Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO2

Seth M Bushinsky[1], Zachary Nachod[1], Andrea J Fassbender[2], Veronica M Tamsitt[3], Yuichiro Takeshita[4], and Nancy William[5]

[1]University of Hawaii at Mānoa
[2]NOAA Pacific Marine Environmental Laboratory
[3]University of South Florida St. Petersburg
[4]Monterey Bay Aquarium Research Institute
[5]No Afilliation

April 01, 2024

Abstract

Profiles of oxygen measurements from Argo profiling floats now vastly outnumber shipboard profiles. Air calibration of a float's oxygen optode upon surfacing enables accurate measurements in the upper ocean but does not necessarily provide similar accuracy at depth. In this study we use a quality controlled shipboard dataset to show that, on average, the entire Argo oxygen dataset is offset relative to shipboard measurements (float minus ship) at pressures of 1450 to 2000 db by -1.9 ± 4.7 µmol kg-1 (95% confidence interval around the mean of {-2.3, -1.5}) and air calibrated floats are offset by -3.1 ± 5.3 µmol kg-1 (95% CI: {-3.7, -2.4}). The difference between float and shipboard oxygen is most likely due to offsets in the float oxygen data and not due to oxygen changes at these depths or biases in the shipboard dataset. In addition to posing problems for the calculation of long-term ocean oxygen changes, these float oxygen offsets impact the adjustment of float nitrate and pH measurements and therefore bias important derived quantities such as the partial pressure of CO2 (pCO2) and dissolved inorganic carbon. Correcting floats with air-calibrated oxygen for the float-ship oxygen offsets changes float pH by 3.2 ± 3.8 mpH and float-derived surface pCO2 by -3.3 ± 4.1 µatm. This adjustment to float pCO2 represents half, or more, of the bias in float-derived pCO2 reported in studies comparing float pCO2 to shipboard pCO2 measurements.

Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO_2

Seth M. Bushinsky[1], Zachary Nachod[1], Andrea J. Fassbender[2], Veronica Tamsitt[3], Yuichiro Takeshita[4], Nancy Williams

[1]Department of Oceanography, School of Earth and Space Science and Technology, University of Hawai‘i at Mānoa, Honolulu, HI, USA
[2]NOAA/OAR Pacific Marine Environmental Laboratory, Seattle, WA, USA
[3]University of South Florida, St. Petersburg, FL, USA
[4] Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

Abstract

Profiles of oxygen measurements from Argo profiling floats now vastly outnumber shipboard profiles. Air calibration of a float’s oxygen optode upon surfacing enables accurate measurements in the upper ocean but does not necessarily provide similar accuracy at depth. In this study we use a quality controlled shipboard dataset to show that, on average, the entire Argo oxygen dataset is offset relative to shipboard measurements (float minus ship) at pressures of 1450 to 2000 db by -1.9 ± 4.7 µmol kg^{-1} (95% confidence interval around the mean of {-2.3, -1.5}) and air calibrated floats are offset by -3.1 ± 5.3 µmol kg^{-1} (95% CI: {-3.7, -2.4}). The difference between float and shipboard oxygen is most likely due to offsets in the float oxygen data and not due to oxygen changes at these depths or biases in the shipboard dataset. In addition to posing problems for the calculation of long-term ocean oxygen changes, these float oxygen offsets impact the adjustment of float nitrate and pH measurements and therefore bias important derived quantities such as the partial pressure of CO_2 (pCO_2) and dissolved inorganic carbon. Correcting floats with air-calibrated oxygen for the float-ship oxygen offsets changes float pH by 3.2 ± 3.8 mpH and float-derived surface pCO_2 by -3.3 ± 4.1 µatm. This adjustment to float pCO_2 represents half, or more, of the bias in float-derived pCO_2 reported in studies comparing float pCO_2 to shipboard pCO_2 measurements.

Plain Language Summary

Oxygen has historically been measured using chemical titrations on water collected by ships at sea. Over the past 20 years, sensors that measure oxygen have been deployed on robotic profiling floats. Measurements by oxygen sensors on profiling floats now greatly exceed those collected by ships. Here we compare all float oxygen data collected to shipboard measurements in deep waters (1450 to 2000 m depth) where we do not expect oxygen to be changing in the ocean. We find a difference between float and shipboard data. If left uncorrected, this difference would give the false impression of a long-term oxygen change. This difference also impacts float-measured pH and float-estimated carbon dioxide, both of which rely on float oxygen measurements. Correcting oxygen, and therefore float pH and carbon dioxide, would largely address a widely-studied bias in float measurements of these parameters.

Key Points

1. Air-calibrated float oxygen measurements are lower than shipboard data by 3.1 μmol kg^{-1} at pressures of 1450 to 2000 db.
2. Correcting float oxygen for this offset would increase float pH by 3.2 mpH and lower float-derived pCO_2 by -3.3 μatm.
3. This float oxygen offset would improve float and ship pCO_2 comparisons, removing most or all of observed biases.

1. Introduction

Measurements of oxygen in the ocean are critical, as the ocean is losing oxygen due to warming, respiration, and stratification-induced changes to ventilation (Helm et al., 2011; Levin, 2018; Schmidtko et al., 2017; Stramma & Schmidtko, 2021). Measurements of ocean oxygen also can be used to understand the balance between photosynthesis and respiration in the surface and deep ocean (Bushinsky & Emerson, 2015; Emerson, 1987; Hennon et al., 2016; Yang et al., 2017). Historically, oxygen was primarily measured by Winkler titrations of water sampled from the ocean (Carpenter, 1965; Dickson, 1994). These titrations require reliable standards and trained operators to acheive high accuracy and precision. Winkler titrations have been supplemented by electrochemical and optical sensors attached to CTDs but have remained the gold standard of accurate oxygen measurement in the ocean.

Oxygen measurements on autonomous profiling floats are transforming our ability to observe the ocean at unprecedented levels of detail. The first Argo oxygen data was from floats deployed in the early 2000s equipped with Clark-cell electrodes (Figure 1) that provided fast response times but could drift rapidly and in unpredictable ways (Gruber et al., 2007). Oxygen optodes that measure the partial pressure of oxygen through an oxygen-sensitive luminophore have now become the standard oxygen sensor deployed on floats (Claustre et al., 2020; Gruber et al., 2009; Körtzinger et al., 2005; Tengberg et al., 2006) (Figure 1). Over 1800 Argo oxygen floats have been deployed as of 2023, with increasing numbers due to a mix of many small deployments by individual research groups and the development of a few projects deploying large numbers of floats (e.g., the Southern Ocean Carbon and Climate Observations and Modeling project, SOCCOM, (Johnson et al., 2017; Sarmiento et al., 2023); Global Ocean Biogeochemistry Array, GO-BGC (Roemmich et al., 2021; Schofield, O. A. et al., 2022)). Optodes drift significantly prior to deployment, but the advent of calibration using atmospheric oxygen has increased accuracy of surface data to better than 1% (Bittig, Körtzinger, et al., 2018; Bushinsky et al., 2016; D'Asaro & McNeil, 2013; Johnson et al., 2015). Floats calibrated using atmospheric data primarily use a single multiplicative gain correction based on the difference between float air measurements and calculated atmospheric oxygen that is then applied to the whole float profile (Figure 2). This approach assumes that the response of the oxygen optode at different oxygen levels and temperatures changes uniformly over time, and that the zero reading of the optode remains unchanged.

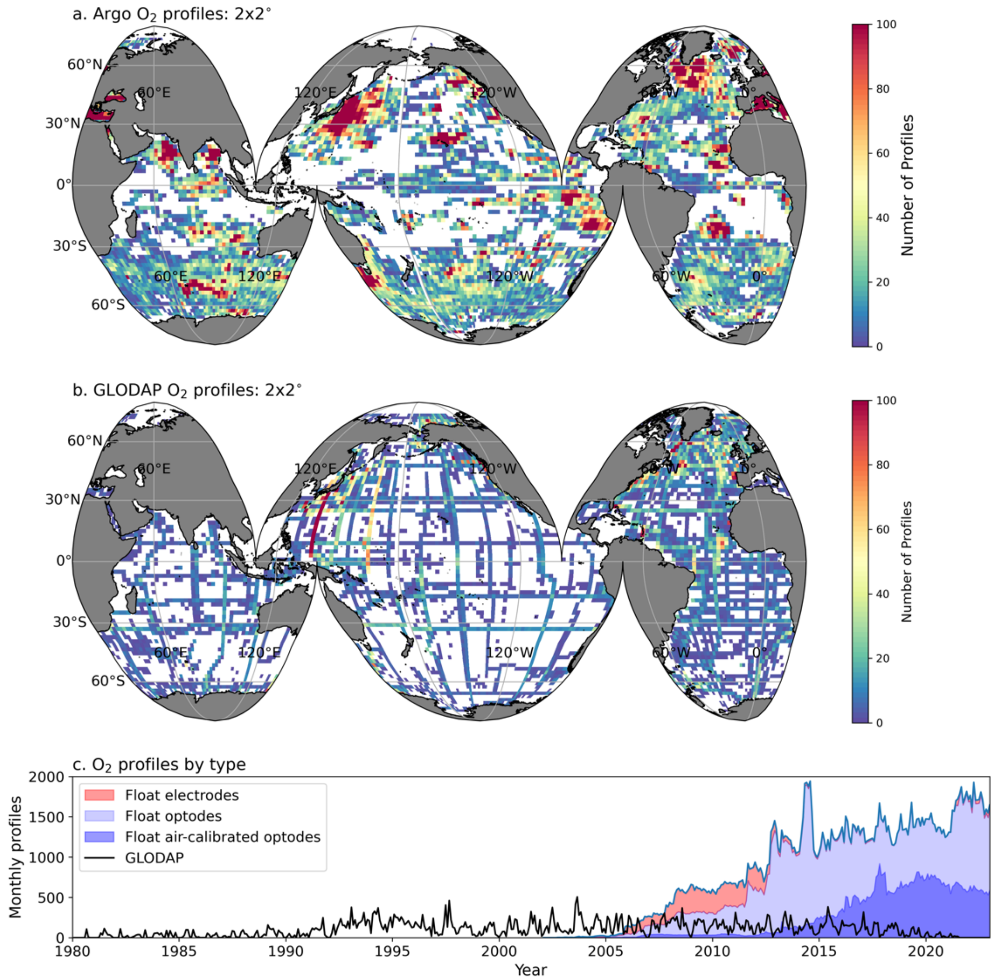

Figure 1. Observational density of float and ship oxygen and change in float sensor type over time. Float (a.) and GLODAP (b.) oxygen profile density on a 2 × 2° grid. (c.) Total number of oxygen profiles per month, with the contribution of float sensor type shown (Clark cell electrodes (red) and optodes (blue), and the subset of optodes calibrated with in-air measurements (dark blue)), and total number of GLODAP profiles that contain oxygen per month (black line).

The Global Ocean Data Analysis Project (GLODAP; Key et al., 2015; Olsen et al., 2016, 2020) is an on-going international synthesis effort that evaluates surface-to-bottom ocean biogeochemical bottle data for outliers and internal consistency, which has been crucial to improving the accuracy and usability of shipboard data. A similar effort, the Surface Ocean CO_2 Atlas (SOCAT) collects and assesses surface pCO_2 measurements for accuracy prior to inclusion in an annual data product (Bakker et al., 2016). These efforts have been instrumental in our understanding of ocean biogeochemistry. There is currently no comparable, on-going, post-

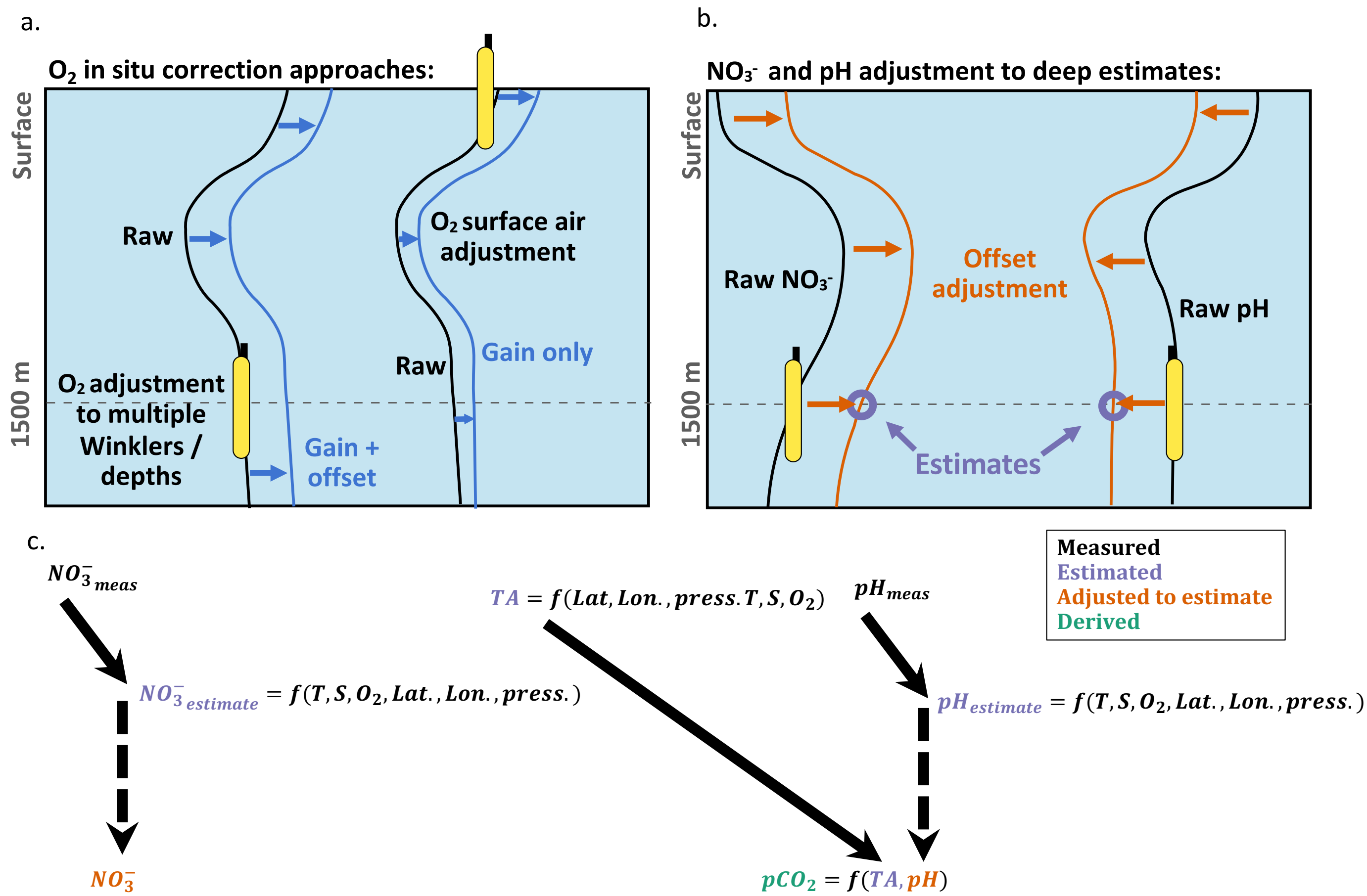

Figure 2. Schematic of the two primary in-situ oxygen calibration approaches and how the calibrated oxygen data are used to adjust float pH and nitrate measurements. (a.) Oxygen data collected on floats are typically calibrated using a gain and offset correction based on in-situ data, typically collected at float deployment, or a gain-only correction based on air-calibration measurements made upon float surfacing throughout its lifetime. (b.) Float nitrate and pH measurements are adjusted to algorithmic estimates of those parameters at 1500 db in a way that impacts the whole profile. (c.) Variable inputs, including oxygen, to the nitrate and pH algorithms, which are trained on shipboard data. Offsets between the float and ship oxygen will propagate through these algorithms into the adjusted nitrate and pH data.

deployment census of float biogeochemical data to assess consistency within the float dataset or between float and shipboard data. Most studies that have used Argo oxygen data and assessed their accuracy have focused on surface data or analyzed changes measured by individual floats, thereby avoiding comparisons between different float measurements (Bushinsky & Emerson, 2015; Johnson et al., 2015; Martz et al., 2008; Wolf et al., 2018). As the float array expands and researchers begin using the entire dataset as a whole, or combining float and shipboard datasets to create global interior ocean oxygen products (e.g., World Ocean Atlas (Garcia et al., 2010); Gridded Ocean Biogeochemistry from Artificial Intelligence – Oxygen (Sharp et al., 2022)), attention must be paid to the accuracy of float oxygen data at depth.

Largescale comparisons of float and ship datasets indicate that, to a first order, float oxygen data have reached a comparable level of accuracy as ship-board measurements (Bushinsky et al., 2017; Johnson et al., 2017; Maurer et al., 2021). However, two float-to-ship comparisons of deep (1500-2000 m) oxygen data (Bushinsky et al., 2016; Drucker & Riser, 2016) indicate that surface calibration may not be a sufficient adjustment for the entire float oxygen profile. There is some indication that this could be due to inadequate calibration of the optode temperature response (Bittig, Körtzinger, et al., 2018), but Bushinsky et al. (2016) re-calibrated the optode temperature response for 11 floats in the lab prior to deployment, which did not address the difference in deep data, so this approach does not seem to fully resolve the observed drift.

The accuracy of oxygen data throughout the water column is critical for understanding long-term changes in ocean oxygen content. At depth, the accuracy of quality-controlled float oxygen data is especially critical due to its use in adjusting other biogeochemical sensor data from Argo floats, which has downstream impacts on derived carbonate system parameters. Nitrate and pH sensors are now being deployed in large numbers (Claustre et al., 2020) and also require in-situ adjustment to correct sensor drift (Maurer et al., 2021). The nitrate and pH measurements are adjusted using estimates of these parameters at 1500 db that are derived from multiple linear regression or neural network algorithms that were trained on the GLODAP shipboard database (Bittig, Steinhoff, et al., 2018; Carter et al., 2016, 2021; Williams et al., 2016). These algorithmic estimates use inputs of temperature, salinity, depth, latitude, longitude, and, importantly, oxygen. Due to sensor response characteristics, adjustments to the nitrate and pH sensor data are applied almost uniformly to the entire profile, so offsets at 1500 m directly impact surface measurements (Maurer et al., 2021).

Float measurements of pH are widely used to estimate $p\mathrm{CO_2}$, and other carbonate system parameters, using an algorithm-based estimate of total alkalinity (TA) and a carbonate system calculator. Many studies apply an additional adjustment to the quality controlled pH data prior to estimating $p\mathrm{CO_2}$ that is meant to correct for an empirical pH-dependent pH bias (Carter et al., 2013, 2018; Williams et al., 2017). The accuracy of float $p\mathrm{CO_2}$ estimates is critical, as studies relying on float $p\mathrm{CO_2}$ have identified significant differences between seasonal cycles of $p\mathrm{CO_2}$, and wintertime air-sea CO_2 fluxes from studies that rely on shipboard pCO_2 alone (e.g. Bushinsky et al., 2019; Gray et al., 2018; Williams et al., 2016). Floats measure year-round, including during winter months when rough weather makes shipboard observations rare, which provides immense observational value, if accurate. A number of studies have directly compared float pCO_2 estimates to shipboard observations (Bushinsky & Cerovečki, 2023; Coggins et al., 2023; Fay et al., 2018; Gray et al., 2018; Williams et al., 2017), while other studies have indirectly compared float pCO_2 and pH through assessment of other carbonate system parameters or CO_2 fluxes (e.g., Long et al., 2021; Mackay & Watson, 2021; Wu et al., 2022). A recent meta-analysis of float pCO_2 accuracy found biases are likely between 6-9 µatm (float pCO_2 high), with direct comparisons yielding lower biases (2-5 µatm) than indirect comparison (Wu & Qi, 2023).

Here we use crossover comparisons between ship and float data to quantify differences in deep oxygen values and calculate the impact those offsets would have on adjusted and derived parameters. We refer to oxygen "offset", without strict attribution of the source of the error. Given the long-history and extensive use of Winkler titrations and the relatively short time-history of oxygen sensors and float-mounted oxygen sensors, we assume that the GLODAP values are likely correct and will present evidence later that supports this assumption.

2. Methods

2.1 Float and shipboard datasets

Biogeochemical float data were downloaded on June 20, 2023 from the Argo Data Assembly Centers (DACs, closest snapshot: 2023-06-09 (Argo, 2023)) according to the list of floats with oxygen, nitrate, or pH sensors in the Argo Global DACs Synthetic-Profile Index file. 1,839 Argo floats deployed between 2002 and 2023 were included in this dataset. Of these floats, 1,830 measured oxygen, of which 151,880 out of 270,654 profiles (56%) contained valid delayed mode "ADJUSTED" oxygen data, indicating the data had been checked and/or corrected post deployment (median difference between uncorrected and adjusted data was 11 µmol kg^{-1}, adjusted data higher). 582 floats measured nitrate, with 43,280 delayed mode, adjusted profiles out of 67,505 total (64%). 492 floats measured pH, of which 19,137 profiles had delayed mode, adjusted data out of 44,407 total (43%). Only delayed mode, adjusted float data flagged as "good" were used for this analysis. Float oxygen sensor calibration type for floats with valid crossovers was determined by reading the "SCIENTIFIC_CALIB_COMMENT" field (198 unique comments, Table S1) in float "Sprof" files and were categorized as: air, non-air, and bad/no calibration.

Shipboard bottle measurements of salinity, temperature, oxygen, nitrate, total CO_2 (DIC), or pH flagged as "good" from the GLODAP v2.2022 (Olsen et al., 2020) were used for comparison with float observations. The GLODAP dataset was chosen because it includes a secondary quality control and adjustment to shipboard data for overall accuracy and internal consistency. Shipboard pH measurements in the GLODAP dataset include a variety of measurement approaches. Carter et al. 2021 and 2018 "homogenized" the GLODAP pH dataset to be consistent with spectrophotometric pH measurements using purified meta-cresol purple (Liu et al., 2011) prior to training of the LIPHR (Locally Interpolated PH Regression)/ESPER (Empirical Seawater Property Estimation Routine) algorithms. To recreate this homogenized dataset, we used ESPER to calculate GLODAP pH on the total scale normalized to 25°C ($pH_{25\text{-}T}$) for every datapoint where a "good" pH measurement was present. By using the algorithms at the same datapoints on which they were trained, we recreated the homogenized dataset and used this data for the pH crossover comparison.

2.2 Crossover comparisons

Criteria for crossover comparisons between float and shipboard measurements were established using distance, pressure, potential density, and spice. For each float profile, we first found all GLODAP bottle measurements within a 100 km radius. Float data from 1450 to 2000 db were interpolated to a 1 db grid, with any gaps of over 125 db between successive float measurements removed from the interpolated profiles. Potential density (σ_θ) and spiciness (τ) relative to 0 db were calculated for both the float and shipboard data using the Gibbs SeaWater Oceanographic Toolbox of TEOS-10 for Python (McDougall & Barker, 2011). For each GLODAP bottle measurement between 1400 and 2100 db within the 100 km distance range, crossovers were determined by looking for interpolated float data with differences of less than ± 0.005 kg m^{-3} σ_θ, ± 0.005 τ, and ± 100 db from the GLODAP sample properties. These difference thresholds were selected by analyzing levels of environmental oxygen noise in comparisons of individual floats against themselves using a range of density, spiciness, and distance thresholds (Text S1, Figure S1). Crossovers from any point in time were allowed. Mean property offsets (e.g., ΔC_{off}, for a given property "C") were calculated for floats with at least 20 oxygen crossovers present and used in the results shown here (Figure S2 for float crossover examples). On average, offsets in the 1450 – 2000 db range did not differ significantly as a function of depth or concentration (Figure S3). While adjusting the filter criteria does impact the number of crossovers found for each float and can impact the mean offset calculated for an individual float, the overall results presented in this manuscript are relatively insensitive to the exact criteria used.

2.3 Impact of oxygen offsets on derived parameters

The impact of oxygen offsets (e.g., ΔC_{imp}, for a given property C) on float nitrate and pH adjustments and float estimated pCO_2 and DIC ($\Delta NO_3^-{}_{imp}$, $\Delta pH_{25\text{-}T,imp}$, $\Delta pCO_{2,imp}$, and ΔDIC_{imp}, respectively) was determined for each float with valid oxygen and nitrate, or pH data and with a valid GLODAP crossover comparison. As previously described, oxygen offsets impact estimated pCO_2 and DIC through the effect of oxygen on the pH adjustment at 1500 m (Figure 2). While oxygen is also used in the algorithmic estimation of total alkalinity, which is required for pCO_2 and DIC estimation, only a surface oxygen offset would impact the surface total alkalinity estimate, and subsequently the surface pCO_2 and DIC estimates, which are the foci of this work. Using temperature, salinity, and oxygen data at 1500 m from each profile as inputs to the calibration algorithms, we calculated pH and nitrate, both with and without adjusting for the mean float oxygen offset relative to GLODAP ($\Delta O_{2,off}$). We then calculated the differences in (impacts on) pH and nitrate with and without the $\Delta O_{2,off}$ correction ($\Delta pH_{25\text{-}T,imp}$ and $\Delta NO_3^-{}_{imp}$) and applied these differences to the adjusted, full float profiles. This approach allows us to determine the impact a mean oxygen correction would have without attempting to replicate any step changes or non-linear adjustments that may have been applied during the original data adjustment procedure (Maurer et al., 2021). For nitrate, a uniform adjustment is applied to the entire nitrate profile. For pH, the adjustment at 1500 m is scaled relative to the difference in temperature between each depth and 1500 m, following the protocol used in the original pH measurement adjustment (Johnson et al., 2023). $\Delta pCO_{2,imp}$ and ΔDIC_{imp} were calculated using CO2SYS first with the original pH as an input and then with mean $\Delta pH_{25\text{-}T,imp}$ applied to the float profile and calculating the difference.

Floats equipped with pH and nitrate sensors have been deployed by many groups throughout the world. The majority have been deployed as part of SOCCOM or GO-BGC and adjusted by the data management teams of SOCCOM and the Monterey Bay Aquarium Research Institute (MBARI). Following adjustment approaches used by SOCCOM/MBARI, total alkalinity was calculated using LIARv2, pH using ESPER (ESPER-mixed, an average of a neural network-based approach and an MLR; Carter et al., 2021) or LIPHR (Carter et al., 2018), and carbonate system calculations using PyCO2SYS (v1.8.1; Humphreys et al., 2022, 2023). Both ESPER and LIPHR were used because SOCCOM pH data at the time of download have been adjusted to one of the two algorithms, depending on when the float was last active. Floats active prior to April 2023 were adjusted to LIPHR pH estimates, while floats active past this date are adjusted to ESPER pH. Figures and results shown in the main text rely on ESPER-based adjustments while the supplement includes results that rely on LIPHR-based adjustments, but average differences between the two are of a second order.

One complication is the presence of a pH-dependent pH correction (Williams et al., 2017) used by the SOCCOM/GO-BGC groups when calculating pCO_2. This correction accounts for the difference between float ISFET measured pH, which has been shown to align with spectrophotometric pH measurements (Takeshita et al., 2020), and pH values calculated from measurements of TA and DIC (Carter et al., 2013; 2018). It is currently unclear how to best apply a global correction similar to the one developed in Williams et al. (2017) and a recently published paper by a working group focused on inter-consistency in carbonate system measurements recommended removing this correction until a suitable correction for all float pH measurements can be developed (Carter et al., 2023). We have avoided dealing with this issue by looking at the difference between calculations with and without the oxygen offset impacts included. Any changes in this pH-dependent bias correction will represent a different, additional impact to float pCO_2 estimates. However, the impacts of the oxygen offset should be very similar for estimated pCO_2 with or without this additional pH-dependent bias correction.

3. Results

For floats where oxygen offsets could be determined, 93% of the offsets were statistically significant using a 1 sample t-test and a p-value threshold of 0.01 (Figures S2A and S2B for examples with significant and non-significant crossovers). Our goal is to quantify the potential impact of float oxygen offsets on other parameters, so we include all oxygen offset estimates, including insignificant ones, when calculating the mean oxygen offset and all floats with pH in the subsequent impact on derived parameters. Only including floats with a statistically significant oxygen offset would overstate the magnitude of the mean dataset offset. However, if these crossover comparisons were used in the future to correct float oxygen data it would be important to only adjust those floats with significant offsets so as not to over-adjust already good data.

