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	89 059	32 396	41.08
2	CORIOLIS Data Centre, France	Coriolis	63 220	33 233	65.09
3	Commonwealth Scientific and Industrial Research Organization, Australia	CSIRO	19 183	3302	23.75
4	Japan Meteorological Agency, Japan	JMA	15 981	11 233	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 Centre, UK	BODC	3533	1905	61.57
9	Korea Ocean Research and Development Institute, South Korea	KORDI	2239	0	0
10	Korea Meteorological Administration, South Korea	KMA	93	0	0

above-depth range, the geographical coordinates of the profile are considered to be in error, and data at all levels are flagged. According to Table 2, about 0.5 % of OSD and CTD profiles fail this check, compared to only 0.08 % for Argo profiles. For each data type, the spatial distribution of profiles failing this test exhibits a rather random pattern (Fig. S1). The highest percentage of OSD outlier profiles is found for the time period before 1946, probably due to less accurate navigation methods during the war (Fig. S1b). CTD profiles exhibit higher outlier scores above 400 m between 200–2014, linked to several cruises. Only 0.077 % of DAC QC-ed Argo profiles fail this check (Fig. S1g–i).

3.2 Global oxygen range check

The test is applied to identify observations that are grossly in error (so-called "blunders"). 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 for depth levels and temperature surfaces because the maximum oxygen solubility depends on temperature. For the construction of overall limits, we use the normalized frequency histograms (Fig. 4). The depth–oxygen histograms are constructed similarly with normalization at each depth level (Fig. 4b). The normalization is done to account for varying numbers of oxygen observations with depth and 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 (indicated by the green lines). The spatial distribution of the OSD and CTD profiles with levels failing this check broadly corresponds to the sampling density (Figs. S2a and d, S3a and d), whereas flagged Argo profiles can be rather linked to distinct floats (Figs. S2g, S3d). The CTD data are characterized by the largest fraction of profiles affected by this check (Figs. S2e, S3e).

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. The histograms of observed oxygen concentration (C_{obs}) versus maximum oxygen solubility (C_{max}) calculated using reported temperature and salinity in different ocean layers depict a close relationship between the mode of observed oxygen distribution and the maximum solubility (Fig. 5a–d). The histograms also show that the distribution mode for the upper-most layer 0–100 m (Fig. 5a) follows the line $C_{obs} = C_{max}$, progressively deviating to lower C_{max} values when $C_{obs} > 300\,\mu\text{mol kg}^{-1}$, suggesting an oxy-

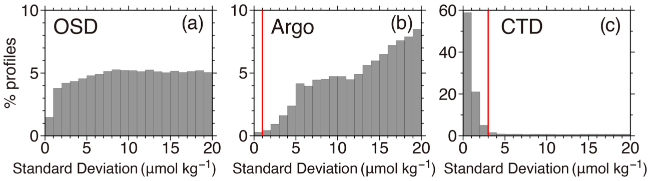

Figure 6. Oxygen profile standard deviation for OSD **(a)**, Argo **(b)**, and CTD **(c)** instrumentation types. Only profiles with at least seven levels of oxygen data are considered. Red vertical lines show the respective threshold values for Argo and CTD profiles.

3.5 Multiple extrema check

The multiple extrema check aims to identify profiles whose shape significantly deviates from the majority of profiles. For each profile with at least seven observed levels (black dots), the number of local extrema and their magnitudes (denoted as M_n in Fig. 7a, defined as oxygen difference between two adjacent oxygen measurements) are calculated. Then, the normalized frequency histograms of oxygen profiles for different combinations of the number of oxygen extrema and of the extremum magnitude are calculated for three instrumentation types separately (Fig. 7b–d). The larger the extremum magnitude, the less frequent the corresponding profiles. Physically, an oxygen profile at a location is not likely to exhibit both too large and too frequent oscillations of oxygen concentrations. Thus, the profiles with many or big extrema are likely erroneous. The histogram for Argo profiles differs from that for OSD and CTD because it is based on profiles already validated by the respective DACs. The multiple extrema check thresholds (black lines in Fig. 7b–d) are defined using the histograms as guidance. The lines crudely correspond to the normalized frequency of 0.01 for OSD and CTD and 0.05 for Argo profiles. The geographical distribution of profiles failing the check is given in Fig. S6a, d, and g. Argo profiles failing the check can be linked to distinct floats (Fig. S6g). The OSD profiles exhibit a higher outlier percentage for the years 1990–2002. The highest rejection rate for the CTD profiles is typical of the years before 2000 (Fig. S6b, e).

3.6 Spike check

Spikes are the values at levels that strongly deviate from the values at the nearest levels above and below. For each observed level 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 an outlier when the test value s exceeds a threshold value. Due to the larger oxygen variability in the upper layers, we set depth-dependent spike thresholds, which are defined for nine depth layers using accumulated histograms for the test value s (Fig. 8a and b for 0–100 and 400–600 m as examples). The threshold profile is defined by the 95 % frequency at each layer (Fig. 8c). The 95 % value is chosen empirically but can be tuned when additional QC-ed benchmark datasets become available. Examples of profiles which failed this check are shown in Fig. 7s. Data from all instrument types are characterized by a rather homogeneous temporal and spatial distribution of outliers.

3.7 Local climatological oxygen range check

The local climatological oxygen range check is one of the most effective QC modules for identifying outliers compared to other checks because the minimum–maximum thresholds are constrained by the local water mass characteristics. For each $1^\circ \times 1^\circ$ latitude–longitude grid point, we calculate min–max thresholds, accounting for the skewness of the data. For calculating climatological ranges, we take 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):

$$[\mathrm{Lf\ Uf}] = [Q1 - 1.5 \cdot \mathrm{IQR} \cdot \exp(-\mathrm{SK} \cdot |\mathrm{MC}|)Q3 + 1.5 \cdot \mathrm{IQR} \cdot \exp(\mathrm{SK} \cdot |\mathrm{MC}|)], \quad (1)$$

TS11 where $Q1$ and $Q3$ are quartiles, $Q2$ is the sample median, and SK is skewness. MC denotes the medcouple, which is defined as MC = median $h(x_i, x_j)$, where $x_i \ll Q2 \ll x_j$, and the kernel function $h(x_i, x_j) = [(x_j - Q2) - (Q2 - x_i)]/(x_j - x_i)$ TS12 (Hubert and Vandervieren, 2008).

The local oxygen ranges are constructed using both the OSD and Argo oxygen profiles. The OSD used to derive the local threshold have undergone the preliminary QC (checks for global oxygen range, spikes, stuck values, multiple extrema), aiming to remove crude outliers to reduce their impact on the local thresholds. This approach is similar to the two-stage thresholding suggested by Yang et al. (2019). The Argo oxygen profiles underwent quality control at the respective DACs.

The local minimum and maximum thresholds were calculated at $1^\circ \times 1^\circ$ grids at a set of 65 depth levels corresponding to the levels implemented for the World Ocean Circulation Experiment – Argo Global Hydrographic Climatology (Gouretski, 2018) using Eq. (1). Examples of the threshold spatial distribution are presented for two depth levels: 98 m (level typically located below the seasonal thermocline, Fig. 9a–c) and 1050 m (level typically located below the main thermocline, Fig. 9d–f). The most striking features are the areas with low minimum oxygen values (oxygen minimum zones, Fig. 9a, b) in the east Pacific, Arabian Sea, Bay of