



OPEN A systematic bias in float pH leads to overestimation of derived $p\text{CO}_2$ and underestimation of carbon uptake by the Southern Ocean

Chuqing Zhang^{1✉}, Yingxu Wu², Peter J. Brown³, David Stappard¹, Amavi N. Silva^{1,4} & Toby Tyrrell¹

The carbon flux estimated from biogeochemical Argo float data indicates a lower annual carbon uptake by the Southern Ocean compared to fluxes derived from other observations (e.g., ship and aircraft measurements). The root cause of this discrepancy remains controversial, with growing evidence suggesting that potential biases in float-derived $p\text{CO}_2$ may be a plausible explanation. Here, we perform a multi-variable comparison of vertical profiles between float- and ship based-data and reveal consistent discrepancies in pH, $p\text{CO}_2$ and dissolved inorganic carbon, which are not found in other variables such as dissolved oxygen, nitrate and total alkalinity. Our findings are consistent with a previously unrecognized negative bias in float pH driving a positive offset in float-derived $p\text{CO}_2$. The float-derived surface $p\text{CO}_2$ is, on average, biased high by $15 \pm 3 \mu\text{atm}$ compared to ship data, representing a larger magnitude of bias than previously recognized. Biases exist in both surface and deep waters, including old deep waters containing minimal anthropogenic carbon. A more sophisticated adjustment for float pH, involving multiple cross-reference depths, may be required for accurate estimation of air-sea CO_2 exchange in the Southern Ocean.

A biogeochemical (BGC) Argo float is an advanced autonomous monitoring platform, equipped with additional sensors (extra to those on standard Argo floats) to measure biogeochemical variables^{1–3} including pH, dissolved oxygen (O_2), and nitrate (NO_3). Under normal operating conditions it will collect data on vertical profiles (0–2000 m), carried out once every 5–10 days⁴. Floats have garnered significant interest for their autonomous sampling capabilities, high-frequency of sampling, and ability to observe remote regions. The Southern Ocean (SO) has historically been one of the least observed ocean basins because its environment is harsh, especially in winter, presenting difficulties for ship-based sampling⁵. However, it plays a critical role in the global heat and carbon cycles as well as the Earth's climate⁶; it accounts for ~40% of heat uptake and ~44% of anthropogenic carbon uptake by the global ocean⁷. To compensate for this, the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) project had deployed over 260 BGC Argo floats in the SO by the end of 2023 to enhance observational capability^{8,9}.

Accurate calculation of air–sea carbon dioxide (CO_2) fluxes has profound implications for understanding the carbon cycle and the ocean carbon sink¹⁰. The CO_2 flux is typically estimated as the product of the gas transfer velocity, the solubility of CO_2 in seawater, and the difference in partial pressure of CO_2 ($p\text{CO}_2$) between the seawater and atmosphere. Previous calculations of CO_2 fluxes were based on high-quality ship data and model estimates^{11–13}. The sparse ship observations in the SO result in larger uncertainties in air-sea CO_2 fluxes than in other ocean basins^{14,15}. The BGC Argo floats deployed by SOCCOM have generated a substantial volume of data with broad spatial and temporal coverage, with great potential for improved estimation of the SO CO_2 flux. Float-based $p\text{CO}_2$ in seawater is calculated from a combination of measured pH and algorithm-estimated total alkalinity (TA)^{16–19}, therefore introducing multiple sources of uncertainty¹⁶.

Float-based CO_2 annual fluxes indicate a considerably lower carbon uptake by the SO than those obtained through other approaches (such as from ship-based measurements or model estimations), thereby challenging conventional views^{18,20}. This discrepancy has been attributed to a scarcity of ship data during winter—when

¹School of Ocean and Earth Science, University of Southampton, National Oceanography Centre Southampton, Southampton, UK. ²Polar and Marine Research Institute, Jimei University, Xiamen, China. ³National Oceanography Centre, Southampton, UK. ⁴GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany. ✉email: Chuqing.Zhang@soton.ac.uk

high-latitude $p\text{CO}_2$ values peak due to intensive upwelling and deep water entrainment^{18,20}. However, two subsequent independent studies have questioned this explanation: Winter outgassing estimated by extrapolation from summertime ship observations is markedly lower than the fluxes derived from float observations²¹. Moreover, estimation using data collected from aircraft of the CO_2 flux south of 45°S suggest an annual ingassing to the ocean of $0.53 \pm 0.23 \text{ Pg C yr}^{-1}$ ²². This aircraft-derived estimate aligns closely with ship-based results^{22–24}, whereas the float-based estimate is an annual outgassing of 0.3 Pg C yr^{-1} . Consequently, the float and aircraft data yield contradictory conclusions of the total annual air–sea CO_2 flux in the SO, with float data indicating net outgassing and aircraft data suggesting net ingassing^{20,22,25}.

One potential explanation for the discrepancy in estimating SO carbon fluxes is that the float pH and derived $p\text{CO}_2$ data may be systematically offset from their true values. Over the past few years, the possibility of such a bias has become a topic of ongoing debate. For instance, a crossover analysis with the Surface Ocean CO_2 Atlas (SOCAT) data reported an average positive bias of $3.6 \mu\text{atm}$ in surface float $p\text{CO}_2$ ¹⁸. Similarly, a meta-analysis of CO_2 and O_2 concentrations identified offsets that resulted in float-derived $p\text{CO}_2$ values being biased high by $\sim 5.8 \mu\text{atm}$ at the surface²⁶. In the Drake Passage, a mean surface $p\text{CO}_2$ discrepancy of $14 \mu\text{atm}$ has been observed, with float $p\text{CO}_2$ being consistently higher than ship observations, a magnitude that cannot be fully attributed to upwelling effects²⁷. Moreover, long-term mooring data from the subantarctic region found that float $p\text{CO}_2$ was $3\text{--}7 \mu\text{atm}$ higher than observed values²⁸. Critically, the current float pH adjustment method has also been questioned, with suggestions that the correction should be pH-dependent rather than a fixed offset²⁹. Incorporating a second reference point closer to the surface in the float pH correction process has been proposed to improve the accuracy of float pH measurements³⁰. Other studies have similarly reported varying degrees of bias in float-derived $p\text{CO}_2$: float-to-reference $p\text{CO}_2$ offsets were reported to span from $-4.6 \pm 4.1 \mu\text{atm}$ to $+8.6 \pm 4.0 \mu\text{atm}$ in subtropical cruises³¹; $+2.6 \pm 12.8 \mu\text{atm}$ during a South Georgia bloom³²; from $+4$ to $+9.5 \mu\text{atm}$ in Subantarctic Mode Water formation regions³³; more than $+11 \mu\text{atm}$ compared with a machine learning reconstruction³⁴ (positive values indicate that the float-derived $p\text{CO}_2$ is higher than the reference, while negative values indicate it is lower).

Previous investigations of the accuracy of float-derived $p\text{CO}_2$ have been hampered by several limitations: the comparisons (i) primarily focused on localized regions, raising concerns about their applicability to the entire SO, (ii) solely concentrated on surface data, thereby neglecting potential biases in subsurface $p\text{CO}_2$ and, (iii) were restricted to instances when a ship and a float are co-located in time and space—typically immediately after float deployment—resulting in the use of less than 1% of the available data. Additionally, (iv) the evaluations have only considered $p\text{CO}_2$ without concurrently assessing the accuracy of other related variables. To address these shortcomings and provide a comprehensive evaluation of float pH and derived $p\text{CO}_2$, we carried out a basin-scale, quantitative, multi-variable, whole-profile comparison between float and ship data across the SO (south of 35°S), using the Global Ocean Data Analysis Project (GLODAP) as reference. GLODAP provides high-quality, internally consistent measurements and is widely used as a benchmark for carbonate system data. While numerous efforts have examined float carbonate system data quality, no consensus has been reached regarding the cause of observed discrepancies. This study seeks to address several unresolved questions: (i) Do float carbonate system data exhibit significant systematic biases? (ii) Are these biases limited to the surface ocean or do they extend to the subsurface? (iii) If present, what is the magnitude of these biases? (iv) What recommendations are proposed for future corrections of float carbonate system data? The results presented here provide new insights into these questions.

Methods

The Southern Ocean is defined as south of 35°S in this study.

Dataset descriptions

Float data

The BGC Argo float data were downloaded from <https://library.ucsd.edu/dc/object/bb51655605>. This snapshot contains all data before 2023-12-20 from SOCCOM floats, GO-BGC (Global Ocean Biogeochemistry Array) floats and also from additional University of Washington/MBARI floats³⁵. The data were subjected to the delayed-mode quality control (DMQC) at MBARI following Maurer et al.³⁶. The float variables that are relevant to this study are: pH (in-situ temperature and pressure, on the total scale)/(25 °C, 0 dbar, on the total scale), $p\text{CO}_2$, dissolved inorganic carbon (DIC), O_2 , NO_3 , in situ temperature (T), practical salinity (S) and TA. Only the data for which quality flags are good (0) for all of these variables were used. 10,003 profiles from 176 floats in the SO were chosen for final analysis. In the following, all pH data are reported on the total scale; temperature and pressure conditions (in-situ or 25 °C, 0 dbar) are used to distinguish them.

The GO-BGC adjustment procedure for float pH data is based on crossover analysis in deep waters between float and shipboard measurements³⁷. Firstly, an empirical algorithm is developed for estimating $\text{pH}_{\text{in-situ}}$ (as a function of T, S, P, and O_2) from shipboard bottle measurements at depths of 1000–2000 m. Then this algorithm is applied to float T, S, P and O_2 in order to generate an expected float $\text{pH}_{\text{in-situ}}$ value at 1500 m depth. Finally, the measured float $\text{pH}_{\text{in-situ}}$ and the expected float $\text{pH}_{\text{in-situ}}$ are compared at 1500 m depth and the resulting $\text{pH}_{\text{in-situ}}$ offset is applied to adjust the entire float profile. This correction approach assumes that the primary float $\text{pH}_{\text{in-situ}}$ bias arises from a pressure-independent offset in the sensor's reference potential, resulting in a consistent error across depths, as supported by previous studies on sensor behavior and deep-ocean stability^{16,36,38}.

