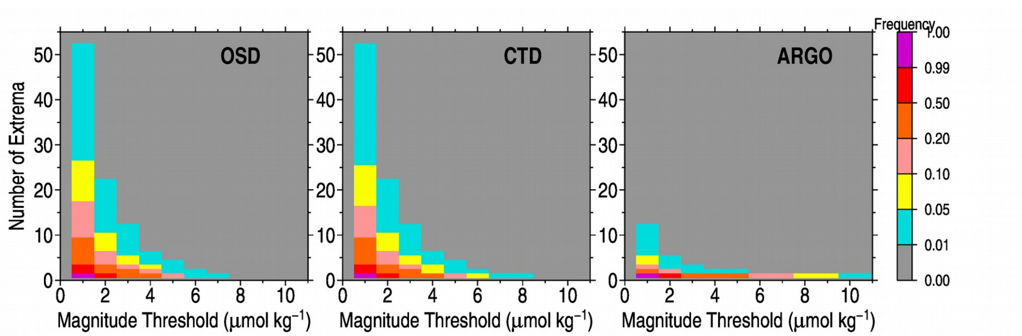

Fig.10. Normalized frequency histograms for multiple extrema check: a) OSD, b) CTD, c) Argo profiles. Grey areas correspond to the oxygen profiles failing the multiple extrema check.

3.7 Oxygen Vertical Gradient check

The oxygen vertical gradient check aims to identify pairs of levels for which the vertical oxygen gradient exceeds a certain threshold. Threshold values (Fig. 11) are calculated for several layers to account for the change of gradient

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Open Access Earth System Science Data Discussions

values with depth. The thresholds are defined using six standard deviation envelopes around the mean gradient value. Due to the nonlinearity of oxygen profiles, vertical gradient values depend on the profile's vertical resolution, e.g., from the gap between two neighbor observed levels. Respectively, oxygen thresholds at each level are calculated for several values of depth gaps, as Tan et al. (2023) did for the quality control of temperature profiles. Calculations are performed separately for positive and negative oxygen gradient values. Because of a high oxygen vertical gradient variability in the near-surface layer and over the entire World Ocean, we conduct vertical gradient check only for levels below 200 m depth.

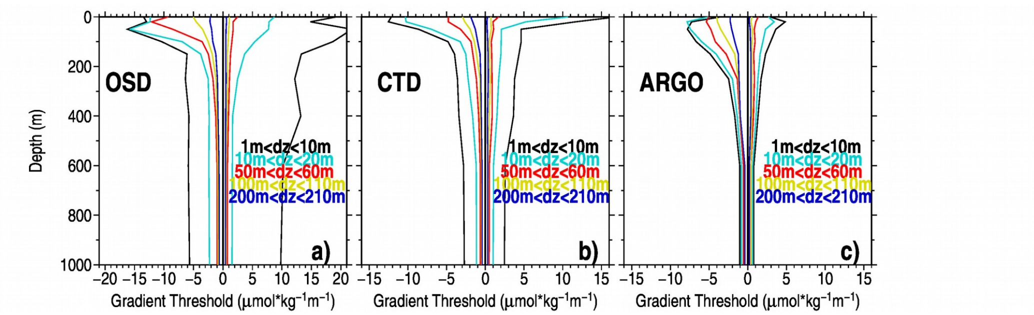

Figure 11. Oxygen vertical gradient threshold versus depth for several depth gaps (dz) between the neighbor observed levels: a) OSD, b) CTD, c) Argo profiles.

3.8 Local Climatological range check

Local climatological oxygen range check is most effective for identifying outliers compared with other checks, because the min/max thresholds constrain the observed values locally. For each 1°×1° latitude/longitude grid point, we calculate min/max thresholds taking into account the skewness of the data. For calculating climatological ranges, we assume the ergodic hypothesis in which the average over time is considered to be equal to the average over the data ensemble within a certain spatial influence radius. Taking into account the skewness of statistical distribution when defining climatological ranges for oceanographic parameters was first suggested by Gouretski (2018), who applied Tukey's box plot method modified for the case of skewed distributions (Hubert and Vandervieren, 2008; Adil and Irshad, 2015). In this method lower (Lf) and upper (Lu) fences are calculated according to formula (1):

$$[Lf\ \ Uf] = [\ Q1 - 1.5*IQR*\exp(-SK*|MC|)\quad Q3 + 1.5*IQR*\exp(SK*|MC|)\ ], \qquad (1)$$

where Q1, Q3 are quartiles, Q2 is sample median, SK is skewness. MC denotes medcouple, which is defined as MC = median $h(x_i,x_j)$, where $x_i << Q2 << x_j$; and the kernel finction $h(x_i,x_j) = [(x_j-Q2)-(Q2-x_i)]/(x_j-xi)$. (Hubert and Vandervieren, 2008).

We note that the local oxygen thresholds are constructed using the data which have undergone the preliminary quality control. This control includes checks for crude range, spikes, stucked value, and multiple extrema, and vertical gradient, aiming to remove most outlying scores in order to reduce their impact on the local thresholds. This approach is similar to the two-stage thresholding suggested by Yang et al. (2019).

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

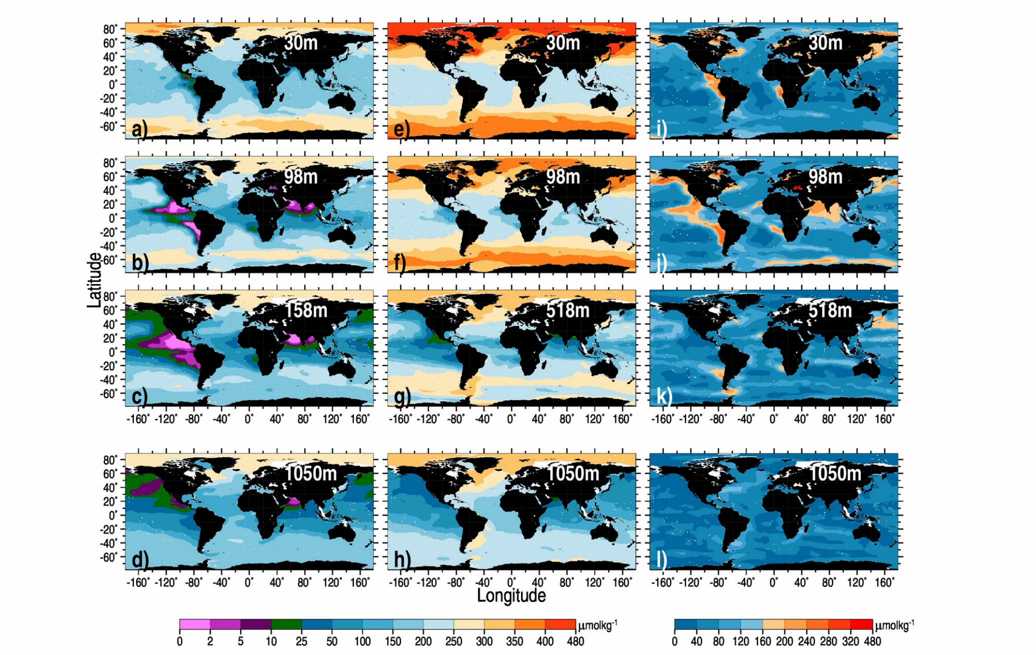

Fig.12. Local climatological lower oxygen threshold for several depth levels: 30m (a), 98 m (b), 158m (c), 1050m (d). (e-h): same but for the upper oxygen threshold; (i-l) same but for the local oxygen range.

