Response to anonymous referee 1:
(https://doi.org/10.5194/egusphere-2023-3143-RC1)

We thank the reviewer for their feedback and constructive comments on our manuscript. Our responses are in blue below.

Summary:

The authors compare the means of pCO2 data measured by ships (1972-2021) with pCO2 data derived from Argo floats (2014-2023) in the Southern Ocean. They find an increase in the float pCO2 values compared to the ship pCO2 and explain the mean difference by seasonality, trends in atmospheric CO2, differences in sampling location, errors in TA, and the choice of carbonate system constants. Consequently, they adjust the pCO2 values by removing the influence of these factors on the mean (e.g., normalizing the data to a reference year). They attribute the remaining difference in mean pCO2 to quality issues. While I appreciate the concept of comparing float pCO2 data with ship pCO2 data, I have major issues with this study. My greatest concerns are as follows:

1. **Content**: The study falls short in making a meaningful contribution to the existing knowledge base. It lacks the presentation of any novel findings. (Higher pCO2 observed in Argo float data than ship data → partially caused by seasonality, different sampling location etc). The conclusion of "bad data quality" appears inadequate given the methodology and is insufficiently discussed).
   Response: we agree with the reviewer that we did not adequately convey the validity of our methods and the novelty of our findings to readers. The reviewer is correct that there have been multiple studies comparing pCO_2 estimates from different observational platforms in surface waters. The possibility of a discrepancy in sea surface pCO_2 data between float-based and ship-based approaches has been considered and investigated in a number of previous studies (Jin et al., 2024, Wimart-Rousseau et al., 2023, Wu and Qi, 2022, Sutton et al., 2021, Mackay and Watson, 2021, Long et al., 2021, Fay et al., 2018). As suspected by some of these authors, we have indeed identified significant deviations that merit further investigation.

   Our novel finding sheds light and provides greater context on the important and controversial question as to why the air-sea CO_2 annual flux calculated from float data is inconsistent with that from other platforms. It remains to be explained why other observations (e.g. from unmanned surface vehicles,

reconstructed pseudo-observations and aircraft-based flux results) are in better agreement with sparse ship observations but differ significantly from float data (Sutton et al., 2021, Mackay and Watson, 2021, Long et al., 2021, Jin et al., 2024). The data collected by these different modalities are broadly consistent, whereas fluxes based on float data suggest significantly lower carbon dioxide uptake. Moreover, sensors on aircraft have not detected the carbon dioxide outgassing at high latitudes in the Southern Ocean predicted by the float data (Long et al., 2021). This study is the first attempt at a novel approach to assessing float data quality and it suggests that the answer to the question is a bias in average float pCO_2. The amount of float data we used to check for discrepancies exceeds that used in previous comparison methods. Further justification of our approach and exploration of uncertainty sources are given below and will be added to an amended version of the manuscript (we concur that it needs further discussion).

2. **Methodology:** I fail to understand the rationale behind comparing data from various time periods, seasons, and sampling locations in the first place, particularly when focusing solely on the mean values. In my opinion, this approach is simply not acceptable, as e.g. ocean biogeochemistry undergoes changes over time, leading to higher CO2 levels in more recent float data. While the authors acknowledge this in their later analyses presented in the discussion section, the results in the results section of the study are therefore not comparable. Additionally, the study does not quantify the sources of uncertainty in float pCO2 data, rendering the conclusions regarding data quality issues questionable.

   Response: in the following, we describe shortcomings of the main method used previously for comparing pCO_2 data from float and ship. We also justify the method applied in this study and explain why it is novel, appropriate and useful.

   The main method that has been used prior to this work to assess float data quality is crossover comparisons, i.e. direct comparison of ship and float data when measurements from both are made at the same place and time. While of course valuable, unfortunately there are limitations to this approach. Firstly, ships are only very rarely in the same place as a float at the same time. Therefore, only a very small proportion of the total amount of data can be used in crossover analyses (less than 1%, up to 2023 December). In contrast, our approach compares the totality of float and ship data. Secondly, nearly all the crossover comparisons are made within 3 days of the time that a float

was deployed (Gray et al., 2018, Johnson et al., 2017), because that is the only time when a ship and a float are likely to be coincident. Crossover comparisons made almost exclusively at time of deployment cannot assess lifetime performance of float pH sensors (and thus the pCO_2 estimates that are derived) whereas our method can. Some additional analyses (part of future work intended for another manuscript, but shown below) suggest that float age dramatically affects the coherence of float pCO_2 data while the oxygen data shows excellent agreement between young and old floats (Figure.1). Although neither overall approach is without shortcomings, our bulk data comparison method is an alternative method of assessing float data quality that is able to assess float sensor performance across whole lifetimes of float deployment. We suggest that it is a valuable complement to crossover analysis.