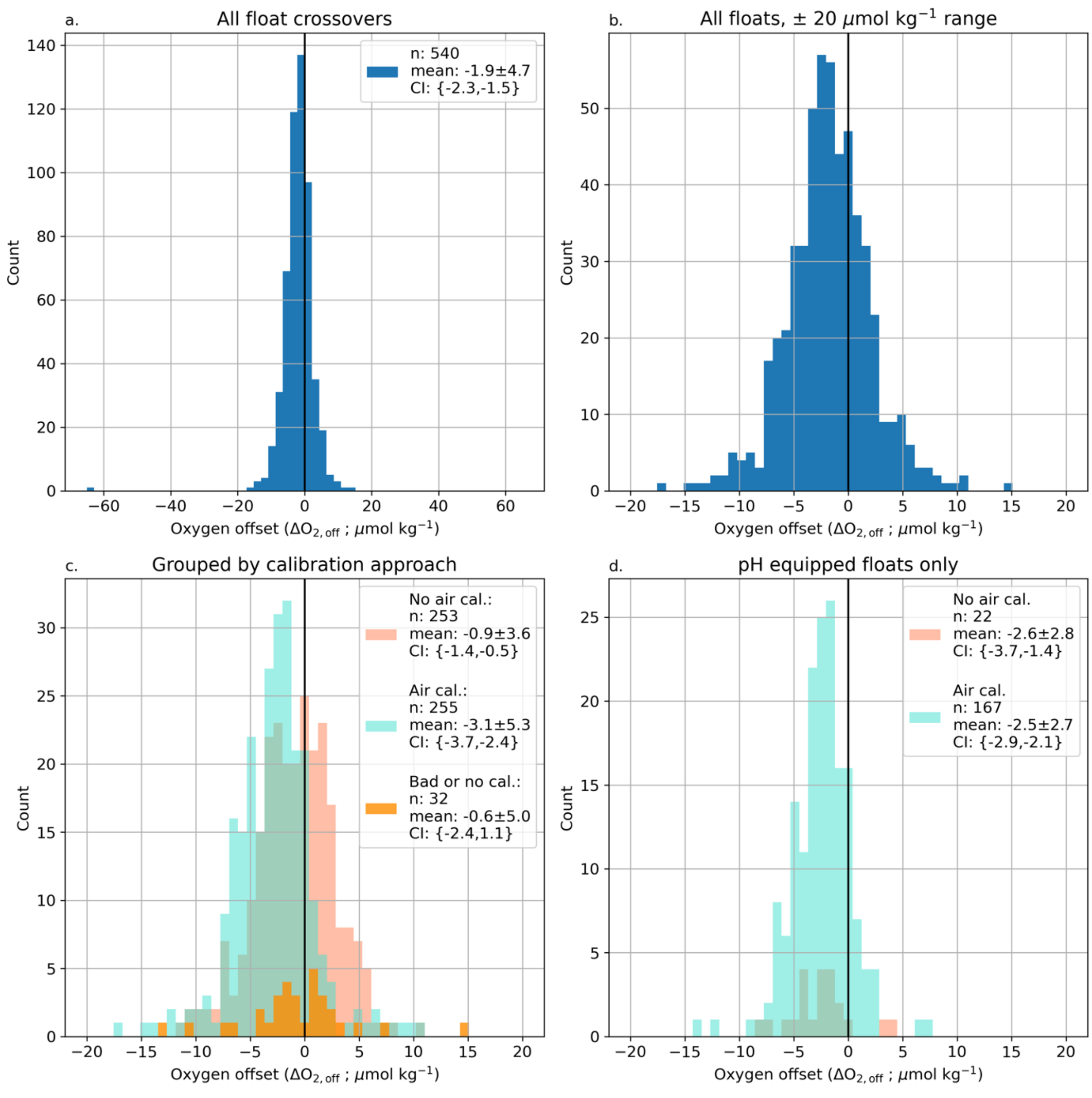

Figure 3. Histograms of float oxygen offsets relative to shipboard measurements. Float minus ship mean offsets between 1450 and 2000 db for all floats (a) with crossovers to GLODAP data and within ± 0.005 kg m^{-3} σ_θ, ± 0.005 τ, and ± 100 db. (b) same as (a) but with a restricted x-axis range of ± 20 μmol kg^{-1}. (c) Offsets grouped by calibration type: non-air calibration listed (light red), air-calibrated (turquoise), and no calibration or bad calibration, though still marked as good data (orange). (d) same as (c) but only showing data from non-air and air calibrated floats equipped with pH sensors. Figure legends list the number of floats, mean offsets ± 1 SD, and 95% confidence intervals for each calibration category.

The mean oxygen offset for all floats is -1.9 ± 4.7 (1 SD) µmol kg^{-1} (n=540, float minus GLODAP, Figure 3, Table S2), with a 95% confidence interval around the mean of -2.3 to -1.5 µmol kg^{-1}. Of the float oxygen sensors with valid crossovers, 253 had a non-air calibration method listed (offset of -0.9 ± 3.6, 95% CI {-1.4, -0.5} µmol kg^{-1}), 255 were air calibrated (offset of -3.1 ± 5.3, 95% CI {-3.7, -2.4} µmol kg^{-1}), and 32 had bad calibration or no calibration listed, though they were still flagged as good data (offset of -0.6 ± 5.0, 95% CI {-2.4, 1.1} µmol kg^{-1}) (Figure 3). For floats equipped with pH, non-air calibrated floats had a mean oxygen offset of -2.6 ± 2.8 µmol kg^{-1} (n=22, 95% CI {-3.7, -1.4} µmol kg^{-1}) and air-calibrated floats had a mean oxygen offset of -2.5 ± 2.7 µmol kg^{-1} (n=167, 95% CI {-2.9, -2.1} µmol kg^{-1}).

Offsets were also calculated between GLODAP and float measurements of nitrate and pH and between float estimates of DIC and GLODAP measurements (Figure 4, blue shaded histograms). The mean difference between float and GLODAP nitrate measurements is 0.2 ± 0.4 µmol kg^{-1} ($\Delta NO_3^-{}_{off}$, n=232, 95% CI {0.1, 0.2} µmol kg^{-1}). The mean difference between float and GLODAP pH (normalized to 25°C, total pH scale) is -4.5 ± 7.7 mpH ($\Delta pH_{25\text{-}T,off}$, n=113, 95% CI {-5.9, -3.1} mpH). The mean difference between DIC estimated from float pH and LIARv2 alkalinity and GLODAP DIC measurements is 2.9 ± 6.1 µmol kg^{-1} (ΔDIC_{off}, n=154, 95% CI {1.9, 4.0} µmol kg^{-1}, full statistics for nitrate, $pH_{25\text{-}T}$, and DIC crossovers in Table S3).

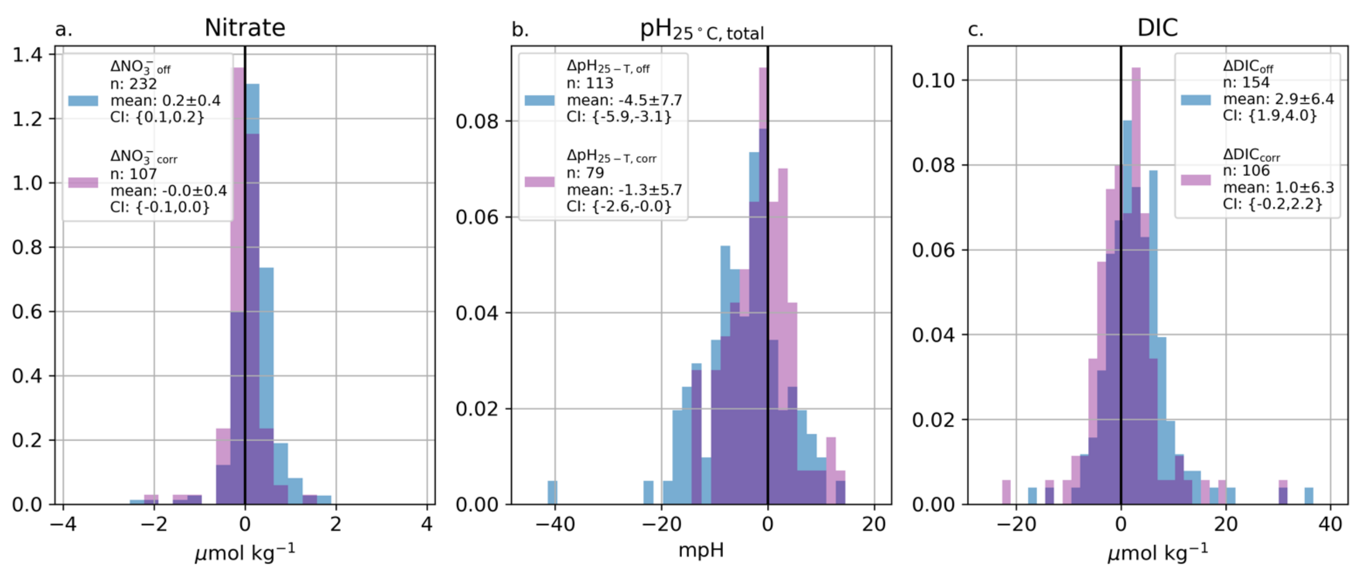

Figure 4. Frequency distributions of original nitrate, pH, and DIC offsets relative to GLODAP crossovers ($\Delta NO_3^-{}_{,off}$, ΔpH_{off}, ΔDIC_{off}, blue shaded histograms) and crossover comparisons after the impact of correcting for the oxygen offset has been applied ($\Delta NO_3^-{}_{,corr}$, ΔpH_{corr}, ΔDIC_{corr}, purple shaded histograms). Correcting for the observed oxygen offset at depth fully corrects the nitrate offset relative to GLODAP and improves the pH and DIC crossover comparisons by 3.2 mpH and 1.9 µmol kg^{-1}, respectively. Full statistics in Tables S3 and S4.

4. Discussion

4.1 Oxygen offsets

The magnitude of the mean offset is larger for air-calibrated floats (-3.1 µmol kg^{-1}) than non-air calibrated floats (-0.9 µmol kg^{-1}). This likely reflects that some non-air calibrated floats are corrected using both deep and near-surface shipboard values or a full oxygen profile from a cast made at the deployment location (Figure 2) (Drucker & Riser, 2016; Takeshita et al., 2013), while air calibrated floats are primarily adjusted using a gain value derived from surface measurements of atmospheric oxygen over the float's lifetime (Bushinsky et al., 2016; Johnson et al., 2015, 2017; Maurer et al., 2021). The offsets shown in Figure 3 are small relative to the original correction from raw to "adjusted" oxygen (median difference between raw and adjusted float oxygen of -11 µmol kg^{-1}), but the mean offset for all comparisons other than the "Bad / no. cal" is significantly different from zero (95% confidence intervals, Figure 3 and Table S2) and, as we will discuss below, can have significant impacts on interpretation and use of this data.

The offset between float and shipboard data could reflect either that the float oxygen measurements are low or ship-based Winkler measurements are high. We will discuss both possibilities. It is important to recognize that we are considering approximately 12 different oxygen sensor models that utilize two different measurement principles made by three different manufacturers. The earliest oxygen sensors deployed on floats were initially Clark-cell electrodes (various SeaBird 43 models) while most current sensors are oxygen optodes made by multiple manufacturers, but with the same basic sensing approach and chemistry (Figure 1). Here we primarily focus on possible offsets for oxygen optodes rather than Clark cell electrodes, and on air-calibrated optodes specifically, as these are the current state of the art sensors and represent the bulk of floats deployed with pH sensors.

If the offset between float and shipboard oxygen is due to low-biased float oxygen measurements, three possible causes of the offsets are: (i) an in-situ difference in ocean oxygen content, (ii) a residual uncorrected pressure response, or (iii) a non-linear concentration-dependent drift in the sensor response. Basin-scale comparisons of float oxygen to shipboard measurements have indicated overall good float sensor accuracy in the surface ocean. A comparison of SOCCOM float oxygen data relative to GLODAP on pressure surfaces above and below the thermocline indicated that the air-calibrated float data at depth may be low by 2-3 μmol kg^{-1} (Bushinsky et al., 2017). Johnson et al. (2017) assessed SOCCOM float data relative to the respective float deployment cruises, across the full depth range, and against GLODAP crossovers within 20 km and below 300 m, finding that float oxygen was lower than GLODAP by 3.7 and 3.2 μmol kg^{-1} at "Midrange", or approximately 250 μmol kg^{-1} [O_2]. Maurer et al. (2021) updated the Johnson et al. (2017) GLODAP comparison, finding a 3.6 to 3.8 μmol kg^{-1} low oxygen offset.

While these earlier studies included shallower waters and compared data on pressure surfaces instead of the combined σ_θ, τ, and pressure criteria of this study, all found similar magnitude and direction of differences. Maurer et al. (2021) attribute the mean [O_2] difference to a mean age difference of 18.6 years between the GLODAP and SOCCOM datasets due to the linear declines in Southern Ocean interior oxygen concentrations found in Helm et al. (2011). In Helm et al. (2011) the Southern Ocean has the largest region of oxygen change in the upper ocean (100-1000 m) with the south Pacific and south Indian basins, seeing changes of up to -0.7 μmol l^{-1} yr^{-1} between 1970 and 1992. However, deeper in the water column where we are conducting our crossover comparison (1500 – 2000 m), changes are of a smaller magnitude and appear to primarily be between 0 to -0.1 μmol l^{-1} yr^{-1} in the Southern Ocean and between -0.2 and +0.2 μmol l^{-1} yr^{-1} across the global ocean. To rule out true oxygen changes as the source of offsets between the float and GLODAP datasets, we can assess the GLODAP dataset for long-term oxygen changes and compare that oxygen change to the magnitude of float oxygen offsets.

If true oxygen changes at depth are responsible for the offsets found in our present study, we would expect the offset for a given float relative to GLODAP to be larger when compared to older cruise data and decrease in magnitude (typically becoming less negative) the closer in time between the float and GLODAP measurements. We therefore fit regressions to the float-GLODAP offsets as a function of time, calculating a 95% confidence interval around the regression, to determine if the regression ± CI included a zero offset at the midpoint time of the float deployment (see Figure S2A/B for examples of the regression and CI). 66 float offsets (12%) could be a result of observed change in oxygen concentration as determined by a trend in the GLODAP dataset. However, in many of these cases, there is simply too much uncertainty in the GLODAP oxygen regression with time, or little to no difference between the float and shipboard data. While actual ocean oxygen change in the 1450 – 2000 db range remains a possible contributing factor to the differences seen between the float and shipboard oxygen, it does not seem to be the main cause.

The optode pressure correction combines a sensor pressure response with the pressure effect on oxygen solubility. Significant effort has been made to determine the appropriate pressure correction between optode response and oxygen partial pressure (e.g., Bittig et al., 2015; Uchida et al., 2008). We primarily focus on the mean oxygen offset for each float, averaging all crossover data from 1450 – 2000 db. However, if we bin the data by depth rather than pressure, we do not see a depth dependency in the float offsets in this depth range. While this may still exist, it does not seem to be a first order factor in the deep oxygen offset (Figure S3). A related issue is the lag in optode oxygen measurements due to the sensor response time (Bittig et al., 2014; Bittig & Körtzinger, 2017). Oxygen gradients at these depths are small, so is not likely to be a first order issue but may contribute to the offset in some regions.

Optodes have been shown to not change response at zero oxygen (Johnson et al., 2015) and multiple calibrations over time have indicated relatively linear drift rates (Bittig & Körtzinger, 2015), lending support for the use of a gain correction across entire oxygen profiles. Bushinsky et al. (2016) measured greater drift rates at lower oxygen concentration and postulated that the faster drift rate at low oxygen concentrations represents a deformation of the oxygen calibration surface, such that neither a gain nor an offset can be used to fully correct the range of oxygen measurements. Drucker and Riser (2016) show data from eight floats indicating that a near-surface gain correction leaves a negative offset at depth, similar to Bushinsky et al. (2016), but they did not provide an explanation. Both Bushinsky et al. (2016) and Drucker and Riser (2016) compared float data to bottle oxygen measurements from deployment casts, so oxygen change at depth did not play a factor. While we are still uncertain as to the mechanism causing low float oxygen values at depth relative to Winkler data, our current results from the entire oxygen Argo dataset indicate that this does not seem to be an issue limited to a small number of floats and the possibility of non-linear drift in the oxygen calibration surface remains.

An alternative to the float oxygen measurement being biased low, is that bottle measurements of oxygen using Winkler titrations are biased high due to contamination by atmospheric oxygen or impurities in reagents (Schmidtko et al., 2017). Sampling of oxygen and subsequent Winkler titration involves careful isolation of the water from atmospheric contamination. Incomplete flushing or trapped bubbles can add a significant bias to Winkler data given that the low solubility of oxygen in water means that in equal volumes of air and water, the air will hold 50 times more oxygen than the water. GLODAP data has been QC'd and adjusted for internal consistency, but it is possible a bias is present in the deep data, especially if, on average, this bias is present in either all shipboard observations or the shipboard observations used as a reference for GLODAP adjustments.

Regardless of the source of the offset or mechanism for its existence, a systematic offset between float and shipboard oxygen measurements presents a problem for the current use of oxygen in float nitrate and pH parameter adjustment and derived parameter calculations. For all mechanisms other than a true change in ocean oxygen content, these offsets also complicate determination of long-term ocean oxygen changes. Given the mean difference in float and GLODAP dataset ages and that float oxygen measurements are now made in far greater numbers than shipboard data, a negative offset in the float data would appear in a combined data product as an increase in the true ocean oxygen loss signal. For the rest of this discussion, we focus on the impact of an oxygen offset on water properties derived from float measured data.

4.2 Impact of oxygen offsets on pH

As described earlier, float oxygen measurements at 1500 m are used to correct for drift in the pH sensor data which is then used with TA estimates to derive $p\mathrm{CO_2}$ (Williams et al., 2017). We only consider pH-equipped floats with air-calibrated oxygen here, with equivalent figures and tables for all pH-equipped floats provided in the supplement. The mean impact of correcting float pH for observed oxygen offsets is 3.2 ± 3.8 mpH (n=119, 95% CI of 2.5 to 3.9 mpH, Figure 5, Table 1). The impact of oxygen offsets on pH corrections can be understood through the relationship between oxygen and inorganic carbon. A negative oxygen offset means that float oxygen is lower than the expected value, so correcting the oxygen by a positive amount would result in a corresponding reduction in the apparent remineralization signature of the water mass. Higher oxygen and less remineralization would then imply lower dissolved inorganic carbon and therefore higher pH.

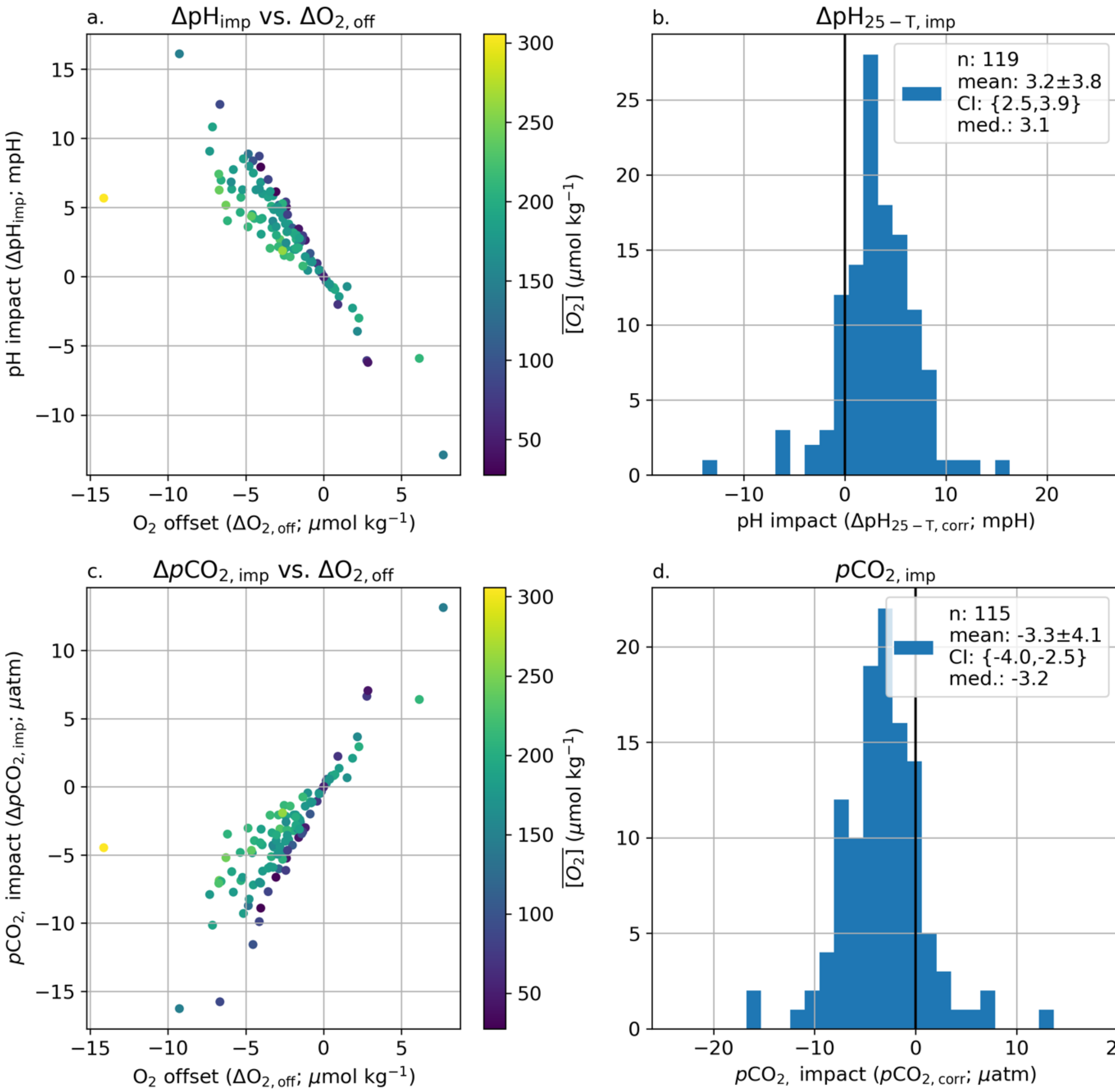

Figure 5. 1500 m pH and surface *p*CO_2 impact from associated oxygen offset. The calculated impact of correcting pH (top panels) and *p*CO_2 (bottom panels) for observed oxygen offsets. (a) and (c) are scatter plots of the change in pH and *p*CO_2 for a given oxygen offset, with the points colored by the mean oxygen concentration of the float crossover comparisons. Correcting a float oxygen sensor that is low of correct would increase the float pH and decrease derived *p*CO_2. (b) and (d) are histograms of the pH and *p*CO_2 impacts, with zero marked with a black line and mean ± 1SD, 95% confidence intervals around the mean, and median values listed in the figure legends. pH impacts shown are at 1500 m, while *p*CO_2 impacts are from surface values. Results shown here are calculated using air-calibrated floats and the ESPER-mixed pH algorithm, equivalent results for LIPHR in Supplemental Figure S4 and for all pH-equipped floats in Figures S5 and S6.

Table 1. Impact on pH at 1500 db and derived surface pCO_2 of observed oxygen offsets.

	Count	Mean	SD	P value[2]	95.0% CI {low, high}	median	min	max
pH impact ($\Delta pH_{25\text{-}T,imp}$, mpH)[1]	119	3.2	3.8	<0.001	{2.5, 3.9}	3.1	-12.9	16.1
pCO_2 impact ($\Delta pCO_{2,imp}$, µatm)	115	-3.3	4.1	<0.001	{-4.0, -2.5}	-3.2	-16.3	13.1

[1]Impacts for pH calculated using ESPER-mixed. Impacts shown here are for air-calibrated floats only. Equivalent numbers for LIPHR in Table S5.

[2]P-value testing the hypothesis that the mean impact is different from 0 at a 95% confidence level.

The sensitivity of pH impact ($\Delta pH_{25\text{-}T,imp}$) to oxygen offset (represented as $\Delta pH_{25\text{-}T,imp}/\Delta O_{2,off}$) is between -0.4 and -2.3 mpH/(µmol kg^{-1}) (Figures 5, 6). The largest $\Delta pH_{25\text{-}T,imp}$ values are observed in the Southern and Pacific Oceans, though the oxygen offsets do not show a similar spatial pattern. Instead, there is a spatial pattern to the $\Delta pH_{25\text{-}T,imp}/\Delta O_{2,off}$ that corresponds to the crossover oxygen concentration. The pH impact sensitivity (slope of the pH impact to O_2 offset) is greatest at low oxygen concentrations (Figure 5), with the mean crossover oxygen concentration explaining 61% of the variance in calculated $\Delta pH_{25\text{-}T,imp}/\Delta O_{2\ Off}$ (Figure 6).

This relationship can be understood by considering the changes in ocean chemistry as respiration takes place in a parcel of water. For example, we can take a parcel of water with initial TA and DIC of newly formed Subantarctic Mode Water from the southeast Pacific Ocean (Carter et al., 2014). If we then calculate the impact of organic matter respiration on oxygen, DIC, and TA, we can calculate the change in pH for every mole of oxygen respired. Initially, the $\Delta pH_{25\text{-}T,imp}/\Delta O_{2\ Off}$ is -2 mpH/(µmol kg^{-1}), but it drops to almost -2.8 mpH/(µmol kg^{-1}) after 250 µmol kg^{-1} of oxygen have been respired (Figure S10), as the buffer capacity of the water is eroded with increasing DIC; a situation analogous to surface ocean acidification. In the real ocean, regional differences in ocean interior biogeochemistry, including the buffer capacity, will cause regional differences in the sensitivity of pH to oxygen change. This likely accounts for larger spread in the relationship between pH impact and oxygen offset for points in the Southern Ocean (Figure 6e), where there is significant mixing of different water masses.

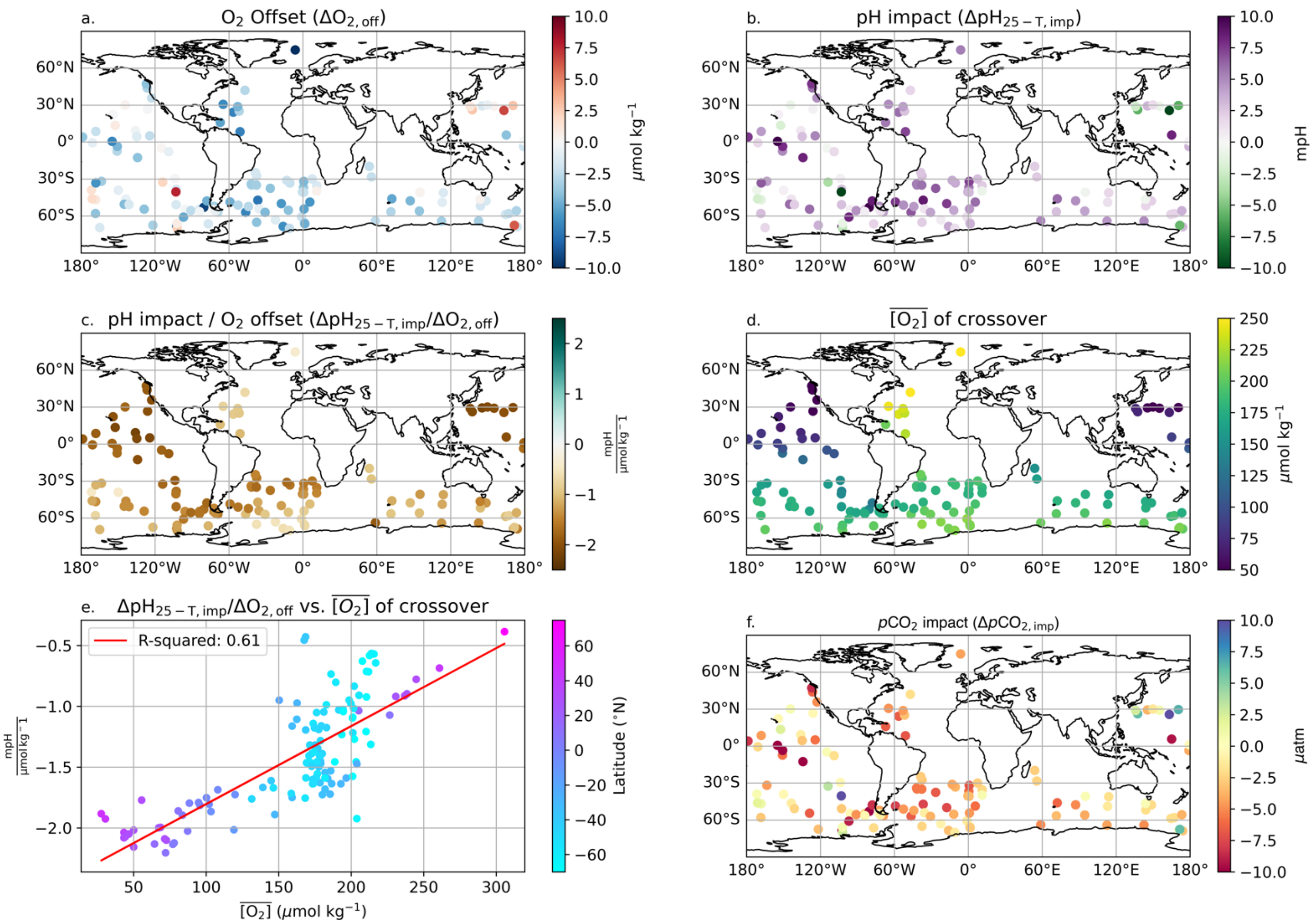

Figure 6. Relationships between oxygen offset, pH impact, mean oxygen concentration at crossover, and pCO_2 impact. Map of O_2 offset (a.) demonstrates no obvious spatial pattern in the magnitude of offsets while the impact of correcting for the O_2 offsets on pH (b.) tends to be greater in the Pacific and subpolar Southern Ocean. The ratio of the pH impact to O_2 offset ($\Delta pH_{25\text{-}T,imp}/\Delta O_{2,off}$, c.) is greatest in the north Pacific and subpolar Southern Ocean, reflecting the mean oxygen concentration at crossover ($\overline{[O_2]}$, d.). Plotting $\Delta pH_{25\text{-}T,imp}/O_{2,off}$ against $\overline{[O_2]}$ (e.) and calculating a linear fit (red line) indicates that variations in $\overline{[O_2]}$ explain ~ 61% of the variance in $\Delta pH_{imp}/\Delta O_{2,off}$. Much of the deviation from the linear fit occurs in the Southern Ocean (light blue points). This response leads to stronger pCO_2 impacts (f.) in the Pacific and subpolar Southern Ocean than in the North Atlantic or polar Southern Ocean. pH impacts shown are at 1500 m, while pCO_2 impacts are from surface values. Results shown here are calculated using floats with air-calibrated oxygen and the ESPER-mixed pH algorithm, equivalent results for LIPHR in Supplemental Figure S7 and for all pH-equipped floats in Figures S8 and S9.

