

## **Technical note: Enhancement of float-pH data quality control methods: A study case in the subpolar northwest Atlantic Ocean**

*In this response, the “original manuscript” refers to the first submitted manuscript that has been evaluated by the reviewer and the “revised manuscript” refers to the manuscript that has been modified according to the reviewer’s comments. Comments from the reviewer are pasted below in black font; our point-by-point responses immediately follow in blue font. Blue italic sentences are those that have been modified / added in the revised manuscript. When indicated, line numbers refer to the new version of the manuscript (Pdf version). We have also submitted a Word document with the “track changes” function activated, which should help the reviewer in figuring out the changes.*

### **Responses to REVIEWER 2 - Anonymous Referee**

This paper explores the impact of different calibration choices on the resulting data accuracy for pH data collected by autonomous sensors on Argo floats in the NW Atlantic including comparisons with independent data in the region. The presentation is detailed and will be of high interest to groups seeking to quality control and use data from pH sensors on floats and other autonomous vehicles. The recommendations for choices to make and issues that remain to be solved are helpful. I have a few suggestions that I think will improve the paper and enhance how useful it is to the broader community.

We acknowledge the referee for his/her constructive and interesting comments about our manuscript and for reviewing it. We thank his/her positive opinion about our work and we greatly appreciate that the referee found our work detailed and helpful for the community. In the revised manuscript, all comments and suggestions made by Reviewer #2 have been fully addressed.

#### ***Moderate comments***

Breakpoint nodes: The existence and use of breakpoint nodes in pH sensor calibration may be well known in the pH community already but not to researchers focusing on other sensors. Some discussion of this phenomenon and its likely causes would help this paper be useful to a broader community who may be coming to pH sensor calibration from other sensors. What is the likely cause of these sudden jumps? Discussion of their likely cause would help support the authors’ recommendation that the jumps should be gradual rather than sudden.

In particular, I think it would be useful to demonstrate that the sudden jumps are directly related to changing sensor response rather than to apparent discontinuities in the reference value derived from CANYON-B or LIR.

The authors states that these swings in the record are mostly interpreted as changes occurring in the sensor, but this could be supported with the available data. Figure A1 may be partly showing this, but I was unsure if the parking depth pH data had already been corrected or was

raw. A figure (perhaps supplemental) showing the raw pH values and the CANYON-B and LIR values (on different scales) at the relevant pressures would make which signal contained the apparent discontinuities clear.

While drifts and offsets for pH sensors often vary linearly over a long time as for conductivity sensors (Owens and Wong, 2009)<sup>1</sup>, calibration jumps are oftentimes observed in the time series, especially over the first few cycles. In the revised manuscript, the reason behind this splitting of the dataset is now more clearly stated “*Indeed, by breaking the time-series sensor record into different segments and fitting each with a linear rate of change in  $k_0$ , the adjustment better represents the sensor behavior over time as both drifts and offsets change independently between segments and oftentimes noticeable jumps occur over the first few cycles in a float’s life (Maurer et al., 2021).*” (L. 186)

This observation has been also reported for nitrate sensors and can be explained by the optics of the nitrate sensor itself which are more sensitive to transient perturbations induced by biofouling. For pH sensors, the sensor  $k_0$  value may drift as the sensor warms up and final equilibration with  $\text{Br}^-$  occurs. The recorded data may also undergo offsets as the system is purged of air bubbles (Johnson et al., 2016)<sup>2</sup>. Thus, Sea-Bird Scientific claims an initial stated accuracy of  $\pm 0.05$  pH and a stability of 0.036 pH/year (Bittig et al., 2019)<sup>3</sup>. In the original manuscript, these changes are discussed in Section 2.3 (L. 171): “*Thus, pH sensors are calibrated in the laboratory using spectrophotometric measurements and are therefore directly related to the laboratory calibration method. Each sensor’s pressure and temperature coefficients, needed to compute the in situ pH, are also determined in the laboratory as described in Johnson et al. (2016). When deployed at sea, temperature changes modify the reference potential of the sensor and in return induce a sensor drift as the Nernst slope that transforms sensor potential to pH depends on temperature (Johnson et al., 2016, 2017).*”. However, the likely cause of these sudden jumps is not well understood so far and we believe that it is somehow out of the scope of our paper. In this manuscript, the purpose of Section 3.1.3. was to discuss the noticeable step-like changes observed with the current correction procedure (i.e. the SAGE method) and to find the best way to represent the smooth sensor drift over time, as observed when looking at the pH time-series recorded at the parking depth (previous Figure A1 in the Supplementary Material). Indeed, in comparison with the pattern of the cycle-by-cycle correction, the high pH changes of ca. 0.01 pH units observed between linear drift phases with the SAGE method appear to be unrealistic, especially when looking at the pH time-series recorded at the parking depth. In our view, the sensor rather shows undulations in response with smooth and less smooth phases and an adaptation of the current correction procedure could be done to better maximize the smoothness of the corrections and to avoid introducing artificial jumps. However, we agree that explanations were missing in the original manuscript and that the original idea to put Figure A1 in the Appendix was not relevant as it is critical for our argument.