OSD validated data were used to construct maps of local minimum and maximum values at a set of depth levels using formula (1) (Fig. 12). The most striking features are the areas with low minimum oxygen values (oxygen minimum zones, Fig. 12a-c) in the East Pacific, Arabian Sea, Bay of Bengal, Black Sea, and Baltic Sea. The maps reflect wide oxygen ranges (Fig. 12 i-l) in several regions of the world's oceans, especially in the East Pacific and the North Indian oceans. The maps also depict higher oxygen variability in the highly dynamic regions of Gulf Stream, Kuroshio, and in the upwelling areas west of Africa. During the quality control, gridded minimum and maximum local oxygen values are interpolated to the profile locations.

3.9 Excessive flagged level percentage check

After applying all distinct quality checks, the percentage of flagged levels for each oxygen profile is calculated. The respective histograms (**Fig. 13**) are used to set thresholds to decide on the quality of the entire profile. Using these histograms as guidance, we set 30%, 10%, and 30% thresholds for OSD, CTD, and Argo profiles, respectively. If the threshold is exceeded, the entire profile is flagged. Both the OSD and the Argo profiles are characterized by a low number of profiles with a high percentage of flagged data. The OSD and Argo groups are characterized by a low percentage of profiles with a high number of rejected levels. In contrast, in the CTD group, most not-dummy oxygen values fail quality checks, resulting in a high percentage of flagged profiles.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

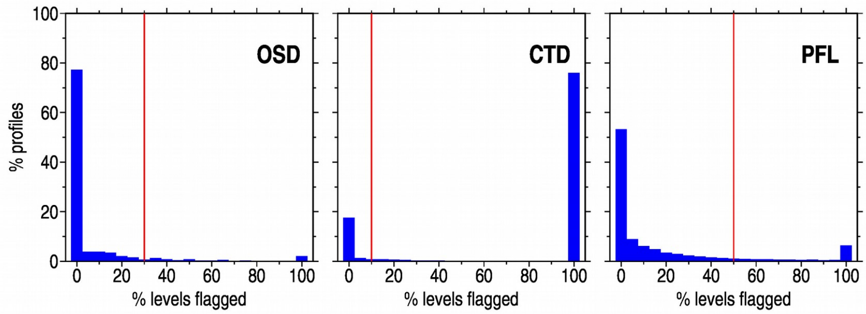

Fig.13. Percentage of oxygen profiles versus percentage of rejected levels per profile for OSD, CTD, and Argo instrument types.

4 Oxygen Data Quality Assessment based on the results of the automated quality control procedure

Table 1 and Fig. 14 summarize the rejection rates for all nine quality checks and the three instrumentation types considered. The Argo oxygen profiles have the lowest overall rejection rate of 6.20%, with Winkler data quality (9.66% outliers) ranking second best. We explain this difference through two factors. First, Winkler profiles cover a century-long period of observations, with a worse data quality in the earlier decades. Secondly, the analyzed Argo oxygen data are represented by adjusted profiles, which have been already quality-controlled.

The CTD oxygen profiles are characterized by the highest percentage of outliers. This is attributed to many high-resolution CTD profiles reporting oxygen values, identified as outliers by multiple quality checks, and probably representing the digital output of the malfunctioning oxygen sensors. The local climatological range check (No.8) results in the highest percentage of flagged observations and the profiles affected by this check. In the case of OSD profiles, about 18% of profiles have at least one flagged measurement. For Argo oxygen profiles, this percentage is about 30%.

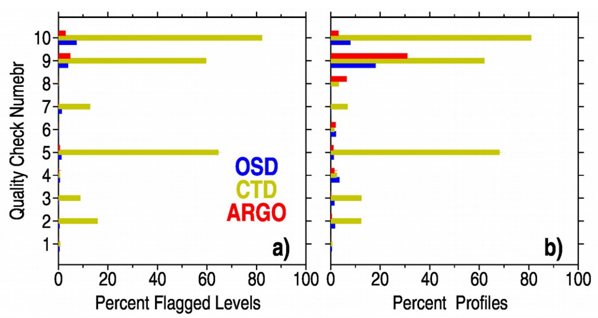

Figure14. Percentage of rejected observations for distinct quality checks and three instrumentation types: a) percent of flagged level; b) percent of profiles affected by the respective check. Check number is indicated in Table 2.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Open Access Earth System Science Data Discussions

Table 2 Outlier score statistics for different instrumentation types

		OSD		CTD		ARGO	
	Quality Check	% flagged observations	% flagged profiles	% flagged observations.	% flagged profiles	% flagged observations	% flagged profiles
1	Location check	0.452	0.515	0.719	0.532	0.092	0.081
2	Global Oxygen Range at depth levels	0.414	1.772	15.797	14.246	0.041	0.420
	Global Oxygen Range on T surfaces	0.272	1.511	8.824	12.381	0.009	0.221
3	Supersaturation check	0.654	3.524	0.635	2.458	0.101	1.496
4	Stucked value check	1.265	1.168	64.596	68.086	0.564	1.060
5	Spike check	0.226	2.151	0.043	1.527	0.012	1.873
6	Multiple extrema check	1.376	0.233	12.846	6.801	0.126	0.057
7	Oxygen vertical gradient check	0.205	1.036	0.536	9.223	0.139	7.671
8	Local climatological range	3.922	18.151	59.663	62.013	4.824	30.915
9	Excessive flagged level percentage profile check	7.397	7.948	82.288	80.912	2.910	3.136
	ALL QC CHECKS	9.66	23.88	82.97	88.69	6.20	36.65

5 Benchmarking of the automated quality control procedure using manually controlled datasets

Good et al. (2022) conducted a comprehensive benchmarking exercise to evaluate the performance of AQC checks on temperature profiles implemented by different research groups, aiming to recommend an optimal set of quality checks. They used several reference datasets with known quality (e.g. evaluated by means of expert control). The procedure does not include the assessment of temperature biases. This study uses a comprehensive set of bottle profile data obtained during the World Ocean Circulation Experiment (WOCE) – the largest international oceanographic experiment ever conducted (Wunsch, 2005). To achieve a high degree of data quality and consistency between the cruises over the entire period of observations, the WOCE Hydrographic Program Office (WHPO) issued operation manuals (WHPO, 1991), where methods and procedures are described. As shown by Gouretski and Jancke (2000) the WHPO quality requirements have been fulfilled with the WOCE hydrographic dataset representing a unique global scale collection of the whole suite of oceanographic parameters. Specifically, the mean inter-cruise oxygen offset was found to be 2.389 µmol kg^{-1}. After completing the WOCE, the GO-SHIP program was established in 2007 to revise the WOCE hydrographic programme (Hood et al, 2010). Applying our quality control procedure to the entire WOCE dataset confirms the high quality of this unique dataset, with only 2.85% of oxygen outliers (Fig. 15a, b) for the entire time period 1990-1998. The QC diagnostics reflect the progressive improvement of the oxygen data quality over the period of WOCE (Fig. 15a). Flagging rate as depicted on oxygen vs depth diagram (Fig. 15a) illustrates the ability of the QC procedure to identify outliers deviating from the main population. The spatial distribution of outliers for the entire time period (Fig. 15c) indicates the majority of WOCE oxygen profiles having a very low percentage of outliers or exhibiting