An additional adjustment¹⁶ is applied when calculating $p\text{CO}_2$ from float pH (Eq. 1) in following order: (1) float measured $\text{pH}_{\text{in-situ}}$ is converted to $\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$ using CO2SYS with LIAR-estimated TA and in-situ temperature, salinity and pressure, following the standard SOCCOM processing³⁵. (2) Using Eq. (1) and the 1500 m depth $\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$ a correction is calculated for each profile. (3) This is added to each pH value in the profile before finally calculating $p\text{CO}_2$. This empirical, pH-dependent correction which is applied for the

calculation of $p\text{CO}_2$ (but not for other carbon variables), is intended to compensate for the systematic bias in $p\text{CO}_2$ calculated from float pH and LIAR-estimated TA relative to $p\text{CO}_2$ derived from shipboard DIC and TA (which is consistent with direct $p\text{CO}_2$ measurements), a bias that arises from internal inconsistencies in inorganic carbon thermodynamics¹⁶.

$$pH_{\text{adjustment}} = -0.034529 \times pH_{25^\circ\text{C}, 0\text{dbar}} + 0.26709 \quad (1)$$

The float $p\text{CO}_2$ was converted from measured-adjusted pH and locally interpolated alkalinity regression (LIAR) algorithm-estimated TA. The LIAR algorithm has a reported uncertainty¹⁹ of $5.6 \mu\text{mol kg}^{-1}$. The conversion from $\text{pH}_{\text{in-situ}}$ and TA to float $p\text{CO}_2$ was calculated using the Matlab version for CO2SYSv3³⁹, based on the following assumptions: (1) pH is on the total scale. (2) K_1 and K_2 dissociation constants were taken from Lueker et al.⁴⁰. (3) The KSO_4 dissociation constant was taken from Dickson⁴¹. (4) The borate to salinity ratio was taken from Lee et al.⁴². (5) The KF dissociation constant was taken from Perez and Fraga⁴³. Silicate and phosphate concentrations were not measured by the floats; however, estimates based on Redfield ratios⁴⁴ improved the carbonate system calculations. If a nitrate value was considered to be of good quality, then silicate = nitrate \times 2.5 and phosphate = nitrate/16; otherwise the best estimate for both was considered to be 0, following standard SOCCOM processing³⁵.

Williams et al. identified a large number of potential sources of float $p\text{CO}_2$ uncertainty, which they grouped into 3 categories¹⁶: (1) uncertainties in pH sensors, (2) uncertainties in derived alkalinity estimates, and (3) uncertainties in carbonate system equilibrium constants, with category (1) thought to have the greatest contribution. The uncertainty of each individual float $p\text{CO}_2$ value was estimated to be 2.7% (11 μatm at $p\text{CO}_2$ of 400 μatm)¹⁶.

Ship data (GLODAP)

Ship data (GLODAP) were obtained from <https://www.glodap.info> (last access: 05 February 2024). The GLODAPv2.2023 dataset contains almost 1.4 million internally-consistent samples of biogeochemical variables, globally collected on 1,108 cruises during the period 1972–2023⁴⁵. The ship data variables that are relevant to this study are: $\text{pH}_{\text{in-situ}}$, $\text{pH}_{25^\circ\text{C}, 0\text{dbar}}$, DIC, O_2 , NO_3 , T, S, silicate, phosphate, and TA. Only the sampling points where all the secondary quality flags for these variables are good (flag value of 1) were selected. 2,862 profiles in the SO containing 42,691 sampling points were included in the final dataset.

Ship $p\text{CO}_2$ was calculated from GLODAP's DIC and TA data, typically directly measured from samples taken from Niskin bottles⁴⁶. The procedure for converting ship DIC and TA to ship $p\text{CO}_2$ also employed the Matlab version of CO2SYSv3³⁹, with the same assumptions as above (i.e., for float $p\text{CO}_2$)^{40–43}. Silicate and phosphate were taken from direct measurements. The uncertainty in each individual ship $p\text{CO}_2$ value is reported to be $\sim 12 \mu\text{atm}$ at $p\text{CO}_2$ of 400 μatm ³⁴.

Gridded data (OceanSODA-ETHZ)

OceanSODA-ETHZ (referred as 'OceanSODA' in this manuscript) is a global gridded data set of the surface ocean carbonate system. This dataset is at a monthly resolution over the period 1985 through 2022 at a spatial resolution of $1^\circ \times 1^\circ$, containing multiple variables in the surface ocean, i.e., DIC, TA, $p\text{CO}_2$, $\text{pH}_{\text{in-situ}}$, and the saturation state with respect to calcium carbonate (CaCO_3). Its $p\text{CO}_2$ is based mainly on data from the Surface Ocean CO_2 Atlas (SOCAT) data^{23,47,48}, with uncertainty of 14 μatm ⁴⁹.

Matched float-ship data

This research focuses on comparing data along the entire vertical profile (0–2000 m, the sampling depth range of floats); subsurface water masses should exhibit little or no seasonality and minimal interannual variation, therefore subsurface matches between float and ship data were established based on depth and distance, disregarding time. The distance match window was 25 km; if the float profile was within 25 km of the ship profile, the two profiles were matched. If multiple ship CTD (Conductivity-Temperature-Depth) casts were matched to a single float profile, only the spatially closest cast was selected. The depth-based matching was carried out as follows: A 10 m depth match window was used for depths shallower than 200 m, while a 100 m window was used for depths deeper than 200 m, to account for larger sampling intervals in the subsurface. If there were multiple matched points, only the point with the smallest depth difference was selected. Finally, we obtained 724 matched profiles containing a total of 11,183 matched float-ship data points (Supplementary Fig. 1). Each point contains the corresponding float and ship variables: $\text{pH}_{\text{in-situ}}$, $\text{pH}_{25^\circ\text{C}, 0\text{dbar}}$, $p\text{CO}_2$, O_2 , NO_3 , T, S and TA. The average depth difference of data points at depths shallower than 200 m was 1.8 m; for those at depths deeper than 200 m, it was 13.5 m. Given the small depth differences, it is reasonable to expect float and ship data to be similar in value, if measured accurately, especially in the subsurface.

Methods for comparing matched float-ship data in subsurface water

Anthropogenic carbon calculation

Float data have generally been collected more recently than ship data, and thus, where in contact with the atmosphere, are expected to contain more anthropogenic carbon. This leads to a potentially confounding effect on our comparisons, mistaking different amounts of anthropogenic carbon for measurement bias. In this study, we therefore paid particular attention to water masses with minimal anthropogenic carbon perturbation (unacidified waters), which serve as a baseline for assessing float carbonate system variables' data quality, independent of long-term ocean acidification trends. These water masses are primarily Upper Circumpolar Deep Waters and Lower Circumpolar Deep Waters (UCDW and LCDW, respectively); the sources of the UCDW are Indian and Pacific Deep Waters (IDW and PDW, respectively) while the LCDW is primarily fed by North Atlantic

Deep Water (NADW). These water masses have been isolated from the atmosphere since preindustrial times and thus are minimally influenced by anthropogenic carbon^{50–52}. This focus on old, deep waters is a key analytical strategy which ensures that any discrepancies observed between float and ship-based data are not attributable to changes in ocean chemistry driven by anthropogenic carbon. Therefore, we used estimates of anthropogenic carbon as one indicator to isolate matched float-ship data that are unaffected by ocean acidification. To establish the accuracy of the calculated anthropogenic carbon in the matched float-ship data, we employed both ΔC^* and TrOCA approaches for estimation.

ΔC^* approach

ΔC^* has been used extensively to calculate anthropogenic carbon^{53–56}. The specific calculation process is as follows:

$$\Delta C^* = C_m - C_{280} - \Delta C_{bio} \quad (2)$$

$$C_{anth} = C_m - C_{280} - \Delta C_{bio} - \Delta C_{dis} = \Delta C^* - \Delta C_{dis} \quad (3)$$

$$C_{280} = CO2SYS(S, T, TA^0, 280) \quad (4)$$

$$TA^0 = 378.1 + 55.22 * S + 0.0716 * PO - 1.236 * \theta \quad (5)$$

$$\Delta C_{bio} = -117/170(O_2 - O_{2,sat}) + 1/2(TA - TA^0 - 16/170(O_2 - O_{2,sat})) - 106/104N_{anom}^* \quad (6)$$

where C_{anth} denotes the anthropogenic carbon concentration; C_m denotes measured DIC; C_{280} denotes the DIC of waters in equilibrium with an atmospheric CO_2 concentration of 280 μatm ; ΔC_{bio} denotes the DIC changes due to remineralization of organic matter and dissolution of calcium carbonate particles in the time since the parcel of water left the surface; TA^0 denotes the preformed TA, based on a multiple linear regression fit of TA values in the Antarctic region⁵⁷; ΔC_{dis} is the air-sea disequilibrium component taken from Sabine et al.⁵⁸. PO is a quasi-conservative tracer⁵⁹. N_{anom}^* is the N^* anomaly from the mean, θ is the potential temperature, and $O_{2,sat}$ is the saturated O_2 concentration.

TrOCA approach

The TrOCA (Tracer combining O_2 , DIC, TA) method is also a widely adopted approach for estimating anthropogenic carbon^{60–62}. The specific calculation process is shown in the following equations:

$$TrOCA = O_2 + 1.279(DIC - 1/2TA) \quad (7)$$

$$TrOCA^0 = e^{(7.511 - (1.087 \times 10^{-2})\theta - 7.81 \times 10^5 / TA^2)} \quad (8)$$

$$C_{anth} = (TrOCA - TrOCA^0) / 1.279 \quad (9)$$

where C_{anth} denotes the anthropogenic carbon concentration, θ is potential temperature, TrOCA is the conservative tracer as originally defined by Touratier and Goyet^{60,62}, and $TrOCA^0$ represents the pre-industrial TrOCA value, derived by fitting a combined dataset of $\Delta^{14}C$ - and CFC-11-based estimates to an exponential function (Eq. 8), using the GLODAP world ocean database as a reference⁵⁶.

Ultimately, the anthropogenic carbon values calculated using both methods show strong agreement (Supplementary Fig. 2) and can be used reliably to identify water masses that are minimally influenced by anthropogenic carbon.

Identification of unacidified subsurface water for matched float-ship data

We applied a three-fold constraint to carefully identify 'old' subsurface waters (UCDW and LCDW) that are largely unaffected by ocean acidification (minimal anthropogenic carbon signal): (1) depth > 200 m. (2) temperature and potential density values characteristic of either UCDW ($2.5^\circ\text{C} < \text{temperature} < 3^\circ\text{C}$, $27.4 < \text{potential density} < 27.7$) or LCDW ($0^\circ\text{C} < \text{temperature} < 2.5^\circ\text{C}$, potential density > 27.7). (3) ship-based $C_{ant} < 10 \mu\text{mol kg}^{-1}$ (For T-S diagrams of ship data and float data, see Supplementary Fig. 3). We did not use float-derived estimates of anthropogenic carbon for screening, as any biases in float-based pH, pCO_2 , and DIC data (as we conclude in this paper) compromise the reliability of their values.