Concerning the discontinuities and Figure A1: Following the recommendation of the two reviewers, Section 3.1.3 has been modified, explanations have been added, and Figure 5 re-drawn: now 4 panels representing differences between raw and corrected float-pH data following the SAGE method (panel

---

<sup>1</sup> Owens, W. B. and Wong, A. P. S. (2009). An improved calibration method for the drift of the conductivity sensor on autonomous CTD profiling floats 565 by  $\theta$ -S climatology, Deep Sea Research Part I: Oceanographic Research Papers, 56(3), 450-457. <https://doi.org/10.1016/j.dsr.2008.09.008>.

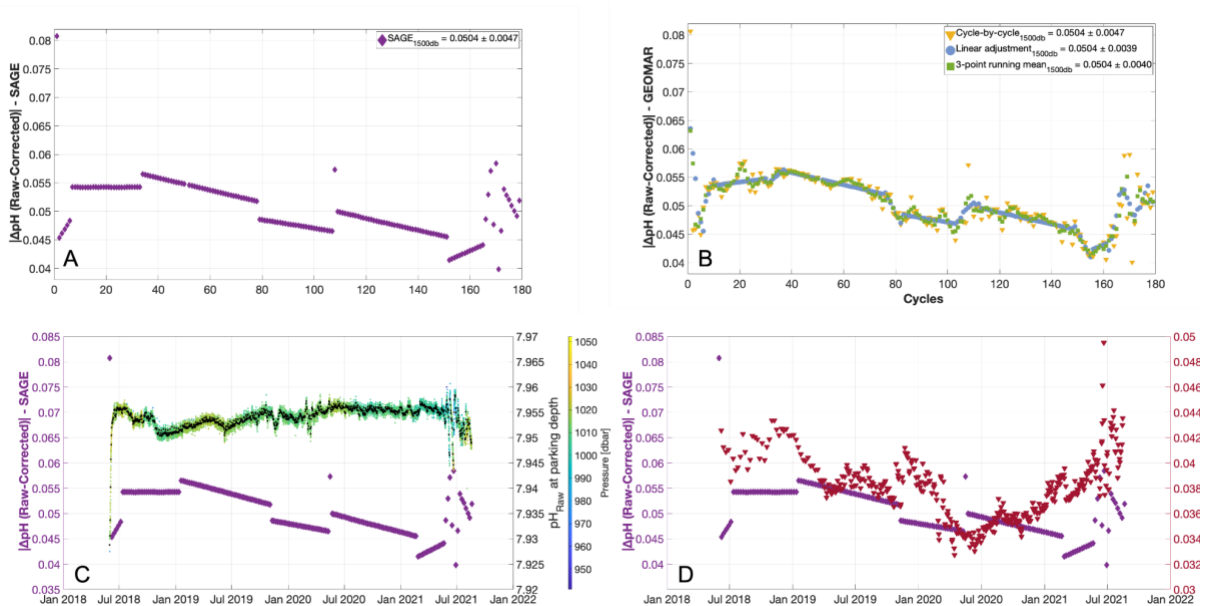
<sup>2</sup> Johnson, K. S., Jannasch, H. W., Coletti, L. J., Elrod, V. A., Martz, T. R., Takeshita, Y., Carlson, R. J., and Connery, J. G. (2016). Deep-Sea DuraFET: A Pressure Tolerant pH Sensor Designed for Global Sensor Networks, Analytical Chemistry, 88(6), 3249-3256. <https://doi.org/10.1021/acs.analchem.5b04653>.

