

Project information	
Project full title	EuroSea: Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans
Project acronym	EuroSea
Grant agreement number	862626
Project start date and duration	1 November 2019, 50 months
Project website	https://www.eurosea.eu

Deliverable information	
Deliverable number	D7.2
Deliverable title	Development of BGC-Argo data quality validation based on an integrative multiplatform approach
Description	
Work Package number	WP7
Work Package title	Ocean climate indicators demonstrator
Lead beneficiary	GEOMAR
Lead authors	Cathy Wimart-Rousseau (GEOMAR), Marine Fourrier (SU/LOV), Björn Fiedler (GEOMAR)
Contributors	Romain Cancouët (Euro-Argo ERIC), Hervé Claustre (SU/LOV), Laurent Coppola (SU/LOV)
Due date	30.11.2022
Submission date	30.11.2022
Comments	



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862626

Table of contents

Executive summary	1
1. Introduction	2
1.1. Chemistry of Marine Carbon Dioxide System	2
1.2. Marine Carbon Dioxide System Parameter way of measurements	2
1.3. Autonomous seawater pH measurements: why and how?	3
2. Strategies and tools to correct pH data	4
2.1. General procedure for float-pH data correction	4
2.2. Independent estimates of pH data accuracy	5
2.2.1. CTD cast at float deployment	6
2.2.2. Matchup between CTD cast with water sampling and float profiles during float's life	6
2.2.3. Crossovers between other platforms and floats	9
3. Application and results acquired in the EuroSea studied region	10
3.1. Float array	10
3.2. Major offset correction	11
3.3. Comparison with profiles at deployment	16
3.4. Crossovers between CTD cast with water sampling and float profiles	16
3.5. Comparison between surface ocean carbon data and float	19
3.6. Ways to go forward with sensor failures?	19
Recommendations	23
Conclusions	25
Data availability statement	26
Acknowledgments	26
References	27

Executive summary

This report presents the results of Task 7.3 on “Development of BGC-Argo data quality validation based on an integrative multiplatform approach”. Observing changes in ocean conditions on the spatiotemporal scales necessary to constrain carbon uptake is a challenge. Defined as an Essential Ocean Variable (EOV) by the Global Ocean Observing System (GOOS, *e.g.*, Tanhua et al., 2019), pH is relevant to assess numerous crucial questions regarding the oceanic evolution in response to the current global changes. However, the large spatiotemporal variability of this carbonate system parameter requires sustained observations to decipher trends and punctual events. Within this scope, numerous pH sensors suitable for deployments both on autonomous observing tools and fixed stations have been developed. Nevertheless, as interpreting changes relies on accurate data, and because offsets or drifts in pH data might appear in response to changes in the sensor k_0 constant, a consistent and rigorous correction procedure to quality-control and process the data has been implemented. This report presents the application of this method to pH data acquired by BGC-Argo floats launched in the Tropical Atlantic area.

Disclaimer: This document represents the situation at the time of data evaluation and writing of the report which is primarily based on data from 2021/2022. As pH-equipped BGC-Argo floats and ASVs are still in operation or have to be reprocessed, more data is coming in and the database is growing daily. This will allow us to improve the statistics of our analyses and hence the robustness of the results. Therefore, the results presented here are based on the *status quo* and are not necessarily the final word on these matters. We therefore point out that further analyses will and need to be carried out. A more detailed and integrative assessment on the quality enhancement of carbon fluxes based on Argo-pH data acquired in the tropical Atlantic will be provided in EuroSea deliverable D7.6, to be submitted in 2023. Moreover, the results shown in this deliverable will be presented and discussed to that the scientific community during the next BGC-Argo community workshop in January 2023.

1. Introduction

1.1. Chemistry of Marine Carbon Dioxide System

Carbon is stored in three main reservoirs linked together by exchange fluxes, and it moves between these reservoirs through a variety of processes. The oceanic carbon cycle is a central process to the global carbon cycle, containing both organic and inorganic carbon. In the ocean, the dissolution of carbon dioxide (CO_2), that reacts with water, follows a series of chemical equilibria. These reactions, governed and connected by equilibrium reactions, give rise to four different chemical species that form the oceanic carbonate system: CO_2 in aquatic solution, carbonic acid, bicarbonate and carbonate (*i.e.*, $\text{CO}_2(\text{aq})$, H_2CO_3 , HCO_3^- , CO_3^{2-}). Nevertheless, individual species of the carbonate system cannot be measured directly. To overcome this issue, four parameters that can be measured at high accuracy have been developed to completely describe (with ancillary information) the CO_2 system in seawater (Dickson et al., 2007).

These four parameters are:

- Total alkalinity (TA),
- Fugacity/partial pressure of CO_2 in gas phase in chemical equilibrium with seawater ($f\text{CO}_2/p\text{CO}_2$),
- Total hydrogen ion concentration (pH),
- Total dissolved inorganic carbon (DIC).

Theoretically, because of the relative consistency of the chemical constituents of seawater, a complete description of the marine CO_2 system can be obtained based on only two of the four measurable carbon parameters, together with the equilibrium constant, temperature, pressure and salinity (Dickson et al., 2007).

1.2. Marine Carbon Dioxide System Parameter way of measurements

To assess long-term changes in ocean chemistry, accurate and sustained time-series datasets are needed to decipher long-term trends to seasonal changes, and to better constrain and predict future changes. Over the last few decades, numerous oceanographic cruises (*e.g.*, PIRATA cruises; Foltz et al. 2019) or fixed stations measurements (buoy, moorings) have been conducted in the Tropical Atlantic area. However, these cruises do not cover full temporal (seasonal) cycles and have low spatial resolution, leading to biased observations. Indeed, in response to biological, physical, and chemical processes, in addition to anthropogenic modifications, the dynamic marine CO_2 system changes on daily to centennial timescales, with seasonal, interannual, and decadal variabilities. In consequence, ship-based observing strategies, being skewed towards peculiar months and regions, especially in some places where current sampling methods are not possible (*e.g.*, oceanic areas where sea ice occurs), cannot capture the dynamic spatiotemporal variability of the carbonate system parameters.

In consequence, datasets based on these historical sampling strategies have “observational gaps” (Tanhua et al., 2019) that have to be filled. To circumvent these gaps and overcome the remoteness of numerous undersampled areas, autonomous platforms such as moorings, profiling floats, underwater gliders, or autonomous surface vehicles have been deployed at regional scales, mainly thanks to the development of autonomous sensors. In comparison to ship-based measurements, these platforms have numerous advantages: the overall cost of deployment, platforms can remain in an area of interest for a long time, and are non-impacted by weather conditions, minimal human resources needed, and measurements closer to the ocean surface.

Recently, autonomous $p\text{CO}_2$ and pH sensors suitable for deployment on autonomous surface vehicles have been developed (Martz et al., 2015), making autonomous *in situ* measurements more accessible to the community. Hereafter, this report will focus on data acquired by autonomous pH sensors and the procedure to correct these data.

1.3. Autonomous seawater pH measurements: why and how?

Defined as an Essential Ocean Variable (EOV) by the Global Ocean Observing System (GOOS¹), pH can be used, among others parameters, to determine oceanic changes in response to anthropogenic impacts and particularly the oceanic acidification. Nevertheless, while pH is depicted as a master variable in marine chemistry, its measurements do not allow direct quantification of derived phenomena such as air-sea CO_2 exchanges. Moreover, it should also be pointed out that numerous tools have been developed to overcome data limitations that can happen in certain areas (ship-time limitations, high cost, human resources, weather conditions). Thus, in the absence of data, a reasonable work-around is to predict TA using algorithms that employ variables such as temperature, salinity, pressure and O_2 together with position in space and time which are also measured on floats. The alkalinity estimation produced by the LIR (the Locally Interpolated Regression method; Carter et al., 2018) tool, which uses data from the Global Ocean Data Analysis Project (GLODAPv2) data set (Olsen et al., 2016), is one of these options. In addition to the LIR algorithmic method, neural-network based methods such as the CANYON neural networks (the Carbonate system and Nutrient concentration from hydrological properties and Oxygen method; Sauzède et al., 2017), and its revised version the CANYON-B approach (Bittig et al., 2018) can be considered to estimate these quantities. Generally, TA predictions have uncertainties ranging between $6 \mu\text{mol kg}^{-1}$ and *ca.* $10 \mu\text{mol kg}^{-1}$ (e.g., Bittig et al., 2018; Williams et al., 2017). As the data coverage of the GLODAPv2 data product is not uniform in space and time (seasonal cycle), the robustness of TA algorithms based on it (or other data products) is not uniform either. Therefore, regional and/or seasonal biases that may compromise the accuracy of DIC and $p\text{CO}_2$ calculated from measured pH and predicted TA are to be expected.

The principle of measurement for seawater pH values obtained by autonomous platforms (such as BGC-Argo floats or gliders) is based on the Ion Sensitive Field Effect Transistor (ISFET) pH sensor, developed by Honeywell Durafet®, which produces highly precise and stable pH measurements in seawater, with a precision better than 0.005 pH units (Martz et al., 2010). Nevertheless, as this sensor is not pressure tolerant, since 2010, significant engineering has been put into these Durafet-based pH sensors to overcome this issue. Thus, based on the Honeywell Durafet ISFET sensor technology, on BGC-Argo floats, two types of pH sensors are implemented: the Deep Sea DURAFET (Johnson et al., 2016), deployed on Apex floats, and the SBE Float Deep SeaFET built by Sea-Bird Scientific. Typically installed on the head of the float, the core of these sensors is an ISFET that responds to proton activity and an Ag/AgCl reference electrode that responds to chloride ion activity (Johnson et al., 2016). pH is derived proportionally from the voltage between the ISFET and the reference electrode (Bittig et al., 2019). For SBE pH sensors, the accuracy ranges from ± 0.05 pH units (manufacturer statement) to ± 0.005 pH units after data adjustment (Johnson et al., 2018). Moreover, it can be noted that significant efforts are underway to also advance optode technology for pH. While earlier versions of the pH optode had significant issues with, for example, drift and response time, recent advances (e.g., AquapHOx platform with pH optode, PyroScience GmbH, Aachen, Germany) are promising, which may warrant dedicated field testing.

¹ <https://www.goosocean.org>

In the framework of the EuroSea project, task 7.3 aims to develop indicators for carbon flux observations in this region based on the improvement of existing components and on the deployment of new observing tools. Thus, this task considers the use of, among other autonomous tools such as moored buoys and autonomous surface vehicles, BioGeoChemical- (BGC) Argo floats for the acquisition of high-quality carbon measurements at a regional scale. In the Eastern Tropical North Atlantic region, a total of 5 pH/O₂ Provor floats from the NKE instrumentation manufacturer have been purchased and deployed in 2021 (Fig. 1).

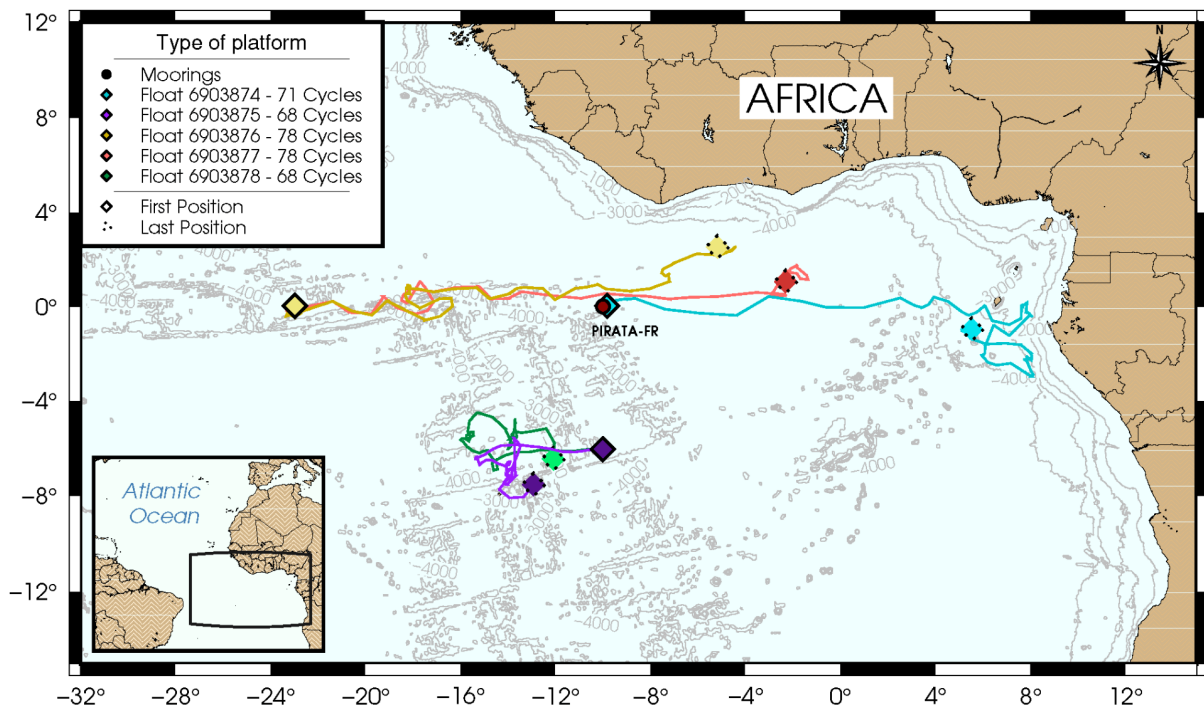


Figure 1. Trajectories of the five EUROSEA pH-equipped BGC-Argo floats. Dotted points show the last locations as of October 17th, 2022. Floats were deployed between March 12th 2021 and April 3rd, 2021.