4.3 Impact of oxygen offsets on derived pCO_2

The mean impact on float pCO_2 ($\Delta pCO_{2,imp}$) for correcting observed oxygen offsets is -3.3 ± 4.1 µatm (air-calibrated floats only, 95% CI of -4.0 to -2.5 µatm, Figure 5, Table 1). A reduction of float derived pCO_2 by this magnitude would account for a large portion of the apparent bias in float pCO_2 described in Wu and Qi (2023). For many of the studies with direct crossover comparisons, a mean adjustment of -3.3 µatm would effectively eliminate the observed differences between float and shipboard pCO_2. It should be noted that the range in impacts is much greater than the mean (a range of pCO_2 impact between -16.3 and 13.1 µatm), with the impact on individual floats and regions differing from the mean impact. Following pH, some of the greatest pCO_2 impacts are found in the Southern Ocean and Pacific (Figure 6). Similar results are found if LIPHR is used to calculate pH impacts instead of ESPER, though with a slightly greater mean pCO_2 impact magnitude (-3.6 ± 4.7, 95% CI {-4.4, -2.7} µatm, Table S5) and differences for individual floats.

4.4 Impacts of oxygen offset on nitrate and derived DIC

As described above, relative to GLODAP float nitrate is offset high (0.2 ± 0.4 µmol/kg), pH is offset low (-4.5 ± 7.7 mpH), and DIC is offset high (2.9 ± 6.4 µmol/kg, Table S3). These offsets are in approximately the correct ratios and directions for a biological signal, either real or due to an offset in the oxygen measurement that is then propagated to pH and DIC. This indicates that the oxygen offset is most likely not due to a problem with the Winkler titrations.

The calculated impact on nitrate of the oxygen offset ($\Delta NO_3^-{}_{imp}$) is small (mean -0.2 ± 0.2, 95% CI {-0.22, -0.14} µmol kg^{-1}, Figure S11, Table S6), in keeping with Redfield stoichiometry of -16 N: 154 O_2 (Hedges et al., 2002) multiplied by an oxygen offset of ~2 µmol kg^{-1}. The impact on DIC at 1500 m of correcting for oxygen offsets is -1 ± 1.2 µmol kg^{-1} (CI {-1.3, -0.8}, Figures S12, S13, Table S7), which is the slightly smaller than multiplying the oxygen offset by a Redfield ratio of -106 C: 154 O_2.

As a check that the mean impacts ($\Delta NO_3^-{}_{imp}$, $\Delta pH_{25\text{-}T,imp}$) calculated from the oxygen offset would, in fact, improve the crossover comparisons with GLODAP data, we also re-ran the crossover comparison after applying the $\Delta O_{2,off}$, $\Delta NO_3^-{}_{imp}$ and $\Delta pH_{25\text{-}T,imp}$ to the original data. The resulting crossovers ($\Delta NO_3^{-}{}_{,corr}$, $\Delta pH_{25\text{-}T,corr,}$ ΔDIC_{corr}, Figure 4, purple shaded histograms, statistics in Table S4) indicate that, on average, correcting the nitrate for the oxygen offset eliminates the entire nitrate offset, 3.2 mpH of the original -4.5 mpH pH offset, and 1.9 µmol kg^{-1} of the original 2.9 µmol kg^{-1} DIC offset.

The improvement in, but not full correction of, float-derived DIC and measured pH relative to the GLODAP crossovers provides an independent assessment that correcting the float data for this deep O_2 offset does, in fact, improve the pH data and subsequent carbonate system derived parameters. Furthermore, the fact that the nitrate bias appears to be fully corrected while the DIC, pH, and pCO_2 impacts only partially correct for the differences between float values and crossover or indirect comparisons gives an indication that the deep oxygen offset is not the only source of bias in the derived float carbonate parameters. This could be due to additional biases in the float pH, biases in the estimated TA, penetration of the ocean acidification signal to these depths, internal consistency issues with the marine carbonate system, or some other factor. It appears that correcting the oxygen may resolve a large fraction of the differences, but additional work remains to fully separate pCO_2, pH, and DIC biases in the float data from true trends.

Float oxygen crossovers with shipboard data are not available for all floats, making it difficult to determine the magnitude of an oxygen offset for all pH-equipped floats at this time. One option to reduce the possibility of an oxygen-induced bias in pH and derived pCO_2 is to correct float pH using an algorithm that does not use oxygen. Removing oxygen from the ESPER algorithm yields a 4.2 ± 6.0 mpH and -4.3 ± 6.4 µatm pCO_2 impact, somewhat larger than the 3.2 ± 3.8 mpH and -3.3 ± 4.1 µatm impact we calculate from applying the observed oxygen offset relative to GLODAP (Figure S14). Plotting the ESPER pH and pCO_2 impacts due to removing oxygen or correcting for the observed offset against one another yields a cluster of points around the 1:1 lines, though with considerable spread (Figure S15). This approach may indeed improve derived parameters for floats with no GLODAP oxygen crossover for comparison, but is less accurate than if an oxygen correction is possible.

5. Conclusions

Here we identify an offset between float and ship-board oxygen measurements between 1450 and 2000 db. The magnitude of this offset is significant for studies assessing long-term ocean oxygen changes and for the use of float oxygen in adjusting pH and nitrate measurements and in subsequent calculated derived carbonate system quantities.

Correcting oxygen offsets of -2.5 µmol kg^{-1} in pH-equipped, air-calibrated floats results in mean pH changes of 3.2 mpH and pCO_2 changes of -3.3 µatm. These differences are of similar direction and magnitude as many of the direct in-situ comparisons between float pCO_2 estimates and underway pCO_2 measurements and therefore represent a first-order bias that needs to be addressed in the biogeochemical float dataset. We do not definitively identify whether the oxygen offset is present in the float or ship-board dataset, though the fact that correcting float oxygen would improve the nitrate, DIC, and pH crossover comparisons to shipboard data are a strong indication that the issue lies with the float observations.

Based on our findings we offer the following recommendations:

1. The oxygen offsets described here represent an empirical correction that must be investigated to determine the underlying mechanism.
2. Float oxygen data at depth should be adjusted to historical shipboard data until our mechanistic understanding of the causes behind the float – ship differences is sufficient such that air calibration or other adjustments do not require a secondary correction.
3. Care should be used in combining float and ship data, with an understanding that small biases may be present in either dataset that could impact the usability of the data compilation to answer some scientific questions. The difference between the average age of the float and GLODAP datasets and the shift to float profile numbers greatly exceeding shipboard profiles in the 2000's mean that any offset between the datasets will appear as a spurious change in ocean oxygen content. Additionally, individual floats may have significantly greater magnitude biases than the mean and data should be assessed prior to any use relying on absolute accuracy.
4. Standardization of float sensor calibration comments and equations will make future studies of overall biogeochemical Argo float performance easier to perform.

Data availability

The "argo_synthetic-profile_index.txt" file used to determine biogeochemical Argo float WMO numbers and data locations was downloaded from ftp.ifremer.fr/ifremer/argo/dac. GLODAP data is available at https://glodap.info/index.php/merged-and-adjusted-data-product-v2-2022/. The analysis and plotting code used for this manuscript are available at: 10.5281/zenodo.10866941 (Bushinsky et al., 2024).

Acknowledgements

SMB was supported by the NOAA Climate Program Office's Climate Observations and Monitoring, Climate Variability and Predictability, and Global Ocean Monitoring and Observation programs (NA21OAR4310260). A.J. Fassbender was supported by NOAA PMEL. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (https://argo.ucsd.edu, https://www.ocean-ops.org). The Argo Program is part of the Global Ocean Observing System (Argo, 2000). Two projects that contributed a significant amount of data used in this study are: The Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project funded by the National Science Foundation, Division of Polar Programs (NSF PLR -1425989 and OPP-1936222), supplemented by NASA, and the Global Ocean Biogeochemistry Array (GO-BGC) Project funded by the National Science Foundation, Division of Ocean Sciences (NSF OCE-1946578). The international effort of GLODAP is foundational to the work described in this manuscript and an essential component of our ocean observing capabilities. This is SOEST contribution number XXXX and PMEL contribution number 5617. We would like to thank an anonymous PMEL technical reviewer for their helpful suggestions.

References

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

Argo. (2023). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) - Snapshot of Argo GDAC of June 09st 2023 [Data set]. SEANOE. https://doi.org/10.17882/42182#103075

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al. (2016). A multi-decade record of high-quality fCO2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/essd-8-383-2016

Bittig, H. C., & Körtzinger, A. (2015). Tackling oxygen optode drift: Near-surface and in-air oxygen optode measurements on a float provide an accurate in-situ reference. *Journal of Atmospheric and Oceanic Technology*, *32*(8), 1536–1543. https://doi.org/10.1175/JTECH-D-14-00162.1

Bittig, H. C., & Körtzinger, A. (2017). Technical note: Update on response times, in-air measurements, and in situ drift for oxygen optodes on profiling platforms. *Ocean Science*, *13*(1), 1–11. https://doi.org/10.5194/os-13-1-2017

Bittig, H. C., Fiedler, B., Scholz, R., Krahmann, G., & Körtzinger, A. (2014). Time response of oxygen optodes on profiling platforms and its dependence on flow speed and temperature. *Limnology and Oceanography: Methods*, *12*, 617–636. https://doi.org/10.4319/lom.2014.12.617

Bittig, H. C., Fiedler, B., Fietzek, P., & Körtzinger, A. (2015). Pressure response of Aanderaa and Sea-Bird oxygen optodes. *Journal of Atmospheric and Oceanic Technology*.

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., et al. (2018). An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian Neural Networks. *Frontiers in Marine Science*, *5*(September), 328. https://doi.org/10.3389/fmars.2018.00328

Bittig, H. C., Körtzinger, A., Neill, C., van Ooijen, E., Plant, J. N., Hahn, J., et al. (2018). Oxygen Optode Sensors: Principle, Characterization, Calibration, and Application in the Ocean. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00429

Bushinsky, S. M., & Cerovečki, I. (2023). Subantarctic Mode Water Biogeochemical Formation Properties and Interannual Variability. *AGU Advances*, *4*(2), e2022AV000722. https://doi.org/10.1029/2022AV000722

Bushinsky, S. M., & Emerson, S. (2015). Marine biological production from in situ oxygen measurements on a profiling float in the subarctic Pacific Ocean. *Global Biogeochemical Cycles*, *29*(12), 2050–2060. https://doi.org/10.1002/2015GB005251

Bushinsky, S. M., Emerson, S. R., Riser, S. C., & Swift, D. D. (2016). Accurate oxygen measurements on modified Argo floats using in situ air calibrations. *Limnology and Oceanography: Methods*, *14*(8), 491–505. https://doi.org/10.1002/lom3.10107

Bushinsky, S. M., Gray, A. R., Johnson, K. S., & Sarmiento, J. L. (2017). Oxygen in the Southern Ocean From Argo Floats: Determination of Processes Driving Air-Sea Fluxes. *Journal of Geophysical Research: Oceans*, *122*(11), 8661–8682. https://doi.org/10.1002/2017JC012923

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., et al. (2019). Reassessing Southern Ocean Air-Sea CO 2 Flux Estimates With the Addition of Biogeochemical Float Observations. *Global Biogeochemical Cycles*, *33*(11), 1370–1388. https://doi.org/10.1029/2019GB006176

Bushinsky, S. M., Tamsitt, V., Nachod, Z., Fassbender, A., & Williams, N. (2024). SethBushinsky/argo_deep_o2_bias: Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO2 analysis and plotting code (Version v0.1.1-submission). Retrieved from https://doi.org/10.5281/zenodo.10866942

Carpenter, J. H. (1965). The Chesapeake Bay Institute Technique for the Winkler Dissolved Oxygen Method. *Limnology and Oceanography*, *10*(1), 141–143. https://doi.org/10.4319/lo.1965.10.1.0141

Carter, B. R., Radich, J. A., Doyle, H. L., & Dickson, A. G. (2013). An automated system for spectrophotometric seawater pH measurements. *Limnology and Oceanography: Methods*, *11*(1), 16–27. https://doi.org/10.4319/lom.2013.11.16

Carter, B. R., Talley, L. D., & Dickson, A. G. (2014). Mixing and remineralization in waters detrained from the surface into Subantarctic Mode Water and Antarctic Intermediate Water in the southeastern Pacific. *Journal of Geophysical Research: Oceans*, *119*(6), 4001–4028. https://doi.org/10.1002/2013JC009355

Carter, B. R., Williams, N. L., Gray, A. R., & Feely, R. A. (2016). Locally interpolated alkalinity regression for global alkalinity estimation. *Limnology and Oceanography: Methods*, (1998). https://doi.org/10.1002/lom3.10087

Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B., & Takeshita, Y. (2018). Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate. *Limnology and Oceanography: Methods*, *16*(2), 119–131. https://doi.org/10.1002/lom3.10232

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y.-Y., et al. (2021). New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, *19*(12), 785–809. https://doi.org/10.1002/lom3.10461

Carter, B. R., Sharp, J. D., Dickson, A. G., Álvarez, M., Fong, M. B., García-Ibáñez, M. I., et al. (2023). Uncertainty sources for measurable ocean carbonate chemistry variables. *Limnology and Oceanography*, lno.12477. https://doi.org/10.1002/lno.12477

Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the Global Ocean with Biogeochemical-Argo. *Annual Review of Marine Science*, *12*(1), 1–26. https://doi.org/10.1146/annurev-marine-010419-010956

Coggins, A., Watson, A. J., Schuster, U., Mackay, N., King, B., McDonagh, E., & Poulton, A. J. (2023). Surface ocean carbon budget in the 2017 south Georgia diatom bloom: Observations and validation of profiling biogeochemical argo floats. *Deep Sea Research Part II: Topical Studies in Oceanography*, *209*, 105275. https://doi.org/10.1016/j.dsr2.2023.105275

D'Asaro, E. A., & McNeil, C. (2013). Calibration and stability of oxygen sensors on autonomous floats. *Journal of Atmospheric and Oceanic Technology*, *30*(8), 1896–1906. https://doi.org/10.1175/JTECH-D-12-00222.1

Dickson, A. G. (1994). Determination of dissolved oxygen in sea water by Winkler titration. *WHP Operations and Methods*, (November), 1–13.

Drucker, R., & Riser, S. C. (2016). In situ phase-domain calibration of oxygen Optodes on profiling floats. *Methods in Oceanography*, *17*, 296–318. https://doi.org/10.1016/j.mio.2016.09.007

Emerson, S. (1987). Seasonal oxygen cycles and biological new production in surface waters of the subarctic Pacific Ocean. *Journal of Geophysical Research*, *92*(C6), 6535–6544. https://doi.org/10.1029/JC092iC06p06535

Fay, A. R., Lovenduski, N. S., Mckinley, G. A., Munro, D. R., Gray, A. R., Landschützer, P., et al. (2018). Utilizing the Drake Passage Time-series to understand variability and change in subpolar Southern Ocean pCO 2. *Biogeosciences Discussions*. https://doi.org/10.5194/bg-2017-489

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., & Johnson, D. R. (2010). *World Ocean Atlas 2009 Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. Noaa Atlas Nesdic* (Vol. 70, p. 344pp). Washington, D. C.: U.S. Government Printing Office. Retrieved from http://www.nodc.noaa.gov/OC5/indprod.html

Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., et al. (2018). Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean. *Geophysical Research Letters*, *45*(17), 9049–9057. https://doi.org/10.1029/2018GL078013

Gruber, N., Scott C. Doney, Steven R. Emerson3, Denis Gilbert4, Taiyo Kobayashi5, Arne Körtzinger6, Gregory C. Johnson7, Kenneth S., & Johnson8, Stephen C. Riser3, and Osvaldo Ulloa. (2007). THE ARGO-OXYGEN PROGRAM: A white paper to promote the addition of oxygen sensors to the international Argo float program.

Gruber, N., Doney, S. C., Emerson, S. R., Gilbert, D., Kobayashi, T., Körtzinger, A., et al. (2009). Adding oxygen to Argo: Developing a global in-situ observatory for ocean deoxygenation and biogeochemistry. *Ocean Obs '09*.

Hedges, J. I., Baldock, J. a., Gélinas, Y., Lee, C., Peterson, M. L., & Wakeham, S. G. (2002). The biochemical and elemental compositions of marine plankton: A NMR perspective. *Marine Chemistry*, *78*(1), 47–63. https://doi.org/10.1016/S0304-4203(02)00009-9

Helm, K. P., Bindoff, N. L., & Church, J. A. (2011). Observed decreases in oxygen content of the global ocean. *Geophysical Research Letters*, *38*(23), 1–6. https://doi.org/10.1029/2011GL049513

Hennon, T. D., Riser, S. C., & Mecking, S. (2016). Profiling float-based observations of net respiration beneath the mixed layer. *Global Biogeochemical Cycles*, *30*(6), 920–932. https://doi.org/10.1002/2016GB005380

Humphreys, M. P., Lewis, E. R., Sharp, J. D., & Pierrot, D. (2022). PyCO2SYS v1.8: marine carbonate system calculations in Python. *Geoscientific Model Development*, *15*(1), 15–43. https://doi.org/10.5194/gmd-15-15-2022

Humphreys, M. P., Schiller, A. J., Sandborn, D., Gregor, L., Pierrot, D., van Heuven, S. M. A. C., et al. (2023, January 19). PyCO2SYS: marine carbonate system calculations in Python (Version v1.8.2). Zenodo. https://doi.org/10.5281/ZENODO.3744275

Johnson, K. S., Plant, J. N., Riser, S. C., & Gilbert, D. (2015). Air oxygen calibration of oxygen optodes on a profiling float array. *Journal of Atmospheric and Oceanic Technology*, *32*(11), 2160–2172. https://doi.org/10.1175/JTECH-D-15-0101.1

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., et al. (2017). Biogeochemical sensor performance in the SOCCOM profiling float array. *Journal of Geophysical Research: Oceans*, *122*(8), 6416–6436. https://doi.org/10.1002/2017JC012838

Johnson, K. S., Maurer, T. L., Plant, J. N., & Takeshita, Y. (2023). *BGC-Argo quality control manual for pH* [Pdf]. [object Object]. https://doi.org/10.13155/97828

Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2015). Global Ocean Data Analysis Project, Version 2 (GLODAPv2). *ORNL/CDIAC-162, NDP-093*. https://doi.org/10.3334/CDIAC/OTG.NDP093_GLODAPv2

Körtzinger, A., Schimanski, J., & Send, U. (2005). High Quality Oxygen Measurements from Profiling Floats: A Promising New Technique. *Journal of Atmospheric and Oceanic Technology*, *22*(3), 302–308. https://doi.org/10.1175/JTECH1701.1

Levin, L. A. (2018). Manifestation, drivers, and emergence of open ocean deoxygenation. *Annual Review of Marine Science*, *10*, 229–260. https://doi.org/10.1146/annurev-marine-121916-063359

Liu, X., Patsavas, M. C., & Byrne, R. H. (2011). Purification and Characterization of meta-Cresol Purple for Spectrophotometric Seawater pH Measurements. *Environmental Science & Technology*, *45*(11), 4862–4868. https://doi.org/10.1021/es200665d

Long, M. C., Stephens, B. B., McKain, K., Sweeney, C., Keeling, R. F., Kort, E. A., et al. (2021). Strong Southern Ocean carbon uptake evident in airborne observations. *Science*, *374*(6572), 1275–1280. https://doi.org/10.1126/science.abi4355

Mackay, N., & Watson, A. (2021). Winter air-sea CO 2 fluxes constructed from summer observations of the Polar Southern Ocean suggest weak outgassing. *Journal of Geophysical Research: Oceans*. https://doi.org/10.1029/2020jc016600

Martz, T. R., Riser, S. C., & Johnson, K. S. (2008). Ocean metabolism observed with oxygen sensors on profiling floats in the South Pacific. *Limnology and Oceanography*, *53*(5, part 2), 2094–2111. https://doi.org/10.4319/lo.2008.53.5_part_2.2094

Maurer, T. L., Plant, J. N., & Johnson, K. S. (2021). Delayed-Mode Quality Control of Oxygen, Nitrate, and pH Data on SOCCOM Biogeochemical Profiling Floats. *Frontiers in Marine Science*, *8*(August), 1–20. https://doi.org/10.3389/fmars.2021.683207

McDougall, T. J., & Barker, P. M. (2011). Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox. SCOR/IAPSO WG127. Retrieved from https://www.teos-10.org/pubs/Getting_Started.pdf

Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2016). The global ocean data analysis project version 2 (GLODAPv2) - An internally consistent data product for the world ocean. *Earth System Science Data*, *8*(2), 297–323. https://doi.org/10.5194/essd-8-297-2016

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Bittig, H. C., Kozyr, A., et al. (2020). GLODAPv2.2020 - the second update of GLODAPv2. *Earth System Science Data Discussions*. https://doi.org/10.5194/essd-2020-165

Roemmich, D., Talley, L., Zilberman, N., Osborne, E., Johnson, K., Barbero, L., et al. (2021). The Technological, Scientific, and Sociological Revolution of Global Subsurface Ocean Observing. *Oceanography*, 2–8. https://doi.org/10.5670/oceanog.2021.supplement.02-02

Sarmiento, J. L., Johnson, K. S., Arteaga, L. A., Bushinsky, S. M., Cullen, H. M., Gray, A. R., et al. (2023). The Southern Ocean carbon and climate observations and modeling (SOCCOM) project: A review. *Progress in Oceanography*, *219*, 103130. https://doi.org/10.1016/j.pocean.2023.103130

Schmidtko, S., Stramma, L., & Visbeck, M. (2017). Decline in global oceanic oxygen content during the past five decades. *Nature*, *542*(7641), 335–339. https://doi.org/10.1038/nature21399

Schofield, O. A., Fassbender, A., Hood, M., Hill, K., & Johnson, K. (2022). A global ocean biogeochemical observatory becomes a reality. *Eos*, *103*. https://doi.org/10.1029/2022EO220149

Sharp, J. D., Fassbender, A. J., Carter, B. R., Johnson, G. C., Schultz, C., & Dunne, J. P. (2022). *GOBAI-O$_2$: temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly two decades* (preprint). ESSD – Ocean/Chemical oceanography. https://doi.org/10.5194/essd-2022-308

Stramma, L., & Schmidtko, S. (2021). Spatial and Temporal Variability of Oceanic Oxygen Changes and Underlying Trends. *Atmosphere-Ocean*, *59*(2), 122–132. https://doi.org/10.1080/07055900.2021.1905601

Takeshita, Y., Martz, T. R., Johnson, K. S., Plant, J. N., Gilbert, D., Riser, S. C., et al. (2013). A climatology-based quality control procedure for profiling float oxygen data. *Journal of Geophysical Research: Oceans*, *118*(10), 5640–5650. https://doi.org/10.1002/jgrc.20399

Takeshita, Y., Johnson, K. S., Coletti, L. J., Jannasch, H. W., Walz, P. M., & Warren, J. K. (2020). Assessment of pH dependent errors in spectrophotometric pH measurements of seawater. *Marine Chemistry*, *223*, 103801. https://doi.org/10.1016/j.marchem.2020.103801

Tengberg, A., Hovdenes, J., Andersson, H. J., Brocandel, O., Diaz, R., Hebert, D., et al. (2006). Evaluation of a lifetime-based optode to measure oxygen in aquatic systems. *Limnology and Oceanography: Methods*, *4*(1964), 7–17. https://doi.org/10.4319/lom.2006.4.7

Uchida, H., Kawano, T., Kaneko, I., & Fukasawa, M. (2008). In Situ Calibration of Optode-Based Oxygen Sensors. *Journal of Atmospheric and Oceanic Technology*, *25*(12), 2271–2281. https://doi.org/10.1175/2008JTECHO549.1

Williams, N. L., Juranek, L. W., Johnson, K. S., Feely, R. A., Riser, S. C., Talley, L. D., et al. (2016). Empirical algorithms to estimate water column pH in the Southern Ocean. *Geophysical Research Letters*, *43*, 3415–3422. https://doi.org/10.1002/2016GL068539

Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., et al. (2017). Calculating surface ocean pCO2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles*, *31*(3), 591–604. https://doi.org/10.1002/2016GB005541

Wolf, M. K., Hamme, R. C., Gilbert, D., Yashayaev, I., & Thierry, V. (2018). Oxygen Saturation Surrounding Deep Water Formation Events in the Labrador Sea From Argo-O 2 Data. *Global Biogeochemical Cycles*, *32*(4), 635–653. https://doi.org/10.1002/2017GB005829

Wu, Y., & Qi, D. (2023). The controversial Southern Ocean air-sea CO2 flux in the era of autonomous ocean observations. *Science Bulletin*, S2095927323006151. https://doi.org/10.1016/j.scib.2023.08.059

Wu, Y., Bakker, D. C. E., Achterberg, E. P., Silva, A. N., Pickup, D. D., Li, X., et al. (2022). Integrated analysis of carbon dioxide and oxygen concentrations as a quality control of ocean float data. *Communications Earth & Environment*, *3*(1), 92. https://doi.org/10.1038/s43247-022-00421-w

Yang, B., Emerson, S. R., & Bushinsky, S. M. (2017). Annual net community production in the subtropical Pacific Ocean from in situ oxygen measurements on profiling floats. *Global Biogeochemical Cycles*, *31*(4), 728–744. https://doi.org/10.1002/2016GB005545

Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO_2

Seth M. Bushinsky[1], Zachary Nachod[1], Andrea J. Fassbender[2], Veronica Tamsitt[3], Yuichiro Takeshita[4], Nancy Williams

[1]Department of Oceanography, School of Earth and Space Science and Technology, University of Hawai‘i at Mānoa, Honolulu, HI, USA
[2]NOAA/OAR Pacific Marine Environmental Laboratory, Seattle, WA, USA
[3]University of South Florida, St. Petersburg, FL, USA
[4] Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

Abstract

Profiles of oxygen measurements from Argo profiling floats now vastly outnumber shipboard profiles. Air calibration of a float's oxygen optode upon surfacing enables accurate measurements in the upper ocean but does not necessarily provide similar accuracy at depth. In this study we use a quality controlled shipboard dataset to show that, on average, the entire Argo oxygen dataset is offset relative to shipboard measurements (float minus ship) at pressures of 1450 to 2000 db by -1.9 ± 4.7 µmol kg^{-1} (95% confidence interval around the mean of {-2.3, -1.5}) and air calibrated floats are offset by -3.1 ± 5.3 µmol kg^{-1} (95% CI: {-3.7, -2.4}). The difference between float and shipboard oxygen is most likely due to offsets in the float oxygen data and not due to oxygen changes at these depths or biases in the shipboard dataset. In addition to posing problems for the calculation of long-term ocean oxygen changes, these float oxygen offsets impact the adjustment of float nitrate and pH measurements and therefore bias important derived quantities such as the partial pressure of CO_2 (pCO_2) and dissolved inorganic carbon. Correcting floats with air-calibrated oxygen for the float-ship oxygen offsets changes float pH by 3.2 ± 3.8 mpH and float-derived surface pCO_2 by -3.3 ± 4.1 µatm. This adjustment to float pCO_2 represents half, or more, of the bias in float-derived pCO_2 reported in studies comparing float pCO_2 to shipboard pCO_2 measurements.

Plain Language Summary

Oxygen has historically been measured using chemical titrations on water collected by ships at sea. Over the past 20 years, sensors that measure oxygen have been deployed on robotic profiling floats. Measurements by oxygen sensors on profiling floats now greatly exceed those collected by ships. Here we compare all float oxygen data collected to shipboard measurements in deep waters (1450 to 2000 m depth) where we do not expect oxygen to be changing in the ocean. We find a difference between float and shipboard data. If left uncorrected, this difference would give the false impression of a long-term oxygen change. This difference also impacts float-measured pH and float-estimated carbon dioxide, both of which rely on float oxygen measurements. Correcting oxygen, and therefore float pH and carbon dioxide, would largely address a widely-studied bias in float measurements of these parameters.

Key Points

1. Air-calibrated float oxygen measurements are lower than shipboard data by 3.1 µmol kg^{-1} at pressures of 1450 to 2000 db.
2. Correcting float oxygen for this offset would increase float pH by 3.2 mpH and lower float-derived $p\mathrm{CO_2}$ by -3.3 µatm.
3. This float oxygen offset would improve float and ship $p\mathrm{CO_2}$ comparisons, removing most or all of observed biases.

1. Introduction

Measurements of oxygen in the ocean are critical, as the ocean is losing oxygen due to warming, respiration, and stratification-induced changes to ventilation (Helm et al., 2011; Levin, 2018; Schmidtko et al., 2017; Stramma & Schmidtko, 2021). Measurements of ocean oxygen also can be used to understand the balance between photosynthesis and respiration in the surface and deep ocean (Bushinsky & Emerson, 2015; Emerson, 1987; Hennon et al., 2016; Yang et al., 2017). Historically, oxygen was primarily measured by Winkler titrations of water sampled from the ocean (Carpenter, 1965; Dickson, 1994). These titrations require reliable standards and trained operators to acheive high accuracy and precision. Winkler titrations have been supplemented by electrochemical and optical sensors attached to CTDs but have remained the gold standard of accurate oxygen measurement in the ocean.

Oxygen measurements on autonomous profiling floats are transforming our ability to observe the ocean at unprecedented levels of detail. The first Argo oxygen data was from floats deployed in the early 2000s equipped with Clark-cell electrodes (Figure 1) that provided fast response times but could drift rapidly and in unpredictable ways (Gruber et al., 2007). Oxygen optodes that measure the partial pressure of oxygen through an oxygen-sensitive luminophore have now become the standard oxygen sensor deployed on floats (Claustre et al., 2020; Gruber et al., 2009; Körtzinger et al., 2005; Tengberg et al., 2006) (Figure 1). Over 1800 Argo oxygen floats have been deployed as of 2023, with increasing numbers due to a mix of many small deployments by individual research groups and the development of a few projects deploying large numbers of floats (e.g., the Southern Ocean Carbon and Climate Observations and Modeling project, SOCCOM, (Johnson et al., 2017; Sarmiento et al., 2023); Global Ocean Biogeochemistry Array, GO-BGC (Roemmich et al., 2021; Schofield, O. A. et al., 2022)). Optodes drift significantly prior to deployment, but the advent of calibration using atmospheric oxygen has increased accuracy of surface data to better than 1% (Bittig, Körtzinger, et al., 2018; Bushinsky et al., 2016; D'Asaro & McNeil, 2013; Johnson et al., 2015). Floats calibrated using atmospheric data primarily use a single multiplicative gain correction based on the difference between float air measurements and calculated atmospheric oxygen that is then applied to the whole float profile (Figure 2). This approach assumes that the response of the oxygen optode at different oxygen levels and temperatures changes uniformly over time, and that the zero reading of the optode remains unchanged.