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

no outliers at all. The high percentage of oxygen outliers is found only for several WOCE lines in the tropical South Atlantic, North-Western Indian Ocean, and the Labrador Sea. µmol kg^{-1}

The WOD database permits data selection for a large number of observational programs using the respective project identification code. The outlier rejection percentage for the data from 128 projects that reported oxygen data is shown in Fig. 16. Several outstanding observational programs like GEOSECS (Geochemical Ocean Sections Study) (Craig, 1974), SAVE (South Atlantic Ventilation Experiment) (Larque et al., 1997), WOCE, CARINA (Carbon dioxide in the Atlantic Ocean) (Falck and Olsen, 2010), and CLIVAR (Climate and Ocean: Variability, Predictability and Change) (Sarachick, 1995) delivered a significant number of high-quality data. We note that the four projects with a median year after 1985 (SAVE, WOCE, CARINA, and CLIVAR) are characterized by the rejection rates lower than the mean. For instance, for the largest WOCE dataset the QC procedure identifies only 2.8% from the total of 354028 oxygen measurements (Fig. 16a) with 79% percent of oxygen profiles without data outliers (Fig. 16b).

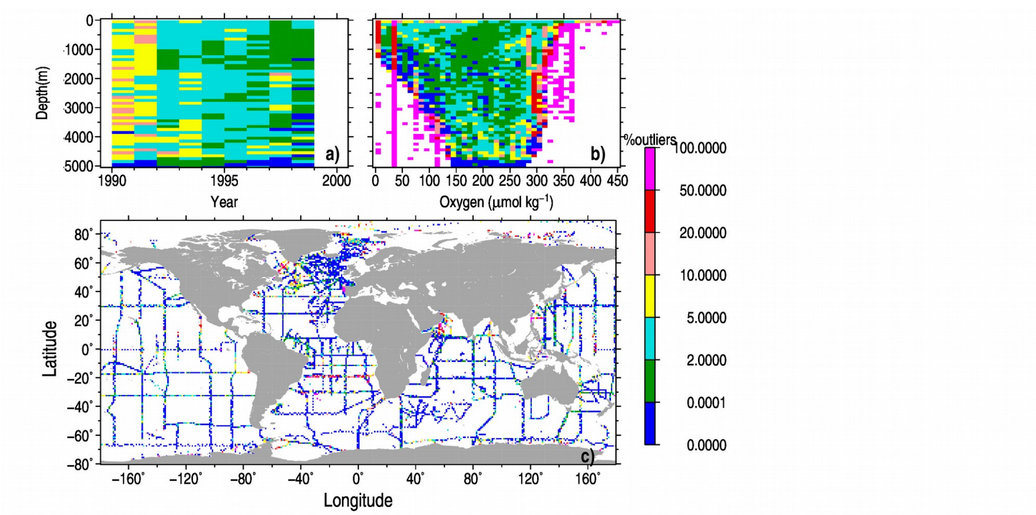

Figure 15. Quality control statistics for WOCE dataset: a) percentage of outliers in year/depth bins; b) percentage of outliers in oxygen/depth bins; percentage of outliers in 1x1-degree squares.

6 Bias assessment for sensor oxygen data

The quality control procedure described in the previous sections is based on the underlying statistics of the data and aims to identify random outliers. The second step in data quality control is estimating the possible systematic errors or biases. These systematic errors may differ depending on the instrumentation type, but the common fcause for systematic errors is the absence of the possibility to calibrate the instrument. A classic example provides temperature data obtained by eXpandable bathythermographs (XBT) where systematic errors are due to the uncertainty in depth, which is calculated from the elapsed time, and the uncertainty in thermistor, which is typically not calibrated (Gouretski and Reseghetti, 2010; Cheng et al. 2014). In the case of dissolved oxygen, only Winkler measurements of discrete samples can be considered to be bias-free because the chemical analysis is based on the KIO_3 standard reference, with the replicate measurements having a precision better than 0.4 µmol kg^{-1} (Thaillandier et al., 2018). In the following, we describe residual biases for CTD and Argo profiles. The term “residual” is used because CTD oxygen profiles are

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

typically adjusted on Winkler bottle samples, and Argo oxygen profiles used in our study undergo adjustment procedures at the respective DACs.

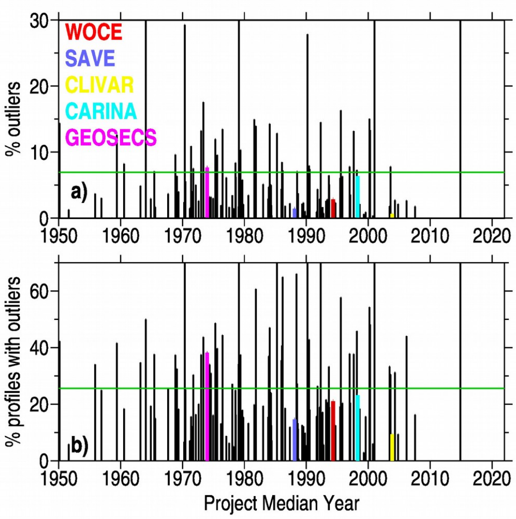

Figure16. Outlier diagnostics for 128 distinct WOD projects (OSD Winkler profiles). a) Overall percent of outliers; b) percent of profiles with oxygen outliers. Acronyms and percentages for selected hydrographic projects described in text are shown in color.

Use of electrochemical and optical oxygen sensors into oceanographic practice has two main aspects. First, these sensors permitted a significantly higher rate of data acquisition and a much finer vertical resolution than bottle data. Secondly, they made the observational process much easier than bottle samples, which need chemical titration in the laboratory. However, like other electronic sensors, oxygen sensors are prone to offsets and drift. Takeshita et al (2013) analyzed data from 130 Argo floats and found a mean bias of -5.0 % O_2 saturation at 100 % O_2 saturation. Bittig et al (2018) explained this negative bias by reducing O_2 sensitivity proportional to oxygen content, with the decrease of sensitivity being on the order of several percent per year. Optode drift characteristics require regular calibration. Use of reference Winkler profiles is possible only for the ship-based CTD oxygen sensors (mostly electrochemical sensors) if CTD rosette water samples are obtained simultaneously with sensor profiles and are analyzed for oxygen during a cruise (Uchida et al., 2010). For unmanned autonomous platforms like Argo, the direct comparison with reference Winkler data is limited to samples from the hydrographic casts conducted during the float deployment. Bittig et al. (2018) recommended adjusting optode data on oxygen partial pressure primarily by the gain (Argo Quality Control Manual, 2021). If no previous delayed-mode adjustment is available, the basic real-time adjustments are performed

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

based on the oxygen saturation maps provided by the WOA digital climatological atlas (Thierry et al., 2021). In case a delayed-mode adjustment is not available after one year, the re-assessment of the gain factor is recommended. Uncertainty in underlying optode calibration and time drift characteristics leads to errors in adjusted data.