<sup>3</sup> Bittig, H. C., Maurer, T. L., Plant, J. N., Schmechtig, C., Wong, A. P. S., Claustre, H., Trull, T. W., Udaya Bhaskar, T. V. S., Boss, E., Dall’Omo, G., Organelli, E., Poteau, A., Johnson, K. S., Hanstein, C., Leymarie, E., Le Reste, S., Riser, S. C., Rupan, A. R., Taillandier, V., Thierry, V. and Xing, X. (2019) A BGC-Argo Guide: Planning, Deployment, Data Handling and Usage. Front. Mar. Sci. 6:502, <https://doi.org/10.3389/fmars.2019.00502>.

A), the GEOMAR methods (panel B), pH data measured at the parking depth (panel C) and pH data measured at the parking depth minus reference (CANYON-B pH data, panel D) are presented. Figure A1 has been modified and is now included in Figure 5 of the revised manuscript. When pH data recorded at the parking depth are plotted, they are uncorrected. This precision has been added in the legend of Figure 5 (in the revised manuscript).

By splitting previous Figure 5 (A and B) into two separate figures, we believe that the new organization of the figures helps the reader to identify the impact of the sensor drift correction used on the final adjustment. While the impact of the correction method on the final corrected dataset is almost non-significant, this section aims to discuss the better representation of the sensor behavior over time and we believe that, by merging Figure A1 (in the original manuscript) to the original Figure 5 in revised manuscript, this purpose has been clarified to the reader. We also thank the referee for his/her suggestion to discuss the pH reference values in order to demonstrate that the jumps are related to the sensor response and the correction method. The new Figure 5D showing the difference between uncorrected pH data recorded at the parking depth and pH values from the algorithm estimate is in agreement with this statement as it shows smoothed transitions and the general pattern does not present noteworthy jumps. It confirms that the sudden jumps are related to changing sensor response rather than to the reference values used to correct float-pH data.

L.327: “The pH sensor behavior when the float drifts at its parking depth is in agreement with this observation (Fig. 5C). In comparison with float-pH data corrected using the SAGE method, no strong visible discontinuities in raw pH data are observed while the float drifts between its measurement’s phases. In our view, the sensor rather shows undulations in response with smooth and less smooth phases over time. In order to test the impact of the reference method on the adjustment pattern, differences between uncorrected float-pH data and CANYON-B pH data derived at the parking depth are presented in Figure 5D. Once again, the pH time-series shows smoothed transitions and the general pattern doesn’t present noteworthy jumps. Such sharp transitions can perhaps be best corrected with our modified GEOMAR segment method or alternatively with a spline fit or a 3-point centered running mean (Fig. 5B).”