2. Strategies and tools to correct pH data

The key to this autonomous platform expansion is the quality of the datasets which relies on sensor pre-deployment rigorous calibration and post-deployment consistent data qualification that can be used globally (Johnson et al., 2018). Indeed, interpreting changes relies on accurate data. For Argo floats, operational procedures for physical data (temperature, salinity, pressure) qualification have been established, from real-time checks to delayed-mode adjustment (Schmechtig et al., 2016; Wong et al., 2022). Indeed, float-pH sensors, like other autonomous sensors such as oxygen optodes, are subject to drift over time. Consequently, numerous delayed-mode assessments for *in situ* chemical sensors measuring pH have been suggested, but having consistent and traceable correction methods is still challenging. In the remainder of this report, corrective procedures detailed in Johnson et al. (2018) will be presented and applied to our dataset.

2.1. General procedure for float-pH data correction

Inside each float equipped with a pH sensor, an internal sensor algorithm converts the measured potential into pH on the total proton scale using laboratory-based calibration coefficients. Thus, pH sensors are

calibrated in the laboratory using spectrophotometric measurements and are directly related to the laboratory calibration method. Sensors pressure and temperature coefficients, needed to compute the *in situ* pH, are also determined in the laboratory as described in Johnson et al. (2016).

Firstly, the float-measured pH undergoes simple automatic Real-Time (RT) Quality-Control (QC) checks that flag as outlier (bad) pH data which do not fall in the range of 7.5 to 8.5. Nonetheless, to be usable as research-quality pH data, an additional adjustment has to be applied to real-time adjusted data. The additional adjustments that are applied to RT data after human verification convert them into the Delayed-Mode (DM) data. Two specific corrections can be observed:

- The offset correction: This correction corrects the significant (generally negative) offset observed with float-based pH measurements. This necessary adjustment relies on an external reference that is used to calculate at-depth (typically around 1500 dbars) offset between measured and estimated reference data. Then, this at-depth anomaly is then applied as a uniform offset correction throughout the water column, based on evidence that drifts and offsets are constant throughout an entire profile (Johnson et al., 2013, 2018). The adjustments necessary to match the sensor pH to the reference results are made by adding a constant offset to the reference potential of the sensor, rather than to pH, directly, because sensor drift results from a reference potential change (Johnson et al., 2016).
- The drift correction: This correction removes the temporal drift observed in the pH data as it emerges from the evolution of the offset correction over a float's lifetime.

Recently, in the framework of the Southern Ocean Carbon and Climate Observations and Modelling project (SOCCOM; Russell et al., 2014), a methodology has been developed for BGC-Argo floats to correct nitrate, pH, and oxygen values from sensor drifts, gains, and offsets in delayed-mode. Two Matlab tools named SAGE (SOCCOM Assessment and Graphical Evaluation) and SAGEO2 have also been created as user interfaces to correct and validate the data. In this procedure, detailed in Maurer et al. (2021), the difference between a selected field of reference and the measured values is first calculated at a depth below 1000 m, where spatiotemporal variability of oceanic components is minimal. Then, for each segment, determined by a cost function, a linear least-square fit is determined, and the modelled correction, computing offsets and slopes, is applied as an offset to the reference potential k_0 of the pH sensor (Johnson et al., 2016). Here, machine learning methods (the CANYON method and its revised version CANYON-B) and algorithmic (*e.g.*, the LIR method or the Multiple Linear Regression equations; Carter et al., 2018; Williams et al., 2016) methods are used as the reference data mandatory to correct the pH datasets. As all of these prediction methods require oxygen as an input variable for computing predicted pH, oxygen must first be corrected (Johnson et al., 2018) as oxygen optodes are known to suffer from storage drift (Johnson et al., 2015). SAGEO2 must therefore be used to correct float-oxygen data before any float-pH data correction. In our case, O_2 from the 5 EUROSEA pH-equipped Argo floats was adjusted following Argo procedures and the adjustments are available in Near-Real Time.

2.2. Independent estimates of pH data accuracy

In addition to the correction of float-pH, further steps are needed to independently estimate pH data accuracy and, if needed, apply additional corrections. In the frame of the EuroSea project, numerous tools, platforms, and cruises have been deployed and carried out, allowing us to do crossovers comparisons.

2.2.1. CTD cast at float deployment

At the time and place of the five floats deployments, CTD casts with discrete water sampling have been performed. For marine carbon dioxide system description, samples for dissolved inorganic carbon and total alkalinity were collected during the PIRATA-FR31 cruise in compliance with the corresponding standard procedure (SOP 1) defined in the “Guide to Best Practices for Ocean CO₂ Measurements” (Dickson et al., 2007). Measurements of DIC and TA were performed simultaneously by potentiometric acid titration using a closed cell following the methods described by Edmond (1970) and Dickson and Goyet (1994). Analyses were performed at the National Facility for Analysis of Carbonate System parameters (SNAPO-CO₂, LOCEAN, Sorbonne University - CNRS, France). The average accuracy of DIC and TA analysis (estimated from repeat measurements of Certified Reference Material provided by Prof. Dickson’s laboratory from the Scripps institution of Oceanography, UC San Diego, USA) was 3.0 and 3.5 $\mu\text{mol kg}^{-1}$, respectively. Then, to allow comparison against float-pH data, seawater carbonate parameters, including *in situ* pH on the total scale, were derived from DIC and TA with the software program CO2SYS-MATLAB (van Heuven et al., 2011). Note that the estimated uncertainty associated with pH derived from DIC and TA computed with CO2SYS (Orr et al., 2018; van Heuven et al., 2011) is 0.0088 pH units.

Table 1. Crossovers between the first profile from pH-equipped BGC-Argo floats and CTD rosette hydrocasts with carbonate parameters sampling in the Eastern Tropical North Atlantic region during the PIRATA-FR31 cruise. The pH difference is the mean absolute difference \pm standard deviation at the matchup depths.

WMO#	Reference hydrocast from PIRATA-FR31	Distance difference (km)	Date difference (days)	pH difference at deployment before adjustment	pH difference at deployment after adjustment
6903874	00049	4.722	0.25	0.0576 \pm 0.0572	0.0361 \pm 0.0245
6903875	00020	4.397	0.76	4 \pm 2.9005	4 \pm 2.9005
6903876	00070	1.824	0.28	0.0201 \pm 0.0179	0.0200 \pm 0.0180
6903877	00070	2.835	0.28	0.0118 \pm 0.0307	0.0192 \pm 0.0219
6903878	00020	5.635	0.76	0.0559 \pm 0.0442	0.0113 \pm 0.0352

For float 6903875, the pH difference at deployment is large as the pH sensor malfunctioned from the start. For the other BGC-Argo floats, the average difference over the entire water column at deployment is of 0.0364 pH units. There is no systematic bias of under- or overestimation from the pH sensor.

2.2.2. Matchup between CTD cast with water sampling and float profiles during float’s life

In situ pH data are generally considered as reference data. Nevertheless, under normal circumstances, it would be critical to obtain specific crossovers between CTD casts and floats profiling during a float’s lifetime without impact field work schedule. In the Eastern Tropical North Atlantic region, successful crossovers have been achieved between data acquisition platforms and pH measurements done during the PIRATA-FR32

cruise² that sampled, for the first time, pH onboard. These successful matchups happened thanks to the cooperation of the PIRATA team onboard the cruise and in a joint effort with Euro-Argo ERIC and EuroSea BGC-Argo floats.

By changing the floats' sampling frequency, onboard pH sampling and analysis close to Argo floats profiles was achieved (Fig. 2, Table 2). The Argo profiles were 1-4 days before or after the PIRATA-FR stations (Argo cycles 58 on March 22nd, 2022 vs. stations st037c01 and st038c01 on March 23rd, 2022 and March 24th, 2022; Argo cycle 57 for float 6903877 on March 12th, 2022 vs station st008c01 on March 08th, 2022). Table 1 details the values of the different variables used for the comparison. The physical parameters show that relatively similar water masses were sampled by both observing platforms with differences mainly due to location. Furthermore, the mean absolute difference over the entire water column between the raw pH from the BGC-Argo floats and the *in situ* reference data is of 0.0622 pH units. This difference is twice as large as the MAD at deployment, confirming the need for correction procedures. In these comparisons, pH from the Argo floats varies between 7.6031 and 8.0173 whereas onboard pH spans from 7.7151 to 8.0101. However, for the matchup around March 22nd, 2022 (first row of Fig. 8), BGC-Argo floats only profiled down to 1000 dbar whereas the *in situ* stations sampled down to 2000 dbar.

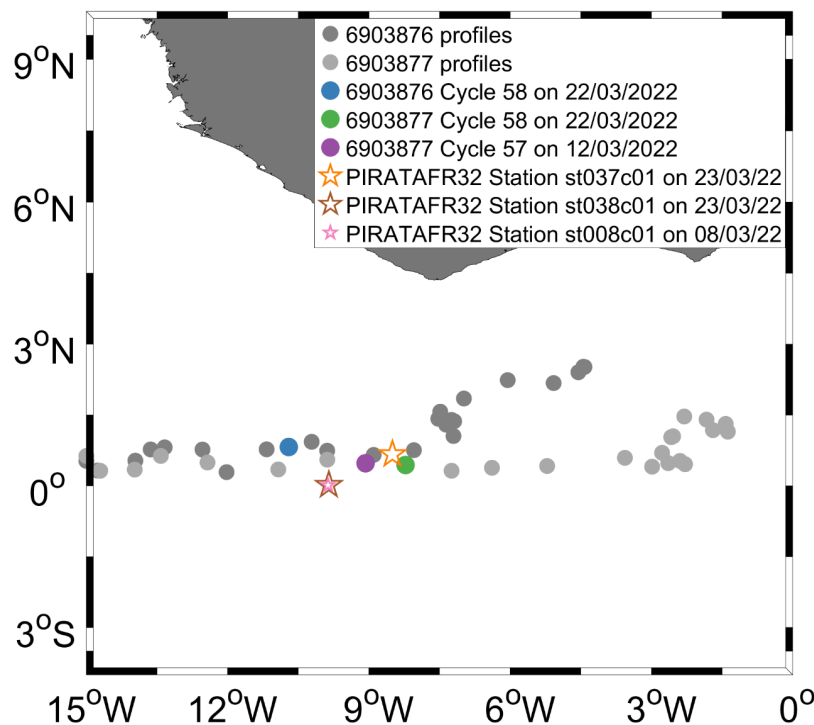


Figure 2. Map of the Argo profiles from floats 6903876 and 6903877 and PIRATA-FR32 stations used for the matchup around March 23rd, 2022. Grey and light-grey dots represent Argo profiles and coloured dots (blue for 6903876 and green and purple for 6903877) are the profiles closest in date to the cruise stations. Orange, pink, and brown stars correspond to the PIRATA-FR32 stations with onboard pH measurements on March 23rd, 2022. Note that the pink star is almost superimposed on the brown star, its size has been reduced to improve visibility.