Methods for identifying float pCO_2 bias in surface water

CORS plot (carbon and oxygen relative to saturation)

The CORS method is a CO_2 - O_2 analysis technique which involves comparing the deviations of O_2 and CO_2 concentrations from their respective saturation levels. Following Wu, et al.²⁶, we adopted an identical approach for both O_2 and CO_2 and, proceeded to compare the observed dissolved concentrations of O_2 and CO_2 ($[O_{2,observed}]$ and $[CO_{2,observed}]$ respectively) in surface seawater with their respective saturation values ($[O_{2,saturation}]$ and $[CO_{2,saturation}]$). The saturation values represent the points at which the net air-sea gas exchange rate should be zero for each gas. Differences from saturation of O_2 and CO_2 are calculated following:

$$\Delta O_2 = [O_{2,observed}] - [O_{2,saturation}] \quad (10)$$

$$\Delta CO_2 = [CO_{2,observed}] - [CO_{2,saturation}] \quad (11)$$

The saturation concentration of O₂ was calculated using Garcia & Gordon's Eqs.^{63,64}; however, while they calculated the O₂ saturation concentration at an assumed 1 atm of atmospheric pressure (Eq. 12), we instead used the local in-situ sea level pressure (SLP, usually < 1 atm in the SO) as given in Eq. (13):

$$[O_{2,saturation}^{1atm}] = K_{O_2} \times pO_2^{1atm} = K_{O_2} \times xO_{2,air} \times (P_{1atm} - P_{Sw}) \quad (12)$$

$$[O_{2,saturation}^{SLP}] = K_{O_2} \times pO_2^{SLP} = K_{O_2} \times xO_{2,air} \times (P_{SLP} - P_{Sw}) \quad (13)$$

Here, the subscripts '1 atm' and 'SLP' indicate two different pressures. K_{O_2} is the solubility of dissolved O₂, $O_{2,saturation}^{1atm}$ is the result based on the Garcia & Gordon's Eqs.^{63,64}. $O_{2,saturation}^{SLP}$ is the result based on our modified equation. P_{Sw} is the water vapor pressure calculated from the surface temperature and salinity⁶⁵. P_{SLP} was taken from the monthly gridded climate data downloaded from National Oceanic and Atmospheric Administration (NOAA) (Gridded Climate: NOAA Physical Sciences Laboratory, last access: 10 July 2023). Combining Eqs. (12) and (13):

$$[O_{2,saturation}^{SLP}] = [O_{2,saturation}^{1atm}] \times (P_{SLP} - P_{Sw}) / (P_{1atm} - P_{Sw}) \quad (14)$$

$[CO_{2,saturation}]$ was calculated with Henry's equation. ($[CO_{2,observed}] = K_{CO_2} \times pCO_{2,equilibrium}$), where $pCO_{2,equilibrium}$ is the partial pressure of CO₂ in seawater at equilibrium with atmospheric CO₂ (Eq. 15).

$$pCO_{2,equilibrium} = xCO_{2,air} \times (P_{SLP} - P_{Sw}) \quad (15)$$

where $xCO_{2,air}$ is the mole fraction of CO₂ (ppm) in dry atmosphere. The monthly mean atmospheric $xCO_{2,air}$ values from the monitoring site at Palmer Station, Antarctica were used for the SO (downloaded from <https://www.esrl.noaa.gov/gmd/ccgg/trends/>, last access: 14 July 2023). The solubility of CO₂ (K_{CO_2}) was calculated from the formula proposed by Weiss⁶⁶.

Matched float-OceanSODA data

Since the surface ocean is defined as shallower than 5 m by both SOCAT and OceanSODA, we selected all float surface data in the top 5 m for which the quality flag is good for pCO_2 . Because of the strong spatial and temporal variability in surface waters, float data and OceanSODA were only matched when from the same month, same year and within 25 km of each other. The final set of matched float-OceanSODA data contains surface data from 594 float profiles and matching OceanSODA data.

Methods for statistical analysis of data

Deming regression

Deming regression⁶⁷ is a technique used to fit a straight line to two-dimensional data where both variables, X and Y, are measured with error. This method contrasts with simple linear regression, where only the response variable, Y, is considered to be subject to measurement error. Applying this method to compare the matched float-ship data takes account of errors in observations in both ship and float data.

Cohen's d

Cohen's d⁶⁸ is a standardized measure of effect size used to quantify the difference between two group means, which scales the difference in the two means to the inherent variability of the data. Specifically in this work, we applied Cohen's d to evaluate the deviation degree of different matched variables. It is calculated as:

$$DD = \frac{M_F - M_S}{\sqrt{SD_F^2 + SD_S^2}} \quad (16)$$

where M_F is the mean value of float data, M_S is the mean value of ship data, SD_F is the standard deviation of float data and SD_S is the standard deviation of ship data. A negative value of DD indicates that the float mean is smaller than the ship mean and vice versa. The absolute value of DD represents the magnitude of this difference, normalized by the variability of the data.

Monte Carlo simulation process

We applied Monte Carlo simulation⁶⁹ to calculate the impact of uncertainties in individual float and ship data on the mean differences of matched float-ship pCO_2 . The uncertainty of an individual float pCO_2 value is 11 μatm ¹⁶ when float pCO_2 is 400 μatm . The uncertainty of ship pCO_2 is 12 μatm ³⁴ at 400 μatm . The probability density function of average float pCO_2 and ship pCO_2 from 1000 Monte Carlo iterations as well as the difference between the two was calculated for an example case. The following procedure was used: (1) the average float pCO_2 was assumed to be 400 μatm while that of the ship pCO_2 was assumed to be 383 μatm (according to the mean ΔpCO_2 in 200–1500 m depth). (2) Where 'G(μ,σ)' is a random number from a normal (Gaussian) distribution with mean of μ and standard deviation of σ , 724 independent float pCO_2 values, each equal to $400 + G(0,11)$ were generated. Similarly, 724 independent ship pCO_2 values, each equal to $383 + G(0,12)$, were also generated. The sample size '724' in both cases is equal to the amount of ship and float profiles used in the study. (3) For each pair of float and ship values, individual pCO_2 difference was calculated. This was then followed by the computation of the average of the 724 differences, (4) The steps (1) to (3) were repeated 1000 times to obtain 1000 average differences, (5) the corresponding frequency distribution plot of the average differences was constructed.

Results

Consistencies of variables between matched float-ship data in subsurface waters with minimal anthropogenic carbon

To minimize the influence of long-term ocean acidification, we focused on deep water masses with minimal anthropogenic carbon, primarily UCDW and LCDW. These unacidified waters serve as a stable reference for assessing discrepancies between matched float-ship data. The approaches for matching float and ship data and for identifying subsurface water masses free of anthropogenic carbon are described in Methods. As a first step, density scatter plots were generated to compare NO_3 , O_2 , $\text{pH}_{\text{in-situ}}$, $p\text{CO}_2$ (Fig. 1) and also T, S, TA, DIC, $\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$ (Supplementary Fig. 4(a), 4(b), 4(c), 4(d), 4(e)) for matched float and ship data collected in old subsurface waters (uninfluenced by anthropogenic carbon). We applied Deming regression to evaluate the consistency of the two datasets. The results show that float-based O_2 , NO_3 , T, S, and TA data agree well with the ship-based values, while $\text{pH}_{\text{in-situ}}$, $\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$, DIC and $p\text{CO}_2$ exhibit noticeable discrepancies (Table 1).

Out of the float-based carbonate system variables, TA is estimated based on a regression algorithm¹⁹, pH is measured by a float sensor, and DIC and $p\text{CO}_2$ are calculated from $\text{pH}_{\text{in-situ}}$ and TA. Out of all the variables in Table 1, $\text{pH}_{\text{in-situ}}$, $\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$, $p\text{CO}_2$ and DIC show the lowest consistency, with best-fit lines having slopes deviating from 1 (i.e., the slope of the 1:1 line) by more than 0.1 (i.e., slope < 0.9 or slope > 1.1). Float $\text{pH}_{\text{in-situ}}$ is on average ~ 0.021 lower than ship $\text{pH}_{\text{in-situ}}$. Float $p\text{CO}_2$ is $\sim 20 \mu\text{atm}$ higher than ship $p\text{CO}_2$. In contrast, variables such as O_2 , NO_3 , T, and S show strong agreement between float and ship data, despite differences in sampling time and season. This consistency supports the validity of using these matched datasets for comparison. There is very good agreement for TA (Supplementary Fig. 4(c), Table 1); the mean difference in TA is $-1 \mu\text{mol kg}^{-1}$ which would lead to no more than a $0.1 \mu\text{atm}$ difference in $p\text{CO}_2$. The positive bias in derived $p\text{CO}_2$ appears to be consistent with the negative deviation in float pH (a lower measured pH results in a higher calculated $p\text{CO}_2$). The degree of deviation increases with decreasing pH and increasing $p\text{CO}_2$ values.