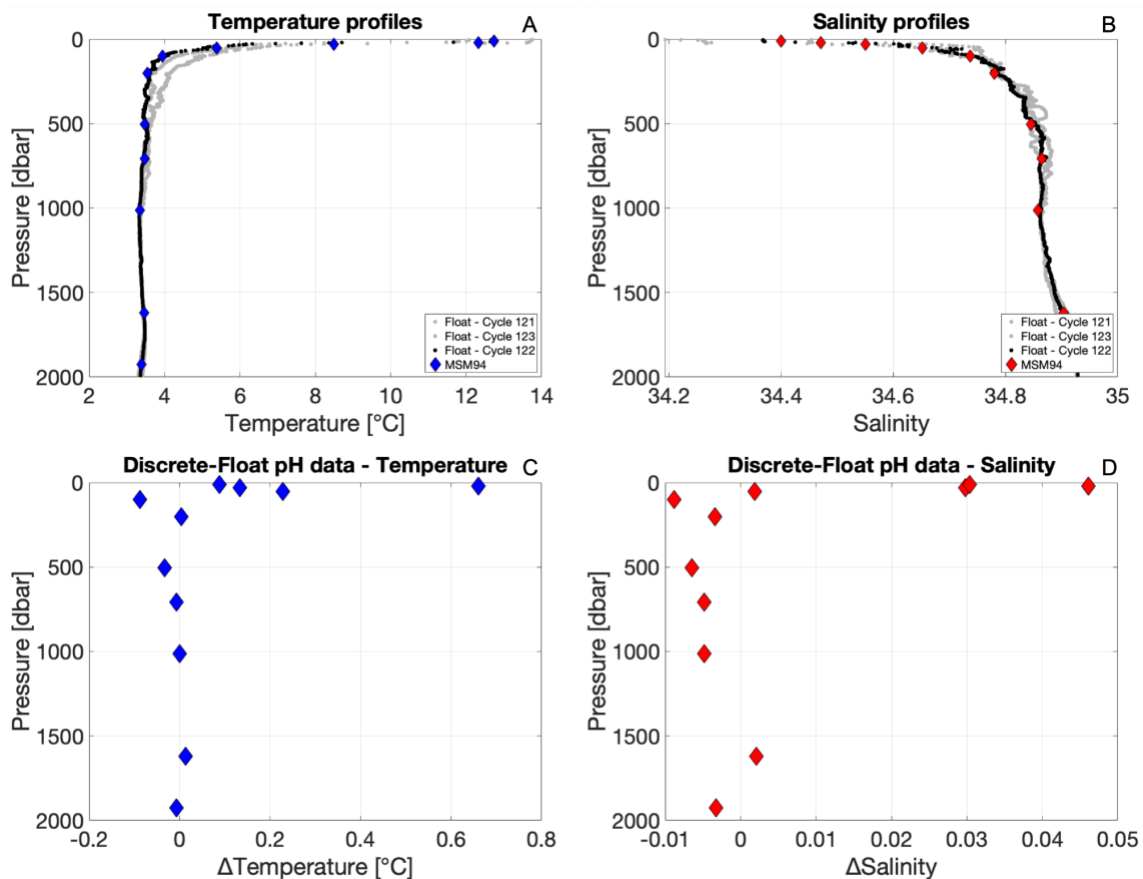


**Figure 1** (Fig. 5 in the revised manuscript). Mean differences between raw float-pH data minus float-pH corrected using the SAGE tool (Fig.5A), the cycle-by-cycle GEOMAR method (yellow dots, Fig.5B), and the linear mean regression

GEOMAR method (blue dots, Fig.5B) and the 3-point centered running mean correction method (green dots, Fig.5B) for float WMO 3901669. In every case, CANYON-B was chosen as a reference method, and 1500 dbar was chosen as the reference depth. Mean differences between raw and corrected float-pH data with the standard deviations are shown in the legend boxes for each reference method. Figure C shows, for comparison with the SAGE correction, the uncorrected pH data measured at the parking depth (right y-axis) with black dots representing mean pH values for each day. The colorbar shows pressure. Figure D shows differences between raw float-pH data minus float-pH corrected using the SAGE tool (purple dots, left y-axis) and differences between uncorrected mean pH data measured at the parking depth minus mean reference CANYON-B pH data calculated using measurements recorded at the parking depth (red dots, right y-axis).

Section 3.2.1: The authors make the point that the crossover between the float and discrete data is extremely close in time and space. I suggest that they compare temperature/salinity between the float and discrete profiles. This would add evidence that the same water masses were sampled by both.

We thank the referee for his/her suggestion. In the revised manuscript, a new figure has been added in the Supplementary Material (Figure 2 below) to support our comparison between float-pH data and discrete pH measurements. A sentence referring to this figure has been also added in the revised manuscript.



**Figure 2** (Fig. A1 in the revised manuscript). (A and B) Vertical profiles of temperature (A) and salinity (B) measured during the MSM94 cruise (diamond dots) and acquired by the float WMO 3901669 during cycles 121, 122 and 123 (gray and black lines, respectively). (C and D) Differences between discrete and float (cycle 122) temperature (C) and salinity (D) data calculated after matching in density space to avoid biases from internal waves.