6.1 Bias assessment method

We aim to assess the magnitude of the possible overall residual bias for CTD profiles and adjusted Argo optode profiles by comparing these profiles with collocated reference discrete samples. The data from 10 national DACs were used for this analysis, for which both unadjusted and adjusted oxygen profiles are available. Data centers and the respective number of oxygen profiles are given in Table 2. We use the Winkler method oxygen profiles available from the World Ocean Database and described in Section 1.1. These profiles are used as reference data for the comparison with collocated Argo optode oxygen profiles.

For the current analysis, we selected a 100 km threshold distance within which two profiles are spatially collocated. To decide upon the choice of the optimal maximum time difference between Argo and reference profiles, we calculated median oxygen offsets increasing threshold value for the time separation between a pair of profiles (Fig.17a). Increasing the temporal collocation bubble leads to the increase of the bias magnitude in agreement with the assumption that the older reference data are richer in oxygen compared to the more recent data. Below 1000 m depth, the difference between the median offsets for the temporal collocation bubble of 5 and 50 years is about 3.5 µmol kg^{-1}, corresponding to a deoxygenation trend of about 0.7 µmol kg^{-1} per decade. This estimate can be compared with 0.75 µmol kg^{-1} per decadedec reported by Gregoire et al. (2021). As Fig. 17c suggests, the overall offset estimate below 1000 m stabilizes after the time difference threshold of 5 years. The extension of the temporal bubble for more than 7 years leads to the progressive increase of the bias magnitude, which we attribute to the impact of the general deoxygenation. Based on these calculations, the 5-year threshold was selected as the maximum time separation between collocated profiles. For this threshold value, the number of collocated pairs below 1000m depth is about 10000 (Fig. 17b). A step-wise decrease of the number of collocated pairs below 950 m is explained by a significant part of reference profiles being limited to the upper 1000-meter layer. These calculations suggest that about 1000 collocated pairs are required for stable offset estimates.

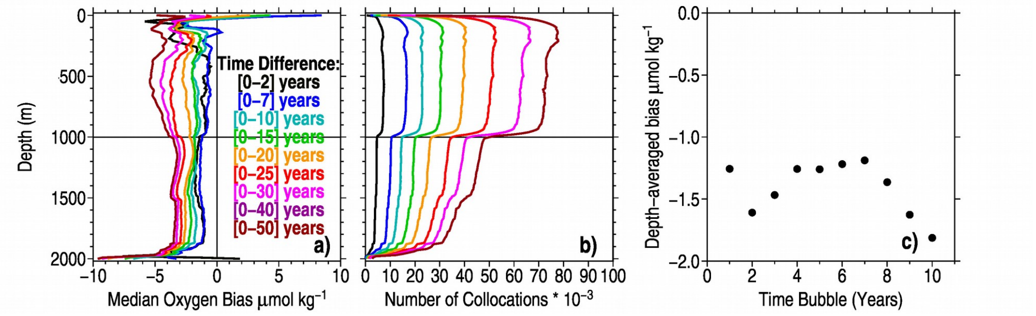

Figure17. a) Overall median oxygen bias versus the size of the temporal collocation bubble; b) number of collocated pairs for different choices of collocation bubbles; c) depth-averaged (1000-1900m) bias versus time bubble size.

The number of Argo profiles having collocations with discrete ship-based Winkler profiles is shown in Table 2. No collocated Winkler profiles are found for the Argo profiles from the two Korean DACs. Profiles from these DACs are restricted within a relatively small area east of the Korean peninsula. The four largest contributors of Argo data (AOML, Coriolis, JMA, and CSIRO) comprise up to XX percent of all Argo profiles having collocations with reference profiles

6.2 Overall bias characteristics

The normalized frequency histograms (Fig. 18) characterize the spread of individual bias estimates around the distribution mode. These histograms are based on all Argo profiles having collocations with reference Winkler data. In these histograms, for each depth bin, the number of values in each bias bin is normalized by the number for the most populated bias bin. The adjustment procedures applied in different DACs reduce the spread of the individual bias estimates and significantly reduce the overall median bias from 10-12 μmol kg^{-1} for unadjusted data to 1-2 μmol kg^{-1} for adjusted data. Based on the histogram, we estimate residual bias using the collocated data below 1000 m depth, where the natural variability is reduced compared to the upper part of the water column.

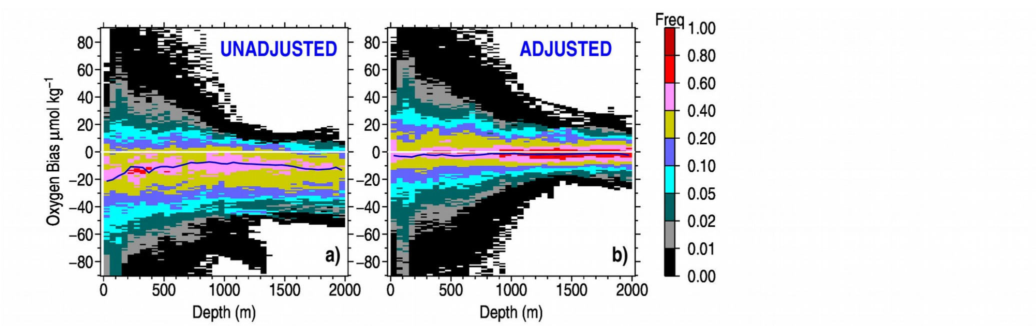

Fig.18. Normalized histograms of the unadjusted (a) and adjusted (b) Argo oxygen offsets versus collocated Winkler profiles. The blue curve shows the median.

6.3 Residual Oxygen Biases for Argo profiles from distinct DACs

According to the Argo Quality Control Manual (Thierry et al., 2021), several adjustment procedures can be applied to unadjusted data (adjustment to climatology, adjustment to nearby Winkler samples, adjustment to in-air data). The adjustment results may depend on many factors, such as the subjective decision of the operator in a DAC, the use of a specific software, the availability of the respective reference data, etc.). If a climatology is used as a reference, the adjusted Argo oxygen values will be adjusted to the median year of a climatology, which can differ by several decades from the year of an Argo profile. In such cases, the long-term deoxygenation trend of the world ocean might bias the results of the adjustment procedure.