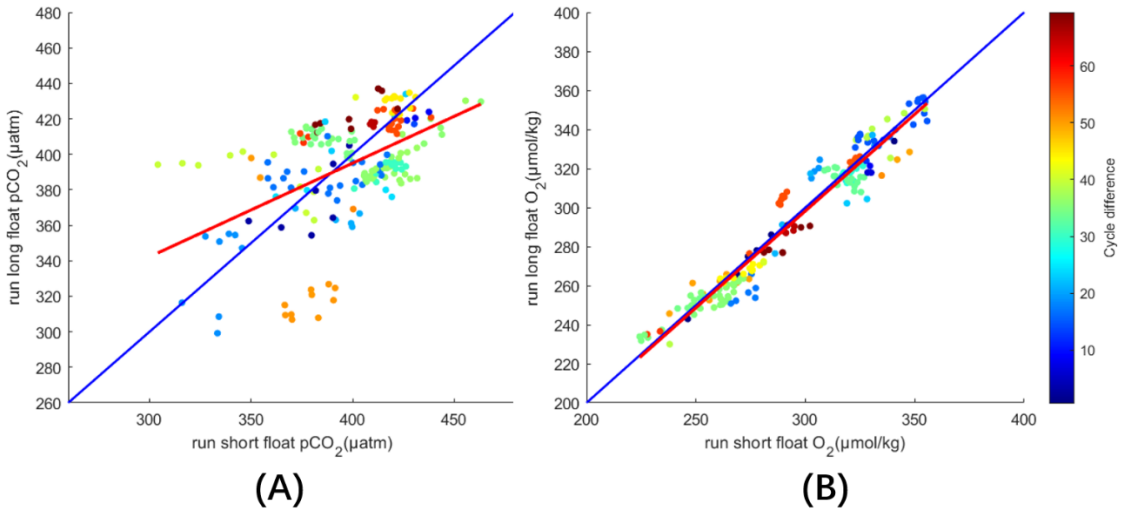

Figure (1): Crossover comparison between different floats when they are coincidentally adjacent in time and space. (A) pCO_2 comparison between floats having conducted < 50 profiles and floats having conducted > 50 profiles. The best-fit line in red is y=0.53x+184 (r=0.51); (B) The same comparison but for O_2 from adjacent floats. The best-fit line in red is y=0.99x+0.78 (r=0.97). Points in both scatterplots are coloured according to the difference in numbers of profiles carried out (as an indication of differences in time since deployment). Two floats are considered adjacent when within 400 km in distance and 7 days in time (Wimart-Rousseau et al., 2023).

We present evidence here that justifies our approach. When two means of measuring the ocean are both measuring correctly then we would expect the large-scale patterns across a basin to agree with each other. We show below that this is true for other parameters measured by floats, although in some

cases only after sampling biases are taken into account. We also show that that discrepancies in pCO_2 remain even after sampling biases are corrected for. The reviewer does make a good point, and we agree that it would have been useful to show the comparison for other parameters and we will add this to the manuscript.

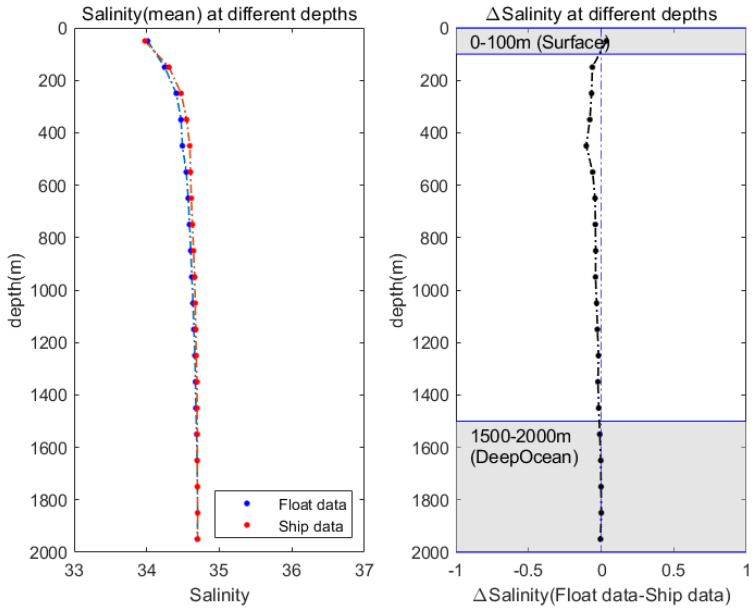

Figure (2): Salinity(mean) at different depths. The left panel shows the average float and ship salinity at different depths; the right panel shows the difference between the two (float salinity minus ship salinity).