² https://www.brest.ird.fr/pirata/pirata_cruises.php, last accessed October 18th, 2022.

Table 2. Mean values of temperature, salinity, dissolved oxygen and pH measurements made by the BGC-Argo floats on the profiles selected for the comparison, the corresponding shipborne data and the difference. SD stands for Standard Deviation. Dissolved oxygen and pH were adjusted following BGC-Argo quality control procedures.

		Float±SD	Shipborne ± SD	Mean absolute difference ± SD	Date difference (days)	Distance difference (km)
6903876 cycle 58 vs PIRATA- FR32 station 37	Temperature	15.76±7.63	13.63±9.07	0.60±0.80	1.7	245.3
	Salinity	35.17±0.49	35.25±0.54	0.18±0.30		
	O ₂	141.67±35.97	153.02±51.36	8.91±7.97		
	pH	7.7518±0.0996	7.8730±0.1012	0.1330±0.0259		
6903876 cycle 58 vs PIRATA- FR32 station 38	Temperature	15.76±7.63	13.75±9.28	0.98±1.28	2.1	129.3
	Salinity	35.17±0.49	35.23±0.63	0.20±0.20		
	O ₂	141.67±35.97	155.79±53.14	13.74±8.68		
	pH	7.7518±0.0996	7.8797±0.1098	0.1411±0.0327		
6903877 cycle 58 vs PIRATA- FR32 station 37	Temperature	15.87±7.42	13.63±9.07	0.57±0.46	1.7	39.8
	Salinity	35.36±0.49	35.25±0.54	0.07±0.08		
	O ₂	143.44±38.24	153.02±51.36	10.32±9.12		
	pH	7.8892±0.1054	7.8730±0.1012	0.0210±0.0079		
6903877 cycle 58 vs PIRATA- FR32 station 38	Temperature	15.87±7.42	13.75±9.28	0.78±0.21	2.1	187
	Salinity	35.36±0.49	35.23±0.63	0.17±0.22		
	O ₂	143.44±38.24	155.79±53.14	9.57±7.50		
	pH	7.8892±0.1054	7.8797±0.1098	0.0201±0.0144		
6903877 cycle	Temperature	14.88±8.02	12.94±8.97	0.66±0.61	4	102.2

		Float±SD	Shipborne ± SD	Mean absolute difference ± SD	Date difference (days)	Distance difference (km)
57 vs PIRATA- FR32 station08	Salinity	35.25±0.49	35.27±0.62	0.11±0.10		
	O ₂	148.27±46.20	164.03±46.51	12.77±7.94		
	pH	7.8859±0.1006	7.8915±0.0963	0.0257±0.0193		

2.2.3. Crossovers between other platforms and floats

A Saildrone (SD) platform equipped with an ASVCO₂ system (PMEL, NOAA) has been deployed in the EuroSea mission operating area and recorded data between September 18th, 2021 and March 8th, 2022. While this autonomous uncrewed surface ocean vehicle was not equipped with a pH sensor, it can be used for comparison with BGC-Argo floats as *p*CO₂ calculated data can be indirectly obtained from surface float-pH measurements and neural network TA. Thanks to the high degree of coordination between the SD team, Euro-Argo and Principal Investigators within this EuroSea task, an intercomparison experiment has been successfully conducted. By changing the Argo floats' cycling frequency (to daily profiles), profiles from 2 pH-equipped BGC-Argo floats (WMOs 6903876 & 6903877) were collected on November 15th and 16th 2021 while the Saildrone was circumnavigating around their estimated next ascent profile location (Fig. 3). Therefore, at the location and time of the crossover between the BGC-Argo floats and the Saildrone, *p*CO₂ was computed using adjusted Argo pH and TA computed using the ESPER (Empirical Seawater Property Estimation Routines; Carter et al., 2021) neural-network based method (specifically using the mixed version, which combines MLR and neural network outputs). Figure 3 indicates locations at which crossovers between these platforms were done.

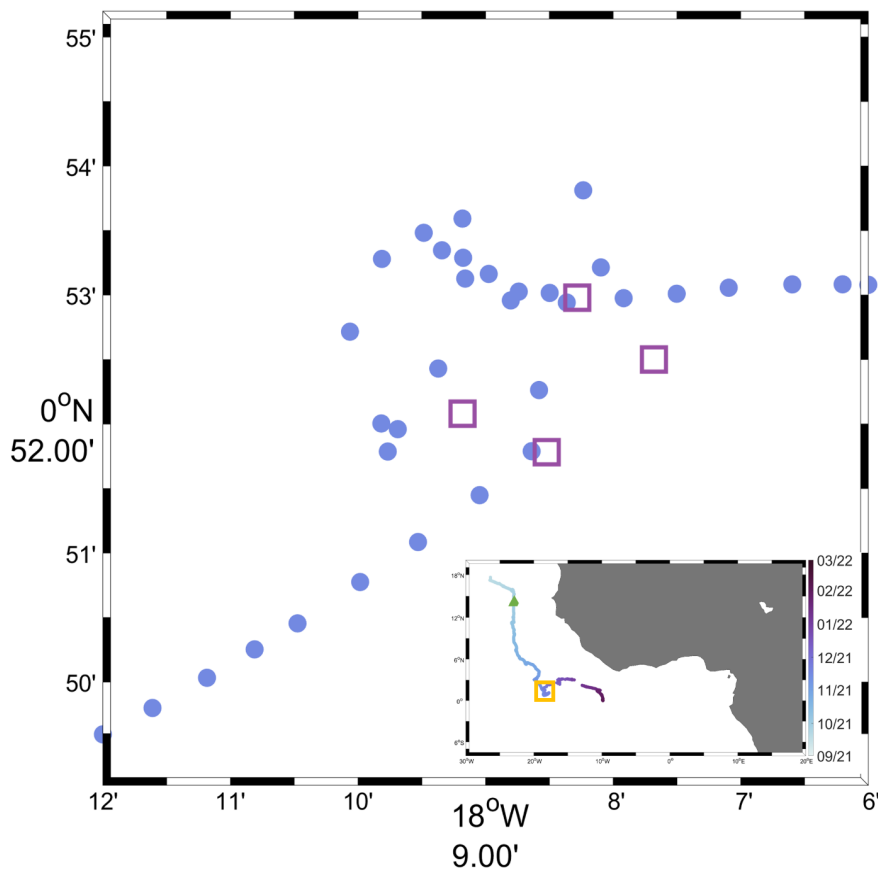


Figure 3. Map of the Saildrone daily measurements (dots) with the Argo profiles (squares) from floats 6903876 and 6903877 used for the matchup around November 16th, 2021. The encased figure on the lower right indicates the location of the focused map in relation to the entire Saildrone mission. The corresponding SD data is mapped according to time. The green triangles correspond to the SOOP data used for the matchup.

Furthermore, the France-Brazil SOOP (Ship of Opportunity Program; Goni et al., 2010) line makes underway carbonate chemistry measurements on the cargo ship Cap San Lorenz. The France-Brazil line began measuring sea surface $p\text{CO}_2$ in July 2014, and LOCEAN/IPSL in France is the responsible institution for this effort, with Dr. Nathalie Lefèvre as PI. $p\text{CO}_2$ surface measurements are performed between Le Havre, France and Santos, Brazil (Watson et al., 2018). The route crossed the Saildrone trajectory, therefore, it can also be used as a comparison platform (green triangles in Figure 3). The SOOP data were matched to Saildrone measurements within 10 km (6.82 ± 2.37 km apart). They are however 65 days apart (65.70 ± 0.26 days apart) from the SD measurements. For comparison purposes, they have been represented at the time of the corresponding SD measurements in Figure 9 to compare at the same location.

3. Application and results acquired in the EuroSea studied region

3.1. Float array

In the framework of the EuroSea project, the emerging BGC-Argo float network has been strengthened in the Eastern Tropical North Atlantic region where a total of 5 pH/O₂ floats from the NKE instrumentation manufacturer have been purchased and deployed (Fig. 1, Table 3). These five BGC-Argo, deployed during the

PIRATA-FR31 cruise between March and April 2021, are still profiling. They are programmed with a 10-day cycle, 1000 dbar parking depth and profile from their parking depth with a monthly 2000 dbar profile.

Table 3. Status and deployment information of the BGC-Argo floats with pH and O₂ sensors deployed in the Eastern Tropical North Atlantic region as part of the EuroSea project.

WMO#	Float Type	Deployment Time (dd/mm/yy yy)	Number of cycles (as of Oct. 18 th 2022)	Status	SEAFET pH sensor SN	First profile maximum depth (db)	Sampling plan
6903874	Provor / NKE	28/03/2021	71	Active	10745	1000	1000 dbar profile every 10 days Monthly 2000 dbar profile
6903875	Provor / NKE	12/03/2021	68	Active, pH sensor failure	10746		
6903876	Provor / NKE	03/04/2021	78	Active, pH drift then pH sensor failure from cycle 60	10754		
6903877	Provor / NKE	03/04/2021	78	Active	10743		
6903878	Provor / NKE	12/03/2021	68	Active, pH drift	10699		

Unfortunately, the deployed floats suffered from an unusual number of manufacturer-related technical issues or failures of the pH sensor itself. Out of those 5 floats, 2 pH sensors experienced drift while two had complete pH sensor failure. For float 6903875 in sensor failure, pH values were aberrant from the 1st cycle onwards (values up to 12). This behaviour might be related to a problem with the reference electrode which was reported to have occurred over the serial number range 10000 to 11117 (Communication Ken Johnson, AST-23 meeting, April 2022) and caused the affected pH sensor to measure wrong pH data relatively shortly after deployment. The pH sensor was therefore turned off after cycle 25 to save battery power. After drifting almost since its deployment, a second float's pH sensor has been remotely turned off because of sensor failure (6903876, cycle 60 onwards) as of September 2022. This has severely compromised the amount of usable pH data. In the end, only a little more than half of the floats will be useful. Currently, only one float's pH sensor is drifting (6903878), but this is correctible in Delayed Mode. Furthermore, as floats 6903876 & 6903877 remained close in space, we can compare their pH during adjustment procedures. In the following, we will illustrate how float-pH data can be corrected in this area and lessons learned from this multi-platform deployment strategy followed in the studied region.

3.2. Major offset correction

Uncorrected float-pH showed a significant offset to the pH from reference profiles, with float-pH data being every time lower and on all profiles. The correction applied in the studied area was carried out using the SAGE GUI software developed at MBARI (see section 2.1). In this code, reference data can be calculated

according to numerous methods (see section 2.1). Hereafter, only the two distinct methods LIR and CANYON-B have been used to calculate reference data as the set of multiple linear regression models proposed by Williams et al. (2016) has been developed to estimate water column pH in the Southern Ocean. The results of the calculation variants are shown for an example profile in Figure 4. In the following we will briefly explain the choices made to correct float-pH data using this tool.