Changes in oxygen sensors over time may cause respective changes in diagnosed biases. Fig. 19 shows the yearly number of observed profiles of AOML-processed Argo floats equipped with different models of optode sensors. Since the beginning of the 2000s, several different models of optodes were implemented in BGC Argo floats, with the most widespread sensors being AANDERAA 3830, implemented between 2004 and 2018, and the following model AANDERAA 4330. Since about 2013, the majority of Argo floats from the two largest AOML and Coriolis datasets

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

have been equipped with this sensor. The AANDERAA 4330 sensor prevails between 2012-2017 for JMA data and after 2020 for CSIRO data (Fig. 19.)

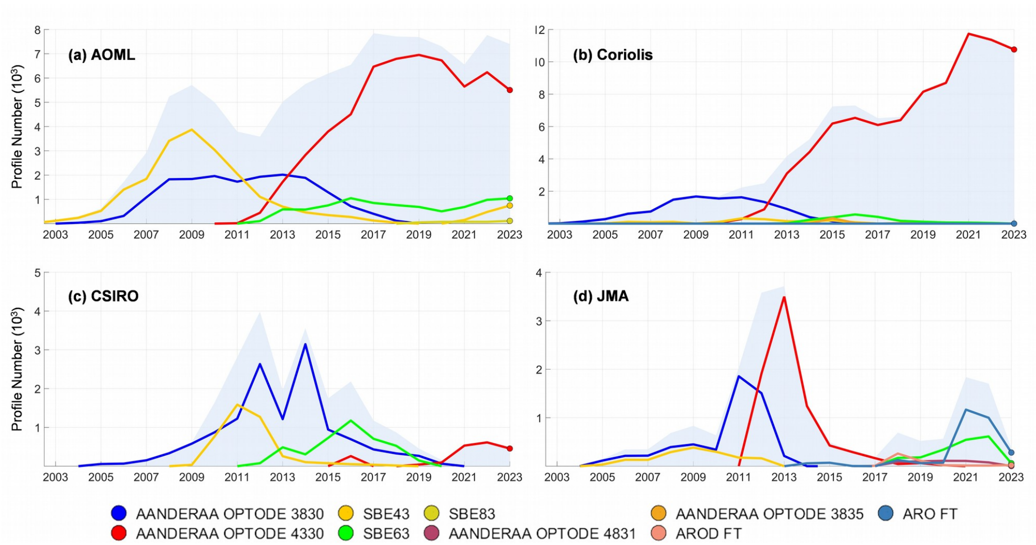

Figure 19. Yearly number of BGC Argo profiles equipped with dfifreent tyees of opotde oxgyen sesnors (colored lines). Light-bulee shading corresponds to the total number of proifles: a) AOLM, b) Coriolis, c) JMA, d) CSIRO

We calculated the resdiual oxygen bias for each depth level as the mean offest between Argo and Winkler oxygen data over all collocated pairs for each DAC (Fig. 20). The offsets for the Korean DACs cannot be estimated due to the lack of the collocated Winkler profiles. The number of available collocations with reference Winkler profiles varies by the order of magnitude for different DACs. Since reference bottle data often cover only part of the upper 2000-meter layer, the number of collocated pairs also changes over depth, with the main step-wise decrease seen around 1000m. A comparison of the vertical bias profiles for other DACs suggests that changes in the number of collocated pairs over depth do not impact the diagnosed bias. Except for CSIRO and MEDS Argo profiles, DAC-adjusted overall median residual bias is negative, ranging between -1.0 to -3.6 µmol kg^{-1}. The residual positive bias for CSIRO and MEDS profiles is typically within the range of 0.4-0.6 µmol kg^{-1} below 1000 m. INCOIS profiles are characterized by the change from negative to positive bias below 1400 m.

For the two largest datasets (e.g., AOML and CORIOLIS), vertical bias profiles exhibit a characteristic hook below about 1900-1950 meters. Such hooks on Argo oxygen profiles were found by (Thallander et al., 2018). The hook can reflect the adjustment of the oxygen sensor at the beginning of the float ascending.

To investigate a possible bias change over time due to the change in the instrumentation (see Fig. 19), we first calculated depth-averaged biases (1000-1900m layer) for each collocation pair. Mean biases within 2°×4° latitude-longitude boxes are shown in Fig. 21, along with the bias histograms for two time periods: 2005-2013 and 2014-2023.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

The choice of these two periods approximately corresponds to the instrumentation change around 2013, which was described above. The calculations were done separately for each DAC. During the first period, the foil-batch calibrated optodes were used predominantly. Bittig et al. (2018) note that differences between batch calibration and individual optode can exist. For the MEDS dataset, only data from the period 2004-2013 are available. For the largest AOML dataset, the 2014-2023 period is characterized by a stronger negative bias of -3.54 μmol kg^{-1} compared to -0.25 μmol kg^{-1} for the time period 2004-2013 (Fig. 20a). We explain this AOML residual bias change by the respective instrumentation change (Fig. 19). However the second largest Coriolis dataset does not show a significant difference between these time periods (Fig. 20b), what could reflect differences in the adjustment procedures implemented by AOML and Coriolis DACs.

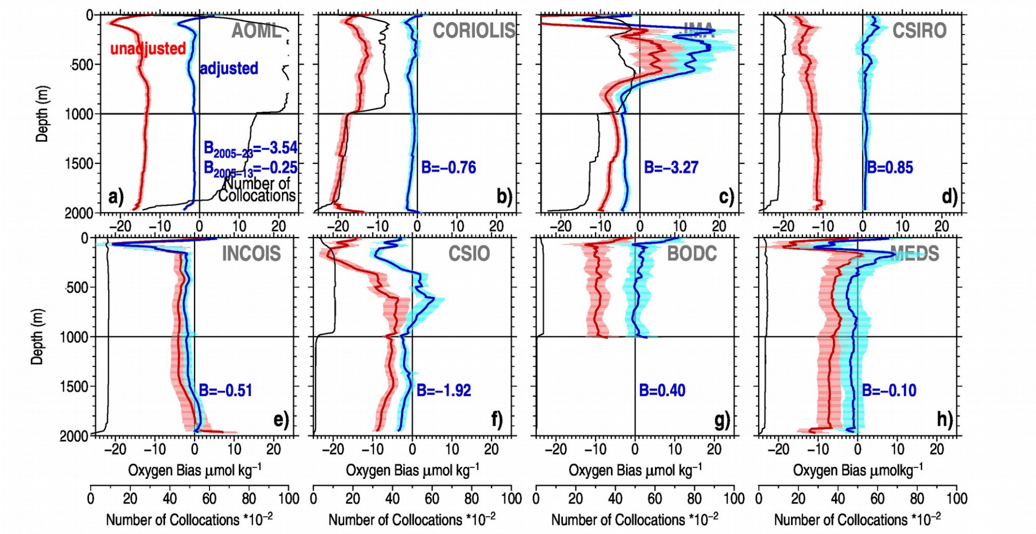

Figure 20. Overall mean Argo oxygen offsets vs Winkler profiles for distinct DACs: a) AOML, b) Coriolis, c) JMA, d)CSIRO, e)INCOIS, f) CSIO, g) BODC, h) MEDS. Offset profiles for unadjusted and adjusted data are shown in red and blue, respectively. Standard error bars (light shading) are calculated using the number of different floats at each level as the number of degrees of freedom. Blue numbers show the depth-averaged residual offsets (μmol kg^{-1}) within the layer 1000-1900m. Black thin lines show the number of collocated pairs at depth levels.