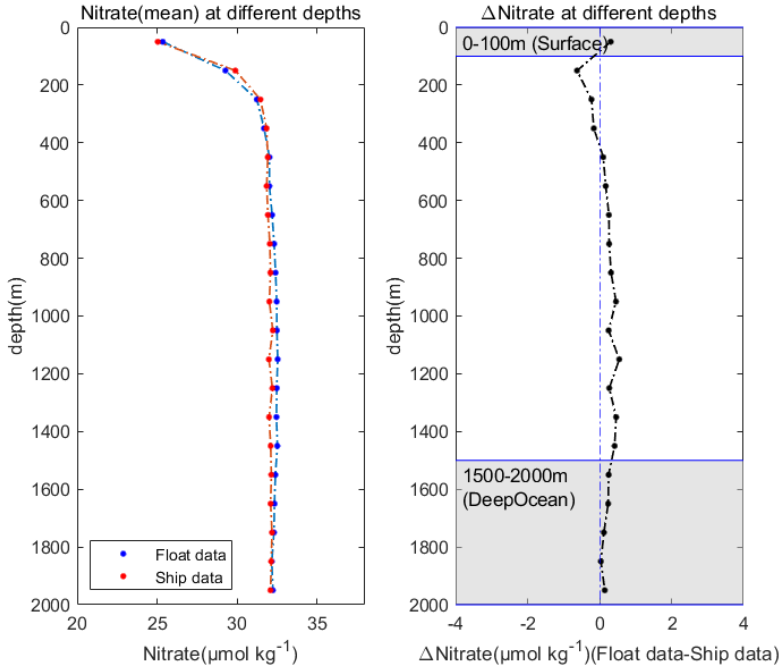

Figure (3): Nitrate(mean) at different depths. The left panel shows the average float and ship nitrate at different depths; the right panel shows the difference between the two (float nitrate minus ship nitrate).

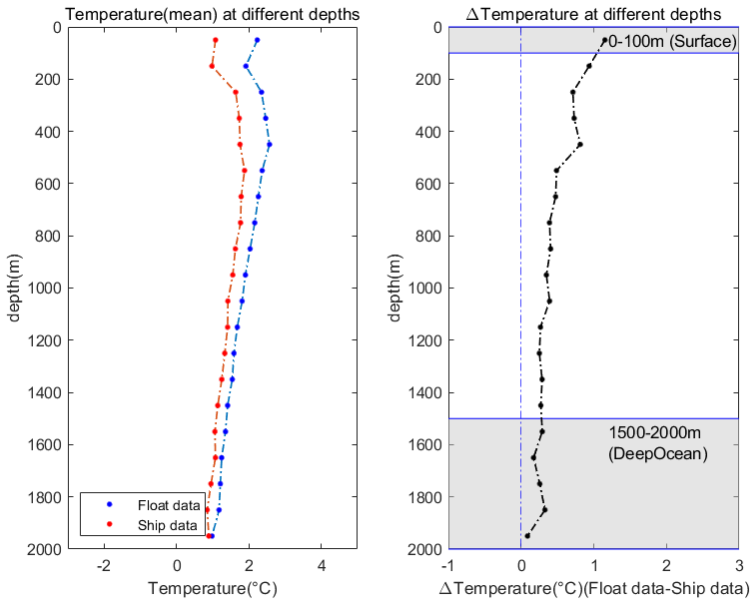

Figure (4): Temperature(mean) at different depths. The left panel shows the average float and ship temperature at different depths; the right panel shows the difference between the two (float temperature minus ship temperature).

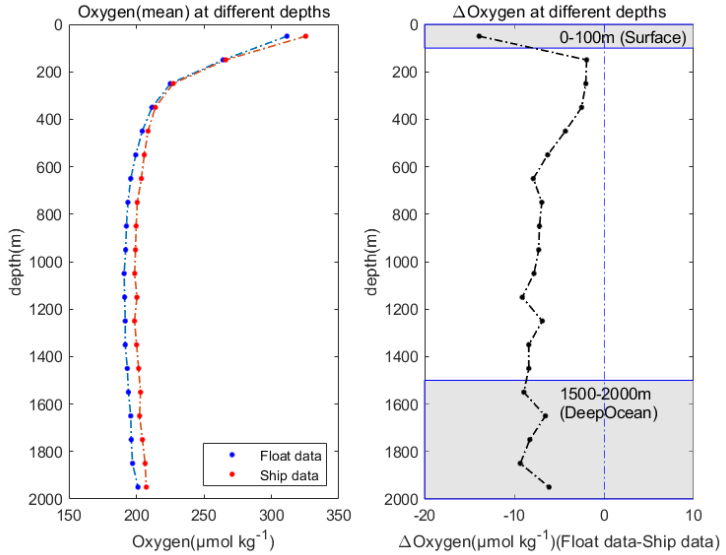

Figure (5): Oxygen(mean) at different depths. The left panel shows the average float and ship oxygen at different depths; the right panel shows the difference between the two (float oxygen minus ship oxygen).