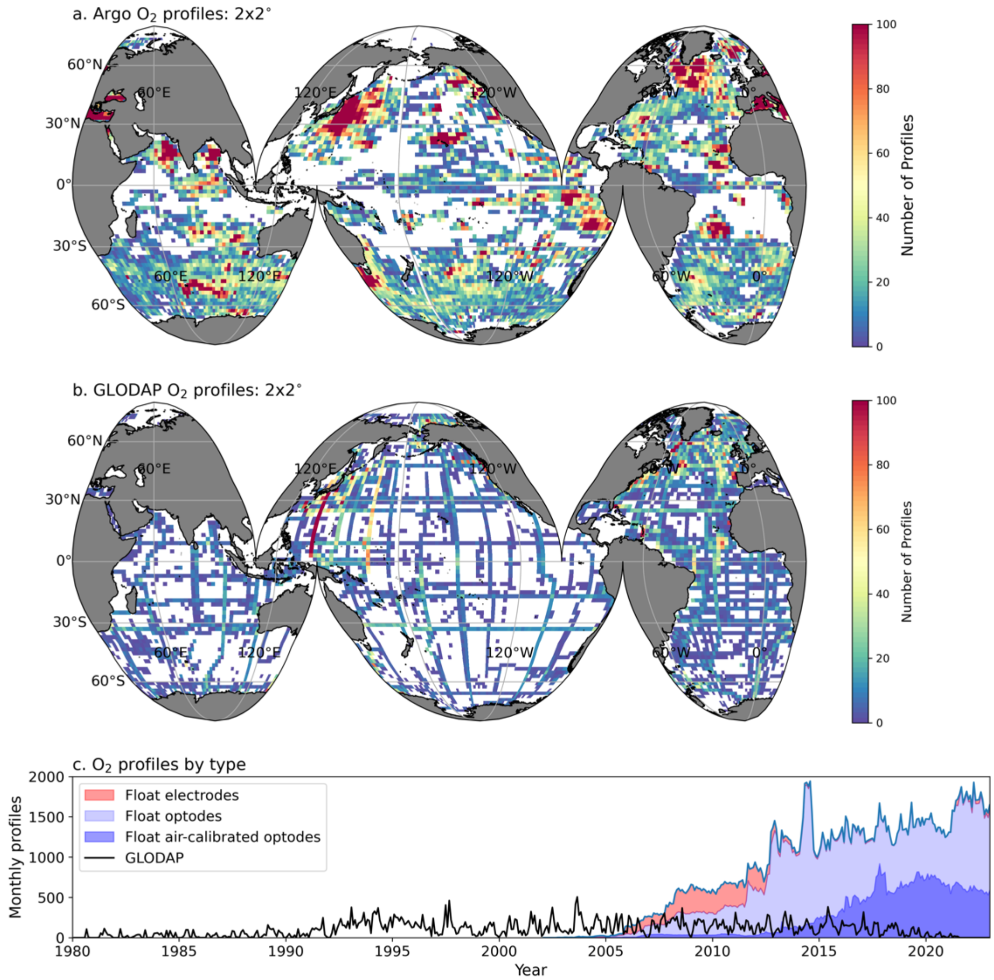

Figure 1. Observational density of float and ship oxygen and change in float sensor type over time. Float (a.) and GLODAP (b.) oxygen profile density on a 2 × 2° grid. (c.) Total number of oxygen profiles per month, with the contribution of float sensor type shown (Clark cell electrodes (red) and optodes (blue), and the subset of optodes calibrated with in-air measurements (dark blue)), and total number of GLODAP profiles that contain oxygen per month (black line).

The Global Ocean Data Analysis Project (GLODAP; Key et al., 2015; Olsen et al., 2016, 2020) is an on-going international synthesis effort that evaluates surface-to-bottom ocean biogeochemical bottle data for outliers and internal consistency, which has been crucial to improving the accuracy and usability of shipboard data. A similar effort, the Surface Ocean CO_2 Atlas (SOCAT) collects and assesses surface pCO_2 measurements for accuracy prior to inclusion in an annual data product (Bakker et al., 2016). These efforts have been instrumental in our understanding of ocean biogeochemistry. There is currently no comparable, on-going, post-

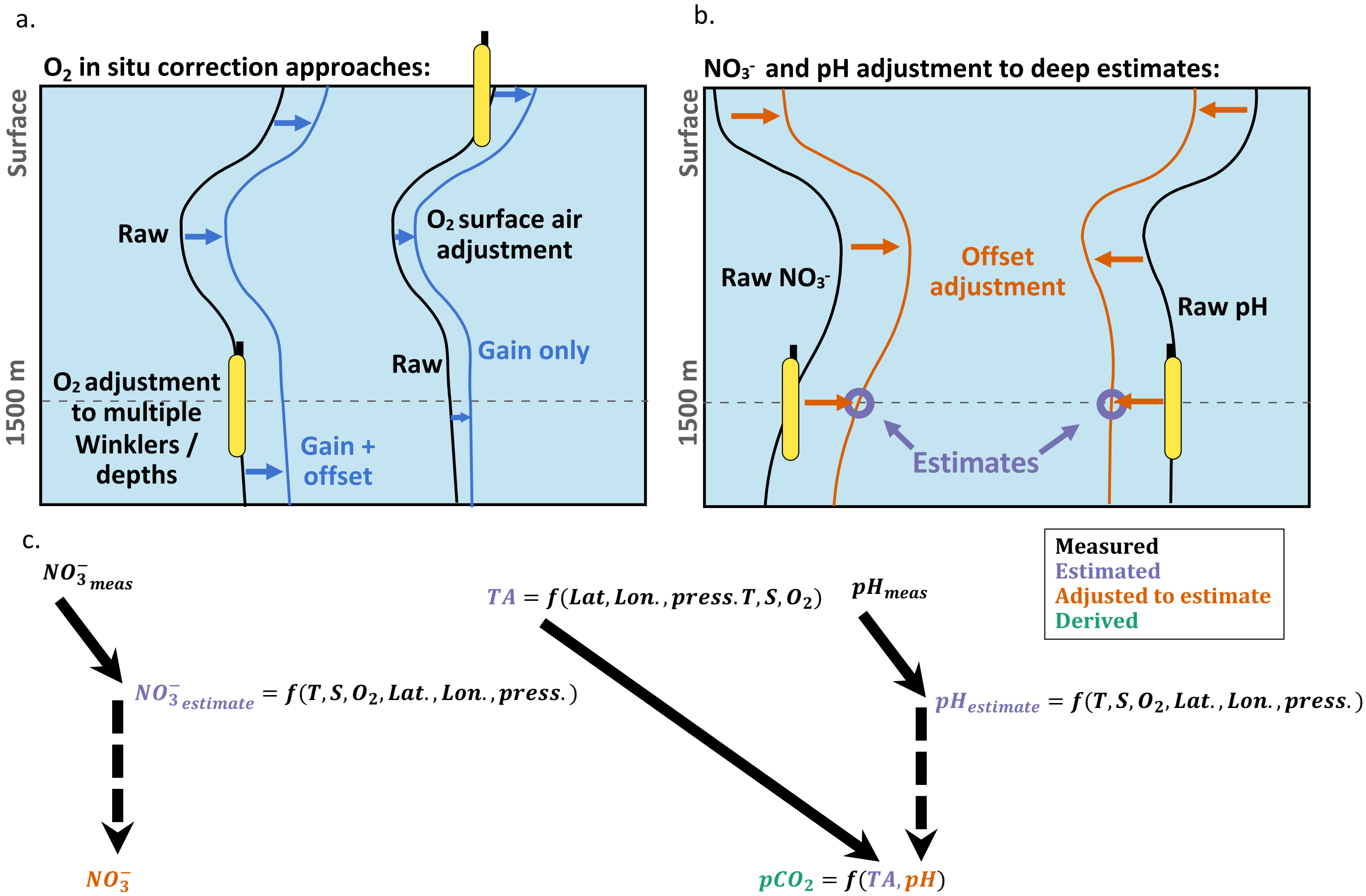

Figure 2. Schematic of the two primary in-situ oxygen calibration approaches and how the calibrated oxygen data are used to adjust float pH and nitrate measurements. (a.) Oxygen data collected on floats are typically calibrated using a gain and offset correction based on in-situ data, typically collected at float deployment, or a gain-only correction based on air-calibration measurements made upon float surfacing throughout its lifetime. (b.) Float nitrate and pH measurements are adjusted to algorithmic estimates of those parameters at 1500 db in a way that impacts the whole profile. (c.) Variable inputs, including oxygen, to the nitrate and pH algorithms, which are trained on shipboard data. Offsets between the float and ship oxygen will propagate through these algorithms into the adjusted nitrate and pH data.

deployment census of float biogeochemical data to assess consistency within the float dataset or between float and shipboard data. Most studies that have used Argo oxygen data and assessed their accuracy have focused on surface data or analyzed changes measured by individual floats, thereby avoiding comparisons between different float measurements (Bushinsky & Emerson, 2015; Johnson et al., 2015; Martz et al., 2008; Wolf et al., 2018). As the float array expands and researchers begin using the entire dataset as a whole, or combining float and shipboard datasets to create global interior ocean oxygen products (e.g., World Ocean Atlas (Garcia et al., 2010); Gridded Ocean Biogeochemistry from Artificial Intelligence – Oxygen (Sharp et al., 2022)), attention must be paid to the accuracy of float oxygen data at depth.

Largescale comparisons of float and ship datasets indicate that, to a first order, float oxygen data have reached a comparable level of accuracy as ship-board measurements (Bushinsky et al., 2017; Johnson et al., 2017; Maurer et al., 2021). However, two float-to-ship comparisons of deep (1500-2000 m) oxygen data (Bushinsky et al., 2016; Drucker & Riser, 2016) indicate that surface calibration may not be a sufficient adjustment for the entire float oxygen profile. There is some indication that this could be due to inadequate calibration of the optode temperature response (Bittig, Körtzinger, et al., 2018), but Bushinsky et al. (2016) re-calibrated the optode temperature response for 11 floats in the lab prior to deployment, which did not address the difference in deep data, so this approach does not seem to fully resolve the observed drift.

The accuracy of oxygen data throughout the water column is critical for understanding long-term changes in ocean oxygen content. At depth, the accuracy of quality-controlled float oxygen data is especially critical due to its use in adjusting other biogeochemical sensor data from Argo floats, which has downstream impacts on derived carbonate system parameters. Nitrate and pH sensors are now being deployed in large numbers (Claustre et al., 2020) and also require in-situ adjustment to correct sensor drift (Maurer et al., 2021). The nitrate and pH measurements are adjusted using estimates of these parameters at 1500 db that are derived from multiple linear regression or neural network algorithms that were trained on the GLODAP shipboard database (Bittig, Steinhoff, et al., 2018; Carter et al., 2016, 2021; Williams et al., 2016). These algorithmic estimates use inputs of temperature, salinity, depth, latitude, longitude, and, importantly, oxygen. Due to sensor response characteristics, adjustments to the nitrate and pH sensor data are applied almost uniformly to the entire profile, so offsets at 1500 m directly impact surface measurements (Maurer et al., 2021).

Float measurements of pH are widely used to estimate $p\mathrm{CO_2}$, and other carbonate system parameters, using an algorithm-based estimate of total alkalinity (TA) and a carbonate system calculator. Many studies apply an additional adjustment to the quality controlled pH data prior to estimating $p\mathrm{CO_2}$ that is meant to correct for an empirical pH-dependent pH bias (Carter et al., 2013, 2018; Williams et al., 2017). The accuracy of float $p\mathrm{CO_2}$ estimates is critical, as studies relying on float $p\mathrm{CO_2}$ have identified significant differences between seasonal cycles of $p\mathrm{CO_2}$, and wintertime air-sea CO_2 fluxes from studies that rely on shipboard pCO_2 alone (e.g. Bushinsky et al., 2019; Gray et al., 2018; Williams et al., 2016). Floats measure year-round, including during winter months when rough weather makes shipboard observations rare, which provides immense observational value, if accurate. A number of studies have directly compared float pCO_2 estimates to shipboard observations (Bushinsky & Cerovečki, 2023; Coggins et al., 2023; Fay et al., 2018; Gray et al., 2018; Williams et al., 2017), while other studies have indirectly compared float pCO_2 and pH through assessment of other carbonate system parameters or CO_2 fluxes (e.g., Long et al., 2021; Mackay & Watson, 2021; Wu et al., 2022). A recent meta-analysis of float pCO_2 accuracy found biases are likely between 6-9 µatm (float pCO_2 high), with direct comparisons yielding lower biases (2-5 µatm) than indirect comparison (Wu & Qi, 2023).

Here we use crossover comparisons between ship and float data to quantify differences in deep oxygen values and calculate the impact those offsets would have on adjusted and derived parameters. We refer to oxygen "offset", without strict attribution of the source of the error. Given the long-history and extensive use of Winkler titrations and the relatively short time-history of oxygen sensors and float-mounted oxygen sensors, we assume that the GLODAP values are likely correct and will present evidence later that supports this assumption.

2. Methods

2.1 Float and shipboard datasets

Biogeochemical float data were downloaded on June 20, 2023 from the Argo Data Assembly Centers (DACs, closest snapshot: 2023-06-09 (Argo, 2023)) according to the list of floats with oxygen, nitrate, or pH sensors in the Argo Global DACs Synthetic-Profile Index file. 1,839 Argo floats deployed between 2002 and 2023 were included in this dataset. Of these floats, 1,830 measured oxygen, of which 151,880 out of 270,654 profiles (56%) contained valid delayed mode "ADJUSTED" oxygen data, indicating the data had been checked and/or corrected post deployment (median difference between uncorrected and adjusted data was 11 µmol kg^{-1}, adjusted data higher). 582 floats measured nitrate, with 43,280 delayed mode, adjusted profiles out of 67,505 total (64%). 492 floats measured pH, of which 19,137 profiles had delayed mode, adjusted data out of 44,407 total (43%). Only delayed mode, adjusted float data flagged as "good" were used for this analysis. Float oxygen sensor calibration type for floats with valid crossovers was determined by reading the "SCIENTIFIC_CALIB_COMMENT" field (198 unique comments, Table S1) in float "Sprof" files and were categorized as: air, non-air, and bad/no calibration.

Shipboard bottle measurements of salinity, temperature, oxygen, nitrate, total CO_2 (DIC), or pH flagged as "good" from the GLODAP v2.2022 (Olsen et al., 2020) were used for comparison with float observations. The GLODAP dataset was chosen because it includes a secondary quality control and adjustment to shipboard data for overall accuracy and internal consistency. Shipboard pH measurements in the GLODAP dataset include a variety of measurement approaches. Carter et al. 2021 and 2018 "homogenized" the GLODAP pH dataset to be consistent with spectrophotometric pH measurements using purified meta-cresol purple (Liu et al., 2011) prior to training of the LIPHR (Locally Interpolated PH Regression)/ESPER (Empirical Seawater Property Estimation Routine) algorithms. To recreate this homogenized dataset, we used ESPER to calculate GLODAP pH on the total scale normalized to 25°C ($pH_{25\text{-}T}$) for every datapoint where a "good" pH measurement was present. By using the algorithms at the same datapoints on which they were trained, we recreated the homogenized dataset and used this data for the pH crossover comparison.

2.2 Crossover comparisons

Criteria for crossover comparisons between float and shipboard measurements were established using distance, pressure, potential density, and spice. For each float profile, we first found all GLODAP bottle measurements within a 100 km radius. Float data from 1450 to 2000 db were interpolated to a 1 db grid, with any gaps of over 125 db between successive float measurements removed from the interpolated profiles. Potential density (σ_θ) and spiciness (τ) relative to 0 db were calculated for both the float and shipboard data using the Gibbs SeaWater Oceanographic Toolbox of TEOS-10 for Python (McDougall & Barker, 2011). For each GLODAP bottle measurement between 1400 and 2100 db within the 100 km distance range, crossovers were determined by looking for interpolated float data with differences of less than ± 0.005 kg m^{-3} σ_θ, ± 0.005 τ, and ± 100 db from the GLODAP sample properties. These difference thresholds were selected by analyzing levels of environmental oxygen noise in comparisons of individual floats against themselves using a range of density, spiciness, and distance thresholds (Text S1, Figure S1). Crossovers from any point in time were allowed. Mean property offsets (e.g., ΔC_{off}, for a given property "C") were calculated for floats with at least 20 oxygen crossovers present and used in the results shown here (Figure S2 for float crossover examples). On average, offsets in the 1450 – 2000 db range did not differ significantly as a function of depth or concentration (Figure S3). While adjusting the filter criteria does impact the number of crossovers found for each float and can impact the mean offset calculated for an individual float, the overall results presented in this manuscript are relatively insensitive to the exact criteria used.

2.3 Impact of oxygen offsets on derived parameters

The impact of oxygen offsets (e.g., ΔC_{imp}, for a given property C) on float nitrate and pH adjustments and float estimated pCO_2 and DIC ($\Delta NO_3^-{}_{imp}$, $\Delta pH_{25\text{-}T,imp}$, $\Delta pCO_{2,imp}$, and ΔDIC_{imp}, respectively) was determined for each float with valid oxygen and nitrate, or pH data and with a valid GLODAP crossover comparison. As previously described, oxygen offsets impact estimated pCO_2 and DIC through the effect of oxygen on the pH adjustment at 1500 m (Figure 2). While oxygen is also used in the algorithmic estimation of total alkalinity, which is required for pCO_2 and DIC estimation, only a surface oxygen offset would impact the surface total alkalinity estimate, and subsequently the surface pCO_2 and DIC estimates, which are the foci of this work. Using temperature, salinity, and oxygen data at 1500 m from each profile as inputs to the calibration algorithms, we calculated pH and nitrate, both with and without adjusting for the mean float oxygen offset relative to GLODAP ($\Delta O_{2,off}$). We then calculated the differences in (impacts on) pH and nitrate with and without the $\Delta O_{2,off}$ correction ($\Delta pH_{25\text{-}T,imp}$ and $\Delta NO_3^-{}_{imp}$) and applied these differences to the adjusted, full float profiles. This approach allows us to determine the impact a mean oxygen correction would have without attempting to replicate any step changes or non-linear adjustments that may have been applied during the original data adjustment procedure (Maurer et al., 2021). For nitrate, a uniform adjustment is applied to the entire nitrate profile. For pH, the adjustment at 1500 m is scaled relative to the difference in temperature between each depth and 1500 m, following the protocol used in the original pH measurement adjustment (Johnson et al., 2023). $\Delta pCO_{2,imp}$ and ΔDIC_{imp} were calculated using CO2SYS first with the original pH as an input and then with mean $\Delta pH_{25\text{-}T,imp}$ applied to the float profile and calculating the difference.

Floats equipped with pH and nitrate sensors have been deployed by many groups throughout the world. The majority have been deployed as part of SOCCOM or GO-BGC and adjusted by the data management teams of SOCCOM and the Monterey Bay Aquarium Research Institute (MBARI). Following adjustment approaches used by SOCCOM/MBARI, total alkalinity was calculated using LIARv2, pH using ESPER (ESPER-mixed, an average of a neural network-based approach and an MLR; Carter et al., 2021) or LIPHR (Carter et al., 2018), and carbonate system calculations using PyCO2SYS (v1.8.1; Humphreys et al., 2022, 2023). Both ESPER and LIPHR were used because SOCCOM pH data at the time of download have been adjusted to one of the two algorithms, depending on when the float was last active. Floats active prior to April 2023 were adjusted to LIPHR pH estimates, while floats active past this date are adjusted to ESPER pH. Figures and results shown in the main text rely on ESPER-based adjustments while the supplement includes results that rely on LIPHR-based adjustments, but average differences between the two are of a second order.

One complication is the presence of a pH-dependent pH correction (Williams et al., 2017) used by the SOCCOM/GO-BGC groups when calculating pCO_2. This correction accounts for the difference between float ISFET measured pH, which has been shown to align with spectrophotometric pH measurements (Takeshita et al., 2020), and pH values calculated from measurements of TA and DIC (Carter et al., 2013; 2018). It is currently unclear how to best apply a global correction similar to the one developed in Williams et al. (2017) and a recently published paper by a working group focused on inter-consistency in carbonate system measurements recommended removing this correction until a suitable correction for all float pH measurements can be developed (Carter et al., 2023). We have avoided dealing with this issue by looking at the difference between calculations with and without the oxygen offset impacts included. Any changes in this pH-dependent bias correction will represent a different, additional impact to float pCO_2 estimates. However, the impacts of the oxygen offset should be very similar for estimated pCO_2 with or without this additional pH-dependent bias correction.

3. Results

For floats where oxygen offsets could be determined, 93% of the offsets were statistically significant using a 1 sample t-test and a p-value threshold of 0.01 (Figures S2A and S2B for examples with significant and non-significant crossovers). Our goal is to quantify the potential impact of float oxygen offsets on other parameters, so we include all oxygen offset estimates, including insignificant ones, when calculating the mean oxygen offset and all floats with pH in the subsequent impact on derived parameters. Only including floats with a statistically significant oxygen offset would overstate the magnitude of the mean dataset offset. However, if these crossover comparisons were used in the future to correct float oxygen data it would be important to only adjust those floats with significant offsets so as not to over-adjust already good data.

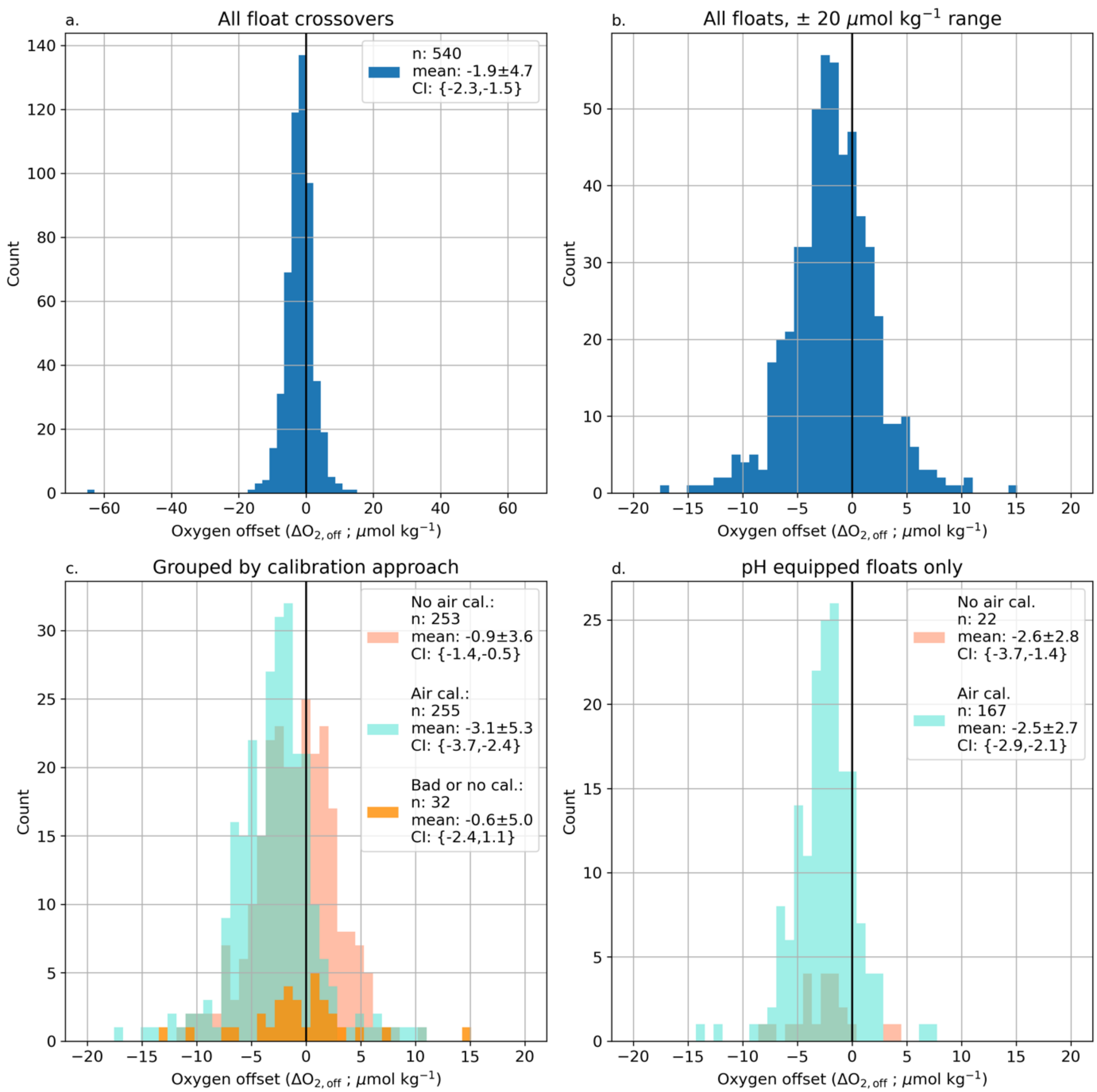

Figure 3. Histograms of float oxygen offsets relative to shipboard measurements. Float minus ship mean offsets between 1450 and 2000 db for all floats (a) with crossovers to GLODAP data and within ± 0.005 kg m^{-3} σ_θ, ± 0.005 τ, and ± 100 db. (b) same as (a) but with a restricted x-axis range of ± 20 μmol kg^{-1}. (c) Offsets grouped by calibration type: non-air calibration listed (light red), air-calibrated (turquoise), and no calibration or bad calibration, though still marked as good data (orange). (d) same as (c) but only showing data from non-air and air calibrated floats equipped with pH sensors. Figure legends list the number of floats, mean offsets ± 1 SD, and 95% confidence intervals for each calibration category.

The mean oxygen offset for all floats is -1.9 ± 4.7 (1 SD) μmol kg^{-1} (n=540, float minus GLODAP, Figure 3, Table S2), with a 95% confidence interval around the mean of -2.3 to -1.5 μmol kg^{-1}. Of the float oxygen sensors with valid crossovers, 253 had a non-air calibration method listed (offset of -0.9 ± 3.6, 95% CI {-1.4, -0.5} μmol kg^{-1}), 255 were air calibrated (offset of -3.1 ± 5.3, 95% CI {-3.7, -2.4} μmol kg^{-1}), and 32 had bad calibration or no calibration listed, though they were still flagged as good data (offset of -0.6 ± 5.0, 95% CI {-2.4, 1.1} μmol kg^{-1}) (Figure 3). For floats equipped with pH, non-air calibrated floats had a mean oxygen offset of -2.6 ± 2.8 μmol kg^{-1} (n=22, 95% CI {-3.7, -1.4} μmol kg^{-1}) and air-calibrated floats had a mean oxygen offset of -2.5 ± 2.7 μmol kg^{-1} (n=167, 95% CI {-2.9, -2.1} μmol kg^{-1}).

Offsets were also calculated between GLODAP and float measurements of nitrate and pH and between float estimates of DIC and GLODAP measurements (Figure 4, blue shaded histograms). The mean difference between float and GLODAP nitrate measurements is 0.2 ± 0.4 μmol kg^{-1} ($\Delta NO_{3}^{-}{}_{off}$, n=232, 95% CI {0.1, 0.2} μmol kg^{-1}). The mean difference between float and GLODAP pH (normalized to 25°C, total pH scale) is -4.5 ± 7.7 mpH ($\Delta pH_{25\text{-}T,off}$, n=113, 95% CI {-5.9, -3.1} mpH). The mean difference between DIC estimated from float pH and LIARv2 alkalinity and GLODAP DIC measurements is 2.9 ± 6.1 μmol kg^{-1} (ΔDIC_{off}, n=154, 95% CI {1.9, 4.0} μmol kg^{-1}, full statistics for nitrate, $pH_{25\text{-}T}$, and DIC crossovers in Table S3).