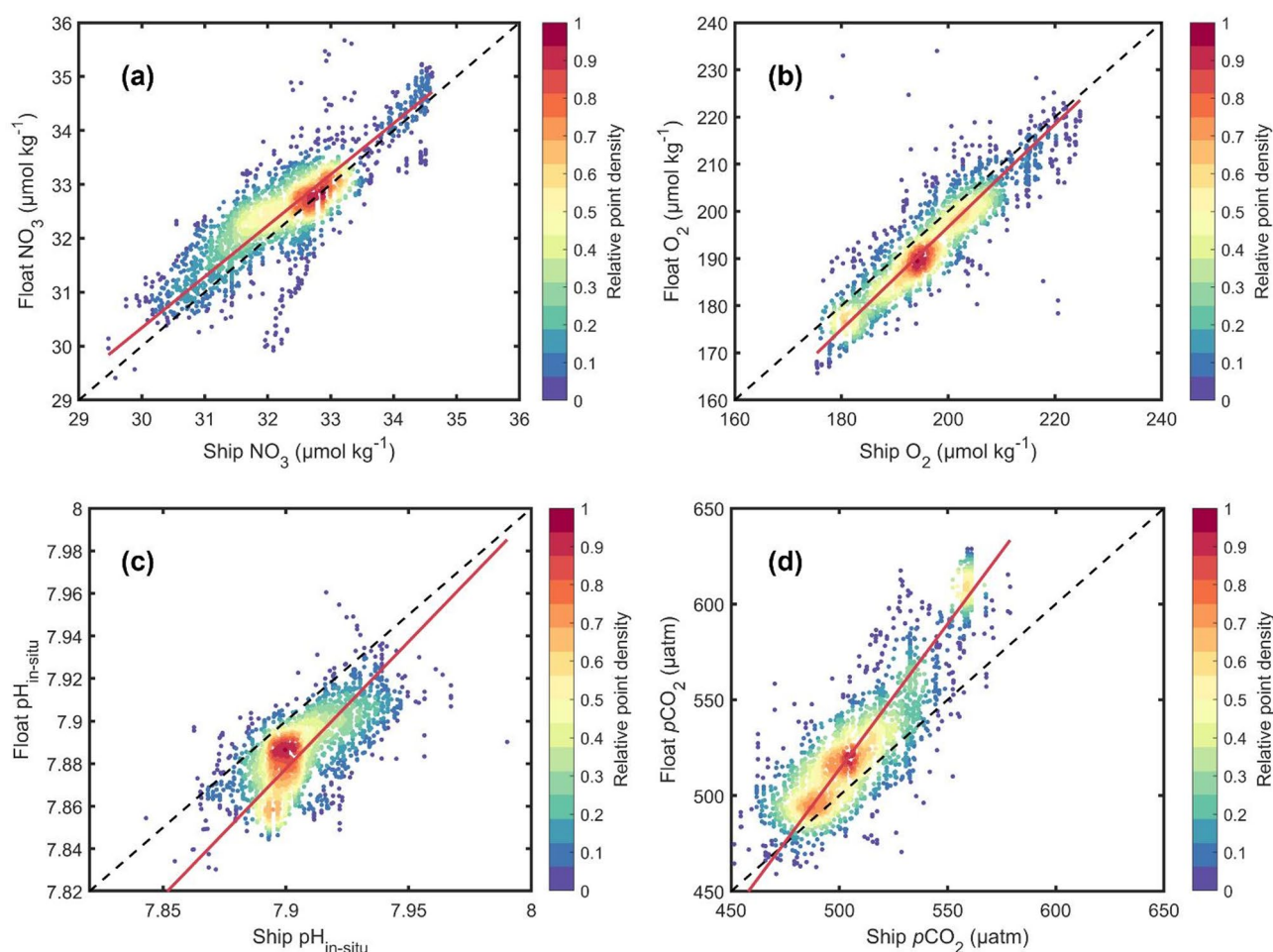


Fig. 1. Density scatterplots of matched float-ship data in subsurface water with minimal anthropogenic carbon. Panels (a)–(d) show NO_3 , O_2 , $\text{pH}_{\text{in-situ}}$ and $p\text{CO}_2$. The dashed line is the 1:1 relationship (perfect agreement) and the red line is the best-fit line calculated using Deming regression. The colour of each point reflects the density of points around it.

Variable	Slope	Standard error of regression estimate*	Mean difference (Float-Ship)	Relative difference (%)**
pH _{in-situ}	1.19 ± 0.06	0.0099	-0.0213	-0.27
pH _{25°C, 0 dbar}	1.48 ± 0.18	0.0075	-0.0187	-0.25
pCO ₂ (µatm)	1.52 ± 0.04	11.0	+ 19.8	3.86
DIC (µmol kg ⁻¹)	1.37 ± 0.05	3.22	+ 5.95	0.26
TA (µmol kg ⁻¹)	0.94 ± 0.03	2.32	-1.09	0.00
O ₂ (µmol kg ⁻¹)	1.08 ± 0.02	3.41	-3.69	-1.89
NO ₃ (µmol kg ⁻¹)	0.95 ± 0.02	0.364	+ 0.221	0.68
T (°C)	1.00 ± 0.01	0.101	0.00	0.00
S (psu)	1.08 ± 0.02	0.0112	0.00	0.00

Table 1. Deming regression of matched float-ship subsurface data in unacidified subsurface water for different variables. * The standard error of regression estimate indicates the average distance between the observed values and the regression line⁷⁰. Its unit is the same as the unit of the dependent variable. It represents a goodness-of-fit measure for regression analysis (the relationship is stronger when the value is smaller). ** Relative difference (%) is the mean percentage difference between float and ship data, calculated as (Float – Ship)/Ship × 100%.

Discrepancies of variables at different depths between matched float-ship data in subsurface water with minimal anthropogenic carbon

To further explore float pH and pCO₂ deviations, we compared matched float-ship data vertical profiles for different variables. Outputs were binned in 100 m vertical intervals, to show mean values of float data and ship data against depth, as well as the mean differences between float and ship data (represented by Δ, e.g., ΔpH_{in-situ}) against depth. Here we mainly compared the results of NO₃, O₂, pH_{in-situ}, TA and pCO₂ (Fig. 2) because of the strong connection between them^{71,72}, (for other variables, T, S, DIC and pH_{25°C, 0 dbar} see Supplementary Fig. 5). O₂ increases/decreases are often accompanied by pCO₂ decreases/increases from photosynthesis and respiration, leading to negative correlations between ΔpCO₂ and ΔO₂.

In the depth plots, it appears that pH_{in-situ} and pCO₂ have a more obvious discrepancy than other variables. Incidentally, the trend and magnitude of discrepancy in pH_{25°C, 0 dbar} is similar to discrepancy in pH_{in-situ} (Supplementary Fig. 5). However, the variables have different units and distribution ranges (e.g., O₂: 170–230 µmol kg⁻¹ and pH_{in-situ}: 7.85–7.90), so the differences in their means cannot be directly compared. To explore the discrepancies more rigorously, we calculated the deviation degree (DD), a version of Cohen's d, which scales the difference in the two means to the inherent variability of the data.

The DDs (Fig. 3) of all variables are generally in line with our depth plots (Fig. 2): pH_{in-situ}, DIC and pCO₂ show the greatest degree of deviation overall. However, there are two points to note: (1) Compared with T, S and TA whose absolute DD is always less than 0.5, O₂ and NO₃ both show larger deviations. This suggests that the current dissolved oxygen and nitrate sensors on floats, as less well-established sensors, may not be quite as accurate as the more established temperature and salinity sensors. (2) The discrepancies in float pH_{in-situ}, DIC, and pCO₂ are large in the 200–1500 m depth range, with mean DD values of -1.1 for pH_{in-situ}, 0.7 for DIC, and 0.7 for pCO₂. In contrast, these deviations are smaller below 1500 m, where the mean DD values are -0.8 for pH_{in-situ}, 0.4 for DIC, and 0.3 for pCO₂.

Estimated surface float pCO₂ bias

Since the calculation of CO₂ fluxes depends on the surface pCO₂, we also estimate the bias present in the surface float pCO₂. We use a CORS plot (see Methods) for this. In situations where biogeochemical and physical processes like photosynthesis and upwelling strongly influence the system, the deviations of O₂ and CO₂ from atmospheric equilibrium demonstrate a coupled behaviour, as photosynthesis and remineralization affect carbon and oxygen in contrasting directions, according to a stoichiometric ratio. If float and ship data are both accurate then they should overlay on a CORS plot; however, while the slopes are very similar for both sets of data (-0.055 and -0.054 respectively) the CORS plot y-intercepts are different (-0.08 and -0.92 µmol kg⁻¹ respectively) (Fig. 4). Converting the [CO₂] offset (0.84 µmol kg⁻¹, float minus ship) to pCO₂ (in µatm), assuming average values for sea surface temperature (1 °C) and salinity (35 psu), gives a difference in surface pCO₂ of 14.1 µatm.

To further investigate the float pCO₂ bias in surface waters, we performed another independent comparison, between float surface data and OceanSODA output. The result is as follows: float surface pCO₂ is generally higher than OceanSODA pCO₂; the range of differences (between the fitting line and the 1:1 line) is 13.8–20.4 µatm, with a mean difference of 17.4 µatm (Fig. 5). This value is close to the bias of 14.1 µatm obtained from the CORS plot. Consequently, our overall estimate of the float pCO₂ bias in surface waters is 14–17 µatm. This estimate is considerably larger than the current generally accepted bias of 3.6 µatm^{18,20}. Here we would like to stress that this bias is the mean bias of 724 float profiles in the whole SO; it does not mean that every float has a deviation of equal magnitude. Comparisons with limited profiles or in localised regions are likely to give different results.

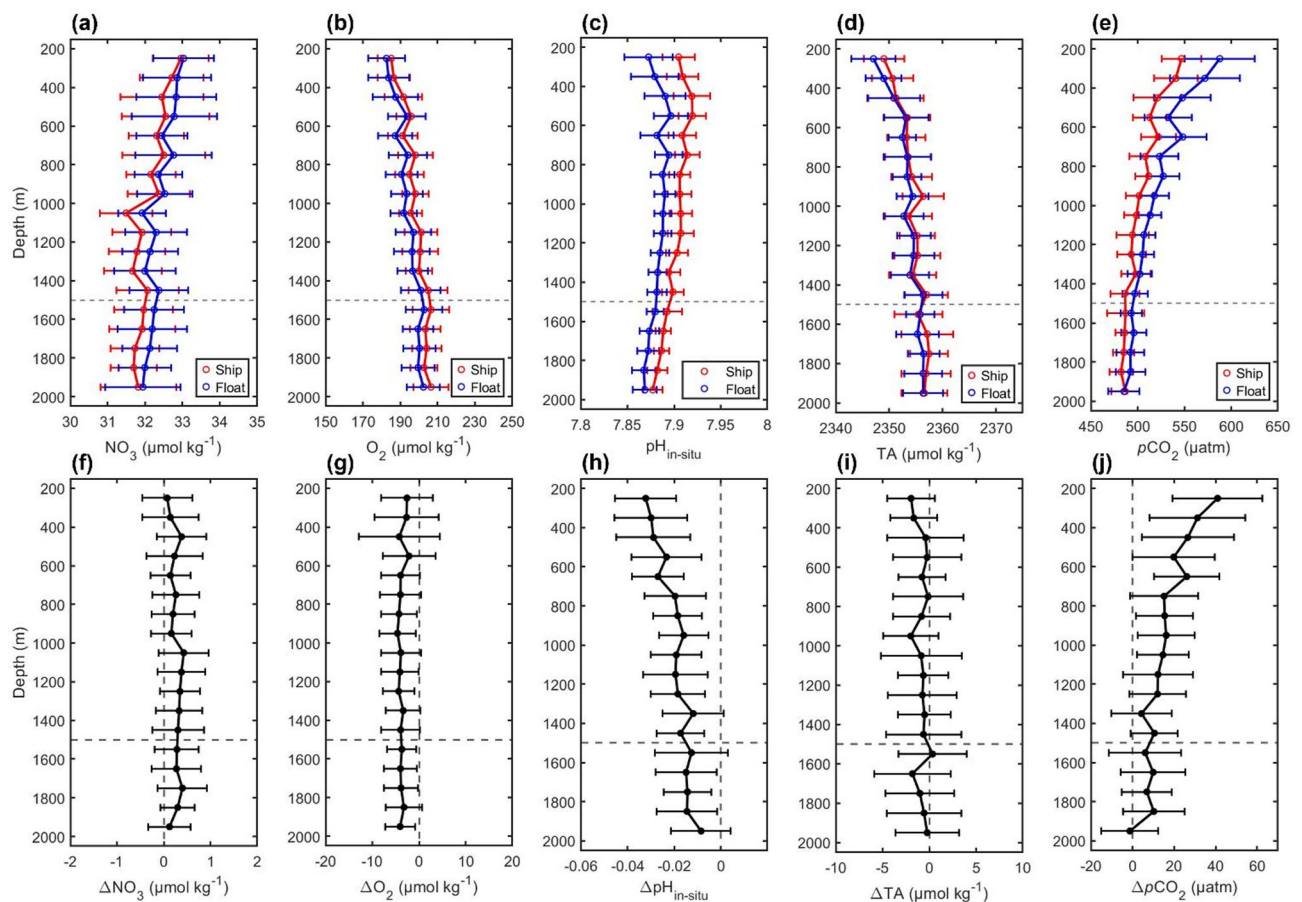


Fig. 2. Depth plots of NO_3 , O_2 , $\text{pH}_{\text{in-situ}}$, TA and pCO_2 from matched float-ship data in subsurface water with no anthropogenic carbon signal. Vertical profiles of mean values: (a) NO_3 , (b) O_2 , (c) $\text{pH}_{\text{in-situ}}$, (d) TA and (e) pCO_2 ; and mean differences (float minus ship): (f) ΔNO_3 , (g) ΔO_2 , (h) $\Delta\text{pH}_{\text{in-situ}}$, (i) ΔTA and (j) ΔpCO_2 . Error bars show standard deviations of all matched float, ship or float minus ship data within each 100 m depth bin.