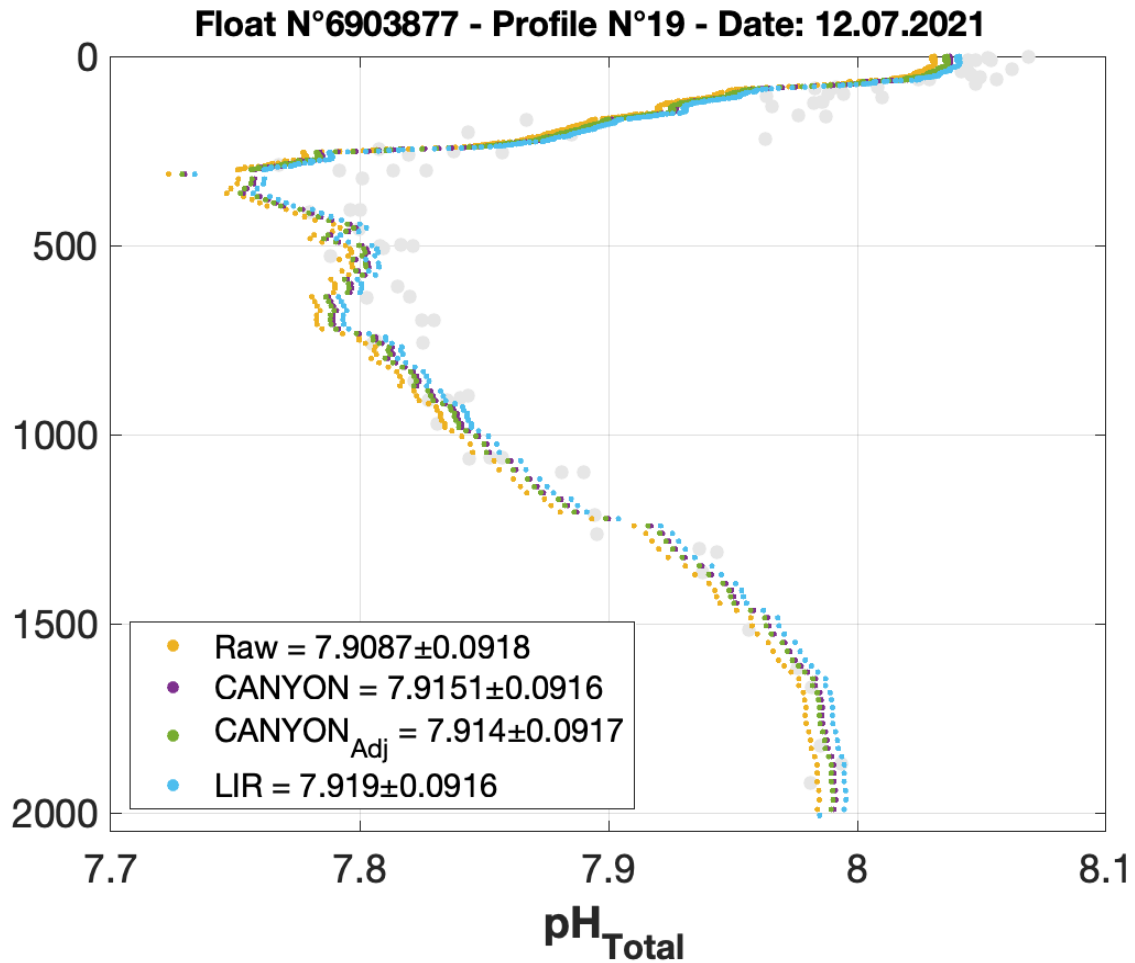


Figure 4. Example float profile (WMO 6903877, cycle 19). Shown are raw pH data (yellow dots), pH corrected by SAGE against CANYON-B (purple dots), pH corrected by SAGE against CANYON-B itself modified according to the adjustment proposed by Carter et al. (2021) (green dots) and pH corrected by SAGE against LIR (cyan dots) superimposed on the GLODAPv2 data corresponding to the studied area (gray dots). Note that the purple and green lines coincide in the plot as the two corrections agree to within 0.0011 pH units.

Reference methods used in the pH data adjustment

In SAGE, in this oceanic region, two distinct reference methods can be employed: the LIR pH regression (LIPHR) method or the CANYON-B method. The LIPHR algorithm uses regression coefficients determined on a 5° latitude and longitude grid box with 33 different depth surfaces determined using multiple linear regression models trained with GLODAPv2 (Olsen et al., 2016). For float-pH data correction, latitude, longitude, depth, salinity, temperature, and dissolved oxygen are used as predictor variables in the SAGE software (LIR regression #7). The uncertainty estimated for pH measurements between 1000 meters and

2000 meters, *i.e.*, the approximate depth range at which pH sensors are corrected, is $0.001 (\pm 0.006)$ pH units (bias \pm RMSE; Carter et al., 2018).

The neural network Bayesian approach, CANYON-B, is a mapping performed in a Bayesian network informed by an ensemble of model components at each stage rather than fixed values. In their publication, Bittig et al. (2018) compare the CANYON-B and LIRPH performances for, among others, the pH against the post-GLODAPv2 validation dataset, and showed global RMSE's for this revised form of the CANYON-B approach and the LIR method equal to 0.013 pH units and 0.016 pH units, respectively. Using the SOCCOM array, Maurer et al. (2021) calculated CANYON-B and LIRPH pH estimates and observed a larger uncertainty toward the surface than at 1500 meters depth, with mean differences (CANYON-B pH minus LIRPH pH data) of -0.025 and -0.001 pH units, respectively. This surface discrepancy can be explained by the difficulty for algorithms to represent well seasonal variability changes and air-sea gas exchanges.

In addition, a final note should be made regarding the LIR methods that also have the limitation that they are unable to capture changes in the relationships between the estimated properties and predictor properties (Carter et al., 2018). Indeed, as the LIR method relies on measurements made over a large time span, oceanic pH changes due to increasing atmospheric carbon dioxide concentrations and /or in areas with strong regional differences in the degree of water mass ventilation and anthropogenic carbon storage are oftentimes under-described by it. While an optional ocean acidification adjustment can be added to the LIR algorithm, this is a static adjustment that does not account for regional differences.

Finally, as stated by Maurer et al. (2021), an independent way to determine which model to use is to use the Bayesian Information Criteria (BIC; Chen and Chen, 2008), provided by the SAGE software, which is a statistical criterion providing a measure of the model performance: the model with the lowest BIC value is, generally, preferred. In our case, BIC values were always lower for the CANYON-B method in comparison with the LIR method (data not shown). For all of these reasons, CANYON-B has been chosen as the reference method to correct float-pH data in this oceanic region.

Reference-pH data adjustment

Over time, changes in ocean pH measurement practices have occurred, leading to a variety of ways to measure and calculate pH. In consequence, heterogeneities in pH data compilations used to train algorithms methods are observed. To overcome this issue, Carter et al. (2018) applied a range of corrections and created the most consistent pH data product presently available. Using this dataset for training (containing pH in line with “purified spectrophotometric pH”), the pH data product in CANYON-B was then adjusted to be in line with pH calculations from TA and DIC (Bittig et al., 2018, Carter et al., 2018). In consequence, pH estimates generated by this method are in line with calculated-pH.

In the SAGE software, an optional CANYON-B pH data adjustment can be applied to bring estimates into alignment with spectrophotometrically pH measurements made using purified dye following Carter et al. (2018, Equation 1). Even if there are still some unresolved issues in homogenising and correcting different pH data observation techniques (Bittig et al., 2018), we have decided to keep this reference-pH data adjustment to correct float pH data as recommended by Carter et al. (2018) and Johnson et al. (2018). In addition, as pH sensors are calibrated in the laboratory using spectrophotometric measurements with purified dyes, sensor measurements should be directly related to the laboratory calibration method. This assessment is reinforced by the recent work from Takeshita et al. (2020) which described measurements

made with purified dyes consistent with measurements made by sensors that have been shown to have the expected Nernstian³ (Bergveld, 2003) response to pH changes.

At-depth correction

In the literature, corrections of float-pH data implied the comparison of raw float data to select reference fields at depths around 1500 m (1480 - 1520 m) where spatial and temporal variability in ocean chemistry is supposed to be minimal (*e.g.*, Maurer et al., 2021). Then, the correction determined at depth is applied to the entire profile. Nonetheless, for the float-pH correction of the five BGC Argo floats launched in this oceanic region, this reference pressure depth couldn't be used as that depth hasn't been reached for all cycles limiting the number of points available for comparison (Fig. 5, Table 3). To overcome this issue, a pressure range of 900 - 940 dbar has been used to correct pH data. It has to be noted that, even if the parking depth has changed for some floats, and thus that a deeper reference pressure depth could have been used, the same depth range has been kept for the entire float array and over their entire deployment time to homogenise the correction method.

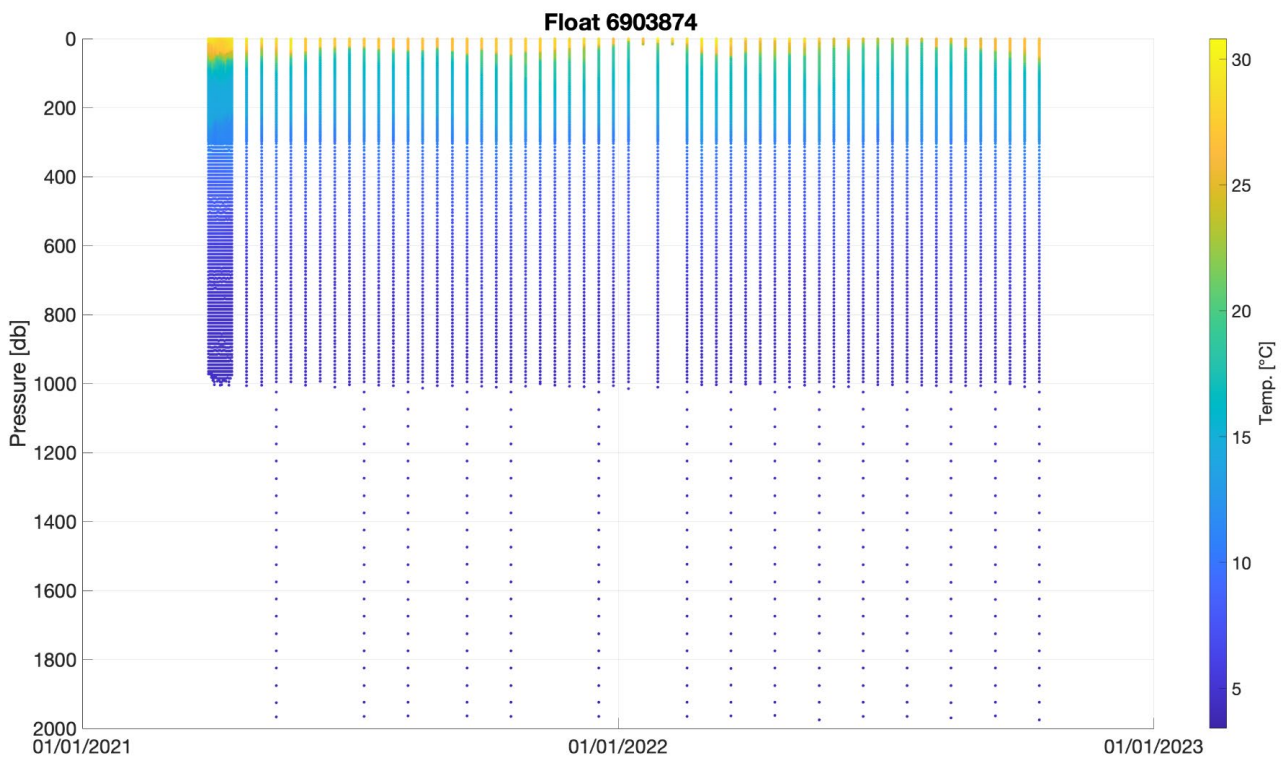


Figure 5. Vertical temperature profiles acquired by the float WMO 6903874 as a function of time. It has to be noted that the same sampling frequency has been used for the 5 floats deployed in this area.

In order to estimate the impact of this reference depth choice on the correction, float-pH data measured by float WMO 7901001 (COMFORT project, Pr. Körtzinger) were corrected using both the “initial” reference

³ Nernst equation describes the potential of an electrochemical cell as a function of concentrations of ions taking part in the reaction. When operated in the constant current mode, the voltage between the reference electrode and source of the ISFET sensor will exhibit a near Nernstian response if the buffer capacity on the insulator over the gate is high enough.

pressure depth (*i.e.*, 1480-1520 dbar) and the “modified” reference pressure depth (*i.e.*, 900-940 dbar). For evaluation, the time period where profiles down to 2000 metres started to occur (cycle 33) has been used (Fig. 6). Whatever the reference method used to correct float-pH data (*i.e.*, LIR or CANYON-B), the observed differences between float-pH data corrected using the 900-940 db pressure range minus float-pH data corrected using the 1480-1520 dbar pressure range are rather stable and constant over time, with a mean difference value higher with the LIR reference method. Nevertheless, considering an average of differences between both depths for the CANYON-B method of 0.00087 pH units (Fig. 6A), we can conclude that the use of this “modified” reference pressure depth is acceptable and does not induce a considerable impact on corrected float-pH data. Indeed, the comparison of the mean difference between raw and corrected (using CANYON-B) float-pH data according to the two reference pressure depths reveals a difference of *ca.* 0.0008 pH units between both depths (Fig. 6B).