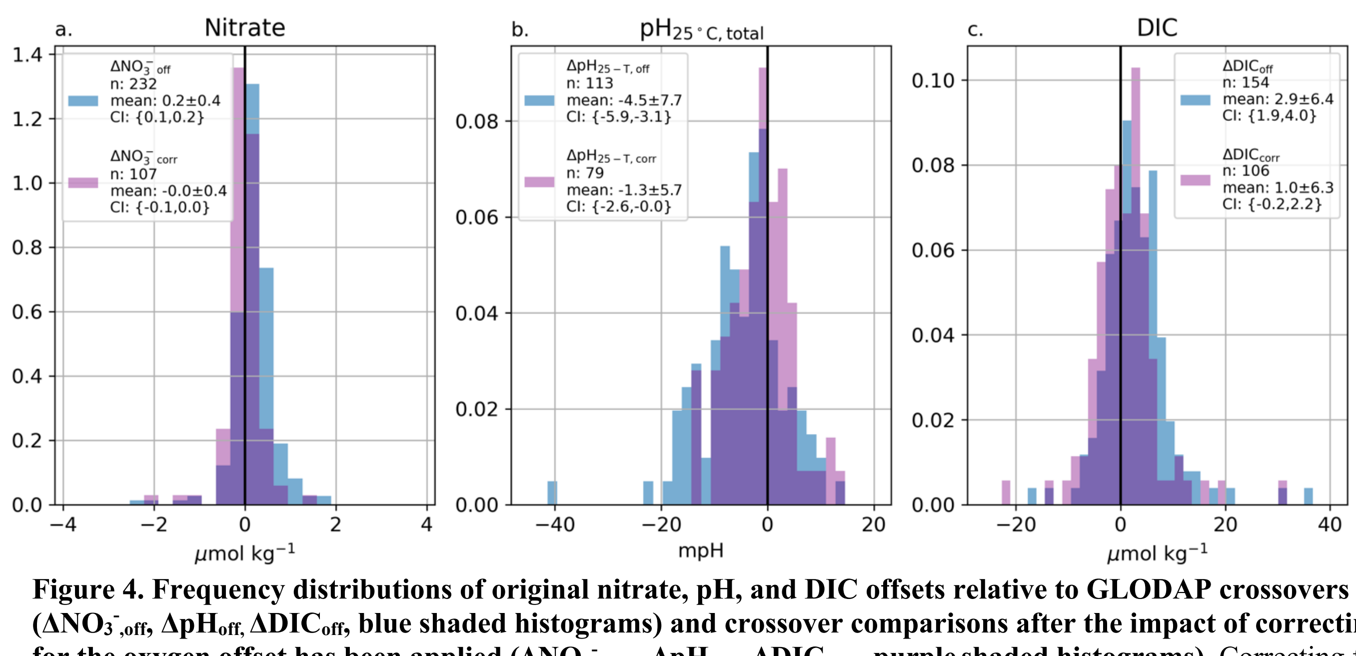

Figure 4. Frequency distributions of original nitrate, pH, and DIC offsets relative to GLODAP crossovers ($\Delta NO_{3}^{-}{}_{,off}$, ΔpH_{off}, ΔDIC_{off}, blue shaded histograms) and crossover comparisons after the impact of correcting for the oxygen offset has been applied ($\Delta NO_{3}^{-}{}_{,corr}$, ΔpH_{corr}, ΔDIC_{corr}, purple shaded histograms). Correcting for the observed oxygen offset at depth fully corrects the nitrate offset relative to GLODAP and improves the pH and DIC crossover comparisons by 3.2 mpH and 1.9 μmol kg^{-1}, respectively. Full statistics in Tables S3 and S4.

4. Discussion

4.1 Oxygen offsets

The magnitude of the mean offset is larger for air-calibrated floats (-3.1 µmol kg^{-1}) than non-air calibrated floats (-0.9 µmol kg^{-1}). This likely reflects that some non-air calibrated floats are corrected using both deep and near-surface shipboard values or a full oxygen profile from a cast made at the deployment location (Figure 2) (Drucker & Riser, 2016; Takeshita et al., 2013), while air calibrated floats are primarily adjusted using a gain value derived from surface measurements of atmospheric oxygen over the float's lifetime (Bushinsky et al., 2016; Johnson et al., 2015, 2017; Maurer et al., 2021). The offsets shown in Figure 3 are small relative to the original correction from raw to "adjusted" oxygen (median difference between raw and adjusted float oxygen of -11 µmol kg^{-1}), but the mean offset for all comparisons other than the "Bad / no. cal" is significantly different from zero (95% confidence intervals, Figure 3 and Table S2) and, as we will discuss below, can have significant impacts on interpretation and use of this data.

The offset between float and shipboard data could reflect either that the float oxygen measurements are low or ship-based Winkler measurements are high. We will discuss both possibilities. It is important to recognize that we are considering approximately 12 different oxygen sensor models that utilize two different measurement principles made by three different manufacturers. The earliest oxygen sensors deployed on floats were initially Clark-cell electrodes (various SeaBird 43 models) while most current sensors are oxygen optodes made by multiple manufacturers, but with the same basic sensing approach and chemistry (Figure 1). Here we primarily focus on possible offsets for oxygen optodes rather than Clark cell electrodes, and on air-calibrated optodes specifically, as these are the current state of the art sensors and represent the bulk of floats deployed with pH sensors.

If the offset between float and shipboard oxygen is due to low-biased float oxygen measurements, three possible causes of the offsets are: (i) an in-situ difference in ocean oxygen content, (ii) a residual uncorrected pressure response, or (iii) a non-linear concentration-dependent drift in the sensor response. Basin-scale comparisons of float oxygen to shipboard measurements have indicated overall good float sensor accuracy in the surface ocean. A comparison of SOCCOM float oxygen data relative to GLODAP on pressure surfaces above and below the thermocline indicated that the air-calibrated float data at depth may be low by 2-3 μmol kg^{-1} (Bushinsky et al., 2017). Johnson et al. (2017) assessed SOCCOM float data relative to the respective float deployment cruises, across the full depth range, and against GLODAP crossovers within 20 km and below 300 m, finding that float oxygen was lower than GLODAP by 3.7 and 3.2 μmol kg^{-1} at "Midrange", or approximately 250 μmol kg^{-1} [O_2]. Maurer et al. (2021) updated the Johnson et al. (2017) GLODAP comparison, finding a 3.6 to 3.8 μmol kg^{-1} low oxygen offset.

While these earlier studies included shallower waters and compared data on pressure surfaces instead of the combined σ_θ, τ, and pressure criteria of this study, all found similar magnitude and direction of differences. Maurer et al. (2021) attribute the mean [O_2] difference to a mean age difference of 18.6 years between the GLODAP and SOCCOM datasets due to the linear declines in Southern Ocean interior oxygen concentrations found in Helm et al. (2011). In Helm et al. (2011) the Southern Ocean has the largest region of oxygen change in the upper ocean (100-1000 m) with the south Pacific and south Indian basins, seeing changes of up to -0.7 μmol l^{-1} yr^{-1} between 1970 and 1992. However, deeper in the water column where we are conducting our crossover comparison (1500 – 2000 m), changes are of a smaller magnitude and appear to primarily be between 0 to -0.1 μmol l^{-1} yr^{-1} in the Southern Ocean and between -0.2 and +0.2 μmol l^{-1} yr^{-1} across the global ocean. To rule out true oxygen changes as the source of offsets between the float and GLODAP datasets, we can assess the GLODAP dataset for long-term oxygen changes and compare that oxygen change to the magnitude of float oxygen offsets.

If true oxygen changes at depth are responsible for the offsets found in our present study, we would expect the offset for a given float relative to GLODAP to be larger when compared to older cruise data and decrease in magnitude (typically becoming less negative) the closer in time between the float and GLODAP measurements. We therefore fit regressions to the float-GLODAP offsets as a function of time, calculating a 95% confidence interval around the regression, to determine if the regression ± CI included a zero offset at the midpoint time of the float deployment (see Figure S2A/B for examples of the regression and CI). 66 float offsets (12%) could be a result of observed change in oxygen concentration as determined by a trend in the GLODAP dataset. However, in many of these cases, there is simply too much uncertainty in the GLODAP oxygen regression with time, or little to no difference between the float and shipboard data. While actual ocean oxygen change in the 1450 – 2000 db range remains a possible contributing factor to the differences seen between the float and shipboard oxygen, it does not seem to be the main cause.

The optode pressure correction combines a sensor pressure response with the pressure effect on oxygen solubility. Significant effort has been made to determine the appropriate pressure correction between optode response and oxygen partial pressure (e.g., Bittig et al., 2015; Uchida et al., 2008). We primarily focus on the mean oxygen offset for each float, averaging all crossover data from 1450 – 2000 db. However, if we bin the data by depth rather than pressure, we do not see a depth dependency in the float offsets in this depth range. While this may still exist, it does not seem to be a first order factor in the deep oxygen offset (Figure S3). A related issue is the lag in optode oxygen measurements due to the sensor response time (Bittig et al., 2014; Bittig & Körtzinger, 2017). Oxygen gradients at these depths are small, so is not likely to be a first order issue but may contribute to the offset in some regions.

Optodes have been shown to not change response at zero oxygen (Johnson et al., 2015) and multiple calibrations over time have indicated relatively linear drift rates (Bittig & Körtzinger, 2015), lending support for the use of a gain correction across entire oxygen profiles. Bushinsky et al. (2016) measured greater drift rates at lower oxygen concentration and postulated that the faster drift rate at low oxygen concentrations represents a deformation of the oxygen calibration surface, such that neither a gain nor an offset can be used to fully correct the range of oxygen measurements. Drucker and Riser (2016) show data from eight floats indicating that a near-surface gain correction leaves a negative offset at depth, similar to Bushinsky et al. (2016), but they did not provide an explanation. Both Bushinsky et al. (2016) and Drucker and Riser (2016) compared float data to bottle oxygen measurements from deployment casts, so oxygen change at depth did not play a factor. While we are still uncertain as to the mechanism causing low float oxygen values at depth relative to Winkler data, our current results from the entire oxygen Argo dataset indicate that this does not seem to be an issue limited to a small number of floats and the possibility of non-linear drift in the oxygen calibration surface remains.

An alternative to the float oxygen measurement being biased low, is that bottle measurements of oxygen using Winkler titrations are biased high due to contamination by atmospheric oxygen or impurities in reagents (Schmidtko et al., 2017). Sampling of oxygen and subsequent Winkler titration involves careful isolation of the water from atmospheric contamination. Incomplete flushing or trapped bubbles can add a significant bias to Winkler data given that the low solubility of oxygen in water means that in equal volumes of air and water, the air will hold 50 times more oxygen than the water. GLODAP data has been QC'd and adjusted for internal consistency, but it is possible a bias is present in the deep data, especially if, on average, this bias is present in either all shipboard observations or the shipboard observations used as a reference for GLODAP adjustments.

Regardless of the source of the offset or mechanism for its existence, a systematic offset between float and shipboard oxygen measurements presents a problem for the current use of oxygen in float nitrate and pH parameter adjustment and derived parameter calculations. For all mechanisms other than a true change in ocean oxygen content, these offsets also complicate determination of long-term ocean oxygen changes. Given the mean difference in float and GLODAP dataset ages and that float oxygen measurements are now made in far greater numbers than shipboard data, a negative offset in the float data would appear in a combined data product as an increase in the true ocean oxygen loss signal. For the rest of this discussion, we focus on the impact of an oxygen offset on water properties derived from float measured data.

4.2 Impact of oxygen offsets on pH

As described earlier, float oxygen measurements at 1500 m are used to correct for drift in the pH sensor data which is then used with TA estimates to derive $p\text{CO}_2$ (Williams et al., 2017). We only consider pH-equipped floats with air-calibrated oxygen here, with equivalent figures and tables for all pH-equipped floats provided in the supplement. The mean impact of correcting float pH for observed oxygen offsets is 3.2 ± 3.8 mpH (n=119, 95% CI of 2.5 to 3.9 mpH, Figure 5, Table 1). The impact of oxygen offsets on pH corrections can be understood through the relationship between oxygen and inorganic carbon. A negative oxygen offset means that float oxygen is lower than the expected value, so correcting the oxygen by a positive amount would result in a corresponding reduction in the apparent remineralization signature of the water mass. Higher oxygen and less remineralization would then imply lower dissolved inorganic carbon and therefore higher pH.

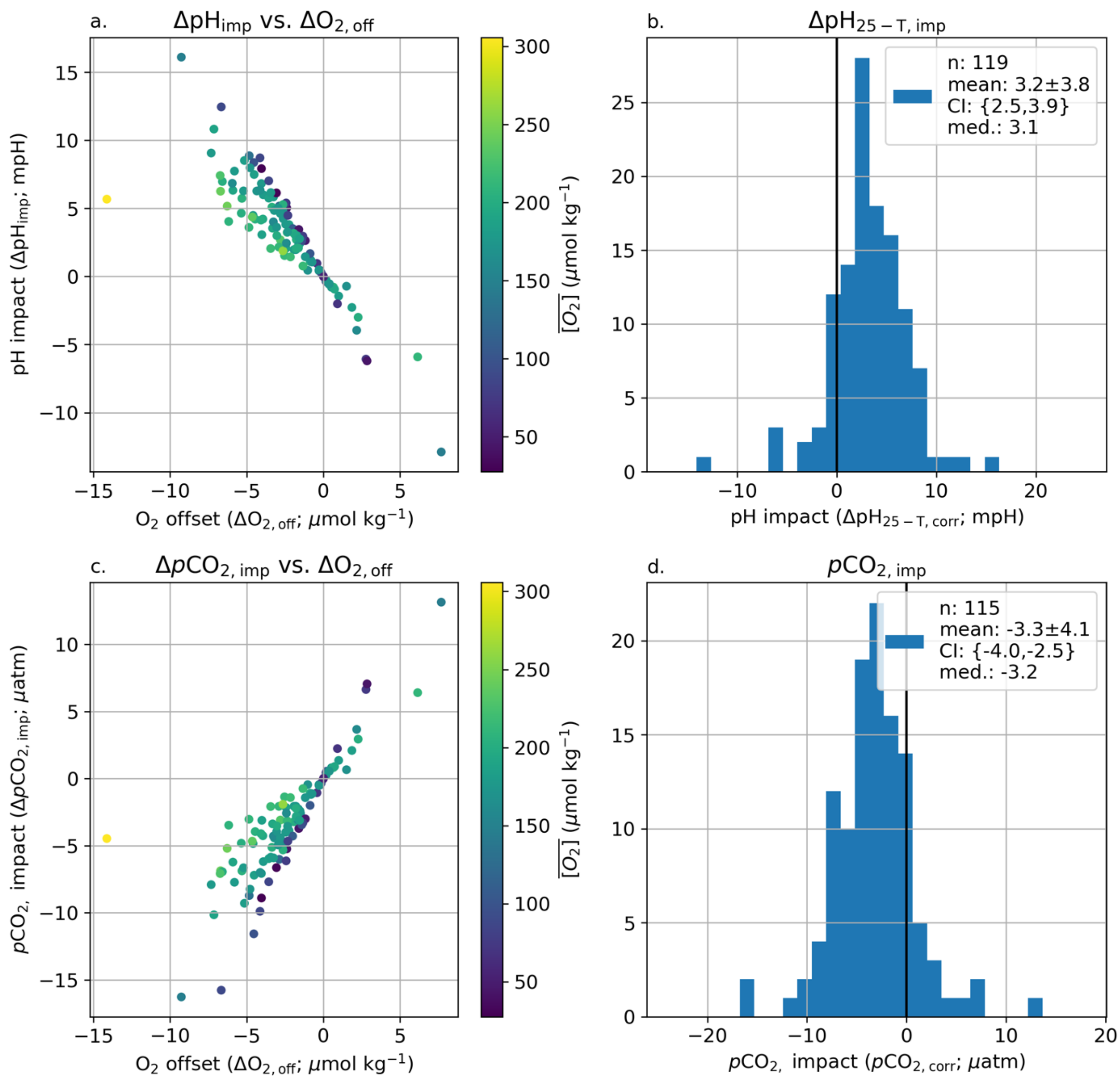

Figure 5. 1500 m pH and surface pCO_2 impact from associated oxygen offset. The calculated impact of correcting pH (top panels) and pCO_2 (bottom panels) for observed oxygen offsets. (a) and (c) are scatter plots of the change in pH and pCO_2 for a given oxygen offset, with the points colored by the mean oxygen concentration of the float crossover comparisons. Correcting a float oxygen sensor that is low of correct would increase the float pH and decrease derived pCO_2. (b) and (d) are histograms of the pH and pCO_2 impacts, with zero marked with a black line and mean ± 1SD, 95% confidence intervals around the mean, and median values listed in the figure legends. pH impacts shown are at 1500 m, while pCO_2 impacts are from surface values. Results shown here are calculated using air-calibrated floats and the ESPER-mixed pH algorithm, equivalent results for LIPHR in Supplemental Figure S4 and for all pH-equipped floats in Figures S5 and S6.

Table 1. Impact on pH at 1500 db and derived surface pCO_2 of observed oxygen offsets.

	Count	Mean	SD	P value[2]	95.0% CI {low, high}	median	min	max
pH impact ($\Delta pH_{25\text{-}T,imp}$, mpH)[1]	119	3.2	3.8	<0.001	{2.5, 3.9}	3.1	-12.9	16.1
pCO_2 impact ($\Delta pCO_{2,imp}$, µatm)	115	-3.3	4.1	<0.001	{-4.0, -2.5}	-3.2	-16.3	13.1

[1]Impacts for pH calculated using ESPER-mixed. Impacts shown here are for air-calibrated floats only. Equivalent numbers for LIPHR in Table S5.

[2]P-value testing the hypothesis that the mean impact is different from 0 at a 95% confidence level.

The sensitivity of pH impact ($\Delta pH_{25\text{-}T,imp}$) to oxygen offset (represented as $\Delta pH_{25\text{-}T,imp}/\Delta O_{2,off}$) is between -0.4 and -2.3 mpH/(µmol kg^{-1}) (Figures 5, 6). The largest $\Delta pH_{25\text{-}T,imp}$ values are observed in the Southern and Pacific Oceans, though the oxygen offsets do not show a similar spatial pattern. Instead, there is a spatial pattern to the $\Delta pH_{25\text{-}T,imp}/\Delta O_{2,off}$ that corresponds to the crossover oxygen concentration. The pH impact sensitivity (slope of the pH impact to O_2 offset) is greatest at low oxygen concentrations (Figure 5), with the mean crossover oxygen concentration explaining 61% of the variance in calculated $\Delta pH_{25\text{-}T,imp}/\Delta O_{2\ Off}$ (Figure 6).

This relationship can be understood by considering the changes in ocean chemistry as respiration takes place in a parcel of water. For example, we can take a parcel of water with initial TA and DIC of newly formed Subantarctic Mode Water from the southeast Pacific Ocean (Carter et al., 2014). If we then calculate the impact of organic matter respiration on oxygen, DIC, and TA, we can calculate the change in pH for every mole of oxygen respired. Initially, the $\Delta pH_{25\text{-}T,imp}/\Delta O_{2\ Off}$ is -2 mpH/(µmol kg^{-1}), but it drops to almost -2.8 mpH/(µmol kg^{-1}) after 250 µmol kg^{-1} of oxygen have been respired (Figure S10), as the buffer capacity of the water is eroded with increasing DIC; a situation analogous to surface ocean acidification. In the real ocean, regional differences in ocean interior biogeochemistry, including the buffer capacity, will cause regional differences in the sensitivity of pH to oxygen change. This likely accounts for larger spread in the relationship between pH impact and oxygen offset for points in the Southern Ocean (Figure 6e), where there is significant mixing of different water masses.

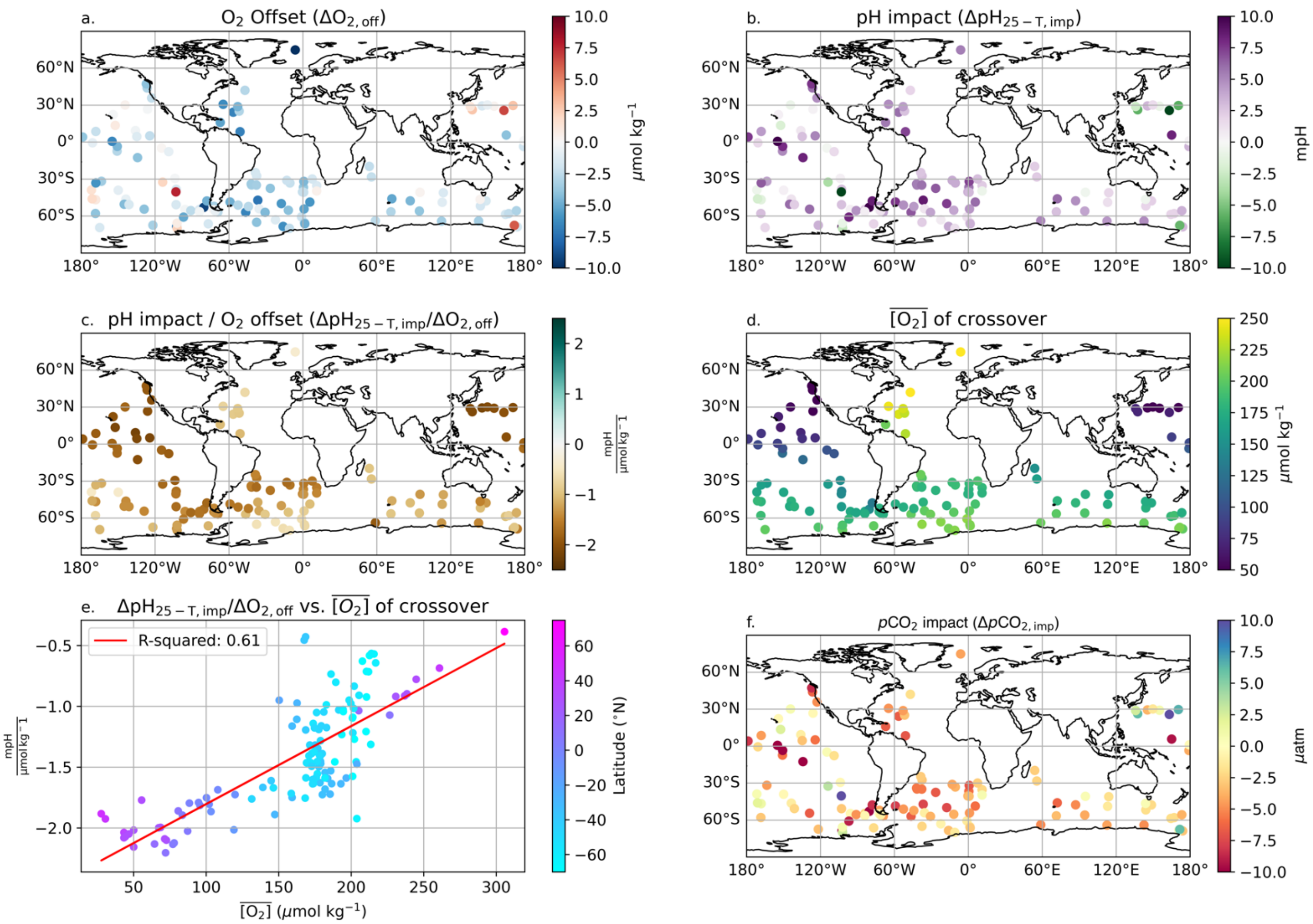

Figure 6. Relationships between oxygen offset, pH impact, mean oxygen concentration at crossover, and *p*CO_2 impact. Map of O_2 offset (a.) demonstrates no obvious spatial pattern in the magnitude of offsets while the impact of correcting for the O_2 offsets on pH (b.) tends to be greater in the Pacific and subpolar Southern Ocean. The ratio of the pH impact to O_2 offset ($\Delta pH_{25\text{-}T,imp}/\Delta O_{2,off}$, c.) is greatest in the north Pacific and subpolar Southern Ocean, reflecting the mean oxygen concentration at crossover ($\overline{[O_2]}$, d.). Plotting $\Delta pH_{25\text{-}T,imp}/O_{2,off}$ against $\overline{[O_2]}$ (e.) and calculating a linear fit (red line) indicates that variations in $\overline{[O_2]}$ explain ~ 61% of the variance in $\Delta pH_{imp}/\Delta O_{2,off}$. Much of the deviation from the linear fit occurs in the Southern Ocean (light blue points). This response leads to stronger *p*CO_2 impacts (f.) in the Pacific and subpolar Southern Ocean than in the North Atlantic or polar Southern Ocean. pH impacts shown are at 1500 m, while *p*CO_2 impacts are from surface values. Results shown here are calculated using floats with air-calibrated oxygen and the ESPER-mixed pH algorithm, equivalent results for LIPHR in Supplemental Figure S7 and for all pH-equipped floats in Figures S8 and S9.

4.3 Impact of oxygen offsets on derived pCO₂

The mean impact on float *p*CO_2 ($\Delta pCO_{2,imp}$) for correcting observed oxygen offsets is -3.3 ± 4.1 µatm (air-calibrated floats only, 95% CI of -4.0 to -2.5 µatm, Figure 5, Table 1). A reduction of float derived pCO_2 by this magnitude would account for a large portion of the apparent bias in float pCO_2 described in Wu and Qi (2023). For many of the studies with direct crossover comparisons, a mean adjustment of -3.3 µatm would effectively eliminate the observed differences between float and shipboard pCO_2. It should be noted that the range in impacts is much greater than the mean (a range of pCO_2 impact between -16.3 and 13.1 µatm), with the impact on individual floats and regions differing from the mean impact. Following pH, some of the greatest pCO_2 impacts are found in the Southern Ocean and Pacific (Figure 6). Similar results are found if LIPHR is used to calculate pH impacts instead of ESPER, though with a slightly greater mean pCO_2 impact magnitude (-3.6 ± 4.7, 95% CI {-4.4, -2.7} µatm, Table S5) and differences for individual floats.

4.4 Impacts of oxygen offset on nitrate and derived DIC

As described above, relative to GLODAP float nitrate is offset high (0.2 ± 0.4 µmol/kg), pH is offset low (-4.5 ± 7.7 mpH), and DIC is offset high (2.9 ± 6.4 µmol/kg, Table S3). These offsets are in approximately the correct ratios and directions for a biological signal, either real or due to an offset in the oxygen measurement that is then propagated to pH and DIC. This indicates that the oxygen offset is most likely not due to a problem with the Winkler titrations.

The calculated impact on nitrate of the oxygen offset ($\Delta NO_3^-{}_{imp}$) is small (mean -0.2 ± 0.2, 95% CI {-0.22, -0.14} µmol kg^{-1}, Figure S11, Table S6), in keeping with Redfield stoichiometry of -16 N: 154 O_2 (Hedges et al., 2002) multiplied by an oxygen offset of ~2 µmol kg^{-1}. The impact on DIC at 1500 m of correcting for oxygen offsets is -1 ± 1.2 µmol kg^{-1} (CI {-1.3, -0.8}, Figures S12, S13, Table S7), which is the slightly smaller than multiplying the oxygen offset by a Redfield ratio of -106 C: 154 O_2.

As a check that the mean impacts ($\Delta NO_3^-{}_{imp}$, $\Delta pH_{25\text{-}T,imp}$) calculated from the oxygen offset would, in fact, improve the crossover comparisons with GLODAP data, we also re-ran the crossover comparison after applying the $\Delta O_{2,off}$, $\Delta NO_3^-{}_{imp}$ and $\Delta pH_{25\text{-}T,imp}$ to the original data. The resulting crossovers ($\Delta NO_3^-{}_{,corr}$, $\Delta pH_{25\text{-}T,corr,}$ ΔDIC_{corr}, Figure 4, purple shaded histograms, statistics in Table S4) indicate that, on average, correcting the nitrate for the oxygen offset eliminates the entire nitrate offset, 3.2 mpH of the original -4.5 mpH pH offset, and 1.9 µmol kg^{-1} of the original 2.9 µmol kg^{-1} DIC offset.

The improvement in, but not full correction of, float-derived DIC and measured pH relative to the GLODAP crossovers provides an independent assessment that correcting the float data for this deep O_2 offset does, in fact, improve the pH data and subsequent carbonate system derived parameters. Furthermore, the fact that the nitrate bias appears to be fully corrected while the DIC, pH, and pCO_2 impacts only partially correct for the differences between float values and crossover or indirect comparisons gives an indication that the deep oxygen offset is not the only source of bias in the derived float carbonate parameters. This could be due to additional biases in the float pH, biases in the estimated TA, penetration of the ocean acidification signal to these depths, internal consistency issues with the marine carbonate system, or some other factor. It appears that correcting the oxygen may resolve a large fraction of the differences, but additional work remains to fully separate pCO_2, pH, and DIC biases in the float data from true trends.

Float oxygen crossovers with shipboard data are not available for all floats, making it difficult to determine the magnitude of an oxygen offset for all pH-equipped floats at this time. One option to reduce the possibility of an oxygen-induced bias in pH and derived pCO_2 is to correct float pH using an algorithm that does not use oxygen. Removing oxygen from the ESPER algorithm yields a 4.2 ± 6.0 mpH and -4.3 ± 6.4 µatm pCO_2 impact, somewhat larger than the 3.2 ± 3.8 mpH and -3.3 ± 4.1 µatm impact we calculate from applying the observed oxygen offset relative to GLODAP (Figure S14). Plotting the ESPER pH and pCO_2 impacts due to removing oxygen or correcting for the observed offset against one another yields a cluster of points around the 1:1 lines, though with considerable spread (Figure S15). This approach may indeed improve derived parameters for floats with no GLODAP oxygen crossover for comparison, but is less accurate than if an oxygen correction is possible.

5. Conclusions

Here we identify an offset between float and ship-board oxygen measurements between 1450 and 2000 db. The magnitude of this offset is significant for studies assessing long-term ocean oxygen changes and for the use of float oxygen in adjusting pH and nitrate measurements and in subsequent calculated derived carbonate system quantities.

Correcting oxygen offsets of -2.5 µmol kg^{-1} in pH-equipped, air-calibrated floats results in mean pH changes of 3.2 mpH and pCO_2 changes of -3.3 µatm. These differences are of similar direction and magnitude as many of the direct in-situ comparisons between float pCO_2 estimates and underway pCO_2 measurements and therefore represent a first-order bias that needs to be addressed in the biogeochemical float dataset. We do not definitively identify whether the oxygen offset is present in the float or ship-board dataset, though the fact that correcting float oxygen would improve the nitrate, DIC, and pH crossover comparisons to shipboard data are a strong indication that the issue lies with the float observations.

Based on our findings we offer the following recommendations:

1. The oxygen offsets described here represent an empirical correction that must be investigated to determine the underlying mechanism.
2. Float oxygen data at depth should be adjusted to historical shipboard data until our mechanistic understanding of the causes behind the float – ship differences is sufficient such that air calibration or other adjustments do not require a secondary correction.
3. Care should be used in combining float and ship data, with an understanding that small biases may be present in either dataset that could impact the usability of the data compilation to answer some scientific questions. The difference between the average age of the float and GLODAP datasets and the shift to float profile numbers greatly exceeding shipboard profiles in the 2000's mean that any offset between the datasets will appear as a spurious change in ocean oxygen content. Additionally, individual floats may have significantly greater magnitude biases than the mean and data should be assessed prior to any use relying on absolute accuracy.
4. Standardization of float sensor calibration comments and equations will make future studies of overall biogeochemical Argo float performance easier to perform.