6.4 Residual Oxygen Biases for distinct Argo floats

In addition to the overall biases described above, we also calculated biases for distinct Argo floats (Fig.22). The scatter diagrams of adjusted versus not-adjusted data show that the overwhelming majority of floats exhibit adjusted residual biases of a smaller magnitude compared to unadjusted data. Among 1020 Argo floats for which at least ten profiles with individual profile offsets between 1000 and 1900m are available, oxygen adjustments for 22 floats (3.1%) were found to be unsuccessful. However, the magnitude of most of these misfits does not exceed 2 μmol kg^{-1} and is well within the uncertainty of bias estimation for individual floats.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

6.5 Residual Oxygen Biases for CTD oxygen sensors

We conducted similar bias calculations for the WOD CTD oxygen profiles, mostly obtained by electrochemical sensors. Unlike Argo profiles, the CTD oxygen sensor data are typically adjusted on the simultaneously obtained bottle samples analyzed in the ship laboratory using the Winkler method (Thaillandier et al., 2018). Therefore the overall mean CTD residual bias is small compared to the residual Argo oxygen bias. Within the layer 1000-1900m, CTD oxygen bias is in the range of 0.2-0.5 µmol kg^{-1}, with the mean value close to 0.3 µmol kg^{-1} (Fig.23). Above the 1000 m level the bias exhibits stronger variability which we explain by the impact of the increased spatial and temporal variability above the main tehrmocline.

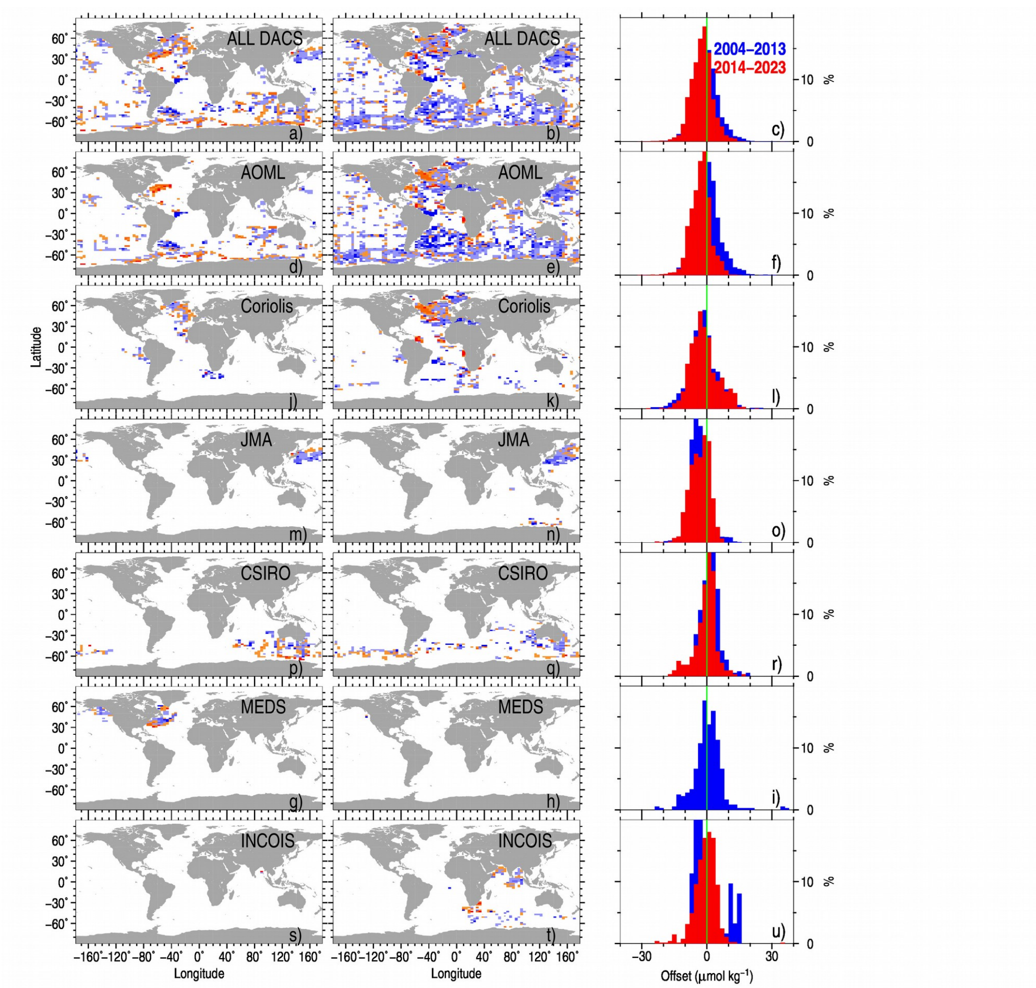

Figure 21. Residual oxygen bias in 2°×4° boxes for DAC-adjusted oxygen values for all collocated pairs: a) years 2004-2013; b) years 2014-2023; c) bias histograms for two periods. d-f) same for AOML data; g-i) same for MEDS data; j-l) same for Coriolis data; m-o) same for JMA data; p-r) same for CSIRO data; s-u) same for INCOIS data.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

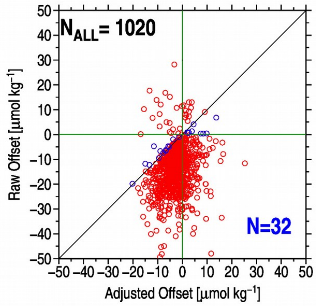

Figure 22. Raw versus adjusted oxygen bias for distinct BGC Argo floats (layer 1000-1900 m). Bad adjustments are shown in blue.

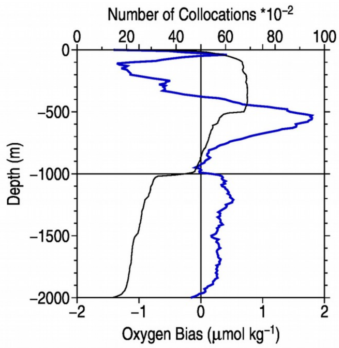

Figure 23. Overall mean oxygen offset CTD-Winkler (blue) and the number of collocated pairs (black)

8 Summary and Conclusions

This study developed a new automated quality control scheme for ocean oxygen profile data and applied it to the oxygen profiles from the World Ocean Database and the Argo float oxygen profiles provided by national DACs. The procedure consists of a suite of nine quality checks, which are based on local or global parameter thresholds. Some checks are conceptually similar to the quality checks used to validate the profiles in the World Ocean Database (Boyer et al., 2018) (for example, global range test, vertical gradient test) and in the Argo data acquisition centers (Thierry et al., 2021) (for example, spike, frozen profile tests), but we provide additional checks (for example, test for the number of local extrema and local climatological range test) which have several advantages to better flag likely erroneous data. For instance, the procedure proves whether an oxygen value falls out of accepted ranges (defined by global or local ranges) or whether an oxygen profile exhibits a very untypical shape. The shape of the profile is characterized by the vertical oxygen gradient, the number and magnitude of local oxygen extrema, and the presence of spikes. The check is also done for the so-called “frozen” profiles occurring when the oxygen sensor stocks and reports the same values throughout the profile.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

The novelty of the proposed quality scheme is that the threshold choice is based on the respective statistics, and the Gaussian distribution is not assumed for the local climatological range check. The accompanying diagnostic tool provides the overview of outlier scores and permits tuning the thresholds. The quality control procedure was benchmarked against several hydrographic datasets known for their outstanding measurement quality, with WOCE experiment data collection being the largest and best documented. Analysis of the outliers and their distribution among distinct hydrographic sections suggests the ability of the procedure to flag outliers retaining the overwhelming majority of good data.