Discussion

Negligible influence of anthropogenic carbon and data uncertainties on detected biases

As described in the Methods section, we used a variety of methods to assess whether biases are present in float pH and therefore also in derived pCO_2 and DIC. Since floats equipped with pH sensors only began to be widely deployed in 2014, their data comes from the most recent decade. The ship data, in contrast, was collected over the last 50 years (1972–2023). A potential concern is whether long-term ocean acidification-induced changes in pH and derived variables (pCO_2 and DIC), stronger in the more recent data, could be misinterpreted as measurement biases, owing to differences in average year of collection of the two datasets. We took steps to address this concern and now explain why the detected biases are not a misattribution due to larger amounts, on average, of anthropogenic carbon in the float data. Any contributions from anthropogenic carbon to the stated deviations are negligible.

When estimating the surface float pCO_2 bias, we utilized both CORS plots and matched float-OceanSODA datasets. In the CORS plots, ΔCO_2 is defined as the difference between the measured pCO_2 and the equilibrium (saturated) pCO_2 in the surface ocean corresponding to the anthropogenic carbon at the time of measurement. While this approach minimizes the impact of long-term trends, we acknowledge that differences in air-sea equilibration timescales, particularly for upwelled waters in the Southern Ocean, may introduce some residual effects. For the matched float-OceanSODA comparisons, OceanSODA data were selected to coincide in year and month with float measurements within a spatial radius of 25 km. Consequently, these paired datasets represent synchronous observations, effectively excluding any influence from anthropogenic carbon.

Another approach to address this concern was to compare ship and float data in old subsurface waters containing little or no anthropogenic carbon. We applied multiple constraints to ensure that these matched float-ship observations were obtained from water masses unaffected, or only lightly affected, by ocean acidification. Specifically, we selected only measurements taken at depths exceeding 200 m, with potential density and temperature characteristics of UCDW or LCDW, and with estimated anthropogenic carbon concentrations (based on ship data) of less than $10 \mu\text{mol kg}^{-1}$. As a final step to ascertain that the differences identified in our selected subsurface matched dataset are not influenced by variability in sampling years, we examined the relationship

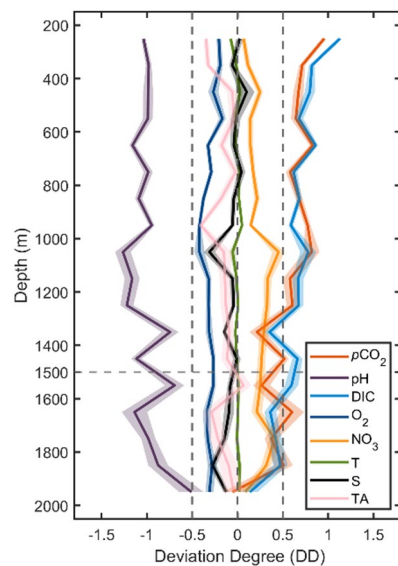


Fig. 3. The deviation degree (DD) of each variable ($p\text{CO}_2$, $\text{pH}_{\text{in-situ}}$, DIC, O_2 , NO_3 , T, S, TA) from matched float-ship data in subsurface water with no anthropogenic carbon in each depth interval. The shading represents ± 1 standard error. Negative values indicate that the float mean is smaller than the ship mean and vice versa. The absolute value of DD represents the magnitude of the average difference scaled by the variability of the data.

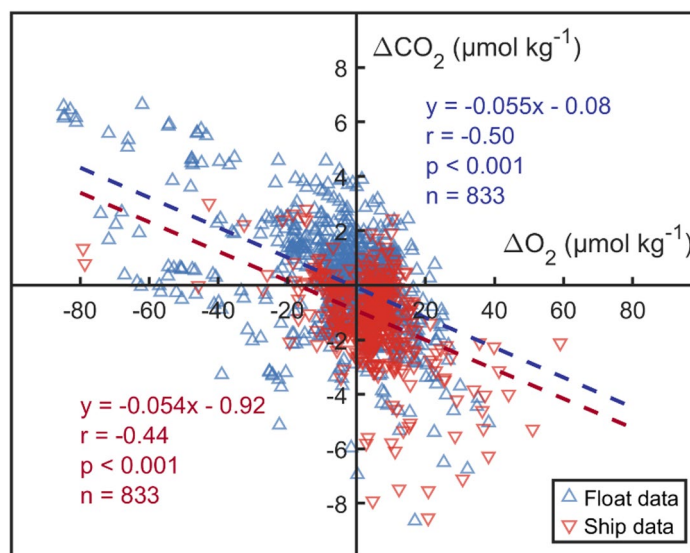


Fig. 4. CORS plot from matched surface float data (blue) and ship data (red). The blue line is the best-fit line to float data and the red line is the best-fit line to ship data. r is the associated Pearson correlation coefficient, p denotes the significance level, and n represents the sample size. Float data (blue) is plotted first and ship data (red) second, masking the locations of many float data.

between the difference in sampling years and the $\Delta p\text{CO}_2$ (float-ship). This was done for the matched float-ship data in unacidified subsurface waters (Fig. 6). The result indicates an extremely weak negative correlation ($r = -0.05$), suggesting that the float $p\text{CO}_2$ deviations are effectively independent of differences in sampling years. Based on the outcomes of all these analyses, we categorically rule out the possibility that the float biases we present are artefacts caused by the float data being more recent than the ship data. It is also important to assess the potential role of seasonal variability. Previous studies suggest that seasonal signals in UCDW and LCDW are very small^{50–52}, but we nonetheless tested this explicitly. Here, we further refined the matched float-ship data by retaining only those pairs where the float and ship observations occurred in the same season (austral summer: December–February; autumn: March–May; winter: June–August; spring: September–November) and then generated a new plot (Supplementary Fig. 6) using this reduced dataset. Compared to Fig. 2 and Supplementary

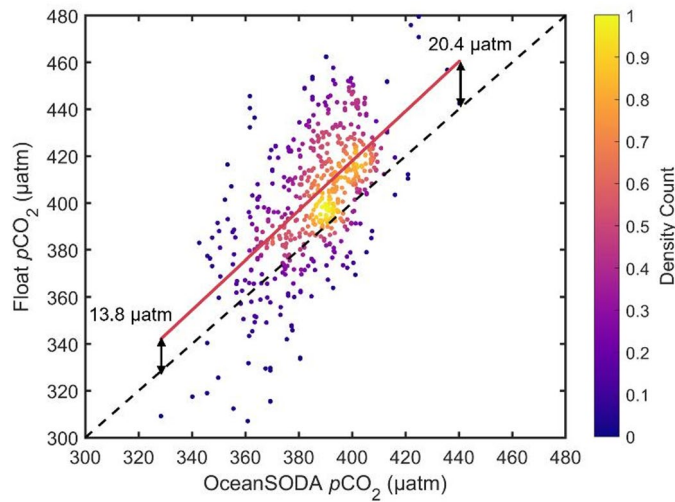


Fig. 5. Density scatter of matched float-OceanSODA $p\text{CO}_2$ surface ocean data. The dashed line is the 1:1 relationship and the red line is the linearly fitted line: $y = 1.06x - 5.6$ ($r = 0.60$). The colour of each point reflects the density of points around it.

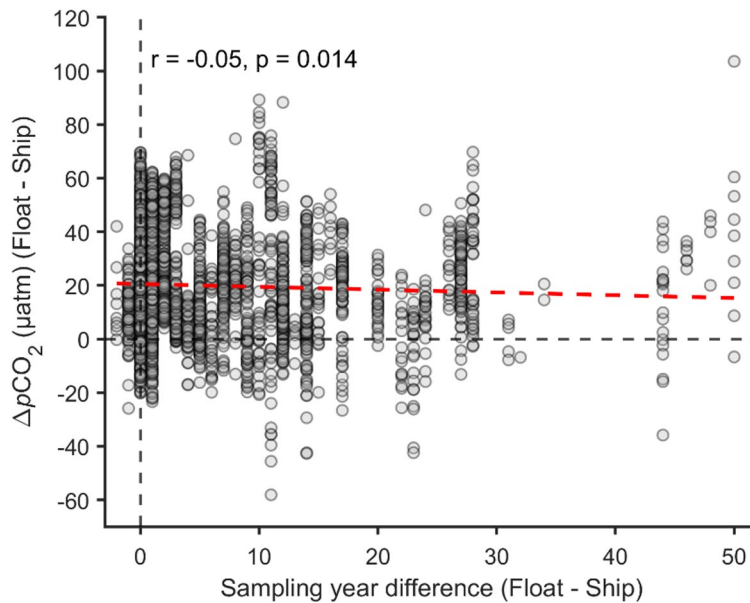


Fig. 6. The relationship between $\Delta p\text{CO}_2$ and sampling year difference (float minus ship) in unacidified subsurface water. The red line is the linearly fitted line: $y = -0.105x + 20.5$ ($r = -0.05$, $p = 0.014$).