Data availability

The "argo_synthetic-profile_index.txt" file used to determine biogeochemical Argo float WMO numbers and data locations was downloaded from ftp.ifremer.fr/ifremer/argo/dac. GLODAP data is available at https://glodap.info/index.php/merged-and-adjusted-data-product-v2-2022/. The analysis and plotting code used for this manuscript are available at: 10.5281/zenodo.10866941 (Bushinsky et al., 2024).

Acknowledgements

SMB was supported by the NOAA Climate Program Office's Climate Observations and Monitoring, Climate Variability and Predictability, and Global Ocean Monitoring and Observation programs (NA21OAR4310260). A.J. Fassbender was supported by NOAA PMEL. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (https://argo.ucsd.edu, https://www.ocean-ops.org). The Argo Program is part of the Global Ocean Observing System (Argo, 2000). Two projects that contributed a significant amount of data used in this study are: The Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project funded by the National Science Foundation, Division of Polar Programs (NSF PLR -1425989 and OPP-1936222), supplemented by NASA, and the Global Ocean Biogeochemistry Array (GO-BGC) Project funded by the National Science Foundation, Division of Ocean Sciences (NSF OCE-1946578). The international effort of GLODAP is foundational to the work described in this manuscript and an essential component of our ocean observing capabilities. This is SOEST contribution number XXXX and PMEL contribution number 5617. We would like to thank an anonymous PMEL technical reviewer for their helpful suggestions.

References

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

Argo. (2023). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) - Snapshot of Argo GDAC of June 09st 2023 [Data set]. SEANOE. https://doi.org/10.17882/42182#103075

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al. (2016). A multi-decade record of high-quality fCO2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/essd-8-383-2016

Bittig, H. C., & Körtzinger, A. (2015). Tackling oxygen optode drift: Near-surface and in-air oxygen optode measurements on a float provide an accurate in-situ reference. *Journal of Atmospheric and Oceanic Technology*, *32*(8), 1536–1543. https://doi.org/10.1175/JTECH-D-14-00162.1

Bittig, H. C., & Körtzinger, A. (2017). Technical note: Update on response times, in-air measurements, and in situ drift for oxygen optodes on profiling platforms. *Ocean Science*, *13*(1), 1–11. https://doi.org/10.5194/os-13-1-2017

Bittig, H. C., Fiedler, B., Scholz, R., Krahmann, G., & Körtzinger, A. (2014). Time response of oxygen optodes on profiling platforms and its dependence on flow speed and temperature. *Limnology and Oceanography: Methods*, *12*, 617–636. https://doi.org/10.4319/lom.2014.12.617

Bittig, H. C., Fiedler, B., Fietzek, P., & Körtzinger, A. (2015). Pressure response of Aanderaa and Sea-Bird oxygen optodes. *Journal of Atmospheric and Oceanic Technology*.

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., et al. (2018). An Alternative to Static Climatologies: Robust Estimation of Open Ocean CO2 Variables and Nutrient Concentrations From T, S, and O2 Data Using Bayesian Neural Networks. *Frontiers in Marine Science*, *5*(September), 328. https://doi.org/10.3389/fmars.2018.00328

Bittig, H. C., Körtzinger, A., Neill, C., van Ooijen, E., Plant, J. N., Hahn, J., et al. (2018). Oxygen Optode Sensors: Principle, Characterization, Calibration, and Application in the Ocean. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00429

Bushinsky, S. M., & Cerovečki, I. (2023). Subantarctic Mode Water Biogeochemical Formation Properties and Interannual Variability. *AGU Advances*, *4*(2), e2022AV000722. https://doi.org/10.1029/2022AV000722

Bushinsky, S. M., & Emerson, S. (2015). Marine biological production from in situ oxygen measurements on a profiling float in the subarctic Pacific Ocean. *Global Biogeochemical Cycles*, *29*(12), 2050–2060. https://doi.org/10.1002/2015GB005251

Bushinsky, S. M., Emerson, S. R., Riser, S. C., & Swift, D. D. (2016). Accurate oxygen measurements on modified Argo floats using in situ air calibrations. *Limnology and Oceanography: Methods*, *14*(8), 491–505. https://doi.org/10.1002/lom3.10107

Bushinsky, S. M., Gray, A. R., Johnson, K. S., & Sarmiento, J. L. (2017). Oxygen in the Southern Ocean From Argo Floats: Determination of Processes Driving Air-Sea Fluxes. *Journal of Geophysical Research: Oceans*, *122*(11), 8661–8682. https://doi.org/10.1002/2017JC012923

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., et al. (2019). Reassessing Southern Ocean Air-Sea CO 2 Flux Estimates With the Addition of Biogeochemical Float Observations. *Global Biogeochemical Cycles*, *33*(11), 1370–1388. https://doi.org/10.1029/2019GB006176

Bushinsky, S. M., Tamsitt, V., Nachod, Z., Fassbender, A., & Williams, N. (2024). SethBushinsky/argo_deep_o2_bias: Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO2 analysis and plotting code (Version v0.1.1-submission). Retrieved from https://doi.org/10.5281/zenodo.10866942

Carpenter, J. H. (1965). The Chesapeake Bay Institute Technique for the Winkler Dissolved Oxygen Method. *Limnology and Oceanography*, *10*(1), 141–143. https://doi.org/10.4319/lo.1965.10.1.0141

Carter, B. R., Radich, J. A., Doyle, H. L., & Dickson, A. G. (2013). An automated system for spectrophotometric seawater pH measurements. *Limnology and Oceanography: Methods*, *11*(1), 16–27. https://doi.org/10.4319/lom.2013.11.16

Carter, B. R., Talley, L. D., & Dickson, A. G. (2014). Mixing and remineralization in waters detrained from the surface into Subantarctic Mode Water and Antarctic Intermediate Water in the southeastern Pacific. *Journal of Geophysical Research: Oceans*, *119*(6), 4001–4028. https://doi.org/10.1002/2013JC009355

Carter, B. R., Williams, N. L., Gray, A. R., & Feely, R. A. (2016). Locally interpolated alkalinity regression for global alkalinity estimation. *Limnology and Oceanography: Methods*, (1998). https://doi.org/10.1002/lom3.10087

Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B., & Takeshita, Y. (2018). Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate. *Limnology and Oceanography: Methods*, *16*(2), 119–131. https://doi.org/10.1002/lom3.10232

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y.-Y., et al. (2021). New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, *19*(12), 785–809. https://doi.org/10.1002/lom3.10461

Carter, B. R., Sharp, J. D., Dickson, A. G., Álvarez, M., Fong, M. B., García-Ibáñez, M. I., et al. (2023). Uncertainty sources for measurable ocean carbonate chemistry variables. *Limnology and Oceanography*, lno.12477. https://doi.org/10.1002/lno.12477

Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the Global Ocean with Biogeochemical-Argo. *Annual Review of Marine Science*, *12*(1), 1–26. https://doi.org/10.1146/annurev-marine-010419-010956

Coggins, A., Watson, A. J., Schuster, U., Mackay, N., King, B., McDonagh, E., & Poulton, A. J. (2023). Surface ocean carbon budget in the 2017 south Georgia diatom bloom: Observations and validation of profiling biogeochemical argo floats. *Deep Sea Research Part II: Topical Studies in Oceanography*, *209*, 105275. https://doi.org/10.1016/j.dsr2.2023.105275

D’Asaro, E. A., & McNeil, C. (2013). Calibration and stability of oxygen sensors on autonomous floats. *Journal of Atmospheric and Oceanic Technology*, *30*(8), 1896–1906. https://doi.org/10.1175/JTECH-D-12-00222.1

Dickson, A. G. (1994). Determination of dissolved oxygen in sea water by Winkler titration. *WHP Operations and Methods*, (November), 1–13.

Drucker, R., & Riser, S. C. (2016). In situ phase-domain calibration of oxygen Optodes on profiling floats. *Methods in Oceanography*, *17*, 296–318. https://doi.org/10.1016/j.mio.2016.09.007

Emerson, S. (1987). Seasonal oxygen cycles and biological new production in surface waters of the subarctic Pacific Ocean. *Journal of Geophysical Research*, *92*(C6), 6535–6544. https://doi.org/10.1029/JC092iC06p06535

Fay, A. R., Lovenduski, N. S., Mckinley, G. A., Munro, D. R., Gray, A. R., Landschützer, P., et al. (2018). Utilizing the Drake Passage Time-series to understand variability and change in subpolar Southern Ocean pCO 2. *Biogeosciences Discussions*. https://doi.org/10.5194/bg-2017-489

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., & Johnson, D. R. (2010). *World Ocean Atlas 2009 Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. Noaa Atlas Nesdic* (Vol. 70, p. 344pp). Washington, D. C.: U.S. Government Printing Office. Retrieved from http://www.nodc.noaa.gov/OC5/indprod.html

Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., et al. (2018). Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean. *Geophysical Research Letters*, *45*(17), 9049–9057. https://doi.org/10.1029/2018GL078013

Gruber, N., Scott C. Doney, Steven R. Emerson3, Denis Gilbert4, Taiyo Kobayashi5, Arne Körtzinger6, Gregory C. Johnson7, Kenneth S., & Johnson8, Stephen C. Riser3, and Osvaldo Ulloa. (2007). THE ARGO-OXYGEN PROGRAM: A white paper to promote the addition of oxygen sensors to the international Argo float program.

Gruber, N., Doney, S. C., Emerson, S. R., Gilbert, D., Kobayashi, T., Körtzinger, A., et al. (2009). Adding oxygen to Argo: Developing a global in-situ observatory for ocean deoxygenation and biogeochemistry. *Ocean Obs '09*.

Hedges, J. I., Baldock, J. a., Gélinas, Y., Lee, C., Peterson, M. L., & Wakeham, S. G. (2002). The biochemical and elemental compositions of marine plankton: A NMR perspective. *Marine Chemistry*, *78*(1), 47–63. https://doi.org/10.1016/S0304-4203(02)00009-9

Helm, K. P., Bindoff, N. L., & Church, J. A. (2011). Observed decreases in oxygen content of the global ocean. *Geophysical Research Letters*, *38*(23), 1–6. https://doi.org/10.1029/2011GL049513

Hennon, T. D., Riser, S. C., & Mecking, S. (2016). Profiling float-based observations of net respiration beneath the mixed layer. *Global Biogeochemical Cycles*, *30*(6), 920–932. https://doi.org/10.1002/2016GB005380

Humphreys, M. P., Lewis, E. R., Sharp, J. D., & Pierrot, D. (2022). PyCO2SYS v1.8: marine carbonate system calculations in Python. *Geoscientific Model Development*, *15*(1), 15–43. https://doi.org/10.5194/gmd-15-15-2022

Humphreys, M. P., Schiller, A. J., Sandborn, D., Gregor, L., Pierrot, D., van Heuven, S. M. A. C., et al. (2023, January 19). PyCO2SYS: marine carbonate system calculations in Python (Version v1.8.2). Zenodo. https://doi.org/10.5281/ZENODO.3744275

Johnson, K. S., Plant, J. N., Riser, S. C., & Gilbert, D. (2015). Air oxygen calibration of oxygen optodes on a profiling float array. *Journal of Atmospheric and Oceanic Technology*, *32*(11), 2160–2172. https://doi.org/10.1175/JTECH-D-15-0101.1

Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., et al. (2017). Biogeochemical sensor performance in the SOCCOM profiling float array. *Journal of Geophysical Research: Oceans*, *122*(8), 6416–6436. https://doi.org/10.1002/2017JC012838

Johnson, K. S., Maurer, T. L., Plant, J. N., & Takeshita, Y. (2023). *BGC-Argo quality control manual for pH* [Pdf]. [object Object]. https://doi.org/10.13155/97828

Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2015). Global Ocean Data Analysis Project, Version 2 (GLODAPv2). *ORNL/CDIAC-162, NDP-093*. https://doi.org/10.3334/CDIAC/OTG.NDP093_GLODAPv2

Körtzinger, A., Schimanski, J., & Send, U. (2005). High Quality Oxygen Measurements from Profiling Floats: A Promising New Technique. *Journal of Atmospheric and Oceanic Technology*, *22*(3), 302–308. https://doi.org/10.1175/JTECH1701.1

Levin, L. A. (2018). Manifestation, drivers, and emergence of open ocean deoxygenation. *Annual Review of Marine Science*, *10*, 229–260. https://doi.org/10.1146/annurev-marine-121916-063359

Liu, X., Patsavas, M. C., & Byrne, R. H. (2011). Purification and Characterization of meta-Cresol Purple for Spectrophotometric Seawater pH Measurements. *Environmental Science & Technology*, *45*(11), 4862–4868. https://doi.org/10.1021/es200665d

Long, M. C., Stephens, B. B., McKain, K., Sweeney, C., Keeling, R. F., Kort, E. A., et al. (2021). Strong Southern Ocean carbon uptake evident in airborne observations. *Science*, *374*(6572), 1275–1280. https://doi.org/10.1126/science.abi4355

Mackay, N., & Watson, A. (2021). Winter air-sea CO 2 fluxes constructed from summer observations of the Polar Southern Ocean suggest weak outgassing. *Journal of Geophysical Research: Oceans*. https://doi.org/10.1029/2020jc016600

Martz, T. R., Riser, S. C., & Johnson, K. S. (2008). Ocean metabolism observed with oxygen sensors on profiling floats in the South Pacific. *Limnology and Oceanography*, *53*(5, part 2), 2094–2111. https://doi.org/10.4319/lo.2008.53.5_part_2.2094

Maurer, T. L., Plant, J. N., & Johnson, K. S. (2021). Delayed-Mode Quality Control of Oxygen, Nitrate, and pH Data on SOCCOM Biogeochemical Profiling Floats. *Frontiers in Marine Science*, *8*(August), 1–20. https://doi.org/10.3389/fmars.2021.683207

McDougall, T. J., & Barker, P. M. (2011). Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox. SCOR/IAPSO WG127. Retrieved from https://www.teos-10.org/pubs/Getting_Started.pdf

Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2016). The global ocean data analysis project version 2 (GLODAPv2) - An internally consistent data product for the world ocean. *Earth System Science Data*, *8*(2), 297–323. https://doi.org/10.5194/essd-8-297-2016

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Bittig, H. C., Kozyr, A., et al. (2020). GLODAPv2.2020 - the second update of GLODAPv2. *Earth System Science Data Discussions*. https://doi.org/10.5194/essd-2020-165

Roemmich, D., Talley, L., Zilberman, N., Osborne, E., Johnson, K., Barbero, L., et al. (2021). The Technological, Scientific, and Sociological Revolution of Global Subsurface Ocean Observing. *Oceanography*, 2–8. https://doi.org/10.5670/oceanog.2021.supplement.02-02

Sarmiento, J. L., Johnson, K. S., Arteaga, L. A., Bushinsky, S. M., Cullen, H. M., Gray, A. R., et al. (2023). The Southern Ocean carbon and climate observations and modeling (SOCCOM) project: A review. *Progress in Oceanography*, *219*, 103130. https://doi.org/10.1016/j.pocean.2023.103130

Schmidtko, S., Stramma, L., & Visbeck, M. (2017). Decline in global oceanic oxygen content during the past five decades. *Nature*, *542*(7641), 335–339. https://doi.org/10.1038/nature21399

Schofield, O. A., Fassbender, A., Hood, M., Hill, K., & Johnson, K. (2022). A global ocean biogeochemical observatory becomes a reality. *Eos*, *103*. https://doi.org/10.1029/2022EO220149

Sharp, J. D., Fassbender, A. J., Carter, B. R., Johnson, G. C., Schultz, C., & Dunne, J. P. (2022). *GOBAI-O_2: temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly two decades* (preprint). ESSD – Ocean/Chemical oceanography. https://doi.org/10.5194/essd-2022-308

Stramma, L., & Schmidtko, S. (2021). Spatial and Temporal Variability of Oceanic Oxygen Changes and Underlying Trends. *Atmosphere-Ocean*, *59*(2), 122–132. https://doi.org/10.1080/07055900.2021.1905601

Takeshita, Y., Martz, T. R., Johnson, K. S., Plant, J. N., Gilbert, D., Riser, S. C., et al. (2013). A climatology-based quality control procedure for profiling float oxygen data. *Journal of Geophysical Research: Oceans*, *118*(10), 5640–5650. https://doi.org/10.1002/jgrc.20399

Takeshita, Y., Johnson, K. S., Coletti, L. J., Jannasch, H. W., Walz, P. M., & Warren, J. K. (2020). Assessment of pH dependent errors in spectrophotometric pH measurements of seawater. *Marine Chemistry*, *223*, 103801. https://doi.org/10.1016/j.marchem.2020.103801

Tengberg, A., Hovdenes, J., Andersson, H. J., Brocandel, O., Diaz, R., Hebert, D., et al. (2006). Evaluation of a lifetime-based optode to measure oxygen in aquatic systems. *Limnology and Oceanography: Methods*, *4*(1964), 7–17. https://doi.org/10.4319/lom.2006.4.7

Uchida, H., Kawano, T., Kaneko, I., & Fukasawa, M. (2008). In Situ Calibration of Optode-Based Oxygen Sensors. *Journal of Atmospheric and Oceanic Technology*, *25*(12), 2271–2281. https://doi.org/10.1175/2008JTECHO549.1

Williams, N. L., Juranek, L. W., Johnson, K. S., Feely, R. A., Riser, S. C., Talley, L. D., et al. (2016). Empirical algorithms to estimate water column pH in the Southern Ocean. *Geophysical Research Letters*, *43*, 3415–3422. https://doi.org/10.1002/2016GL068539

Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., et al. (2017). Calculating surface ocean pCO2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles*, *31*(3), 591–604. https://doi.org/10.1002/2016GB005541

Wolf, M. K., Hamme, R. C., Gilbert, D., Yashayaev, I., & Thierry, V. (2018). Oxygen Saturation Surrounding Deep Water Formation Events in the Labrador Sea From Argo-O 2 Data. *Global Biogeochemical Cycles*, *32*(4), 635–653. https://doi.org/10.1002/2017GB005829

Wu, Y., & Qi, D. (2023). The controversial Southern Ocean air-sea CO2 flux in the era of autonomous ocean observations. *Science Bulletin*, S2095927323006151. https://doi.org/10.1016/j.scib.2023.08.059

Wu, Y., Bakker, D. C. E., Achterberg, E. P., Silva, A. N., Pickup, D. D., Li, X., et al. (2022). Integrated analysis of carbon dioxide and oxygen concentrations as a quality control of ocean float data. *Communications Earth & Environment*, *3*(1), 92. https://doi.org/10.1038/s43247-022-00421-w

Yang, B., Emerson, S. R., & Bushinsky, S. M. (2017). Annual net community production in the subtropical Pacific Ocean from in situ oxygen measurements on profiling floats. *Global Biogeochemical Cycles*, *31*(4), 728–744. https://doi.org/10.1002/2016GB005545

Global Biogeochemical Cycles

Supporting Information for

Offset between Argo and shipboard oxygen observations at depth imparts bias on derived pH and pCO_2

Seth M. Bushinsky[1], Zachary Nachod[1], Andrea J. Fassbender[2], Veronica Tamsitt[3], Yuichiro Takeshita[4], Nancy Williams

[1]Department of Oceanography, School of Earth and Space Science and Technology, University of Hawai‘i at Mānoa, Honolulu, HI, USA

[2]NOAA/OAR Pacific Marine Environmental Laboratory, Seattle, WA, USA

[3]University of South Florida, St. Petersburg, FL, USA

[4] Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

Contents of this file

Text S1 to S3
Figures S1 to S15
Tables S1 to S7

Introduction

Figures in this supplement include additional detail on the crossover analysis and alternative algorithms described in the main text.

Text S1. Offsets between float data and GLODAP measurements.

To help determine the optimal criteria to select crossovers with shipboard observations, we analyzed the relationship between different selection criteria and changes in float oxygen data. For each float with oxygen data, the profile closest in time to the mean float lifetime was selected as a reference profile. For each float oxygen measurement between 1500 and 2000 db, differences in measured oxygen were calculated between the reference profile and all other float profiles using a range of distance, potential density (σ_θ) and spiciness (τ) criteria. Potential density and spiciness were best able to reduce the “environmental” data noise while still retaining adequate numbers of crossovers (Figure S1).

Crossovers relative to GLODAP were analyzed to determine if differences are significant and whether they could be due to true change in ocean oxygen (Figures S2A/B). Offsets show no average significant relationship with pressure Figure S3.

Oxygen calibration approaches were grouped into “air”, “non-air”, and “no/bad calibration” according to the calibration comment (Table S1).

Text S2. Results for alternative pH algorithms and float selection.

Two pH algorithms are used in this manuscript. Results for ESPER (Carter et al., 2021) are presented in the main text, while results using LIPHR are included here (Figures S4, S6, S7, S9, S13). Additionally, crossovers and impacts presented in the main text focus on air calibrated floats. Results for all floats are presented here (Figures S5, S6, S8, S9).

Oxygen offset relative to nitrate impact and nitrate impact histograms (Figure S11). DIC is derived from float pH and estimated TA. Oxygen impacts on derived DIC are presented here for surface and 1500 db estimates, and for both LIPHR and ESPER pH algorithms (Figures S12 and S13).

We also tested the impact of removing oxygen entirely from the ESPER algorithm (Figure S14). This would allow a mode of correcting float pH when no crossover is available with which to assess float oxygen sensor accuracy. The impact of removing float oxygen vs. correcting float oxygen on pH and pCO_2 is shown in Figure S15.

Text S3. Example calculation of relationship between respiration and sensitivity of pH to incremental changes in DIC.

The pH impact of a given change in DIC will vary according to the current carbonate chemistry of a given water mass. To test this, we took water properties from newly formed Subantarctic Mode Water (Carter et al., 2014) and calculated the DIC and TA changes due to organic matter respiration in units of oxygen. We assumed a stoichiometry of DIC changes of -106 C: 154 O_2 and TA changes of -16 TA: 106 DIC. For each mol of O_2 consumed during respiration, we calculated a change in pH per mol O_2 (Figure S10). While initial starting chemistries would be different in different water masses and the fraction of carbonates produced will also influence the carbonate system, this provides an example of the changing relationship between pH impact and oxygen offset as a function of accumulated respiration signal.

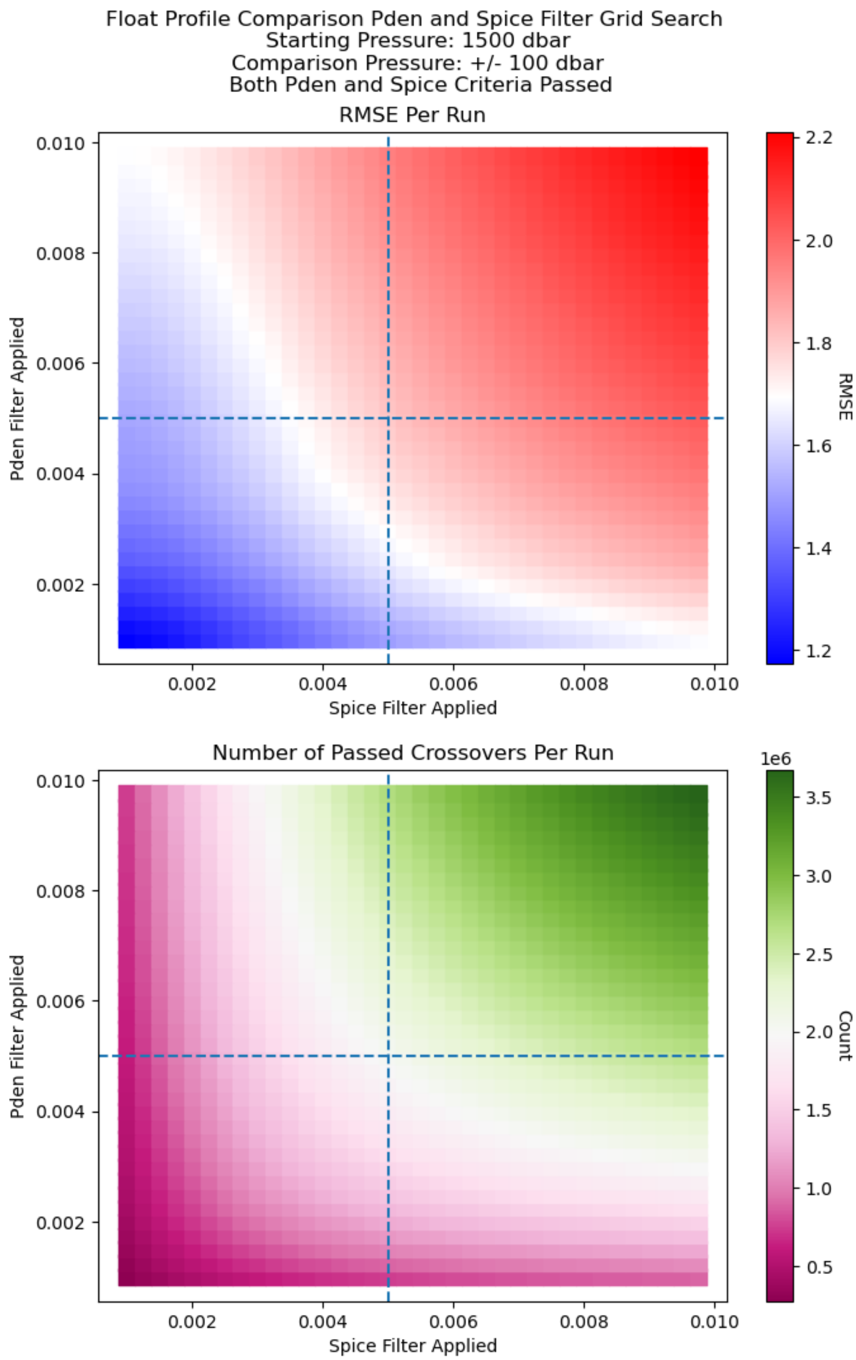

Figure S1. RMSE (µmol kg^{-1}) and amount of data remaining for different combinations of density and spice filters for float oxygen data. The "filter applied" level indicates the allowable differences (±) between measurements to include them as a crossover.

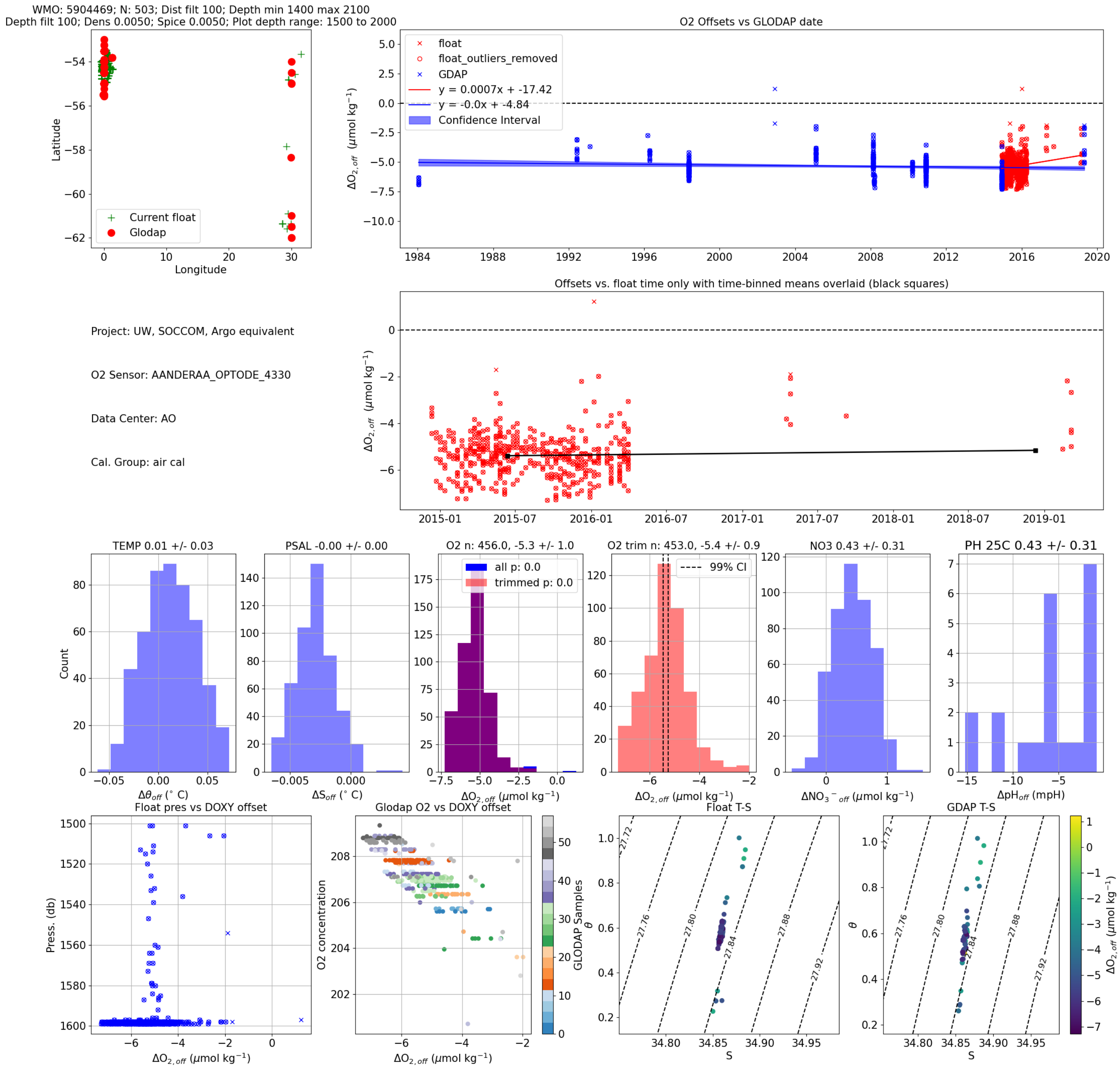

Figure S2A. Example crossover comparison between float 5904469 and GLODAP observations of oxygen, nitrate, and pH. From top left to bottom right: float and ship crossover locations; oxygen offsets (float minus ship) relative to GLODAP measurement times (blue x's) and float measurement times (red x's); oxygen offsets relative to float measurement times only (red x's); histograms of offsets for temperature, salinity, oxygen, oxygen with outliers removed (trimmed), nitrate, pH at 25C on the total scale; oxygen offset vs. float pressure; oxygen offset vs. GLODAP oxygen concentration, colored to indicate individual GLODAP measurements; oxygen offsets on T-S diagrams for float and Glodap data. Regression of oxygen offsets relative to GLODAP data (blue line and shading) indicate that the offset for this float is not likely due to true changes in ocean oxygen concentrations. Confidence intervals for the mean oxygen offset with outliers removed do not overlap with zero, indicating that this float oxygen is likely different from the GLODAP data.