Further, we estimated possible residual oxygen biases in the delayed-mode adjusted Argo oxygen profiles. The bias estimates are based on analyzing the collocated Argo and discrete water sample ship-based profiles. The latter represents reference measurements as the bottle samples are analyzed by means of the Winkler chemical method. The size of the collocation bubble (e.g., the maximum distance between two profiles and the maximum time difference) have been set by 100 km and 5 years, respectively, after conducting several experiments for different bubble sizes. Residual biases relative to the Winkler reference data are represented by the difference at an isobaric level between the Argo sensor oxygen value and the Winkler oxygen, with the overall residual bias at each level being defined by the average overall individual differences.

Our calculations find a small negative residual oxygen bias in the range -1 to -4 µmol kg^{-1} for all individual DAC datasets except CSIRO and MEDS. The residual positive bias for CSIRO and MEDS profiles is typically within the range of 0.5-1.0 µmol kg^{-1} below 1000 m. Calculations suggest at least an order of 1000 collocations is needed for the stable residual bias estimation. This number of collocations is available only for AOML, Coriolis, JMA, CSIRO, and INCOIS datasets. Further, we found a change in the diagnosed residual oxygen bias around 2014 for the largest AOML dataset, possibly related to the instrumentation change, when the AANDERAA optode A4330 became the primary sensor type used on Argo floats. However, this change of the residual bias could not be diagnosed for the second largest Coriolis dataset. Analysis of the residual bias for 1020 Argo floats having at least ten profiles with collocations confirmed bias reduction for 97% of the floats (compared to the unadjusted data) due to the adjustments conducted by DACs .

Diagnosed residual biases for the quality-controlled CTD oxygen sensor profiles revealed a high degree of agreement between the CTD and Winkler reference data, with the bias being in the range 0.2-0.5 µmol kg^{-1}. We explain this low bias as the result of the adjustment of CTD oxygen sensor data on simultaneous discrete samples analyzed by the Winkler method.

In summary, this study proposed a new QC approach to process CTD, bottle, and Argo data and investigated the consistency between the three primary instrumental data for ocean oxygen. Our investigations ensure the consistency between the three datasets and provide a solid basis for merging the three datasets into a single, integrated, and homogeneous oxygen database. Therefore, the database obtained in this study supports next-step assessment and understanding of the change in ocean oxygen levels.

7 Data availability

The quality control procedure described above was applied to the OSD and CTD oxygen profiles between 1920 and 2023 from the World Ocean Database and to the oxygen profiles from the BGC Argo floats. The resulting dataset comprises observed level data with quality flags and data interpolated on 10-meter levels. The data are in NetCDF format and also include the metadata information. The complete dataset (Gouretski et al., 2023) can be found at http://dx.doi.org/10.12157/IOCAS.20231208.001.

Author contributions.

LC and VG – conceptualization, supervision, methodology; VG – software, formal analysis, data validation, visualization, and writing (original draft preparation, final version, and editing); JD., XX, FC – methodology, data curation; LC –funding acquisition.

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

Acknowledgements. We are thankful to the institutions and organizations that provided the data used in this paper (the specific datasets and sources are as described in the text). We also thank all anonymous reviewers for their detailed and constructive comments. The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (ARGO, 2000). The Argo Program is part of the Global Ocean Observing System.

Financial support. This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences [grant number XDB42040402], the National Natural Science Foundation of China [grant numbers 42122046 and 42076202], and the Youth Innovation Promotion Association, CAS [grant number 2020-077]. The author also acknowledges the support from the new Cornerstone Science Foundation through the XPLORER PRIZE, Youth Innovation Promotion Association, Chinese Academy of Sciences.

References

Adil, I. H. and Irshad, A. R.:A modified approach for detection of outliers, Pak. J. Stat. Oper. Res., XI, 1, 91-102.

Argo (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182

Bittig, H. C., Maurer, T.L., Plant, J. N., Schmechtig, C., Wong,A. P. S. , Claustre,H., Trull, T. W. , Udaya Bhaskar, T. V. S., Boss, E., Dall'Olmo, G., Organelli, E., Poteau, A., Johnson, K. S., Hanstein, Leymarie,C., E., Le Reste, S., Riser, S. C., Rupan, A., Taillandier, V., Thierry, V. and Xing, X. : A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage, Front. Mar. Sci., 6:502, doi: 10.3389/fmars.2019.00502, 2018.

Boyer, T. P., Baranova, O.K., Coleman, C., Garcia, H. E. , Grodsky, A., Locarnini, R. A.,Mishonov, A. V., Paver, C. R.,Reagan, J. R. , Seidov, D., Smolyar, I. V., Weathers, K., Zweng, M.M.: World Ocean Database 2018, A. V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87, 2018.

Carpenter, J.H. :The accuracy of the Winkler method for dissolved oxygen analysis, Limnology and Oceanography, 10, 1, p. 135-140. https://doi.org/10.3419/lo.1965.10.1.0135, 1965.

Cheng, L. J., Zhu, J., Cowley, R., Boyer, T. and Wijffels, S.: Time, probe type and temperature variable bias corrections to historical expendable bathyther-mograph observations, J. Atmos. Ocean. Technol.,31, 1793–1825, doi:10.1175/JTECH-D-13-00197.1, 2014.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Clark, L. C.: Cellophane/Platinum electrode for blood PO2, J. Appl. Physiology, 6, 189, 1953.

Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the global ocean with biogeochemical-Argo, Annual Review of Marine Science, 12, 23–48, 2020.

Coppola, L., Salvetat, F., Delauney, L., Machoczek, D., Larstensen, J., Sparnocchia, S.,Thierry, V., Hydes, D., Haller, M.,Nair, R., Lefevre, D.: White paper on dissolved oxygen measurements: scientific needs and sensors accuracy, 22 pages, 2013.