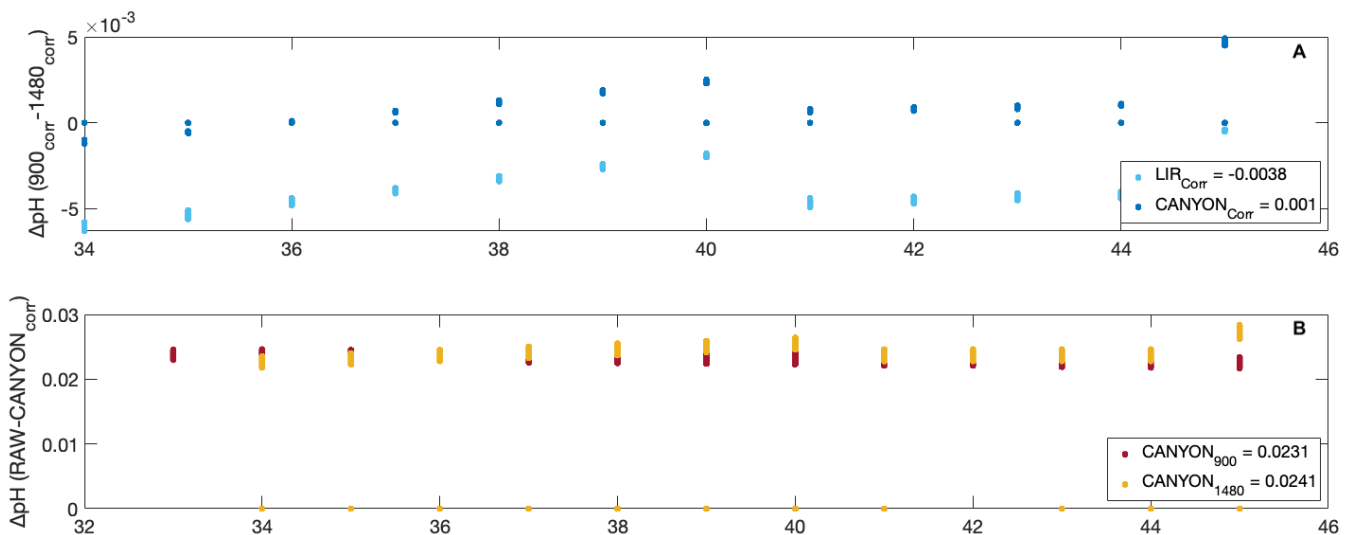


Figure 6. (A) Differences between float-pH data (from cycles 33 to 45) corrected using the 900-940 db pressure range minus float-pH data corrected using the 1480-1520 db pressure range, for each corrective method (LIR or CANYON-B), (B) Differences between raw float-pH data (from cycles 33 to 45) minus float-pH data corrected using the CANYON-B method and both the 900-940 and 1480-1520 db pressure ranges.

Sensor drifting adjustment

Conceptually, the pH correction has to be done by adjusting the sensor reference potential (k_0) as that is the one drifting over time (Johnson et al., 2016). By normalising the adjustment along the profile to the temperature at which the adjustment was derived, the modelled correction is applied as an offset to the reference potential (k_0) in SAGE. To agree with the reference pressure depth used to correct float-pH data, and as the reference pressure depth of 1500 dbar hasn't been reached during some cycles for all of the floats present in this area, it should be noted that the same depth (*ca.* 900 dbar) has been used to obtain at-depth temperature reference and then calculate the offset. To follow this modification, the file entitled “apply_GUIQC_corr_GLT.m” has been slightly modified. Nevertheless, as stated by Maurer et al. (2021) and implemented in Johnson et al. (2018), results are nearly identical when the correction is directly applied to the pH rather than to k_0 .

3.3. Comparison with profiles at deployment

There is a mean offset of 0.0216 pH units (Fig. 7) between reference casts at the deployment site and the Argo's adjusted first pH profiles over the water column. There seems to be a good agreement in near surface waters that are highly variable. As stated earlier, float-pH data is systematically lower than the *in situ* values, especially deeper than 300 m. In the case of float 6903875, the sensor malfunctioned from the first profile (Fig. 7B) explaining the abnormal values around 12 plotted in red in Figure 7B.

The larger offset at depth might be due to the fact that the correction was done at 900-940 m due to the maximum profile depth of 1000m of the BGC-Argo float. A study by Williams et al., (2017) revealed a pH-dependent discrepancy between spectrophotometrically measured pH using purified meta-cresol purple (pH_{spec}) and pH calculated from DIC and TA (pH_{calc}) from full water column hydrographic data with discrepancies ranging from -0.018 to 0.014 in the Southern Ocean. This discrepancy was also present when the analysis was extended to the GLODAPV2 data set (Lauvset et al., 2016), confirming that this is a global phenomenon (Carter et al., 2018). As they highlighted this has numerous implications for the study of the marine carbonate system, and especially for calibration protocols to adjust and quality control pH from sensors mounted on BGC-Argo profiling floats. Based on this finding, Takeshita et al., (2020) conducted a series of lab-based experiments to assess the magnitude of pH-dependent errors for spectrophotometric pH measurements in seawater. First, they showed that the cause of the pH-dependent discrepancy is probably caused by a combination of biases in our understanding of the marine inorganic carbon system, such as the presence of an additional acid/base species contributing to TA and not accounted for. Second, another source of error could come from the spectrophotometric measurements itself through a problem in wavelength accuracy, spectral resolution, dye purity or even dye perturbation calculations (DeGrandpre et al., 2014). Lastly, they also highlighted the potential impacts of light-absorbing impurities in the dye used in the analyses. Therefore, their overall difference between pH_{spec} and pH_{calc} was 0.003, which is an order of magnitude lower than our average post-adjustment error of 0.0216. However, it is close to the range reported by Williams et al., (2017). Therefore, while not being perfect adjustments, we feel confident that our methods are consistent and coherent with other pH-adjustments on these platforms in the oceanographic community.

3.4. Crossovers between CTD cast with water sampling and float profiles

Furthermore, other comparisons can be done between data acquisition platforms. In March 2022, the PIRATA-FR32 cruise sampled, for the first time, *in situ* pH onboard. By changing the floats' sampling frequency, onboard pH sampling and analysis close to Argo floats profiles was achieved (Figure 7; 1 to 4 days difference, mean distance 141 km). Indeed, two sampling stations for pH (st037c01 and st038c01) occurred close to cycle 58 of floats 6903876 and 6903877 around March 22nd, and another station (st008c01) was sampled on March 08th, 2022 and so, 4 days before cycle 57 of float 6903877 on March 12th. Figure 10 presents the temperature, salinity, O₂ and adjusted pH measured by the floats during these cycles and the corresponding measurements from PIRATA-FR32 stations. Table 4 details the values of the different variables used for the comparison. The physical parameters show that relatively similar water masses were sampled by both observing platforms with greater differences in surface waters, mainly due to location. Furthermore, there is a good agreement between the corrected pH from the Argo floats and the *in situ* reference data. The mean absolute difference is 0.026 over the water column. In these comparisons, pH from the Argo floats varies between 7.67 and 8.06 whereas onboard pH spans from 7.72 to 8.01. However, for the matchup around March 22nd, 2022 (first row of Fig. 10), only profiled down to 1 000 dbar whereas the *in situ* stations sampled down to 2000 dbar.

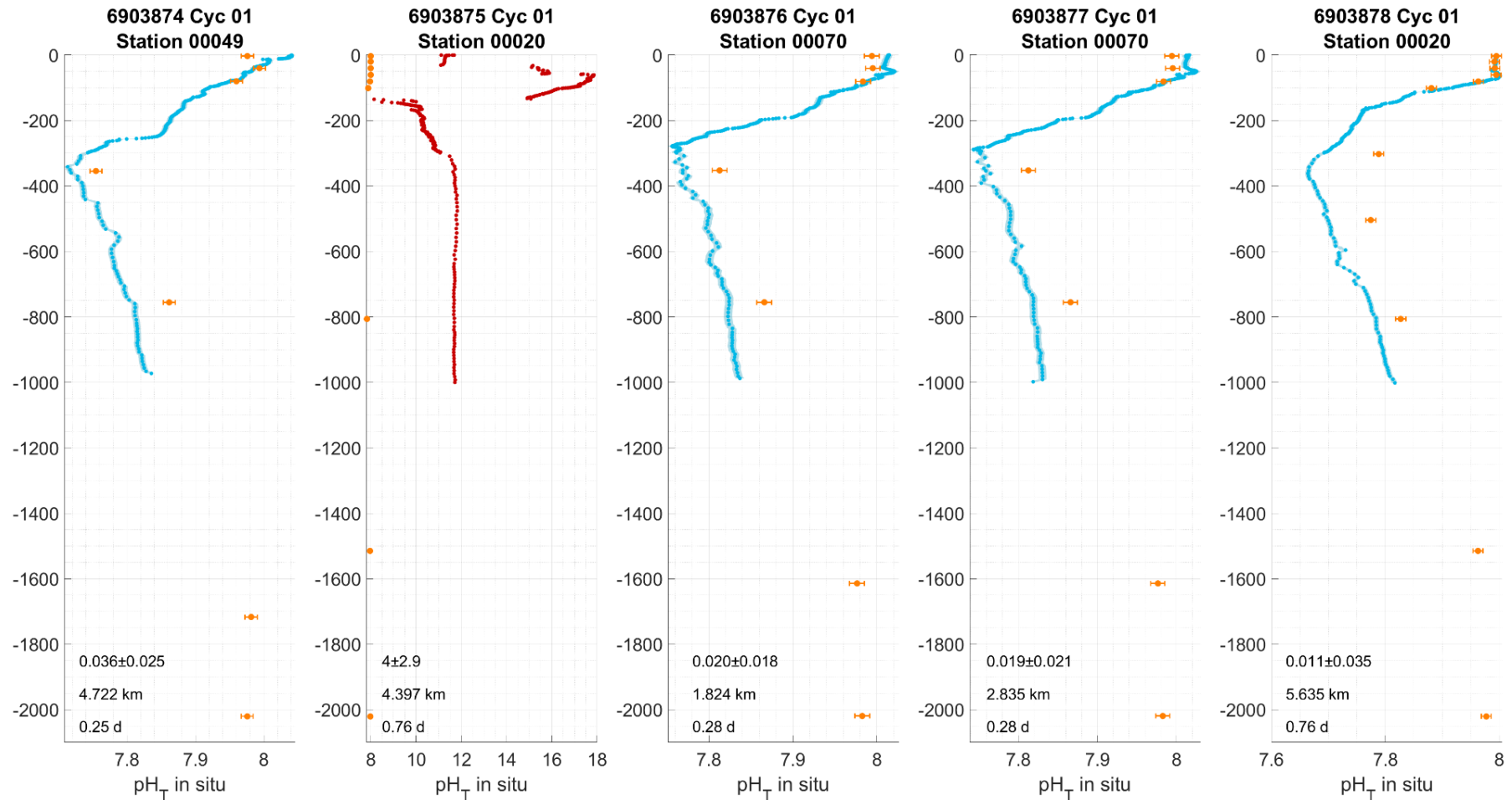


Figure 7. Adjusted first pH profile from Argo floats (blue) and pH recomputed from TA and DIC (orange) sampled near the Argo's location during PIRATA-FR31 cruise. Note that the pH sensor from float 6903875 malfunctioned from the first profile, hence its being presented in red. In the bottom left corner of each panel are the mean absolute difference at the comparison depths, and the difference in space and time (as detailed in Table 1).