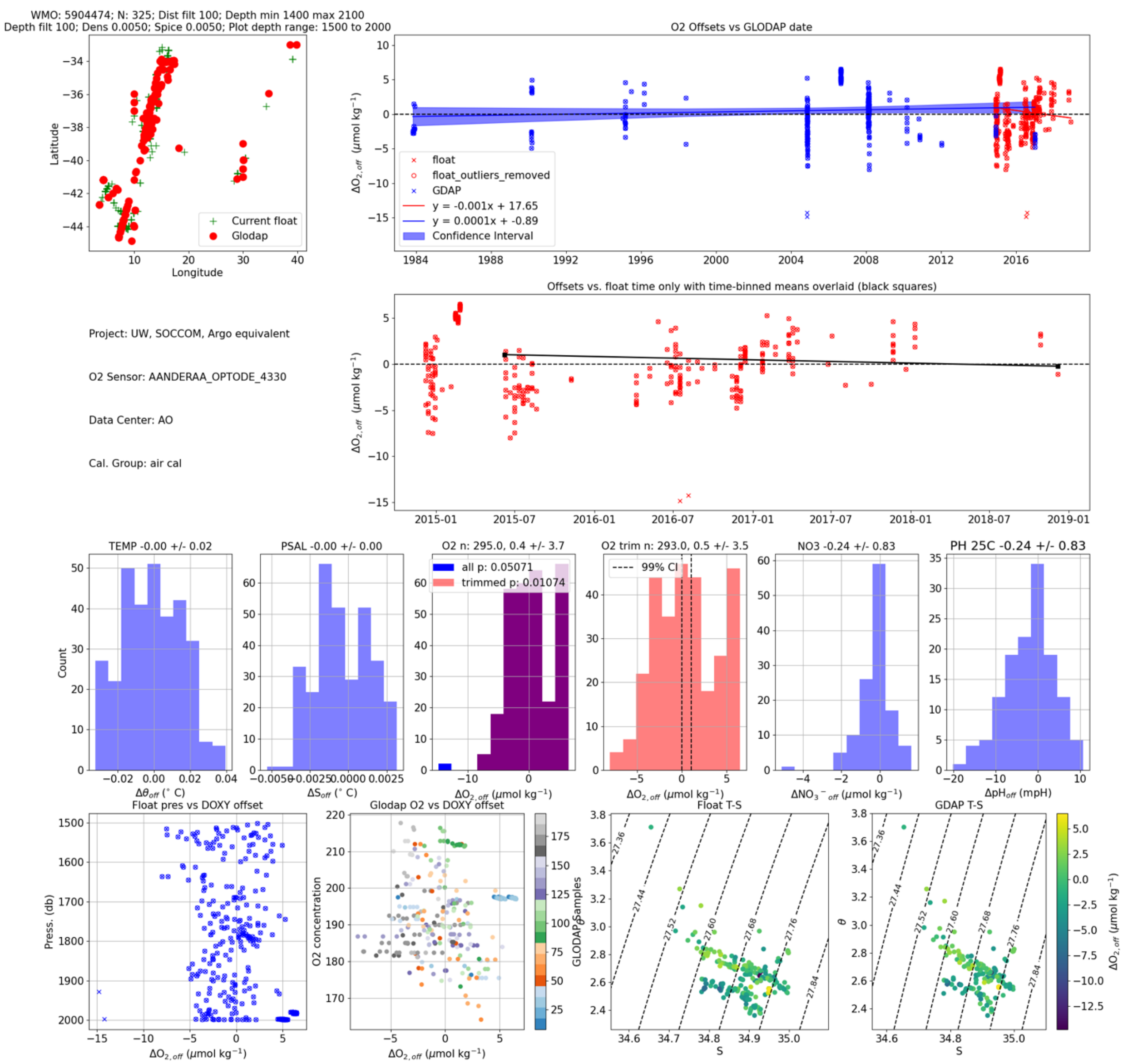

Figure S2B. Same as Figure S2A, but for float 5904474. Regression of oxygen offsets relative to GLODAP data (blue line and shading) indicate that the offset for this float may be due to true changes in ocean oxygen concentrations. Confidence intervals for the oxygen offset with outliers removed overlap with zero, indicating that this difference is not significant.

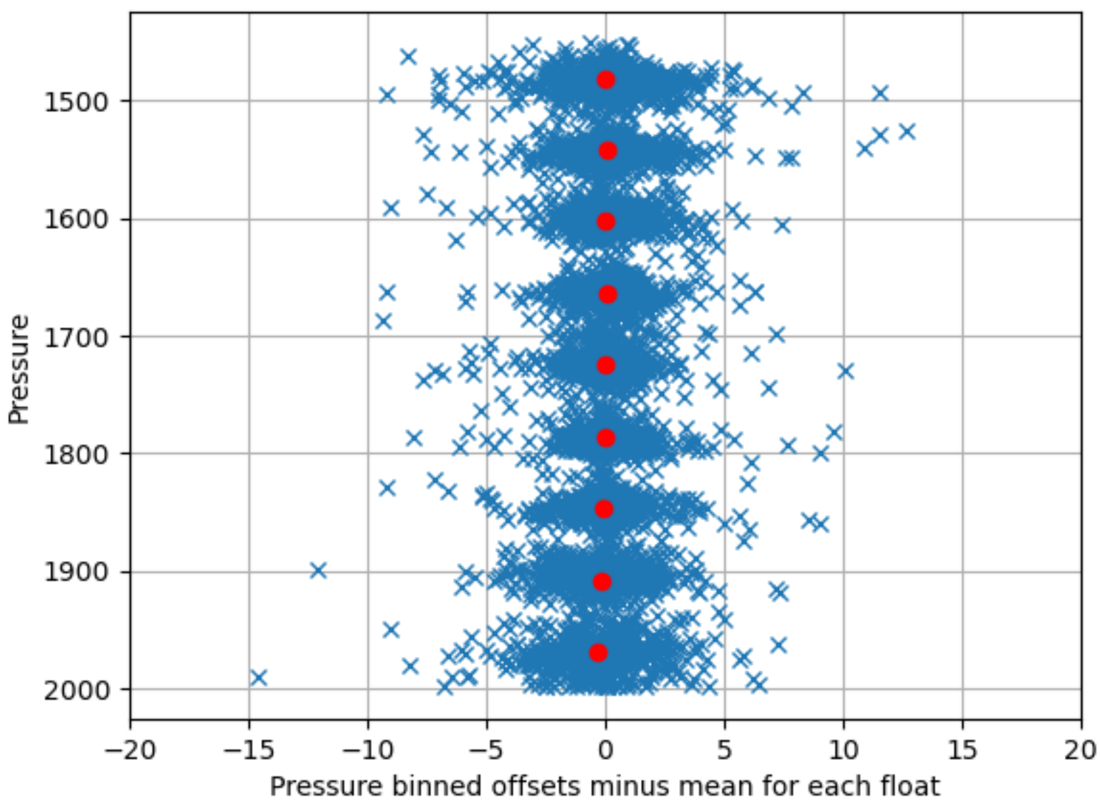

Figure S3. Float oxygen offsets (µmol kg^{-1}) binned by pressure (db), with mean float offset removed. Red dots indicate 50 db bin means.

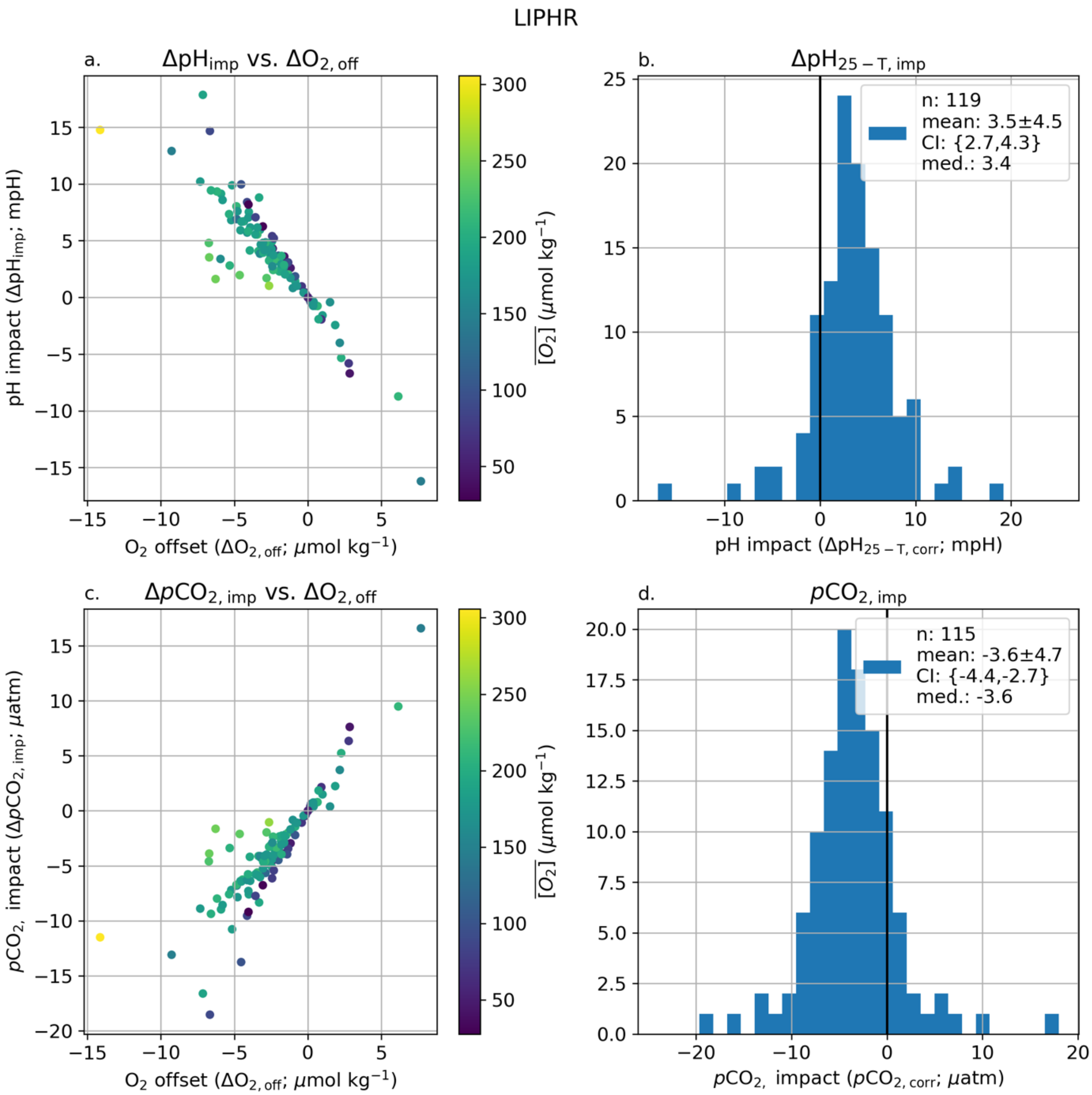

Figure S4. Equivalent to manuscript Figure 5 but using the LIPHR pH MLR algorithm instead of the ESPER mixed (neural network/MLR average). Data shown are for air calibrated floats only. Full statistics in Table S4. Figures S5 and S6 show results for all pH equipped floats.

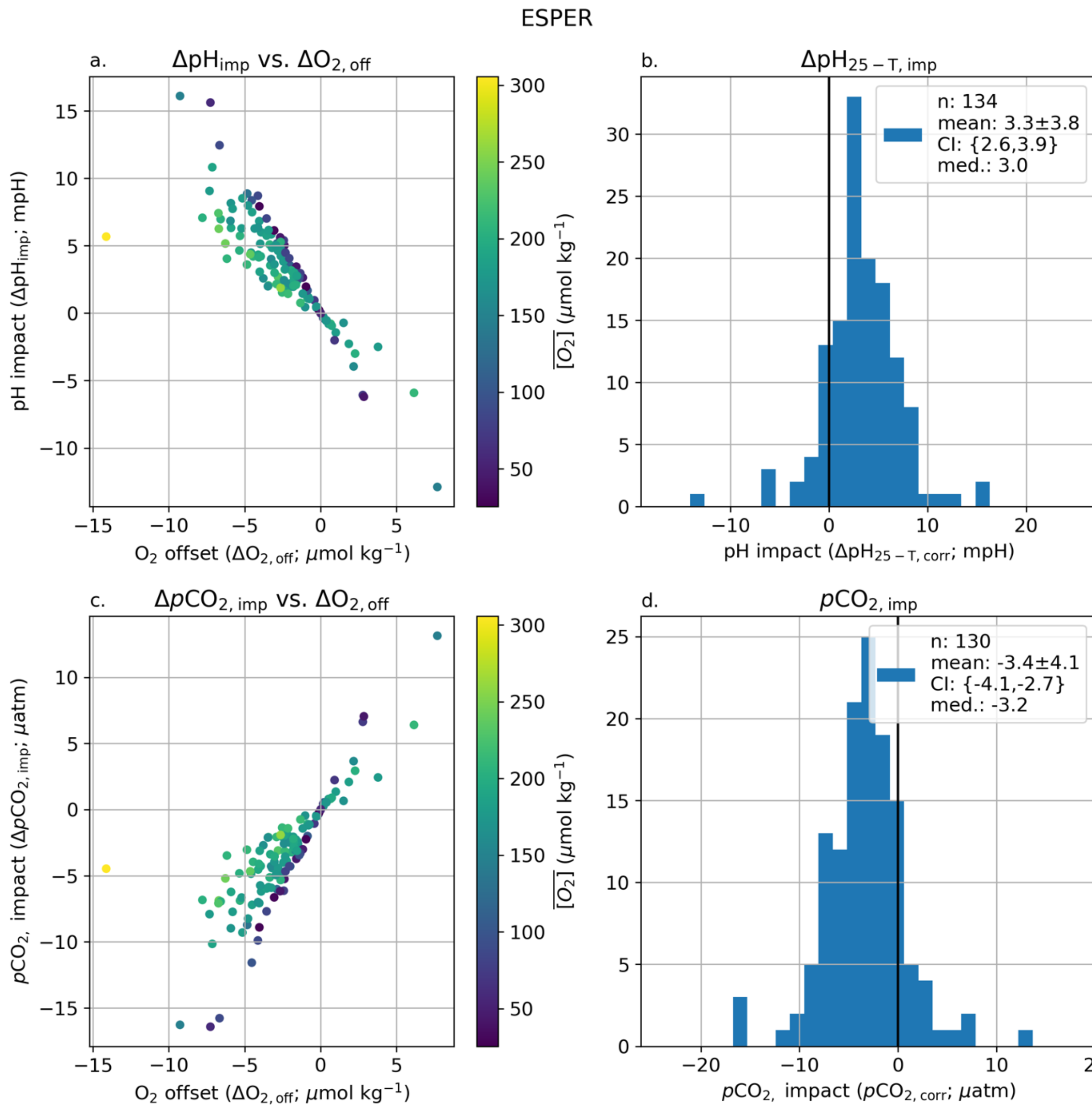

Figure S5. Impact scatter / histograms on pH and pCO_2 for all floats (not just air) – ESPER.

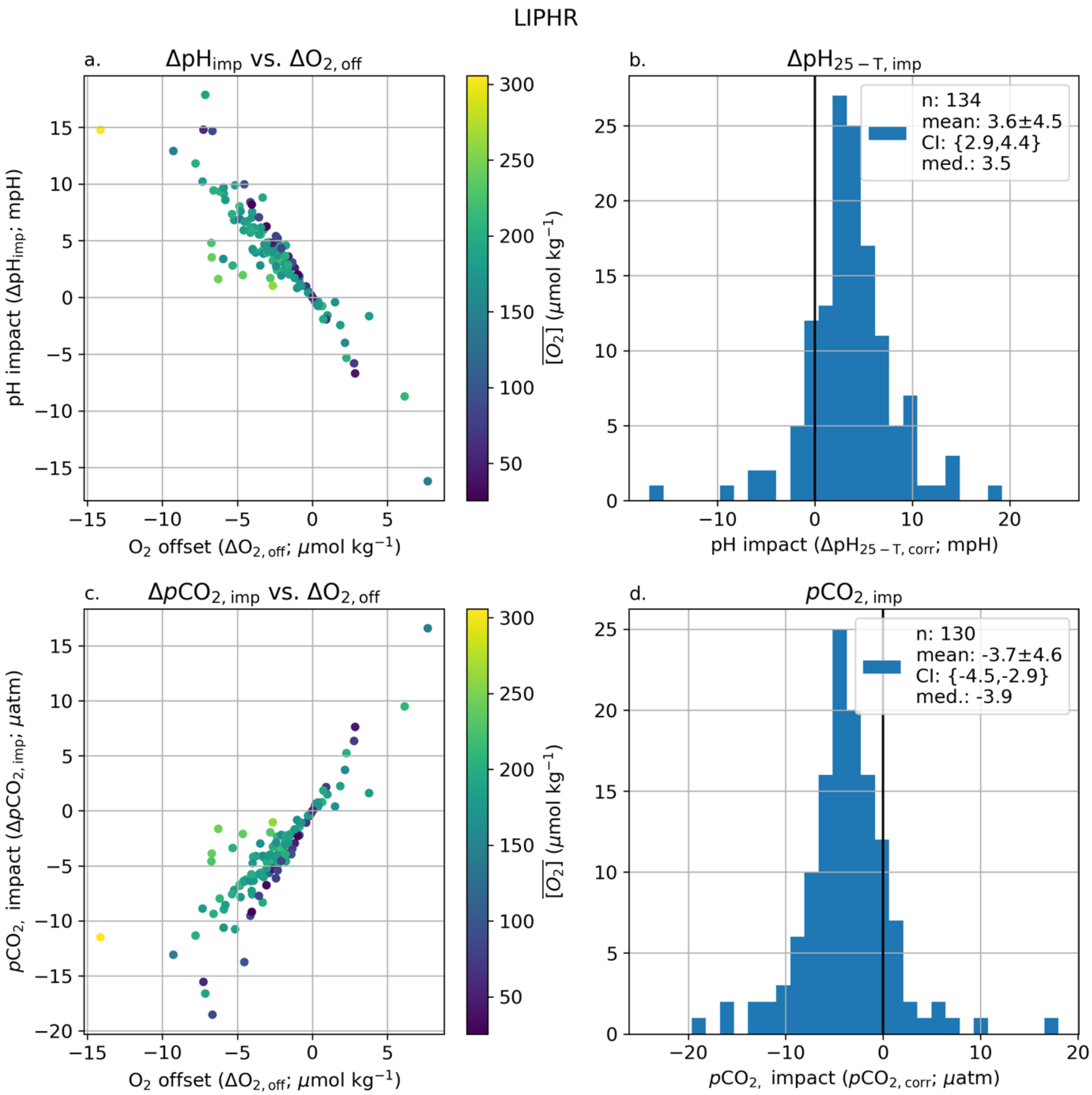

Figure S6. Impact scatter / histograms on pH and pCO_2 for all floats (not just air) – LIPHR.

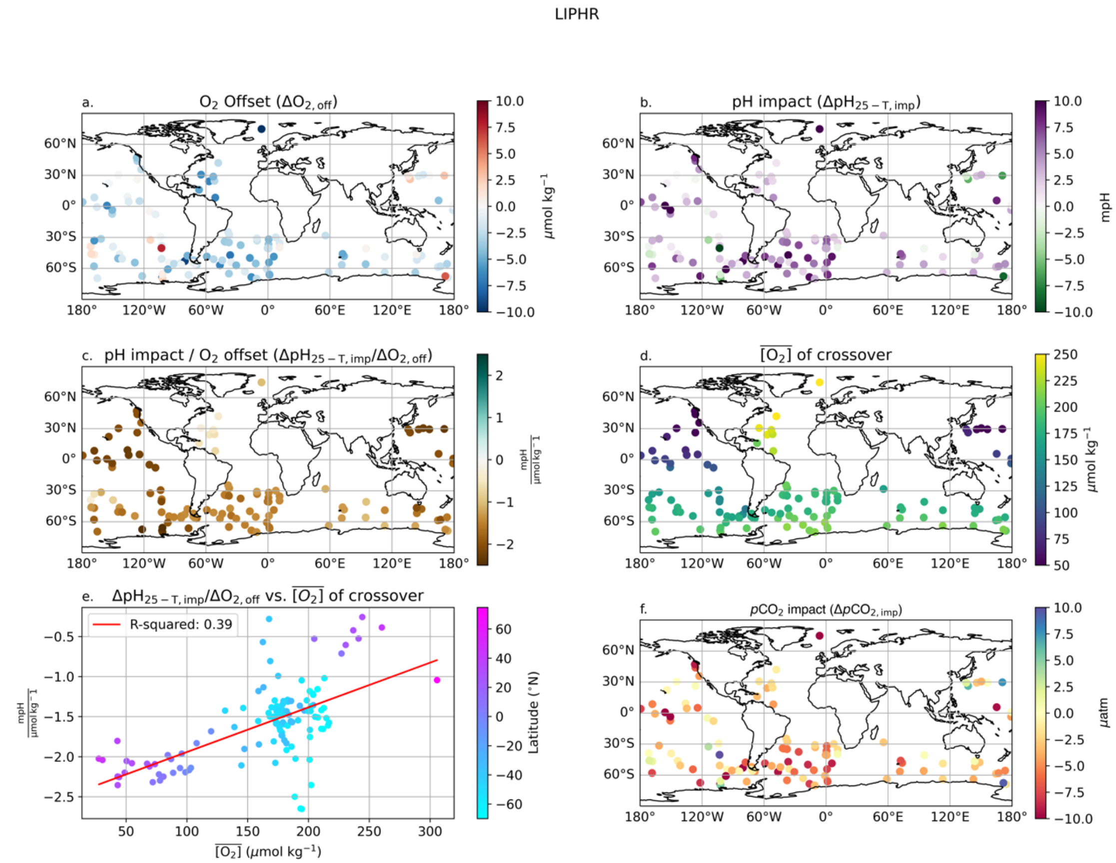

Figure S7. Same as manuscript figure 6 but using the LIPHR pH algorithm.

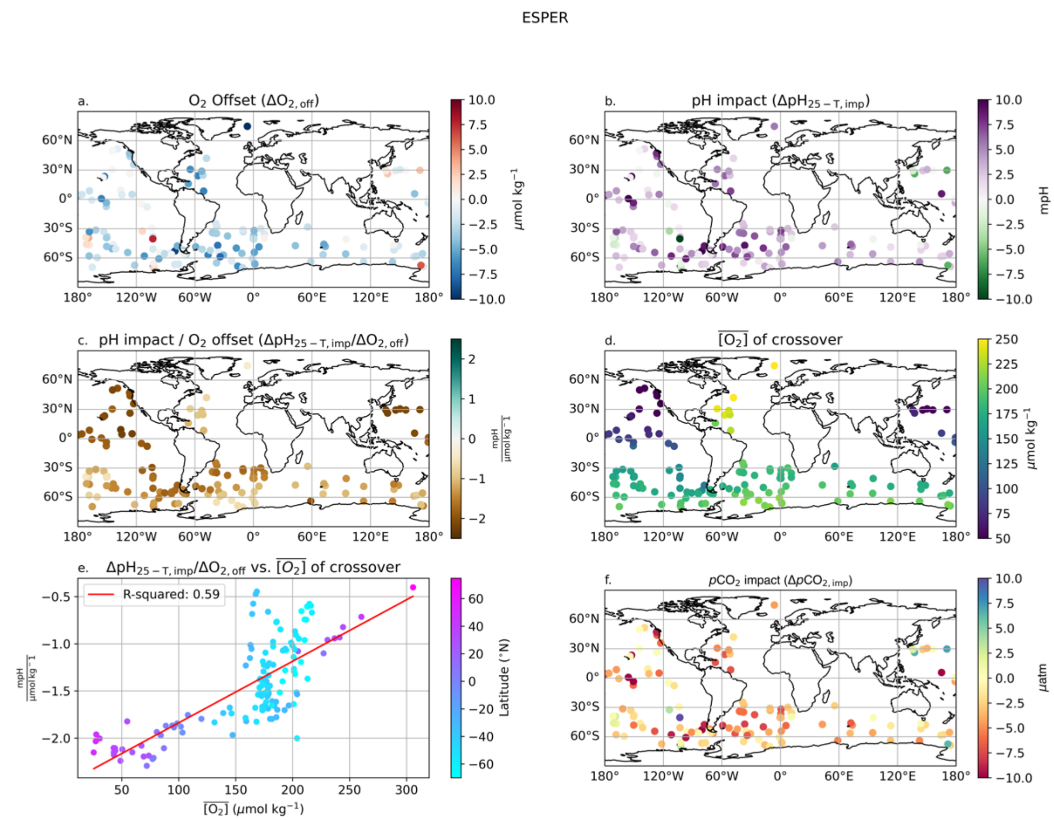

Figure S8. Same as manuscript figure 6, but showing all floats equipped with pH.

LIPHR

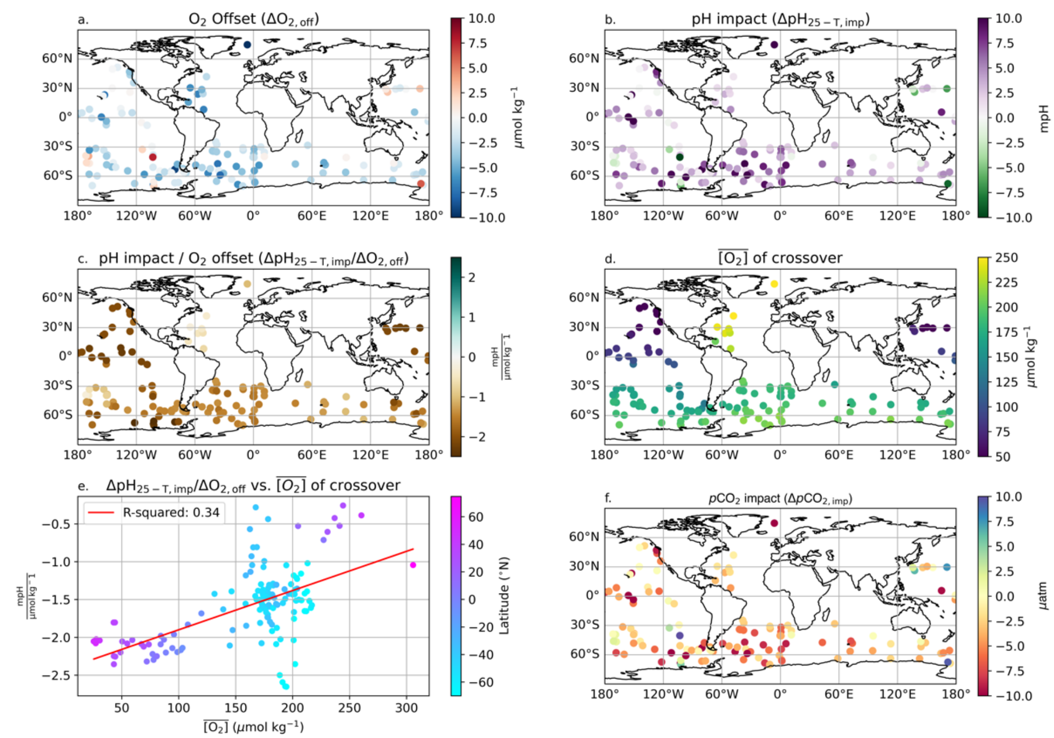

Figure S9. Same as manuscript figure 5, but showing all floats equipped with pH and using the LIPHR pH algorithm.