Cowley, R., Killick, R. E., Boyer, T., Gouretski, V., Reseghetti, F., Kizu, S., Palmer, M. D., Cheng, L., Storto, A., Le Menn, M., Simoncelli, S., Macdonald, A. M., and Domingues, C. M.: International Quality-Controlled Ocean Database (IQuOD) v0.1: The Temperature Uncertainty Specification, Front. Mar. Sci. 8:689695. doi: 10.3389/fmars.2021.6896, 2021

Craig, H.: The GEOSECS program: 1972-1973: Earth Planetary Science Letters, v 23, p 63–64., 1974

Falck, E. and Olsen, A.: Nordic Seas dissolved oxygen data in CARINA, Earth Syst. Sci. Data, 2, 123–131, https://doi.org/10.5194/essd-2-123-2010, 2010.

Garcia, H. E., Weathers, K., Paver, C. R., Smolyar, I.,Boyer, T. P., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Baranova, O. K., Seidov, D., and Reagan, J. R.: World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. A. Mishonov Technical Ed.; NOAA Atlas NESDIS 83, 38pp, 2018.

Good., S., Mills, B., Boyer T., Bringas, F., Castelão, G., Cowley, R., Goni, G., Gouretski, V. and Domingues, C. M.: Benchmarking of automatic quality control checks forocean temperature profilesand recommendations for optimal sets, Front. Mar. Sci., DOI 10.3389/fmars.2022.1075510, 2022.

Gouretski, V., Cheng, L., Du, J., Xing, X., Chai, F.: A quality-controlled and bias-adjusted global ocean oxygen profile dataset, Marine Science Data Center of the Chinese Academy of Sciences, http://dx.doi.org/10.12157/IOCAS.20231208.001

Gouretski, V.: World Ocean Circulation Experiment – Argo Global Hydrographic Climatology, Ocean Sci., 14, 1127-1146, https://doi.org/10.5194/os-14-1127-2018, 2018.

Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global database, Deep-Sea Res., I, 57, 812-833, 2010.

Gouretski, V. V., and Jancke, K.:Systematic errors as the cause for an apparent deep water property variability: global analysis of the WOCE and historical hydrographic data., Prog. Oceanogr., 48, 4, 337-402, 2000.

Gregoire, M. et al.: 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, 1-29, https://doi.org/10.3389/fmars.2021.724913 , 2021.

Hood, E.M., Sabine, C.L., and .M. Sloyan, B.M., eds.:The GO-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidelines, IPCC Report Number 14, ICPO Publication Series Number 134, 2010.

Hubert, M. and Vandervieren, E.: An Adjusted boxplot for skewed distributions, Comput. Stat. Data Anal., 52, 5186–5201, 2008.

Keeling, R.F., Koetzinger, A., and Gruber, N.: Ocean Deoxygenation in a Warming world, Annu. Rev. Mar. Sci., 2, 199-229. doi:10.1146/annurev.marine.010908.163855, 2010.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Koertzinger, A., Schimanski, J., and Send, U.: High quality oxygen measurements from profiling floats: A promising new technique, J. Atmos. Ocean. Technol., 22, 302-308, 2005.

Langdon, C.: Determination of Dissolved Oxygen in Seaweater By Winkler Titration using Amperometric Technique, The GO-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidelines, Version 1, (eds Hood, E.M., C.L. Sabine, and B.M. Sloyan), 18pp.. (IOCCP Report Number 14; ICPO Publication Series Number 134), DOI: https://doi.org/10.25607/OBP-1350, 2010.

Larqué, L., Maamaatuaiahutapu, K., Garçon, V.: On the intermediate and deep water flows in the South Atlantic Ocean. Journal of Geophysical Research, 102,C6, https://doi.org/10.1029/97JC00629, 1997.

Marks, R.: Dissolved oxygen supersaturation and its impact on bubble formation in the southern Baltic Sea, Hydrol. Res., 39,3, 229-236, 2008.Monhor, D. and Takemoto, S.: Understanding the concept of outlier and its relevance to the assessment ofdata quality: Probabilistic background theory, Earth Planets Space, 57, 1009–1018, 2005.

Sarachik, E.S.: CLIVAR: A Study of Climate Variability and Predictability: Science Plan. World Climate Research Programme Report 89, WMO Technical Document No 690. 157 pp, 1995.

Taillander, V., Wagener, T., D'Ortenzio, F., Mayot, N., Legoff, Ras, H. J., Coppola, L., De Fommervault, O. P., Schmechtig, C., Diamond, E., Bittig, H., Lefevre, D., Leymarie, E., Poteau, A., and Prieur A.: Hydrography and biogeochemistry dedicated to the Mediterranean BGC-Argo network during a cruise with RV Tethys 2 in May 2015, Earth Syst. Sci. Data, 10, 627-641, 2018, https://doi.org/10.5194/essd-10-627-2018, 2018.

Takeshita, Y., Martz, O. P., Johnson, K. S., Plant, J. N., Gilbert, D., Riser, S. C., Neil, C., and Tilbrook, B.: A climatology-based quality control procedure for plotting float oxygen data, J. Geoph, Res: Oceans, 118, 1-11, doi:10.1002/jgr.20399, 2013.

Tan, Z., Cheng, L., Gouretski, V., Zhang, B., Wang, Y., Li, F., Liu, Z., Zhu, J.: A new automatic quality control system for ocean profile observations and impact on ocean warming estimate, Deep-Sea Res. Part I: Oceanographic Research Papers, 194, https://doi.org/10.1016/j.dsr.2022.103961, 2023.

Tengberg, A., Hovdenes, J., Andersson, H. J., Brocandel, O., Diaz, R., Hebert, D., Arnerich, T., Huber, C ., Körtzinger, A., Khripounoff, A., Rey, F., Rönning, C., Schimanski, J., Sommer, S. and Stangelmayer, A. : (2006): Evaluation of a lifetime-based optode to measure oxygen in aquatic systems. Limnol. Oceanogr.: Methods, 4, 7-17, 2006.

Thierry, V., Bittig, H., and the Argo-BGC team:Argo quality control manual for dissolved oxygen concentration, Version 2.1, Argo Data Management, doi: https://dx.doi.org/10.13155/46542, 2021.

Uchida, H., Johnson, G. C., and McTaggart, K. E.: CTD Oxygen sensor calibration procedures. The Go-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidlines. IOCCP Report No. 14, ICPO Publication Series No. 134, Version 1, 17p., 2010.

WHPO: WOCE Operations Manual, Section 3.1.3: WHP operations and methods,WOCE report no. 69/91, WHPO 91-1. 80 pp., 1991.

Winkler, L.: Die Bestimmung des in Wasser gelosten Sauerstoffes. Berichte der Deutschen Chemischen Gesellschaft. 21 (2): 2843–2855. doi:10.1002/cber.188802102122, 1888.

https://doi.org/10.5194/essd-2023-518
Preprint. Discussion started: 3 January 2024

Wunsch, C.: Towards the World Ocean Circulation Experiment and a bit of aftermath, Springer-Verlag, Jochum, M., Murtugudde, R., Phys. Oceanogr., 181-201, 2006.

Yang, J., Rahardja, S., and Fränti, P.: Outlier Detection: How to Threshold Outlier Scores?, AIIPCC '19: Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, December 2019. Article No.: 37, p. 1–6 https://doi.org/10.1145/3371425.3371427, 2019.