Fig. 5, over 200–1500 m, the mean biases then become $17.1 \mu\text{atm}$ for $\Delta p\text{CO}_2$ (compared to $18.9 \mu\text{atm}$ when all matched pairs are used without regard to season), $4.9 \mu\text{mol kg}^{-1}$ for ΔDIC ($6.0 \mu\text{mol kg}^{-1}$ with all matches), -0.0201 for $\Delta\text{pH}_{\text{in-situ}}$ (-0.0216 with all matches), and -0.0178 for $\Delta\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$ (-0.0188 with all matches). Thus, restricting the dataset to same-season pairs changes the subsurface biases by $\sim 1.8 \mu\text{atm}$ in $p\text{CO}_2$, $\sim 1 \mu\text{mol kg}^{-1}$ in DIC, ~ 0.0015 in $\text{pH}_{\text{in-situ}}$ and ~ 0.001 in $\text{pH}_{25^\circ\text{C}, 0 \text{ dbar}}$ (Supplementary Fig. 6) are nearly indistinguishable from those obtained when all matched pairs are used. We therefore conclude that seasonal variability has only a minor effect on the diagnosed subsurface bias. Although our analysis focused on the mean bias, we acknowledge the broad distribution of float–ship $p\text{CO}_2$ differences (Fig. 6). This spread reflects the combined effect of multiple factors: sensor variability, spatial–temporal heterogeneity etc. While the average bias is robust, the wide scatter suggests that not all floats behave similarly, and that environmental and methodological uncertainties play a role. A more detailed float-by-float evaluation is a suitable topic for future work.

A second potential concern is whether random errors in individual measurements could be responsible for the biases seen. Data uncertainty reflects the variability or error associated with a single measurement, arising from factors such as instrument precision and random disturbances. While random errors are unavoidable,

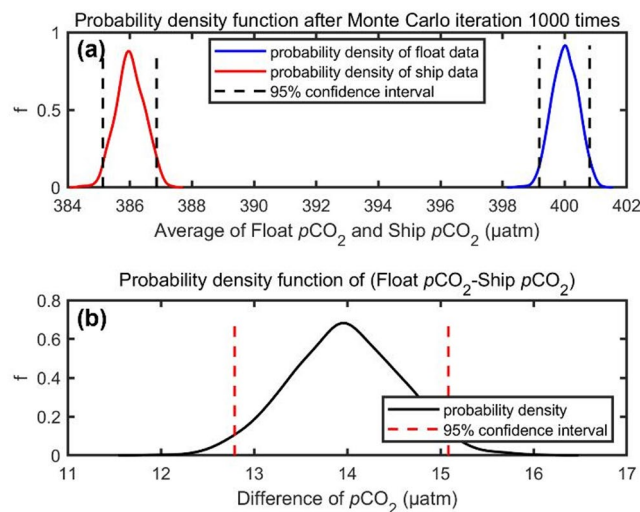


Fig. 7. Results of Monte Carlo calculations to test the effect of uncertainties in individual float $p\text{CO}_2$ and ship $p\text{CO}_2$ on the average difference between float and ship $p\text{CO}_2$. Assessment of (a) the impact of uncertainty in individual float $p\text{CO}_2$ (11 μatm) and ship $p\text{CO}_2$ (12 μatm) data on respective averages, and (b) the impact of uncertainty in individual values on the overall value of $\Delta p\text{CO}_2$ (float minus ship).

the key issue for air-sea CO_2 fluxes is the degree of accuracy of averaged float estimates of $p\text{CO}_2$ rather than the degree of accuracy of each individual observation^{33,73,74}. We used a Monte Carlo technique (see Methods) to quantify the probable effect on overall $\Delta p\text{CO}_2$ (float minus ship) from uncertainties in individual float and ship $p\text{CO}_2$ values. Figure 7a presents the probability density functions of average float $p\text{CO}_2$ and ship $p\text{CO}_2$. The effects of uncertainty in each single point of float and ship $p\text{CO}_2$ data (11 and 12 μatm respectively) on mean $\Delta p\text{CO}_2$ are minor. At the 95% confidence level, the effect on $\Delta p\text{CO}_2$ (Fig. 7b) is less than 1.3 μatm , i.e., an order of magnitude lower than the detected bias. These results are based on the assumption that individual errors are random and independent. It does not hold for systematic biases, but that is the subject of this study. In conclusion, it is clear that, for the large datasets used in this study, the observed systematic bias cannot be attributable to the uncertainty in individual values. Based on these simulations, the confidence interval for the final estimated surface bias in float $p\text{CO}_2$ is approximately ± 2 μatm . Accordingly, given the observed bias of 14–17 μatm and this associated confidence interval, we report the final surface bias as 15 ± 3 μatm .

Revisiting data quality and bias of float pH and derived $p\text{CO}_2$

pH and derived $p\text{CO}_2$ data (hereafter collectively referred to as float carbon data) have been more frequently questioned than other variables obtained by BGC Argo floats. This is partly due to the low return rate of high-quality data^{75,76}, from the raw measurements of the pH sensor, and partly due to the complexity of the calibration and calculation processes for pH and $p\text{CO}_2$, some of which rely on empirical formulas without a full understanding of the underlying principles^{16,76}. While a host of exciting discoveries have been made using float carbon data⁷³, some concerns remain about its overall reliability.

Our results provide evidence that float carbon data exhibits considerable biases, not only within the surface layer but also throughout the entire vertical profile. The surface $p\text{CO}_2$ bias found by this research 15 ± 3 μatm is significantly larger than the previously recognized bias (3.6 μatm) suggested by Gray et al.¹⁸, but close to a regional comparison result (14 μatm)²⁷. It is also much larger than biases of 0.003 in float pH and 3.2 μatm in float $p\text{CO}_2$ attributed to a lag in the response of the float O_2 sensors⁷⁷ – our analyses suggest that correction of this lag, while helpful, would only remove a small fraction of the total biases we detect in float pH and $p\text{CO}_2$. The observed subsurface $p\text{CO}_2$ bias (mean $\Delta p\text{CO}_2$ of 18 μatm between 200 and 1500 m depth, Fig. 2) is even more pronounced than the surface bias. The pH and $p\text{CO}_2$ biases we have found in the subsurface (200–1500 m depth) imply that there may be issues with the current calibration process (described in detail in the Methods section) for float pH data. Using the mismatch between float and algorithm pH at 1500 m depth to adjust the whole profile pH data may therefore be insufficient or not always appropriate. The pH and $p\text{CO}_2$ consistency in the 1500–2000 m depth range indicates that the correction procedure works well below 1500 m depth. However, the biases identified in the 200–1500 m depth range may occur because the difference between float measured pH and actual pH over the whole profile is not always the same as at 1500 m depth. Our recommendation is to add a series of cross-reference points above 1500 m depth for the float pH correction, and to use the differences at multiple depths to adjust pH, rather than only the difference at 1500 m, which aligns with the conclusion presented by Wimart-Rousseau³⁰. Further examination and improvement of the float pH adjustment procedure is likely to be necessary in order to obtain more accurate pH and derived $p\text{CO}_2$ values.

Implications for flux calculations

If the surface $p\text{CO}_2$ bias found by this research 15 ± 3 μatm is considered in the calculation of SO carbon fluxes based on float $p\text{CO}_2$ data, the recalculated SO net annual carbon uptake will be much higher. In this context,

negative flux values represent CO₂ uptake by the ocean (ingassing), and positive values represent CO₂ release to the atmosphere (outgassing). Recalculation of the SO fluxes is beyond the scope of this study, but we give an estimate here, based on the results from Bushinsky et al.^{20,22}. They found that a 4 μatm reduction in the SOCCOM product led to an increase of $-0.21 \text{ Pg C yr}^{-1}$ in the estimated Southern Ocean carbon sink²⁰. Assuming that the effect of bias on flux is proportional, our conclusion of a bias of $15 \pm 3 \mu\text{atm}$ implies a reduction of the float-based flux from $+0.35 \text{ Pg C yr}^{-1}$ to between $-0.28 \text{ Pg C yr}^{-1}$ and $-0.60 \text{ Pg C yr}^{-1}$, closer to flux estimates based on ship data ($-0.38 \text{ Pg C yr}^{-1}$) and aircraft data ($-0.53 \text{ Pg C yr}^{-1}$)^{22–24}. This rough estimate suggests that the bias reported in this study could explain a large fraction of the current large discrepancy between fluxes based on float data versus fluxes based on other observations. This suggests that previous studies using float data have underestimated carbon uptake by the SO due to the biases in float pH and derived $p\text{CO}_2$.

Despite the current biases in the BGC Argo float pH and derived $p\text{CO}_2$ data, the value of the approach is without question. Float data expands the area and time coverage of data^{3,78}. The significant advances in observational capabilities with float data have contributed to an expanded understanding of processes acting at smaller scales⁹. Incorporating the insights from our research into the design of a future calibration process for float pH will further improve the quality of the pH data. At the same time, our results emphasize the value of continuing to collect high quality ship-based measurements as a reference for the larger volumes of autonomously collected data. Although various autonomous observation platforms have been developed in recent years, their data is less certain and will continue to need to be calibrated against ship observations. A combination of high-accuracy ship data and improved high-volume, wide-coverage float data will provide a deeper understanding of carbon dioxide uptake/release in the SO and a better quantification of the role of SO in the global carbon cycle.

Limitations and future directions

Our findings are based on data analysis and are subject to several inherent limitations. In the following, we describe some of these limitations and suggest directions for future research that could complement and build upon the work described here.

A first limitation is that our analysis is carried out within the inorganic carbonate system which can be affected by internal thermodynamic inconsistency⁷⁹. The subsurface comparisons rely on GLODAP as a reference dataset, within which pH is most often calculated from measured DIC and TA instead of being measured directly using spectrophotometric techniques⁷⁹. Previous studies have revealed a pH-dependent offset between DIC&TA-derived pH and spectrophotometrically measured pH^{17,29,80,81}. This suggests a scenario in which the entire offset found in our study arises not from bias in float pH but instead from biases in matched GLODAP pH, because the latter are computed incorrectly from measured DIC and TA.

However, two independent lines of evidence argue against this being the main reason for the biases we find. First, float-derived DIC is systematically higher than ship DIC according to the subsurface comparison (Fig. 3 and supplementary Figs. 4 d and 5 d) while float TA agrees closely with ship TA. The study by Garcia-Ibáñez et al. has shown that when high-quality spectrophotometric pH and accurate TA are used, DIC calculated from pH and TA is internally consistent with measured DIC and exhibits little systematic bias⁸². It would seem therefore that the sizeable DIC offsets we find cannot be explained solely by errors in GLODAP pH, because if that were the case then float and ship DIC should match. Second, our estimate of the surface $p\text{CO}_2$ bias relies not only on comparison to GLODAP but also on comparison to the OceanSODA data set, where the latter is constrained by direct surface CO₂ measurements in the SOCAT database²³ (i.e. does not involve any calculations of pH from DIC and TA). We obtain similar high biases (both $\sim 15 \mu\text{atm}$) from the comparisons to GLODAP and OceanSODA. This indicates that the observed discrepancies are not solely due to carbonate system thermodynamic inconsistency. Future work could include comparisons of float pH to matched spectrophotometrically measured ship pH sometime after float deployment, for instance by determining which cruises in the GLODAP dataset include direct measurements of pH and then looking for crossovers with float tracks.