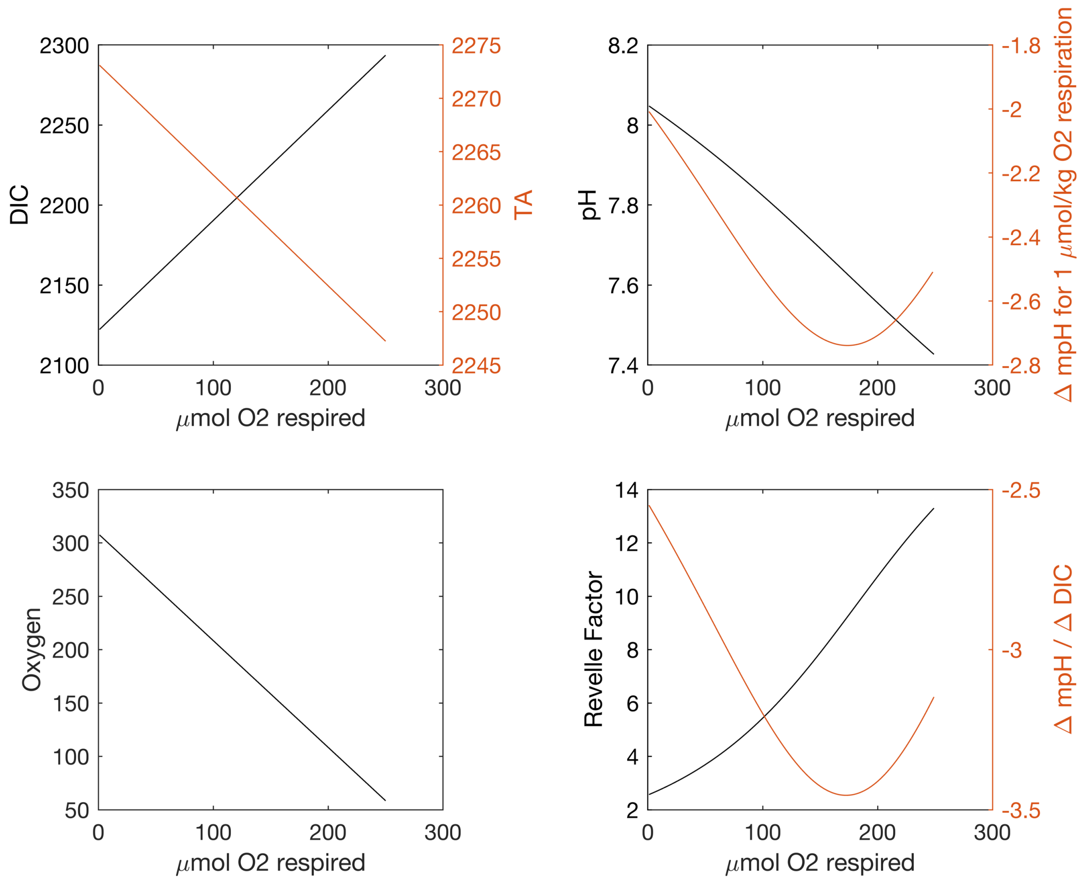

Figure S10. Example of changes in carbonate chemistry responses with accumulated respiration. Changes in DIC, TA, pH, ΔmpH/(Δμmol kg^{-1} O_2 respired), Revelle Factor, and ΔmpH/ΔDIC for a parcel of water starting with properties of newly formed Subantarctic Mode Water from the southeast Pacific. For mol of oxygen respired by the respiration of organic matter, each parameter is calculated to demonstrate the sensitivity of pH to a given change in oxygen.

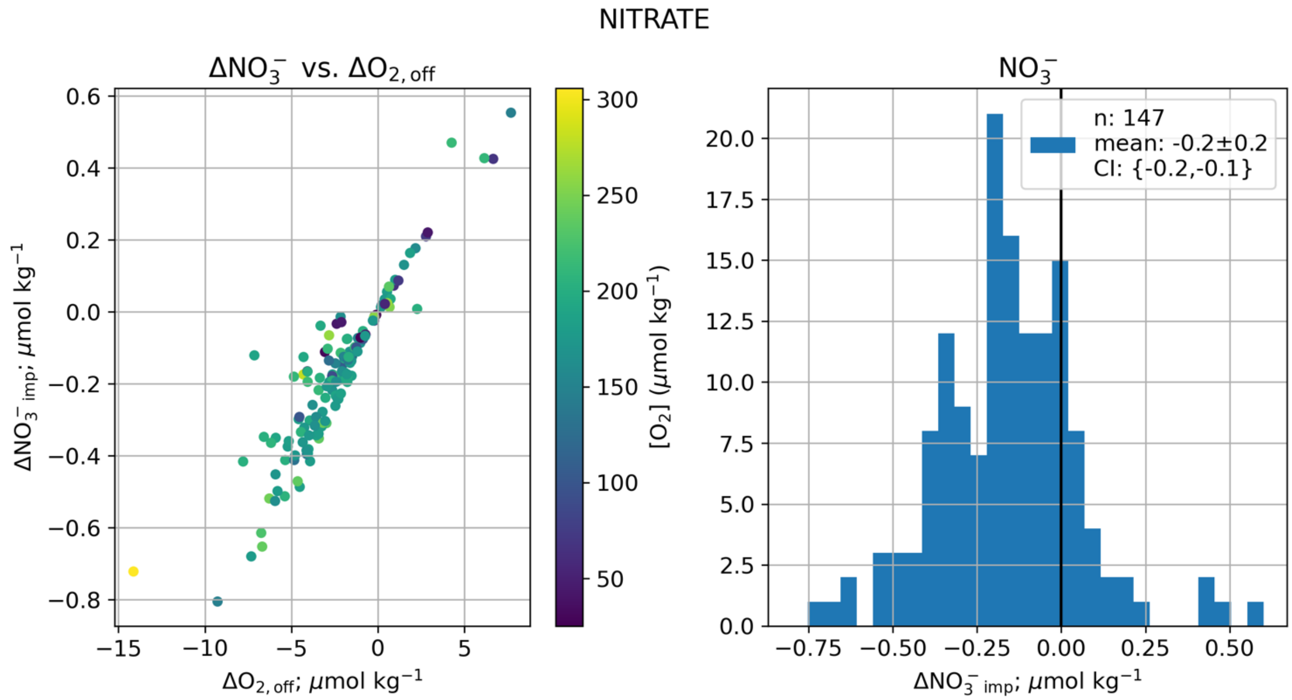

Figure S11. Impact of correcting O2 on nitrate.

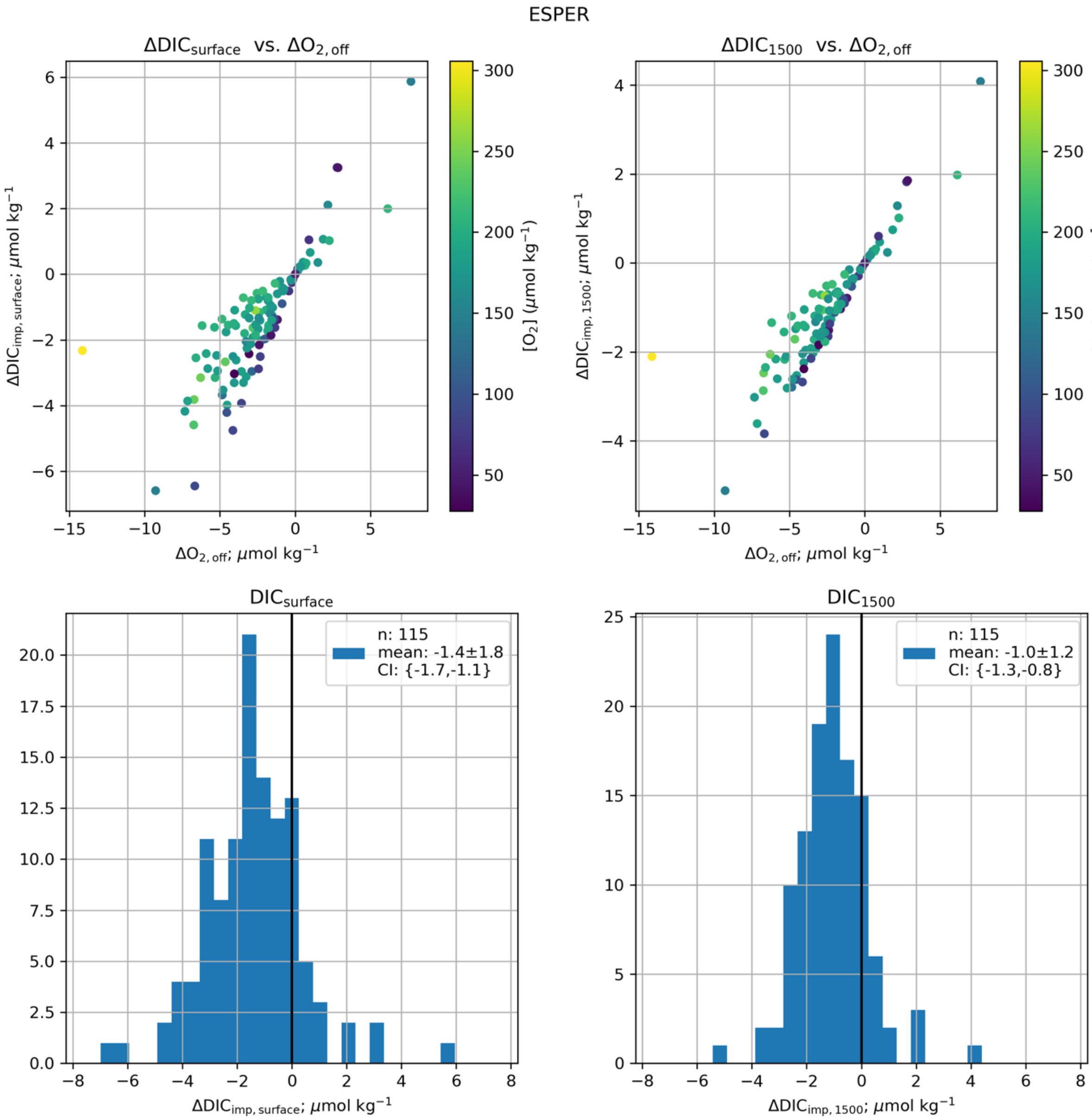

Figure S12. ESPER DIC impact. LIPHR DIC impact in Figure S13.

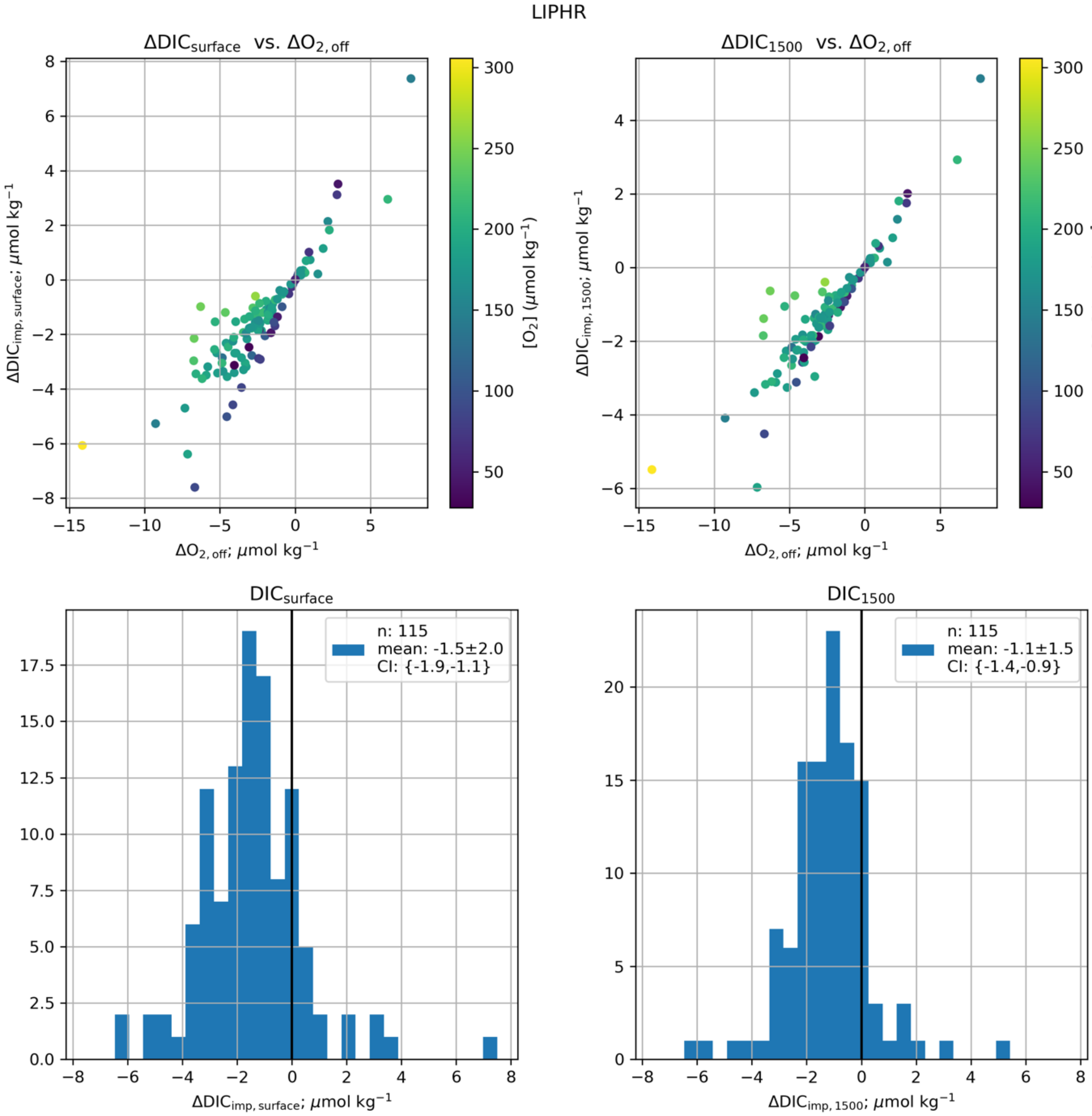

Figure S13. Impact of correcting O2 on DIC LPIHR

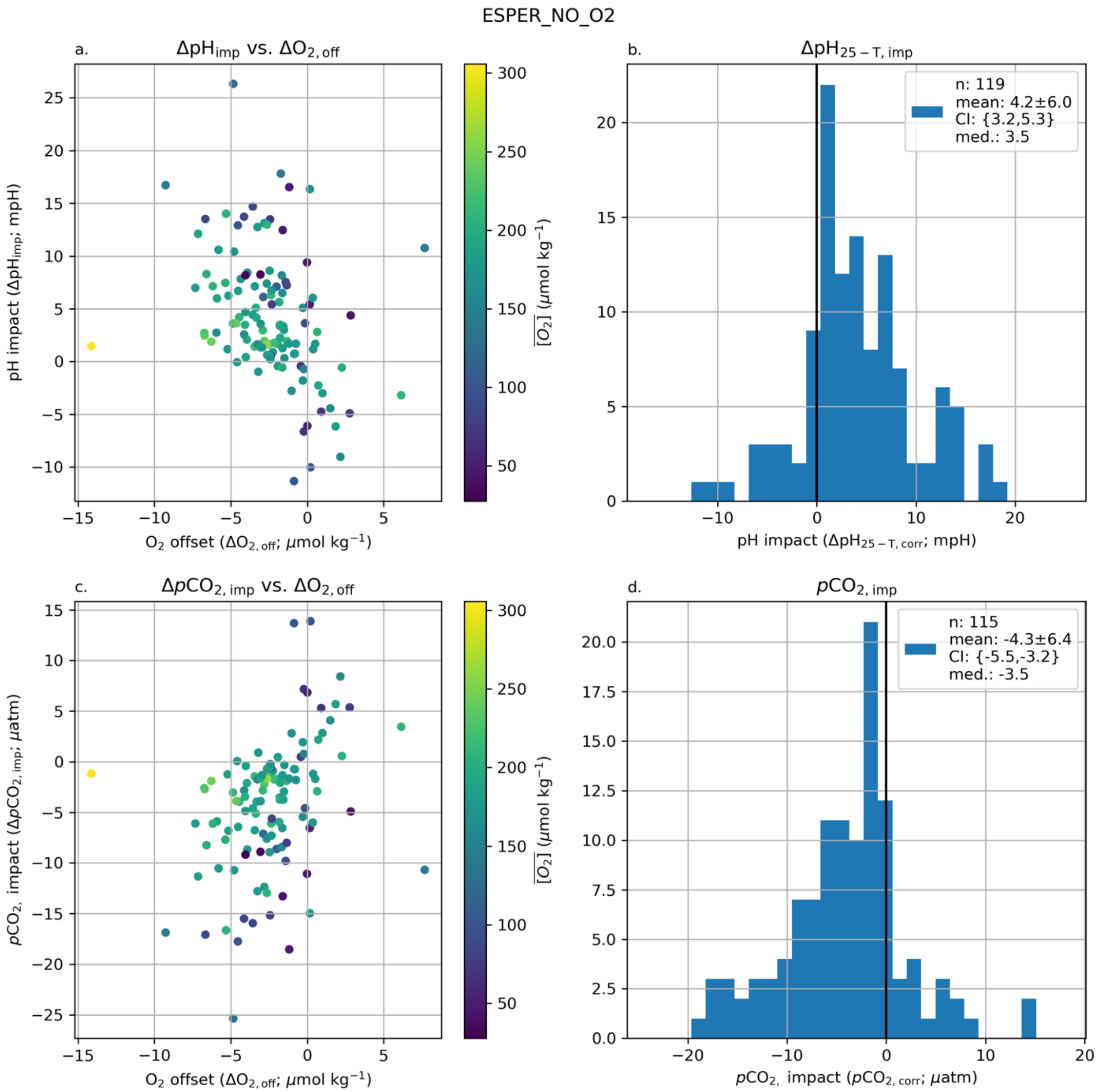

Figure S14. Impact of removing oxygen from the ESPER algorithm used to correct float pH.

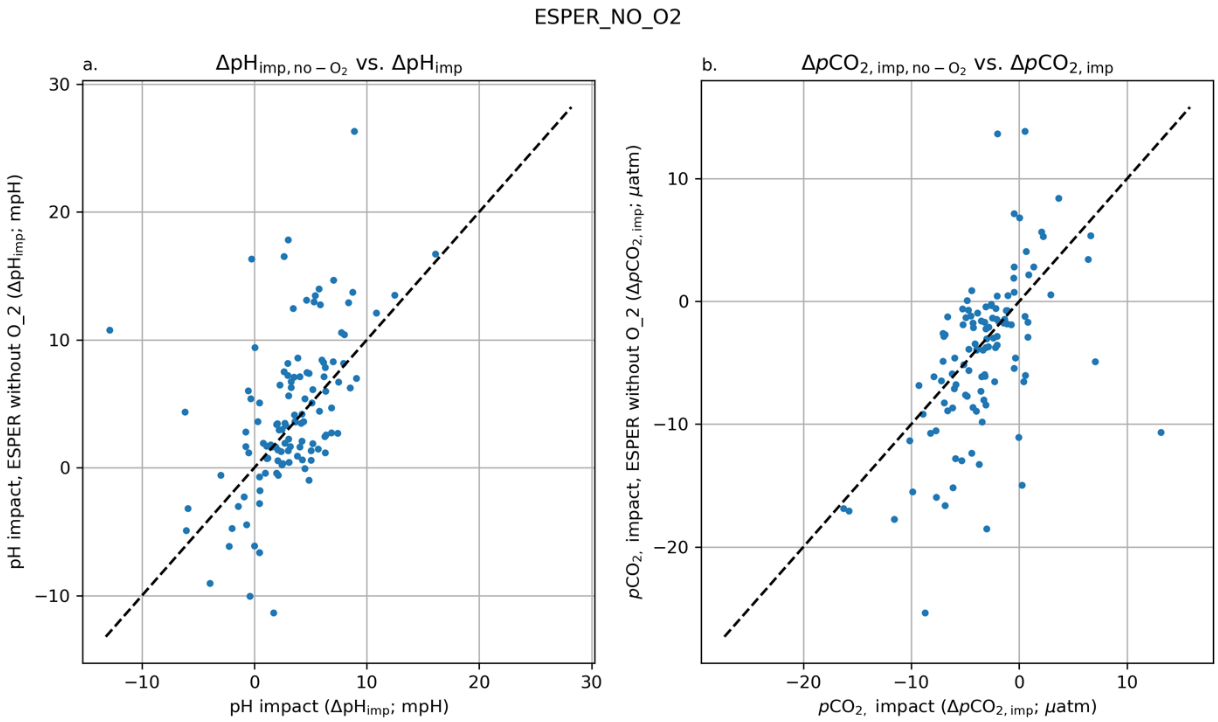

Figure S15. pH and pCO_2 impacts calculated from observed oxygen offsets relative to GLODAP (x axes) vs. those calculated from removing oxygen from the ESPER algorithm (y axes).

Table S1. Oxygen sensor calibration comments

Comment number	**Calibration Comment (**[SCIENTIFIC_CALIB_COMMENT] field in Sprof.nc files)	**Number of WMOs w/ Comment[1]**
1	[blank]	381
2	SVU Foil calibration coeficients were used. G determined from float measurements in air. See Johnson et al.,2015,doi:10.1175/JTECH-D-15-0101.1	253
3	Polynomial calibration coeficients were used. G determined by surface measurement comparison to World Ocean Atlas 2009.See Takeshita et al.2013,doi:10.1002/jgrc.20399	186
4	DOXY_ADJUSTED is computed from an adjustment of in water PSAT or PPOX float data at surface by comparison to woaPSAT climatology or woaPPOX{woaPSAT,floatTEMP,floatPSAL} at 1 atm, DOXY_ADJUSTED_ERROR is computed from a PPOX_ERROR of 10 mbar +1mb/year	96
5	DOXY_ADJUSTED is estimated from an adjustment of in water PPOX float data at surface by comparison to WOA2018 PPOX {woaPSAT, floatTEMP, floatPSAL} at 1 atm; DOXY_ADJUSTED_ERROR recomputed from a PPOX_DOXY_ERROR = 10 mbar increasing by 1 mbar per year	96
6	Bad data; not adjustable	94
7	Polynomial calibration coeficients were used. G determined from float measurements in air. See Johnson et al.,2015,doi:10.1175/JTECH-D-15-0101.1	86
8	DOXY_ADJUSTED is computed from an adjustment of in water PSAT or PPOX float data at surface by comparison to woaPSAT climatology or WOA PPOX in using woaPSAT and floatTEMP and PSAL at 1 atm, DOXY_ADJUSTED_ERROR is computed from a PPOX_ERROR of 10 mbar	73
9	none	72
10	DOXY_ADJUSTED corrected based on the WOA 2018 climatology as described in Johnson et al. (2015)	71
11	DOXY_ADJUSTED corrected using continuous in-air measurements as in Johnson et al. (2015)	69
12	Percent saturation corrected as a linear function of PSAT modified from Takeshita et al (2013); PSAT converted from DOXY and DOXY_ADJUSTED converted from PSAT_ADJUSTED; oxygen corrected as a linear function of DOXY by comparison to a climatology (CARS09)	60
13	no adjustment is performed because of issues in CTD	51

14	DOXY_ADJUSTED computed using Stern-Volmer equation with coeffs refit from foil calibration data & WOD data as in Drucker & Riser (2016). The quoted error was computed via comparisons with WOA09 data, interpolated to float location, depth, and season.	47
15	SVU Foil calibration coeficients were used. G determined by surface measurement comparison to World Ocean Atlas 2009.See Takeshita et al.2013,doi:10.1002/jgrc.20399	44
16	DOXY_ADJUSTED is computed from an adjustment of in water PSAT or PPOX float data at surface by comparison to woaPSAT climatology or woaPPOX{woaPSAT,floatTEMP,floatPSAL} at 1 atm, DOXY_ADJUSTED_ERROR is computed from a PPOX_ERROR of 10.0 mbar +1mb/year	37
17	optode multi calibration, adjusted with median of the annual WOA climato at surface	36
18	GAIN determined from WOA2013 O2sat along the five initial float cycles	35
19	DOXY_ADJUSTED is estimated from an adjustment of in water PSAT or PPOX float data at surface by comparison to WOA PSAT climatology or WOA PPOX in using PSATWOA and TEMP and PSALfloat at 1 atm, DOXY_ADJUSTED_ERROR is estimated from a PPOX_ERROR of 10 mbar.	34
20	No QC available for DOXY	34
21	DOXY_QCs are modified during visual check.	33
22	1-point multiplicative correction using WOD at 1800 dbar. The quoted error was computed via comparisons with monthly or annual climatology data, interpolated to float location, depth, and season, from WOA09.	26
23	optode multi calibration, adjusted with median of the monthly WOA climato at surface	26
24	RT adjustment by comparison of float surface O2sat to WOA2018 surface O2sat using automatically generated gain from MBARI (jplant@mbari.org)	24
25	DOXY_ADJUSTED is estimated from an adjustment of in air PPOX float data by comparison to NCEP reanalysis, DOXY_ADJUSTED_ERROR is recomputed from a PPOX_ERROR = 10 mbar	17
26	Data are unadjustable owing to dissolved oxygen sensor problem.	17
27	G determined by surface measurement comparison to World Ocean Atlas 2009. See Takeshita et al.2013,doi:10.1002/jgrc.20399	17
28	RT adjustment based on extrapolation from previous DMQC adjustment; DOXY adjusted by gain-only factor on DOXY	17

29	RT adjustment by comparison of float surface O2sat to WOA18 surface O2sat; DOXY adjusted by gain-only factor on DOXY	17
30	Adjustment via comparison of float surface O2sat to WOA2018 surface O2sat using automatically generated gain from MBARI (jplant@mbari.org). All profiles individually inspected and compared against WOA2018 for accuracy	16
31	Adjusted PPOX at zero for the minimum and estimate the gain with Monthly WOA2018 (OMZ)	14
32	No significant oxygen drift detected - Calibration error is manufacturer specified accuracy	13
33	Percent saturation corrected as a linear function of PSAT; Comparison to a single reference profile (isobaric match as in Takeshita et al. (2013)) on cycle 0; PSAT converted from DOXY and DOXY_ADJUSTED converted from PSAT_ADJUSTED	13
34	Percent saturation corrected as a linear function of PSAT; Comparison to a single reference profile (isobaric match as in Takeshita et al. (2013)) on cycle 1; PSAT converted from DOXY and DOXY_ADJUSTED converted from PSAT_ADJUSTED	13
35	not applicable	13
36	Adjusted on WOA monthly climatology at surface	12
37	Adjusted with SAGEO2 with in-air measurements (Johnson et al., 2015)	12
38	Partial pressure corrected as a linear function of PPOX using in-air measurements as in https://doi.org/10.3389/fmars.2021.683207, PPOX converted from DOXY and DOXY_ADJUSTED converted from PPOX_ADJUSTED, ERROR calculated as 2*std(SLOPE)*205mbar+2mbar	12
39	RT adjustment by comparison of float surface O2sat to WOA13 surface O2sat; DOXY adjusted by gain-only factor on DOXY	10
40-198	159 additional comments represented in 7 or fewer floats	267

[1]Total number of WMOs per comment is greater than the total number of floats analyzed for this study because comments are listed per profile and some floats have multiple comments included throughout their deployment.

Table S2. Statistics for oxygen offsets between 1450 and 2000 db shown in Figure 3.

Float cal. type	Count	Mean	SD	p_value	95.0% CI low	95.0% CI high	Median	Min.	Max.
All	540	-1.9	4.7	<0.001	-2.3	-1.5	-1.8	-63.9	15.0
No air cal.	253	-0.9	3.6	<0.001	-1.4	-0.5	-0.7	-11.2	11.0
Air cal.	255	-3.1	5.3	<0.001	-3.7	-2.4	-2.6	-63.9	10.5
No cal.	32	-0.6	5.0	0.493	-2.4	1.1	-0.7	-13.5	15.0
pH - No cal.	22	-2.6	2.8	<0.001	-3.7	-1.4	-2.2	-7.8	3.8
pH - Air cal.	167	-2.5	2.7	<0.001	-2.9	-2.1	-2.4	-14.1	7.7

All columns other than count and p value have units of µmol kg^{-1}.

Table S3. Statistics for Nitrate, pH 25C total, and DIC crossovers between 1450 and 2000 db.

Parameter	count	mean	std	p_value	95.0% CI low	95.0% CI high	median	min	max
NITRATE_ADJUSTED	232	0.2	0.4	<0.001	0.1	0.2	0.2	-2.2	1.8
pH_25C_TOTAL_ADJUSTED	113	-4.5	7.7	<0.001	-5.9	-3.1	-3.3	-40.0	13.9
DIC	154	2.9	6.4	<0.001	1.9	4.0	2.3	-16.1	36.6

All columns other than count and p value have units of µmol kg^{-1} or mpH.

Table S4. Statistics for Nitrate, pH 25C total, and DIC crossovers between 1450 and 2000 db after oxygen correction impacts have been applied.

Parameter	count	mean	std	p_value	95.0% CI low	95.0% CI high	median	min	max
NITRATE_ADJUSTED	107	-0.0	0.4	0.501	-0.1	0.0	-0.0	-2.2	1.4
pH_25C_TOTAL_ADJUSTED	79	-1.3	5.7	0.049	-2.6	-0.0	-1.3	-13.2	14.4
DIC	106	1.0	6.3	0.103	-0.2	2.2	0.7	-22.5	31.7

All columns other than count and p value have units of µmol kg^{-1} or mpH.

Table S5. Stats for pH and pCO_2 impact LIPHR:

	Count	Mean	SD	p_value	95.0% CI low	95.0% CI high	median	min	max
pH_impact	119	3.5	4.5	<0.001	2.7	4.3	3.4	-16.2	17.9
pCO2_impact	115	-3.6	4.7	<0.001	-4.4	-2.7	-3.6	-18.5	16.6

Table S6. Stats for NO3 impact between 1450 and 2000 db.

Parameter	count	mean	SD	p_value	95.0% CI low	95.0% CI high	median	min	max
no3_impact	110	-0.2	0.2	<0.001	-0.22	-0.14	-0.2	-0.8	0.6

Table S7. Stats for DIC impact ESPER and LIPHR

ESPER

Parameter	count	mean	SD	p_value	95.0% CI low	95.0% CI high	median	min	max
dic_impact_surf	115	-1.4	1.8	<0.001	-1.73	-1.08	-1.4	-6.6	5.9
dic_impact_1500	115	-1.0	1.2	<0.001	-1.27	-0.81	-1.0	-5.1	4.1

LIPHR

Parameter	count	mean	SD	p_value	95.0% CI low	95.0% CI high	median	min	max
dic_impact_surf	115	-1.5	2.0	<0.001	-1.86	-1.14	-1.5	-7.6	7.4
dic_impact_1500	115	-1.1	1.5	<0.001	-1.42	-0.88	-1.1	-6.0	5.1

Carter, B. R., Talley, L. D., & Dickson, A. G. (2014). Mixing and remineralization in waters detrained from the surface into Subantarctic Mode Water and Antarctic Intermediate Water in the southeastern Pacific. *Journal of Geophysical Research: Oceans*, *119*(6), 4001–4028. https://doi.org/10.1002/2013JC009355

Carter, B. R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, Y.-Y., et al. (2021). New and updated global empirical seawater property estimation routines. *Limnology and Oceanography: Methods*, *19*(12), 785–809. https://doi.org/10.1002/lom3.10461