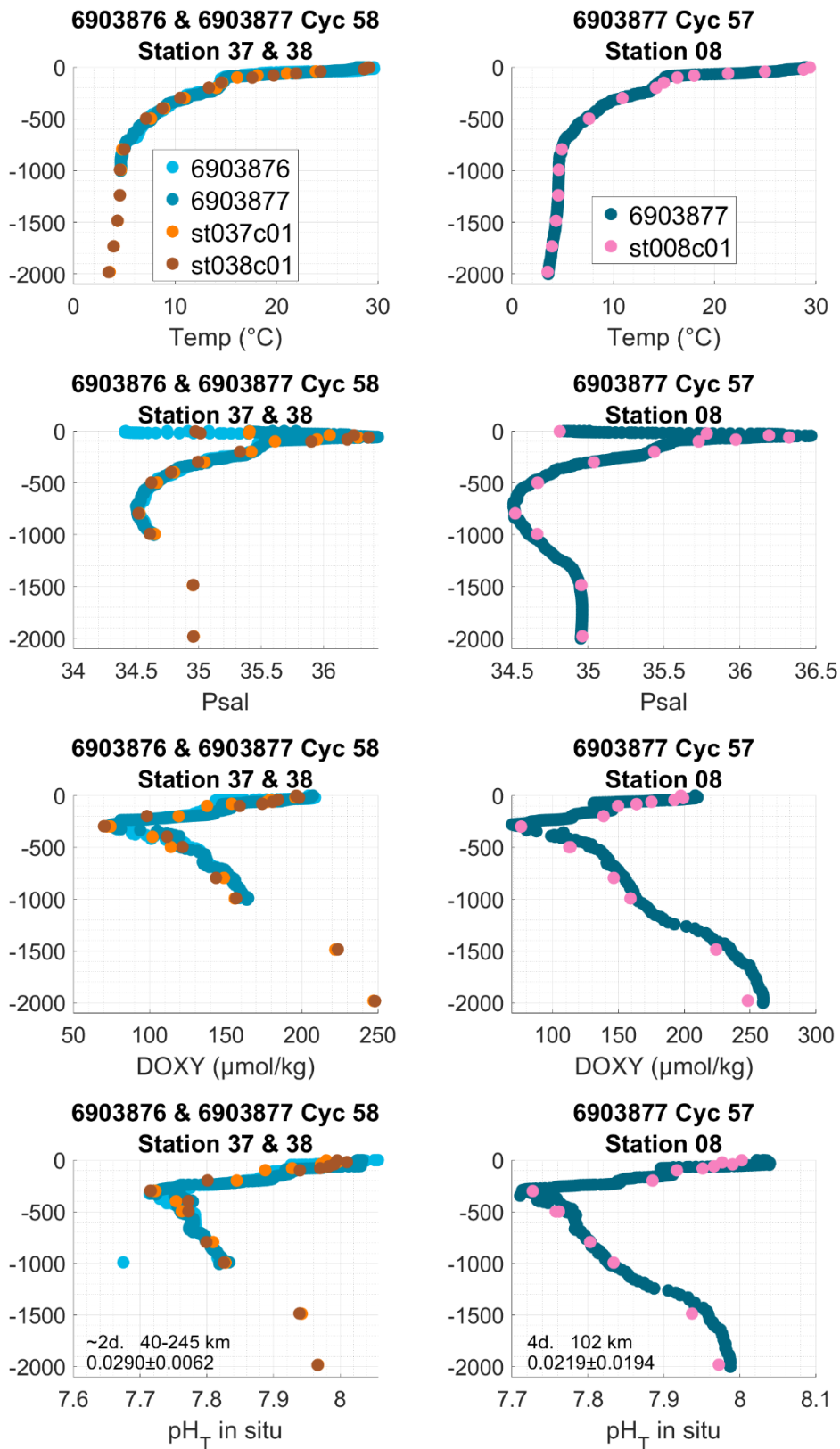


Figure 8. Profiles of temperature, salinity, dissolved oxygen and corrected in situ pH from Argo floats cycles and stations from the PIRATA-FR32 cruise.

3.5. Comparison between surface ocean carbon data and float

Figure 9 presents the comparison between the Argo derived $p\text{CO}_2$ and the Saildrone's $p\text{CO}_2$ time series, together with seawater $p\text{CO}_2$ measurements from the France-Brazil SOOP line. Even though two months apart (see section 2.2.3), the SOOP data is in the range of the SD data and their error. The BGC-Argo derived $p\text{CO}_2$ matches the Saildrone's values. There was a high variability in a relatively small-time frame in the Saildrone's $p\text{CO}_2$ and this is also reproduced in the BGC-Argo measurements. This might be due to the sharp decrease in temperature at that time (loss of 1°C in SST in a few days). Note that the estimated uncertainty associated with the $p\text{CO}_2$ derived from BGC-Argo float's pH together with neural-network based TA (ESPER; Carter et al., 2021) computed with CO2SYS (van Heuven et al., 2011) is $12 \mu\text{atm}$ (as shown with the purple error bars).

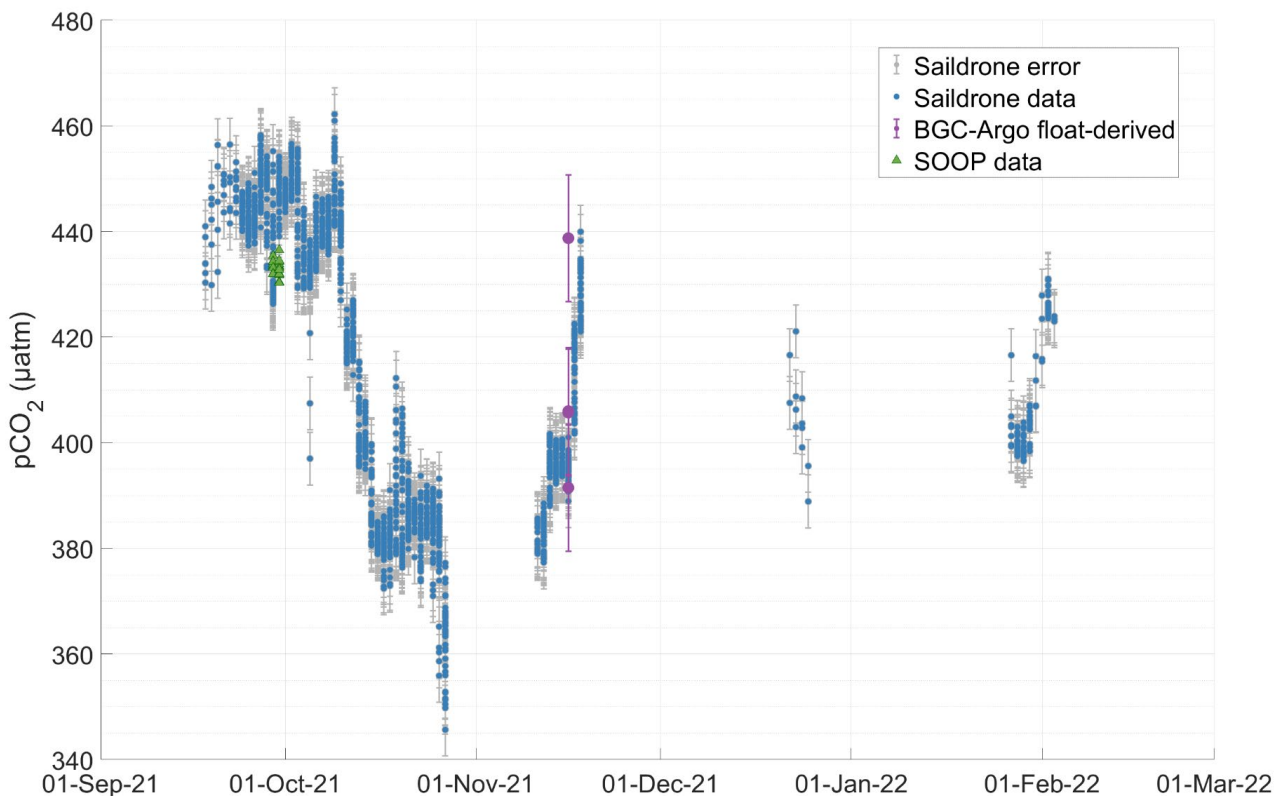


Figure 9. Saildrone seawater $p\text{CO}_2$ time series (blue) with error (grey), SOOP data (green) and $p\text{CO}_2$ derived from BGC-Argo float's pH and neural-network based TA (purple) with error.

Therefore, while $p\text{CO}_2$ can be derived from Argo float data, reliable high-quality surface ocean carbon data, such as SOOP or Saildrone remains essential and provides data with a spatio-temporal resolution like no other. Furthermore, the high propagation of uncertainties is to be considered.

3.6. Ways to go forward with sensor failures?

Nowadays, machine learning methods such as neural networks are being used more and more for oceanographic applications to virtually densify the limited number of measurements that can be done by autonomous platforms. Some of them are implemented in the SAGE GUI software to correct pH, such as CANYON-B (Bittig et al., 2018). Others more recent like ESPER (Carter et al., 2021) are not implemented yet. However, all of these methods allow for the prediction of carbonate system variables with a given accuracy,

relying only on temperature, salinity, O_2 together with the position in time and space. The correction of pH from Argo floats is an essential step and can, in some cases, lead to large amounts of high-quality pH data reaching accuracies of 0.005 (Johnson et al, 2018, SOCCOM floats). Furthermore, the objective in OneArgo/GO-BGC is to have (add detailed objective for pH accuracy but also for coverage). Therefore, on top of correcting pH, we can also rely on these neural network methods to derive additional carbonate system variables, namely pH, requiring only one biogeochemical sensor: O_2 . And in cases such as ours, where pH sensors do fail, we can predict pH and fill the gaps left by bad sensors.

Building on the comparisons shown in Figure 8, BGC-Argo float 6903876's sensor drifted between cycles 1 and 60 (Figure 10B). At cycle 58, the raw pH profile (in grey in Figure 10C) had a MAD of 0.1330 and 0.1411 with reference stations 37 and 38 from PIRATA-FR32 over the water column. After adjustment, the MAD was reduced to around 0.0290 (blue profile in Figure 10C). Comparing this corrected pH and the *in situ* pH from the reference stations with neural-network based estimates (ESPER, in purple), they are in agreement within the range of the neural network predicted error (purple shadowed area). Therefore, this is a confirmation that, while not replacing precious *in situ* data, neural networks provide pH estimates and errors coherent with adjusted pH data from the floats.

For example, for floats 6903875 (all cycles) and 6903876 (cycle>60) that went into complete sensor failure, Figure 11 highlights the inexploitability of the sensor pH. However, by applying the ESPER neural network method (Carter et al, 2021) on temperature, salinity and O_2 from the profiling floats, we can derive "predicted pH" (Fig. 11B and D). This virtual dataset of pH comes with a given error (shown in the purple shaded area) but can help augment our datasets and go beyond sensor failures.

Finally, it should also be mentioned that several modelling products exist for pH and/or carbonate system parameters. They are gridded products and their resolution is often limited to surface monthly 1° grids. For example, the MULTI_OBS_GLO_BIO_CARBON_SURFACE_REP_015_008 product from Chau et al., (2022) provides surface pH estimates with a pH uncertainty of 0.0348. Second, the OceanSODA-ETHZ product from Gregor and Gruber (2021) has a pH uncertainty of 0.023. While these products exist, we feel that in order for BGC-Argo data to be used to validate or calibrate models, 1) the data must absolutely have gone through quality control procedures and adjustments and 2) the error associated with the correction must be taken into account by the modelers.