Lines 356: “Differences between discrete and float temperature and salinity data add confidence in the density space matching performed in this study (Fig. A1).”

First paragraph of section 3.1.1: Very interesting discussion, but I think you could be clearer about whether the main reason for the big offsets between using 1500 and 1960 dbar as the reference depth in this region is due to the occasional large convection depths or just because there are offsets in the reference pH values between these two depths, i.e., depth based accuracy differences in CANYON-B or LIR.

The correction scheme for float-pH data involves a comparison of raw float-pH data against selected reference fields where concentrations are relatively stable and can be predicted using interpolation methods (multiple linear regressions or neural networks) based on shipboard observations. In the subpolar North Atlantic Ocean, inter-annually varying deep convection, water mass formation, as well as decadal variability affect water masses at a depth greater than 1500 dbar. The latter shows that the deep ocean correction at an arbitrarily chosen depth around 1500 dbar is not straightforward, and in this case, induced an uncertainty of at least 0.005 pH units. This may to some extent be a special characteristic of the Labrador Sea with its extreme winter convection depth of up to 1500 m. In this particular case, this speaks for a deeper reference level as: (1) raw float-pH data at this perturbed reference might be biased by some units and could not be used to correct the entire profile and, (2) the reference pH values are calculated using physical and biogeochemical properties at the selected pressure range and might change depending on the choice of the reference depth. In the revised manuscript, this statement is now more clearly expressed: “The deepest mixed layer depth estimated from the float time-series was at 1937 dbar, showing that the entire water column covered by the float profiles is probably affected. In this regard, the subpolar North Atlantic region with its deep-reaching anthropogenic CO<sub>2</sub> imprint is certainly a most difficult area for the unambiguous choice of a stable and unperturbed reference depth as both float-pH data and reference pH values could vary noticeably at the classical reference depth.” (L.238)

We also would like to point out that, in the revised manuscript, a presentation of pH data estimates with the ESPERs routines has been included.

L.284: “The new ESPERs methods attempt to resolve the issues encountered with existing routines (especially the OA estimate) by expanding their functionality and being trained on a larger data product. In comparison with the LIR-pH estimates, large differences are observed in the SNWA region and might be attributable to the OA adjustment as well as the omission of depth as a predictor variable from ESPER-LIR (Carter et al., 2021). Updated global algorithms (i.e., ESPERs) show comparable estimates in the SNWA area with ESPER-LIR pH estimates slightly higher than pH data estimated with CANYON-B or ESPER-NN. In the dynamic and strongly human-impacted studied region, the lack of coordinate information as a predictor variable in the ESPER-LIR routine could also be argued as an explanation of the observed differences. However, according to Carter et al. (2021), regional assessment statistics obtained in the Northern Atlantic indicate almost similar biases for both the ESPERs and the CANYON-B methods, with a better RMSE statistic for CANYON-B.”

Line 125: That deployment cast pH measurements are an ineffective means of calibrating float pH sensors seems very important to me. Suggest that you support this statement more strongly. The two papers cited are on oxygen optodes rather than pH sensors. How large a drift rate is typically observed at the start of deployment? When does this change? Could this be overcome with a conditioning period before deployment?

We thank the referee for his/her useful comment and we agree that, in the original manuscript, information about pH sensor behavior during the first cycles were missing.

Recently, Maurer et al. (2021)<sup>6</sup> tested the sensor performance upon deployment by considering the offsets associated with the first and second segments. Indeed, as each segment is treated independently, the value of any subsequent offset can provide information on sensor health over time when viewed relative to the first offset. In the Southern Ocean, a bias of -0.032 pH units has been observed for the first segment. However, similar to nitrate sensors, pH sensors are relatively stable after a few cycles and the second offset was null. The negative skew of the pH first offset distribution demonstrates that the majority of SOCCOM pH sensors are biased low upon deployment within the array. Maurer et al. (2021)<sup>6</sup> stated that this behavior is not surprising; oftentimes the largest anomaly is observed on the first cycle as the sensor re-conditions to an aqueous environment. Continued exposure to seawater at 1500 m helps to stabilize the sensors, particularly the pH sensor. After float-pH data adjustment, median differences between shipboard pH measurements and float-pH data at the time of deployment ranged between 0.006 pH units (Johnson et al., 2017)<sup>4</sup> and 0.002 pH units (Maurer et al., 2021)<sup>6</sup>, respectively. In the current literature (Johnson et al., 2018)<sup>5</sup>, it is therefore advised to perform the first delayed-mode quality control and adjustment after at least five cycles.