A second limitation concerns the way surface and subsurface biases are linked in this study. Because the spatial–temporal coverage of co-located ship measurements is too limited to apply the same float–ship matching strategy at the surface and at depth, we diagnosed the bias from a subsurface float–ship comparison in old water with minimal anthropogenic carbon signal and then use an independent surface analysis based on CORS and OceanSODA to quantify how this bias manifests at the surface. Considering that the standard SOCCOM processing applies a single offset, estimated near 1500 m, to the entire pH profile, it is reasonable to expect surface and subsurface pH biases to be related rather than independent; our results are consistent with this picture, but variable concentrations of anthropogenic CO₂ and seasonality in surface waters make it harder to analyse biases there. Additional co-located ship–float observations with full-depth carbonate-system measurements (including ship-measured pH) would provide the most direct test of the bias patterns inferred here.

A third limitation is that, although we have explicitly examined and discussed a wide range of potential contributors to the observed discrepancies, it is impossible to prove an absence and so we cannot definitively exclude other unanticipated contributors to the diagnosed biases. Our analyses considered: (i) anthropogenic carbon, (ii) measurement uncertainties in both ship and float data, (iii) the accuracy of algorithm-estimated float TA, and (iv) effects of interannual variability as well as seasonality. None of these factors were found able to explain the magnitude and vertical structure of the pH, DIC and $p\text{CO}_2$ offsets. However, the inorganic carbon system and the float $p\text{CO}_2$ processing are complex, and with the currently available data we cannot separate sensor-related effects (e.g. drift, ageing, pressure response) from limitations of the adjustment procedure or from other as yet unidentified sources of bias. Clarifying the underlying mechanisms will require targeted laboratory and field experiments on the float pH sensor. Further development and intercomparison of alternative float pH adjustment schemes is also recommended.

Conclusions

This study investigated the accuracy in the SO of pH and derived $p\text{CO}_2$ data from BGC Argo floats by comparing it to high-quality ship data from the GLODAP dataset. A first comparison, restricted to old subsurface waters in order to exclude effects of anthropogenic carbon, identified systematic differences in float pH and derived $p\text{CO}_2$. On average, float pH was ~ 0.021 units lower than ship-based values, $p\text{CO}_2$ was ~ 20 μatm higher, whereas other variables (T, S, NO_3 , O_2) showed good agreement. A second comparison used a CORS plot to investigate float $p\text{CO}_2$ bias in surface waters, yielding a result of 14 μatm . A similar offset (range: 14–20 μatm , mean 17 μatm) was obtained by an independent method, comparing float surface $p\text{CO}_2$ to OceanSODA results for the same location, month and year. Float $p\text{CO}_2$ values are derived from both float pH and algorithm-estimated TA, but float-ship TA comparison showed that the influence of TA bias on discrepancies can only be very small (order 0.1 μatm). We consider a previously unrecognized bias in float pH to be the most plausible explanation for the observed $p\text{CO}_2$ offsets. This bias is larger above 1500 m depth and smaller below 1500 m depth, which suggests that the current float pH correction process (adjusting the entire float pH profile by comparing float-measured to algorithm-estimated pH at 1500 m depth) may not fully account for some depth-dependent offsets. Improving the existing float pH correction method could lead to more accurate pH and $p\text{CO}_2$ data, and thus more robust quantification of carbon uptake in the SO. Further work is necessary to clarify the mechanisms underlying this pH bias. Overall we estimate that float surface $p\text{CO}_2$ is biased high by 15 ± 3 μatm . Simple scaling arguments based on published flux sensitivities suggest that reducing float surface $p\text{CO}_2$ downward by this amount would largely account for the currently debated large discrepancy between float-based SO carbon flux estimates and those derived from other platforms.

Data availability

The ship (GLODAP) data were obtained from <https://www.glodap.info> (last access: 05 February 2024). The float data were downloaded from <https://library.ucsd.edu/dc/object/bb51655605>, 2023-12-20 snapshot. The OceanSODA data were downloaded from <https://www.ncei.noaa.gov/data/oceans/ncei/ocads/data/0220059/>.

Code availability

Matlab code for this study is available upon request to C.Z.

Received: 18 April 2025; Accepted: 6 March 2026

Published online: 17 March 2026

References

- Bittig, H. C. et al. A BGC-Argo guide: Planning, deployment, data handling and usage. *Front. Mar. Sci.* **6**, 502 (2019).
- Organelli, E. et al. Two databases derived from BGC-Argo float measurements for marine biogeochemical and bio-optical applications. *Earth Syst. Sci. Data.* **9**, 861–880 (2017).
- Claustre, H., Johnson, K. S. & Takeshita, Y. Observing the global ocean with biogeochemical-Argo. *Annual Rev. Mar. Sci.* **12**, 23–48 (2020).
- Claustre, H., Johnson, K. S. & Takeshita, Y. Observing the global ocean with biogeochemical-Argo. *Annual Rev. Mar. Sci.* **12**, 23–48 (2020).
- Smith, W. O. Jr, Anderson, R. F., Moore, J. K., Codispoti, L. A. & Morrison, J. M. The US southern ocean joint global ocean flux study: an introduction to AESOPS. *Deep Sea Res. Part II.* **47**, 3073–3093 (2000).
- Rintoul, S. R. The southern ocean in the earth system. Science diplomacy: Antarctica, science, and the governance of international spaces (2011).
- Williams, R. G., Ceppi, P., Roussenov, V., Katavouta, A. & Meijers, A. J. The role of the Southern Ocean in the global climate response to carbon emissions. *Philosophical Trans. Royal Soc. A.* **381**, 20220062 (2023).
- Sarmiento, J. L. et al. The Southern Ocean carbon and climate observations and modeling (SOCCOM) project: A review. *Prog. Oceanogr.* **219**, 103130 (2023).
- Gray, A. R. The four-dimensional carbon cycle of the Southern Ocean. *Annual Rev. Mar. Sci.* **16**, 163–190 (2024).
- Bauer, J. E. et al. The changing carbon cycle of the coastal ocean. *Nature* **504**, 61–70 (2013).
- Lenton, A. et al. Sea-air CO₂ fluxes in the Southern Ocean for the period 1990–2009. *Biogeosciences* **10**, 4037–4054 (2013).
- Metzl, N., Brunet, C., Jabaud-Jan, A., Poisson, A. & Schauer, B. Summer and winter air-sea CO₂ fluxes in the Southern Ocean. *Deep Sea Res. Part I.* **53**, 1548–1563 (2006).
- McNeil, B. I., Metzl, N., Key, R. M., Matear, R. J. & Corbiere, A. An empirical estimate of the Southern Ocean air-sea CO₂ flux. *Glob. Biogeochem. Cycles.* **21**, GB3011 (2007).
- Takahashi, T. et al. Climatological mean and decadal change in surface ocean pCO₂, and net sea-air CO₂ flux over the global oceans. *Deep Sea Res. Part II.* **56**, 554–577 (2009).
- Landschützer, P., Gruber, N., Bakker, D. C. & Schuster, U. Recent variability of the global ocean carbon sink. *Glob. Biogeochem. Cycles.* **28**, 927–949 (2014).
- Williams, N. L. et al. Calculating surface ocean pCO₂ from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Glob. Biogeochem. Cycles.* **31**, 591–604 (2017).
- Carter, B. et al. Updated methods for global locally interpolated estimation of alkalinity, pH, and nitrate. *Limnol. Oceanography: Methods.* **16**, 119–131 (2018).
- Gray, A. R. et al. Autonomous biogeochemical floats detect significant carbon dioxide outgassing in the high-latitude Southern Ocean. *Geophys. Res. Lett.* **45**, 9049–9057 (2018).
- Carter, B., Williams, N., Gray, A. & Feely, R. Locally interpolated alkalinity regression for global alkalinity estimation. *Limnol. Oceanography: Methods.* **14**, 268–277 (2016).
- Bushinsky, S. M. et al. Reassessing Southern Ocean air-sea CO₂ flux estimates with the addition of biogeochemical float observations. *Glob. Biogeochem. Cycles.* **33**, 1370–1388 (2019).
- Mackay, N. & Watson, A. Winter air-sea CO₂ fluxes constructed from summer observations of the polar southern ocean suggest weak outgassing. *J. Geophys. Research: Oceans.* **126**, e2020JC016600 (2021).
- Long, M. C. et al. Strong Southern Ocean carbon uptake evident in airborne observations. *Science* **374**, 1275–1280 (2021).
- Bakker, D. C. et al. An update to the Surface Ocean CO₂ Atlas (SOCAT version 2). *Earth Syst. Sci. Data.* **6**, 69–90 (2014).
- Landschützer, P., Gruber, N. & Bakker, D. C. Decadal variations and trends of the global ocean carbon sink. *Glob. Biogeochem. Cycles.* **30**, 1396–1417 (2016).