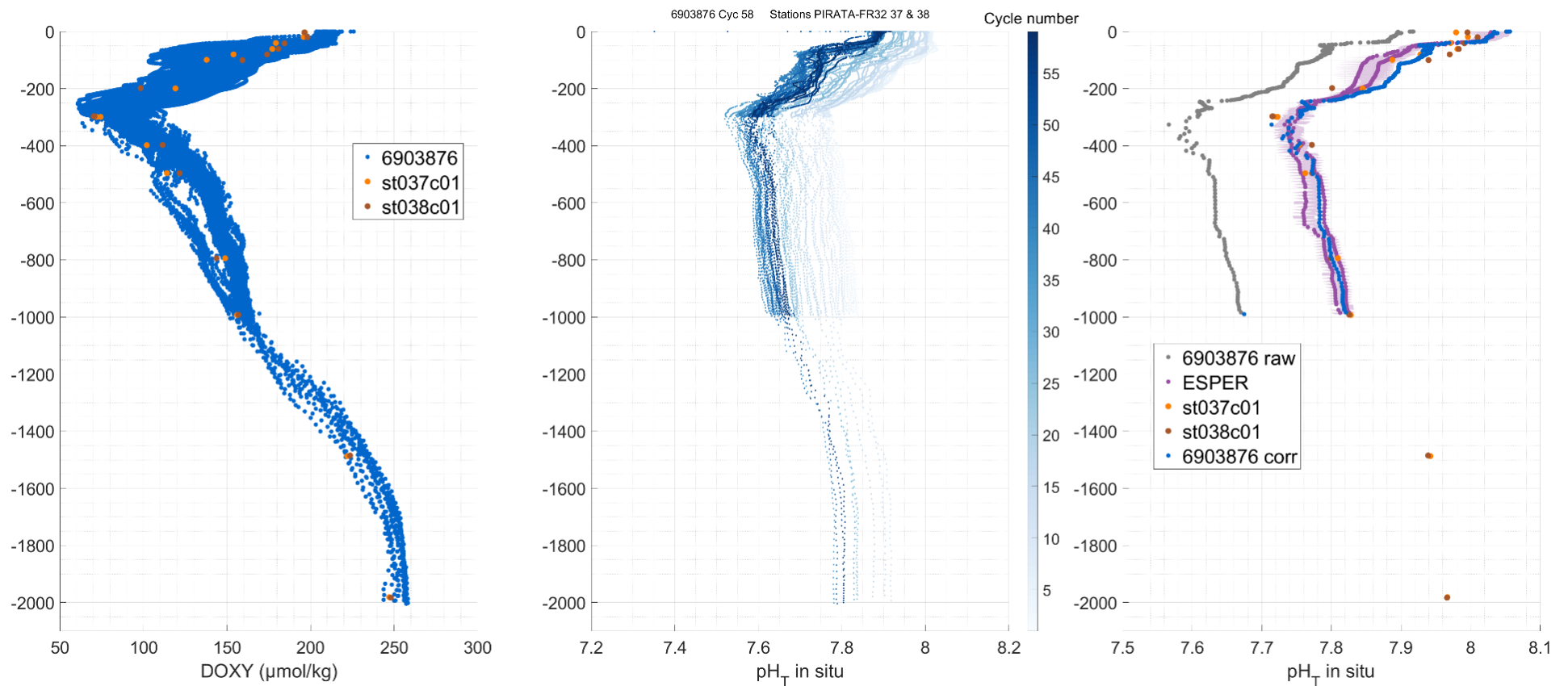


Figure 10. (A) BGC-Argo float 6903876's O_2 (blue), reference Winkler measurements from stations 37 (orange) and 38 (brown) of cruise PIRATA-FR32. (B) Scatter plots of raw pH from BGC-Argo float 6903876 (cycle ≤ 60) according to cycle number showing sensor drift. (C) BGC-Argo float 6903876 cycle 58's raw pH (grey), adjusted pH (blue), pH from reference stations 37 (orange) and 38 (brown) from PIRATA-FR32, together with ESPER-derived pH (purple with predictor error in shadowed area).

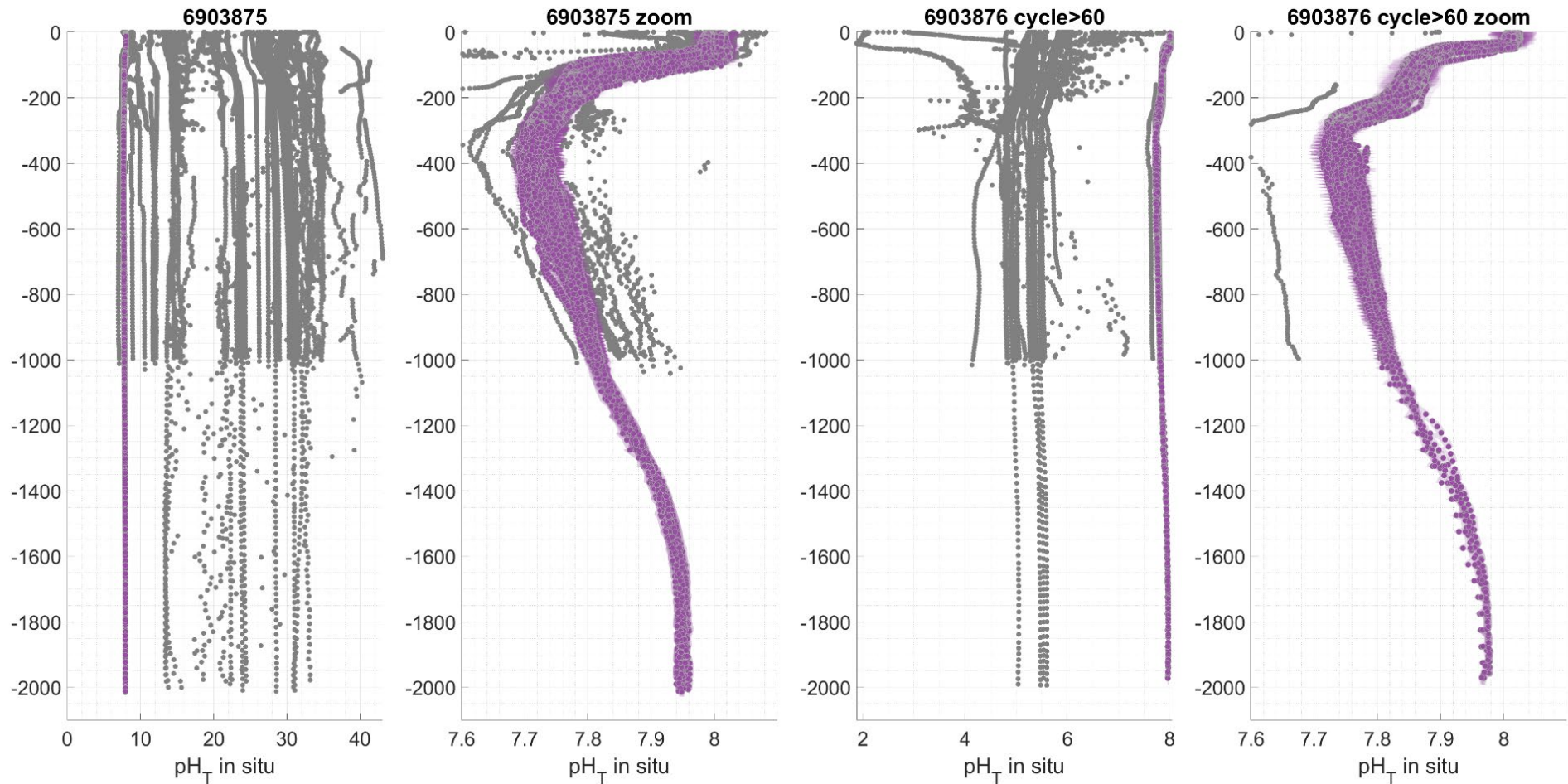


Figure 11. (A) BGC-Argo float 6903875's pH_T and neural-network based TA (purple) with error. (B) Zoom of panel (A). (C) BGC-Argo float 6903876 cycle 61 onwards's pH_T and neural-network based TA (purple) with error. (D) Zoom of panel (C).

Recommendations

Although the results and observations presented in this report only represent a limited number of floats, few crossovers and a relatively small regional domain, final conclusions and suggestions can already be drawn and summarised as:

- *The first profile: 1st profile at 2000 dbar*

From high-quality co-located and synchronous *in situ* DIC and TA observations realized over the entire water column, pH data have been calculated at relatively high quality to serve as a direct comparison for BGC-Argo floats at the deployment location (see section 3.3). Results of these direct comparisons point at significant differences and issues with pH data quality, at least under the first hundred meters of the water column. Accuracy assessments based on crossovers point at mean differences in excess of *ca.* 0.02 pH units, which correspond to about 20 μatm uncertainty in $p/f\text{CO}_2$. While this observed difference has been previously reported in the literature and can be explained by some factors, the robustness of the conclusions to be drawn from our comparison is limited given the reduced profile depth associated with the first profiles. We, therefore propose Argo floats to have a systematic first deep profile at 2000 dbar to better understand and, if needed, correct this at-depth the observed variability. Better constraining and understanding these differences are particularly important in the context of converting surface ocean pH into $p\text{CO}_2$ for the purpose of calculating the air-sea CO_2 flux.

- *The sampling plan: More frequent 2000 dbar profiles*

Based on the comparison between calculated pH data (from *in situ* DIC and AT data) and corrected float-pH data (see section 3.4 and Table 2), results found for each float indicate that float pH may be biased by up to several hundredths of a pH unit at depth. In fact, two of the five floats show a virtually identical *ca.* 0.02 pH units mean bias, *i.e.*, float-pH is systematically higher on average by this amount when compared to *in situ* calculated pH data. These early findings, therefore, warrant further and more sophisticated analyses to better constrain float pH at depth and highlight the need to implement the sampling plan with more frequent profiles reaching 200 dbars to cover the entire water column. With regards to these observations, these differences may mean that the established at-depth correction (*ca.* 900 dbars) used does not seem to yield adequate pH accuracy at depth. This uncertainty may partly be incurred by the complication of finding a reliable at-depth reference with the initial sampling strategy.

- *The float correction procedure*

To correct float-pH data, the established at-depth correction currently used relies on the difference between float-based pH measurements and a selected field of reference calculated at a depth below *ca.* 1500 meters, where spatiotemporal variability of oceanic components is assumed to be minimal. In this study, because of the parking depth associated with the first profiles, an at-depth reference pressure of around 900 dbars has been used. In this region, we point out that the choice of the reference depth (900 vs. 1500 dbar) changed the pH correction in a negligible way as the resulting uncertainty is on the order of a few thousand of pH units (0.0008 pH units), which is tolerable. By also modifying the at-depth temperature used to mimic the k_0 sensor drift, this pressure range can be used for float datasets. Nevertheless, the choice of reference depth may have a noticeable impact on corrected float-pH data in some oceanic regions as for example in the peculiar subpolar North

Atlantic area where complex water masses meet during deep convection events and where deep penetrations of anthropogenic CO₂ occur. In consequence, we suggest keeping, at least, this original at-depth reference to correct float-pH data and overcome the uncertainty resulting from this reference pressure depth. In consequence, the sampling plan should be adjusted to allow a float-pH correction with this pressure range, and this from the first profiles on. Without that, a potentially very significant uncertainty associated with the resulting accuracy of fully corrected delayed mode pH data could be observed.

- Ancillary data

The general philosophy for float-pH correction employs empirical algorithms, such as the CANYON-B method, requiring *in situ* oxygen measurements (together with other ancillary data such as temperature and salinity measurements) as an input variable to then derive the predicted pH. Therefore, the use of these correction tools cannot be done without high-quality oxygen data, and consequently meticulous O₂ delayed-mode quality control (DMQC) procedures. As the robustness and accuracy of the corrected float-pH datasets depend on these oxygen values, we warn about a continuous attention to the quality of O₂ values and their passing through DMQC procedures.

- Crossovers and matchups

When a BGC-Argo float is deployed and then acquires data at-sea, it is desirable to establish further measurements to independently estimate pH data accuracy and ideally also to apply corrections. In the context of float data quality control, measurements that can be considered as necessary are CTD casts with discrete water sampling performed at the time and place of the float deployment. This would require an established reference method – in this case for discrete pH measurements – which itself can be traced back to an external reference. Indeed, while DIC and TA discrete samples are calibrated against certified CRMs, an inevitable error is associated with the pH calculated from DIC and TA measurements. Thus, if possible, discrete *in situ* pH measurements should be done over the entire water column. Then, with the help of the oceanographic campaign leaders, and potentially by changing the floats' sampling frequency, this study highlights the crucial need to develop integrated observation systems to obtain specific crossovers between CTD casts and floats profiling during the floats's lifetimes without impact on the field work schedule.

- Sampling the ocean

While autonomous platforms such as BGC-Argo floats are useful tools to observe the ocean and circumvent many issues encountered with historical classical sampling strategies (low spatio-temporal resolution, high human resources request, weather dependency), it should also be pointed out that oceanographic cruises and data acquired through classical observing systems (including SOOP-line) remain the reference tool to obtain high-quality data to then compare datasets from autonomous platforms. Float-pH correction methods rely on empirical algorithms, themselves trained using a hydrography database (GLODAPv2; Olsen et al., 2016). Nevertheless, as the data coverage of the GLODAPv2 data product is far from perfect, both in terms of spatial and annual variability (seasonal cycle) the robustness algorithms based on it are globally not uniform and regional and/or seasonal biases are therefore to be expected. Moreover, the comparison between corrected float-pH data and *in situ* discrete pH (calculated and/or measured) data is a robust tool to achieve an independent quality control and perhaps correction of float-pH data. In consequence, this

study underlines the major interest in maintaining and continuing oceanographic campaigns and other classic sampling methods in order to feed the databases but also to compare and correct the data acquired by autonomous platforms such as floats.