In the revised manuscript, additional information have been added. “*In the Southern Ocean, Maurer et al. (2021) reported an offset value for the first segment of -0.32 pH units, illustrating the sensor performance upon deployment caused by the lack of conditioning in some of the pH sensors as well as the sensor re-conditions to an aqueous environment. However, after float-pH data adjustments, Johnson et al. (2017) and Maurer et al. (2021) showed median shipboard bottle- minus-float differences of 0.006 pH units and 0.002 pH units, respectively.*” (L. 137)

### **Minor comments**

Internal waves: The authors take care to remove the effect of internal waves on the discrete – float data comparison. I suspect that this effect is implicitly removed in the calculation of reference pH at a specific depth by CANYON-B and LIR, but it would be useful to state that explicitly.

Internal waves can create mismatches in water properties at a given depth over time between a hydrographic cast and a float profile. As they exist only when the water body consists of layers of different densities, a matching in density space has been performed in our study to allow the comparison between discrete pH measurements and float-pH data. In the current float-pH data adjustment procedure, detailed in Maurer et al. (2021)<sup>6</sup>, the difference between a selected field of reference and the measured values is first calculated at nearly 1500 m where spatiotemporal variability of oceanic components is minimal. Thus, the correction scheme for sensor data depends on having accurate estimates of deep chemical concentrations, where concentrations are relatively stable and can be

---

<sup>4</sup> Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., et al. (2017). Biogeochemical sensor performance in the SOCCOM profiling float array. *J. Geophys. Res. Oceans* 122, 6416–6436. <https://doi.org/10.1002/2017JC012838>.

<sup>5</sup> Johnson, K. S., Plant, J. N., and Maurer, T. L. (2018). Processing BGC-Argo pH data at the DAC level, v1.0, Argo data management, <https://doi.org/10.13155/57195>.

<sup>6</sup> Maurer, T. L., Plant, J. N., and Johnson, K. S. (2021) Delayed-Mode Quality Control of Oxygen, Nitrate, and pH Data on SOCCOM Biogeochemical Profiling Floats, *Frontiers in Marine Sciences*, 8, 683207. <https://doi.org/10.3389/fmars.2021.683207>.

predicted using interpolation methods (multiple linear regressions or neural networks) based on shipboard observations.

Line 60: suggest changing “qualification” to “quality control”.

The word has been modified as suggested.

Line 62: suggest adding citations for the suggested procedures.

Citations have been added in the revised manuscript.

L. 67: *“For pH, numerous delayed-mode procedures have been suggested (Williams et al., 2016; Johnson et al., 2017) but a uniform, fully tested and globally-proven correction method is still missing.”*

Line 99: suggest rephrasing. Your paper is evaluating the accuracy range after data adjustment.

This sentence has been rephrased as suggested: *“The initial pH accuracy “claimed” by the manufacturer is of  $\pm 0.05$  pH units. Data adjustment can bring accuracies varying between  $\pm 0.005$  pH units (Johnson et al., 2017) and  $\pm 0.007$  pH units (Maurer et al., 2021).”* (L.107)

Line 106: suggest adding a few more details. Were the adjustments based on air-calibration or surface climatologies or ...?

Two references as well as some information about the adjustment have been added in the revised manuscript.

*“In our case, O<sub>2</sub> from the 10 pH-equipped Argo floats was adjusted following Argo procedures (Bittig et al., 2018a; Thierry et al., 2022) with in-air measurements and the adjustments are available in near-real time.”* (L. 116)

Line 200: “not” should be “note”

The sentence has been modified.

Lines 246-247: suggest clarifying what is meant by “keep the adjustment”. I wasn’t sure if you meant that you did apply the optional CANYON-B pH data adjustment to align with spectrophotometric data or you did not.