25. Landschützer, P., Bushinsky, S. M. & Gray, A. R. *A combined globally mapped carbon dioxide (CO₂) flux estimate based on the Surface Ocean CO₂ Atlas Database (SOCAT) and Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) biogeochemistry floats from 1982 to 2017 (NCEI Accession 0191304)* (NOAA National Centers for Environmental Information, 2019).
26. Wu, Y. et al. Integrated analysis of carbon dioxide and oxygen concentrations as a quality control of ocean float data. *Commun. Earth Environ.* **3**, 92 (2022).
27. Wu, Y. & Qi, D. Inconsistency between ship-and Argo float-based pCO₂ at the intense upwelling region of the Drake Passage, Southern Ocean. *Front. Mar. Sci.* **9**, 1002398 (2022).
28. Wynn-Edwards, C. A. et al. Subantarctic pCO₂ estimated from a biogeochemical float: comparison with moored observations reinforces the importance of spatial and temporal variability. *Front. Mar. Sci.* **10**, 1231953 (2023).
29. Álvarez, M. et al. Global ocean spectrophotometric pH assessment: consistent inconsistencies. *Environ. Sci. Technol.* **54**, 10977–10988 (2020).
30. Wimart-Rousseau, C., Steinhoff, T., Klein, B., Bittig, H. & Körtzinger, A. Enhancement of float-pH data quality control methods: A study case in the Subpolar Northwestern Atlantic region. *Biogeosciences Discuss.* **2023**, 1–26 (2023).
31. Takeshita, Y., Johnson, K. S., Martz, T. R., Plant, J. N. & Sarmiento, J. L. Assessment of autonomous pH measurements for determining surface seawater partial pressure of CO₂. *J. Geophys. Research: Oceans.* **123**, 4003–4013 (2018).
32. Coggins, A. et al. Surface ocean carbon budget in the 2017 south Georgia diatom bloom: observations and validation of profiling Biogeochemical Argo floats. *Deep Sea Res. Part II.* **209**, 105275 (2023).
33. Bushinsky, S. M. & Cerovečki, I. Subantarctic mode water biogeochemical formation properties and interannual variability. *Agu Adv.* **4**, e2022AV000722 (2023).
34. Gregor, L., Lebehoh, A. D., Kok, S. & Scheel Monteiro, P. M. A comparative assessment of the uncertainties of global surface ocean CO₂ estimates using a machine-learning ensemble (CSIR-ML6 version 2019a)—have we hit the wall? *Geosci. Model Dev.* **12**, 5113–5136 (2019).
35. Riser, S. C. T. et al. (2024). Alison. SOCCOM and GO-BGC float data - Snapshot 2023-12-20.
36. Maurer, T. L., Plant, J. N. & Johnson, K. S. Delayed-mode quality control of oxygen, nitrate, and pH data on SOCCOM biogeochemical profiling floats. *Front. Mar. Sci.* **8**, 683207 (2021).
37. Williams, N. L. et al. Empirical algorithms to estimate water column pH in the Southern Ocean. *Geophys. Res. Lett.* **43**, 3415–3422 (2016).
38. Johnson, K. S. et al. Deep-Sea DuraFET: A pressure tolerant pH sensor designed for global sensor networks. *Anal. Chem.* **88**, 3249–3256 (2016).
39. Sharp, J. D. et al. & Wallace. CO2SYSv3 for MATLAB (v3.2.1). Zenodo (2023).
40. Lueker, T. J., Dickson, A. G. & Keeling, C. D. Ocean pCO₂ calculated from dissolved inorganic carbon, alkalinity, and equations for K₁ and K₂: validation based on laboratory measurements of CO₂ in gas and seawater at equilibrium. *Mar. Chem.* **70**, 105–119 (2000).
41. Dickson, A. G. Standard potential of the reaction: AgCl (s) + 12H₂ (g) = Ag (s) + HCl (aq), and the standard acidity constant of the ion HSO₄⁻ in synthetic sea water from 273.15 to 318.15 K. *J. Chem. Thermodyn.* **22**, 113–127 (1990).
42. Lee, K. et al. The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans. *Geochim. Cosmochim. Acta.* **74**, 1801–1811 (2010).
43. Perez, F. F. & Fraga, F. Association constant of fluoride and hydrogen ions in seawater. *Mar. Chem.* **21**, 161–168 (1987).
44. Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem. Cycles.* **8**, 65–80 (1994).
45. Lauvset, S. K. et al. GLODAPv2. Global Ocean Data Analysis Project version 2.2023 (2023). (2023).
46. Olsen, A. et al. An updated version of the global interior ocean biogeochemical data product, GLODAPv2. 2020. *Earth Syst. Sci. Data.* **12**, 3653–3678 (2020).
47. Bakker, D. C. et al. A multi-decade record of high-quality fCO₂ data in version 3 of the Surface Ocean CO₂ Atlas (SOCAT). *Earth Syst. Sci. Data.* **8**, 383–413 (2016).
48. Pfeil, B. et al. A uniform, quality controlled Surface Ocean CO₂ Atlas (SOCAT). *Earth Syst. Sci. Data.* **5**, 125–143 (2013).
49. Gregor, L., Gruber, N. & OceanSODA-ETHZ A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data Discussions* 1–42 (2020). (2020).
50. Dinniman, M. S., Klinck, J. M. & Smith, W. O. Jr A model study of Circumpolar Deep Water on the West Antarctic Peninsula and Ross Sea continental shelves. *Deep Sea Res. Part II.* **58**, 1508–1523 (2011).
51. Sloyan, B. M. Antarctic bottom and lower circumpolar deep water circulation in the eastern Indian Ocean. *J. Geophys. Research: Oceans.* **111**, C02006 (2006).
52. Peterson, R. G. & Whitworth, I. I. The Subantarctic and Polar Fronts in relation to deep water masses through the southwestern Atlantic. *J. Geophys. Research: Oceans.* **94**, 10817–10838 (1989).
53. Gruber, N. Anthropogenic CO₂ in the Atlantic Ocean. *Glob. Biogeochem. Cycles.* **12**, 165–191 (1998).
54. Sabine, C. et al. Distribution of anthropogenic CO₂ in the Pacific Ocean. *Glob. Biogeochem. Cycles.* **16** (31-), 30 (2002).
55. Sabine, C. L. et al. The oceanic sink for anthropogenic CO₂. *science* **305**, 367–371 (2004).
56. Zhang, S. et al. Transport of anthropogenic carbon from the Antarctic shelf to deep Southern Ocean triggers acidification. *Glob. Biogeochem. Cycles.* **37**, e2023GB007921 (2023).
57. Lo Monaco, C., Metzl, N., Poisson, A., Brunet, C. & Schauer, B. Anthropogenic CO₂ in the Southern Ocean: Distribution and inventory at the Indian-Atlantic boundary (World Ocean circulation experiment line I6). *J. Geophys. Research: Oceans* **110**, 2004JC002643 (2005).
58. Sabine, C. et al. Anthropogenic CO₂ inventory of the Indian Ocean. *Glob. Biogeochem. Cycles.* **13**, 179–198 (1999).
59. Broecker, W. S. NO₃⁻, a conservative water-mass tracer. *Earth Planet. Sci. Lett.* **23**, 100–107 (1974).
60. Touratier, F. & Goyet, C. Definition, properties, and Atlantic Ocean distribution of the new tracer TrOCA. *J. Mar. Syst.* **46**, 169–179 (2004).
61. Yool, A., Oschlies, A., Nurser, A. & Gruber, N. A model-based assessment of the TrOCA approach for estimating anthropogenic carbon in the ocean. *Biogeosciences* **7**, 723–751 (2010).
62. Touratier, F. & Goyet, C. Applying the new TrOCA approach to assess the distribution of anthropogenic CO₂ in the Atlantic Ocean. *J. Mar. Syst.* **46**, 181–197 (2004).
63. Garcia & Gordon. Oxygen solubility in seawater - better fitting equations (VOL 37 & PG 1308., *Limnology and Oceanography* **38**, 656–656 (1993). (1992).
64. Garcia, H. E., Gordon, L. I. Oxygen solubility in & seawater - better fitting equations. *Limnol. Oceanogr.* **37**, 1307–1312 <https://doi.org/10.4319/lo.1992.37.6.1307> (1992).
65. Weiss, R. F., Price, B. A., Nitrous-oxide solubility & in water and seawater. *Mar. Chem.* **8**, 347–359 [https://doi.org/10.1016/0304-4203\(80\)90024-9](https://doi.org/10.1016/0304-4203(80)90024-9) (1980).
66. Weiss, R. F. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Mar. Chem.* **2**, 203–215 (1974).
67. Deming, W. E. Statistical adjustment of data. (1943).
68. Cohen, J. *Statistical power analysis for the behavioral sciences* (routledge, 2013).
69. Mooney, C. Z. *Monte carlo simulation* (Sage, 1997).
70. Gonçalves, S. & White, H. Bootstrap standard error estimates for linear regression. *J. Am. Stat. Assoc.* **100**, 970–979 (2005).

71. Carrillo, C. J., Smith, R. C. & Karl, D. M. Processes regulating oxygen and carbon dioxide in surface waters west of the Antarctic Peninsula. *Mar. Chem.* **84**, 161–179 (2004).
72. Manning, A. & Keeling, R. F. Global oceanic and land biotic carbon sinks from the Scripps atmospheric oxygen flask sampling network. *Tellus B: Chem. Phys. Meteorol.* **58**, 95–116. <https://doi.org/10.1111/j.1600-0889.2006.00175.x> (2006).
73. Chen, H., Haumann, F. A., Talley, L. D., Johnson, K. S. & Sarmiento, J. L. The deep ocean's carbon exhaust. *Glob. Biogeochem. Cycles* **36**, e2021GB007156 (2022).
74. Prend, C. J. et al. Indo-Pacific sector dominates Southern Ocean carbon outgassing. *Global Biogeochemical Cycles* **36**, e2021GB007226 (2022).
75. Stoer, A. C. et al. A census of quality-controlled Biogeochemical-Argo float measurements. *Front. Mar. Sci.* **10**, 1233289 (2023).
76. Johnson, K. S. et al. Biogeochemical sensor performance in the SOCCOM profiling float array. *J. Geophys. Research: Oceans.* **122**, 6416–6436 (2017).
77. Bushinsky, S. M. et al. Offset between profiling float and shipboard oxygen observations at depth imparts bias on float pH and derived pCO₂. *Glob. Biogeochem. Cycles* **39**, e2024GB008185 (2025).
78. Riser, S. C., Swift, D. & Drucker, R. Profiling floats in SOCCOM: Technical capabilities for studying the Southern Ocean. *J. Geophys. Research: Oceans.* **123**, 4055–4073 (2018).
79. Lauvset, S. K. et al. The annual update GLODAPv2. 2023: the global interior ocean biogeochemical data product. *Earth Syst. Sci. Data.* **16**, 2047–2072 (2024).
80. Takeshita, Y. et al. Assessment of pH dependent errors in spectrophotometric pH measurements of seawater. *Mar. Chem.* **223**, 103801 (2020).
81. Fong, M. B. & Dickson, A. G. Insights from GO-SHIP hydrography data into the thermodynamic consistency of CO₂ system measurements in seawater. *Mar. Chem.* **211**, 52–63 (2019).
82. García-Ibáñez, M. I. et al. Gaining insights into the seawater carbonate system using discrete fCO₂ measurements. *Mar. Chem.* **245**, 104150 (2022).

Acknowledgements

We acknowledge the generosity of the Swire Educational Trust in funding this work. We acknowledge the efforts of the data providers and scientists who conducted data collection, quality control, and synthesis, which ultimately resulted in the creation of the datasets used in this study. We also appreciate the other researchers who provided comments and constructive feedback on this work.

Author contributions

C.Z. and T.T. conceived the original idea and came up with methods of research. C.Z. performed the analytic calculations, coding and writing. T.T. supervised the project. Y.W, P.B., D.S. and A.S provided significant comments and specific help in coding, analysis and writing.

This research has been supported by the Swire Educational Trust.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-43863-4>.

Correspondence and requests for materials should be addressed to C.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026