- The reference method

In SAGE, numerous tools and methods can be used as reference models to adjust and correct float-pH data. Recently, a new reference algorithm named ESPER (Carter et al., 2021) has been developed and provides, at intermediate-depths, pH estimations with biases ranging between -0.001 and -0.002 pH units (depending on the input variables) using the Mixed method. Indeed, as proposed by Bittig et al. (2018), averaging the estimates from the LIR and CANYON-B methods seems to improve the global average prediction statistics. Regarding this global average ESPER performance which is as good or better than older algorithms, we think it would be wise to implement this reference algorithm into SAGE in the future.

Conclusions

This deliverable describes the successful float-pH data correction strategy followed in the Eastern Tropical North Atlantic region in response to the growing and crucial need for accurate and precise float-pH data to better constrain ocean acidification and derive oceanic carbon data and their variability regarding current climate change. To correct float-pH data, the established and generally used at-depth correction presented in the SAGE tool has been used. By modifying some parameters such as the reference pressure depth or the temperature at-depth used to correct data, we have been able to correct pH from the 5 BGC-Argo floats launched in the studied region. Moreover, this study points out that, while deep profiles always have to be preferred, the choice of the reference depth changes the pH correction in a negligible way in this region as the resulting uncertainty is on the order of a few thousand of pH units (0.0008 pH units), which is tolerable. An independent quality assessment of corrected float-based pH data has been done with comparisons against *in situ* calculated pH data. These crossovers were carried out in a synergistically combined effort between numerous institutions and scientists.

Besides highlighting the indubitable need to carry out reference data sampling strategies based on classical ways of sampling (*e.g.*, vessel-based hydrographic surveys, SOOP lines, autonomous high-quality $p\text{CO}_2$ sensor measurements on drones or moorings), these comparisons reveal overall differences of approx. 0.02 pH units, which correspond to about 20 μatm uncertainty in $p/f\text{CO}_2$, *i.e.*, two times higher than the limit for inclusion into the SOCAT database. Thus, this study illustrates the several ongoing challenges regarding the QC procedure for float-based pH measurements and questions the purpose of this data correction process. Indeed, depending on whether the correction applied is to compare and correct the pH-sensor data to pH values spectrophotometrically measured, or to correct float-pH data to subsequently derive the parameters of the carbonate system, the final statement on the accepted accuracy may be different due to the present incomplete internal consistency in CO_2 estimations.

Moreover, while this result is in agreement with the literature, the robustness of the conclusions to be drawn is also limited given the reduced number of best-performing pH sensors and hence a much smaller amount of highest-quality data. This warrant continued attention to the quality of float-based pH in the ocean. Indeed, this study suffered tremendously from manufacturing-related problems. Particularly, the recent serious problems with the pH sensors' reference electrodes are likely to greatly reduce sensor lifetimes and

increase the variability of the drift characteristics. This points to an urgent need to improve manufacturing quality and encourages further work on alternative pH sensors, both in terms of manufacturers as well as in terms of sensor principles.

This deliverable also proposes, based on this regional focus but also limited dataset, some suggestions that the scientific community may explore in a systematic way to harmonise both the way of measuring but also correcting float-pH data. The joint drafting of reference guides, such as Standard Operating Procedures or Best Practices, by the entire community of users of these tools, could help to answer these questions. This report further illustrates the noticeable uncertainty and lack of additional reference points to compare and possibly correct float-based pH datasets.

Data availability statement

Fully processed and finalised surface $p\text{CO}_2$ data will be submitted in 2023 to the Surface Ocean CO_2 Atlas (SOCAT) for community-based quality control and final ingestion into global carbon synthesis products and assessments. Argo data are available at <http://doi.org/10.17882/42182#96550> or at <ftp://ftp.ifremer.fr/ifremer/argo/dac/coriolis>. 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. Atmospheric data recorded at the CVAO station are available at <https://catalogue.ceda.ac.uk/uuid/81693aad69409100b1b9a247b9ae75d5>. Data from the french PIRATA cruises are available on the SEANOE website (<https://www.seanoe.org>).

Acknowledgments

First and foremost, we would like to thank the funding agencies who made this mission possible, namely the European Commission and the German Federal Ministry of Education and Research (BMBF; FKZ 03F0885AL1). We wish to thank the crew members of the R/V “La Thalassa” (IFREMER) for making the PIRATA-FR31 and -32 cruises possible. We acknowledge the “Flotte Oceanographique Française”, FOF, and the crew of the R/V “La Thalassa” (IFREMER) for their help in the PIRATA-FR31 and -32 sampling. The many researchers responsible for the collection of data and quality control are thanked for their contribution. For pH seawater sampling and analyses, particular thanks to Pierre Rousselot, Thierry Cariou and Sandrine Hillion (IRD, UAR191, Instrumentation, Moyens Analytiques, Observatoires en Géophysique et Océanographie (IMAGO), Technopôle de Brest-Iroise, Plouzané, France). For Argo deployments in the area during PIRATA cruises, we thank Bernard Bourlès. We also thank Nathalie Lefèvre for her valued contribution to the FR-SOOP-France-Brazil line (LOCEAN/IPLS).

References

Bergveld, P., 2003. Thirty years of ISFETOLOGY: What happened in the past 30 years and what may happen in the next 30 years. *Sens. Actuators, B* 2003, 88,1–20.

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. <https://doi.org/10.3389/fmars.2018.00328>

Bittig, H. C., Maurer, T. L., Plant, J. N., Schmechtig, C., Wong, A. P. S., Claustre, H., et al. (2019). A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. *Frontiers in Marine Science*, 6, 502. <https://doi.org/10.3389/fmars.2019.00502>

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, Brendan R., Bittig, H. C., Fassbender, A. J., Sharp, J. D., Takeshita, Y., Xu, 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>

Chau, T. T. T., Gehlen, M., & Chevallier, F. (2022). A seamless ensemble-based reconstruction of surface ocean CO₂ and air–sea CO₂ fluxes over the global coastal and open oceans. *Biogeosciences*, 19(4), 1087–1109. <https://doi.org/10.5194/bg-19-1087-2022>

Chen, J. & Chen, Z., 2008. Extended Bayesian Information Criteria for Model Selection with Large Model Spaces. *Biometrika*, 95 (3): 759–771.

DeGrandpre, M.D., Spaulding, R.S., Newton, J.O., Jaqueth, E.J., Hamblock, S.E., Umansky, A.A., Harris, K.E., 2014. Considerations for the measurement of spectrophotometric pH for ocean acidification and other studies. *Limnology and Oceanography: Methods*, 12, 830–839. <https://doi.org/10.4319/lom.2014.12.830>.

Dickson, A. G., C. L. Sabine, and J. R. Christian (Eds.), (2007). Guide to Best Practices for Ocean CO₂ Measurements. PICES Special Publication 3, 191 pp.

Dickson, A. G., and Goyet, C. (1994). ORNL/CDIAC-74 (Ed.), Handbook of Methods for the Analysis of the Various Parameters of the Carbon Dioxide System in Sea Water. Version 2, no 74. Washington, DC: US Department of Energy.

Edmond, J. M. (1970). High precision determination of titration alkalinity and total carbon dioxide content of sea water by potentiometric titration. *Deep Sea Res. Oceanogr.* Abstract. 17:73750.

Foltz, G. R., Brandt, P., Richter, I., Rodríguez-Fonseca, B., Hernandez, F., Dengler, M., et al. (2019). The Tropical Atlantic Observing System. *Frontiers in Marine Science*, 6, 206. <https://doi.org/10.3389/fmars.2019.00206>

Gregor, L., & Gruber, N. (2021). 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*, 13, 32. <https://doi.org/10.5194/essd-13-777-2021>

Goni G. J., Roemmich, D., Molinari, R., Meyers, G., Sun, C., Boyer, T., Baringer, M., Gouretski, V., DiNezio, P., Reseghetti, F., Vissa, G., Swart, S., Keeley, R., Garzoli, S., Rossby, T., Maes, C. and Reverdin, G. (2010), The ship of opportunity program, in Proceedings of the OceanObs'09: Sustained Ocean Observations and Information for Society, edited by J. Hall, D. E. Harrison, and D. Stammer, ESA Publication WPP-306.

Johnson, K. S., Coletti, L., Jannasch, H., Sakamoto, C., Swift, D., and Riser, S. (2013). Long-term nitrate measurements in the ocean using the *in situ* ultraviolet spectrophotometer: Sensor integration into the apex profiling float. *J. Atmos. Oceanic Technol.* 30, 1854–1866. <https://doi.org/10.1175/JTECH-D-12-00221.1>

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., Jannasch, H. W., Coletti, L. J., Elrod, V. A., Martz, T. R., Takeshita, Y., et al. (2016). Deep-Sea DuraFET: A Pressure Tolerant pH Sensor Designed for Global Sensor Networks. *Analytical Chemistry*, 88(6), 3249–3256. <https://doi.org/10.1021/acs.analchem.5b04653>

Johnson, K. S., Plant, J. N. & Maurer, T. L. (2018). Processing BGC-Argo pH data at the DAC level. Argo data management. <http://doi.org/10.13155/57195>

Lauvset, S.K., Key, R.M., Olsen, A., et al., 2016. A new global interior ocean mapped climatology: the 1° × 1° GLODAP version 2. *Earth System Science Data* 8, 325–340. <https://doi.org/10.5194/essd-8-325-2016>.

Martz, T. R., Connery, J. G., and Johnson, K. S. (2010). Testing the honeywell durafet® for seawater pH applications. *Limnol. Oceanogr. Methods* 8, 172–184.

Martz, T., Daly, K., Byrne, R., Stillman, J. and Turk, D. (2015). Technology for Ocean Acidification Research: needs and availability. *Oceanography*, 25:40–7.

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, 683207. <https://doi.org/10.3389/fmars.2021.683207>

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>.

Orr, J. C., Epitalon, J.-M., Dickson, A. G., & Gattuso, J.-P. (2018). Routine uncertainty propagation for the marine carbon dioxide system. *Marine Chemistry*, 207, 84-107. <https://doi.org/10.1016/j.marchem.2018.10.006>

Russell, J., Sarmiento, J., Cullen, H., Hotinski, R., Johnson, K., Riser, S., & Talley, L. (2014). The Southern Ocean Carbon and Climate Observations and Modeling Program (SOCCOM). 28.

Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., & Johnson, K. S. (2017). Estimates of water-column nutrient concentrations and carbonate system parameters in the global ocean: a novel approach based on neural networks. *Frontiers in Marine Science*, 4. <https://doi.org/10.3389/fmars.2017.00128>

Schmechtig, C., Thierry, V., & Team, T. B. A. (2016). Argo quality control manual for biogeochemical data (1.0) [Pdf]. Bio-Argo group. <http://doi.org/10.13155/40879>.

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>.

Tanhua, T., McCurdy, A., Fischer, A., Appeltans, W., Bax, N., Currie, K., et al. (2019). What we have learned from the framework for ocean observing: evolution of the global ocean observing system. *Frontiers in Marine Sciences*, 6: 471. <https://doi.org/10.3389/fmars.2019.00471>.

van Heuven, S., Pierrot, D., Rae, J. W. B., Lewis, E. & Wallace, D. W. R. (2011). MATLAB Program Developed for CO₂ System Calculations. ORNL/CDIAC-105b. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, doi: 10.3334/CDIAC/otg.CO2SYS_MATLAB_v1.1.

Watson, A., Lefèvre, N., Smythe, T., Hartman, S., Reverdin, G., Gonzalez-Davila, M. & Fietzek, P. (2018). Deliverable D2.4: SOOP Network Enhancement Report, AtlantOS – 633211 project.

Williams, N. L., Juranek, L. W., Johnson, K. S., Feely, R. A., Riser, S. C., Talley, L. D., Russell, J. L., Sarmiento, J. L., & Wanninkhof, R. (2016). Empirical algorithms to estimate water column pH in the Southern Ocean. *Geophysical Research Letters*, 43(7), 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>

Wong, A., Keeley, R., Carval, T., & Team, A. D. M. (2022). Argo Quality Control Manual for CTD and Trajectory Data(3.6) [Pdf]. <http://dx.doi.org/10.13155/33951>.