Carter et al. (2018)<sup>7</sup> have proposed an optional CANYON-B pH data adjustment that can be applied to data derived using this neural network method to align estimates with spectrophotometric pH measurements. As the LIR-pH training dataset consists of values either measured or calculated but adjusted using the same purified-dye adjustment (Equation 1; Carter et al., 2018<sup>7</sup>), we have decided to keep this adjustment to then compare the corrected datasets. In the revised manuscript, this is now more clearly stated: *“In this study, we have decided to include this reference-pH data adjustment to correct*

---

<sup>7</sup> Carter, B. R., Feely, R. A., Williams, N. L., Dickson, A. G., Fong, M. B., and 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>.

*float-pH data: a linear transformation was applied to CANYON-B pH estimates to bring estimates back into alignment with spectrophotometrically measured pH.” (L. 266)*

Line 332: I think “as” should be “than”

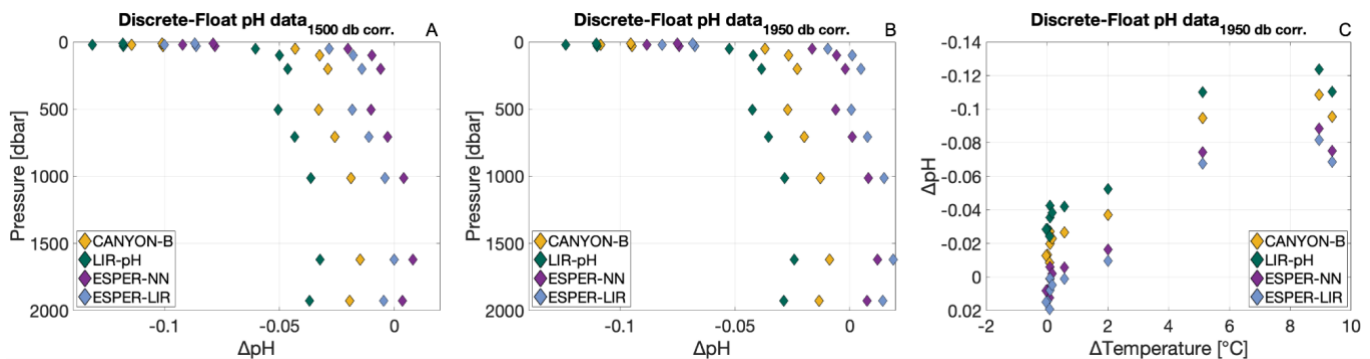
The word has been changed.

Line 394: remove comma after both

The comma has been removed.

Figure 6d: Add y-axis labels and negative signs to all y-axis labels.

We thank the referee for pointing out that Figure 6D was incomplete in the original manuscript. In the revised manuscript, Figure 6 (Page 16 in the revised manuscript) has been re-drawn, reduced and now presents differences between discrete pH measurements and float-pH data along the water column and according to two distinct reference pressure levels (1500 dbar Fig. 6A and 1950 dbar Fig. 6B) and  $\Delta pH$  (discrete pH measurements minus float-pH data corrected at the reference depth level 1950 dbar, Fig. 6C) as a function of the difference between discrete water temperature and temperature values recorded at the reference depth of 1950 dbar (i.e. 3.3733°C). On Fig. 6C, y-axis labels have been corrected.



**Figure 3** (Fig.6 in the revised manuscript). (A and B) Differences between discrete and float-pH data (for the cycle 122) calculated after matching in density space to avoid biases from internal waves and corrected using corrected reference levels of 1500 dbar (Fig. 6A) and 1950 dbar (Fig. 6B). (C)  $\Delta pH$  (discrete pH measurements minus float-pH data corrected at the reference depth level 1950 dbar) as a function of the difference between discrete water temperature (i.e., the temperature measured in situ at the time of bottle triggering at sea) and temperature values recorded at the reference depth of 1950 dbar. The color code refers to the reference method used to correct float-pH data: CANYON-B (yellow diamonds), LIR-pH (green diamonds), ESPER-NN (purple diamonds) or ESPER-LIR (blue diamonds).