The nitrate and salinity profiles show very good consistency between ship data and float data, while the temperature and oxygen profiles are not well aligned. The misalignment in temperature is explained by a latitudinal gradient in temperature and a sampling bias between ships and floats. More ship data comes from areas further south than does float data (Figure.6).

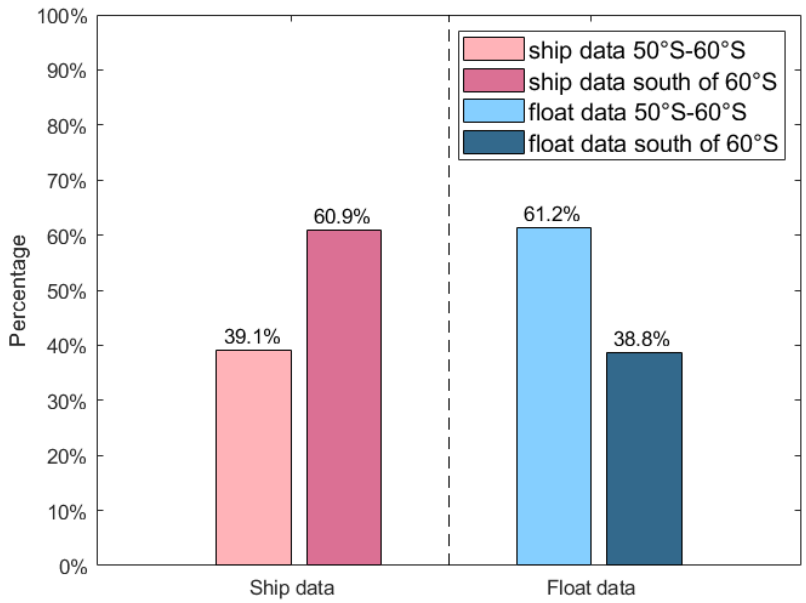

Figure (6): Proportions of float and ship data from between 50°S and 60°S versus south of 60°S.

The plots below show a comparison between ship temperature and float temperature when the data is separated into 2 regions: between 50°S-60°S and south of 60°S.

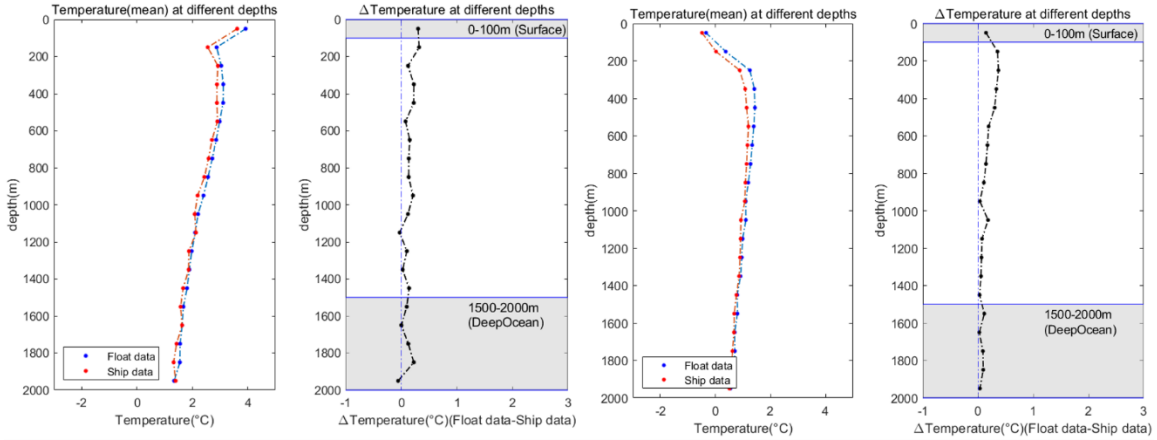

Figure (7): Temperature(mean) and difference at different depths in different regions. (A) float and ship data located between 50°S and 60°S; (B) float and ship data south of 60°S region.

The temperature discrepancies are explained by latitudinal effects (Figure.7). This raises the question as to whether a latitudinal effect could also explain the pCO_2 discrepancy. The plots below show the effect of latitude on the pCO_2 discrepancy (Figure.8). The pCO_2 difference in surface waters is 7.0 μ

atm between 50-60°S and 15 μatm south of 60°S. The discrepancy exists in both regions and is in line with the average discrepancy derived in our manuscript. We recognise the necessity of adding a discussion of the effect of latitude on average pCO_2 discrepancy to the next version of the manuscript. This will be discussed in addition to the other possible sampling biases already considered (seasonal, spatial and temporal).

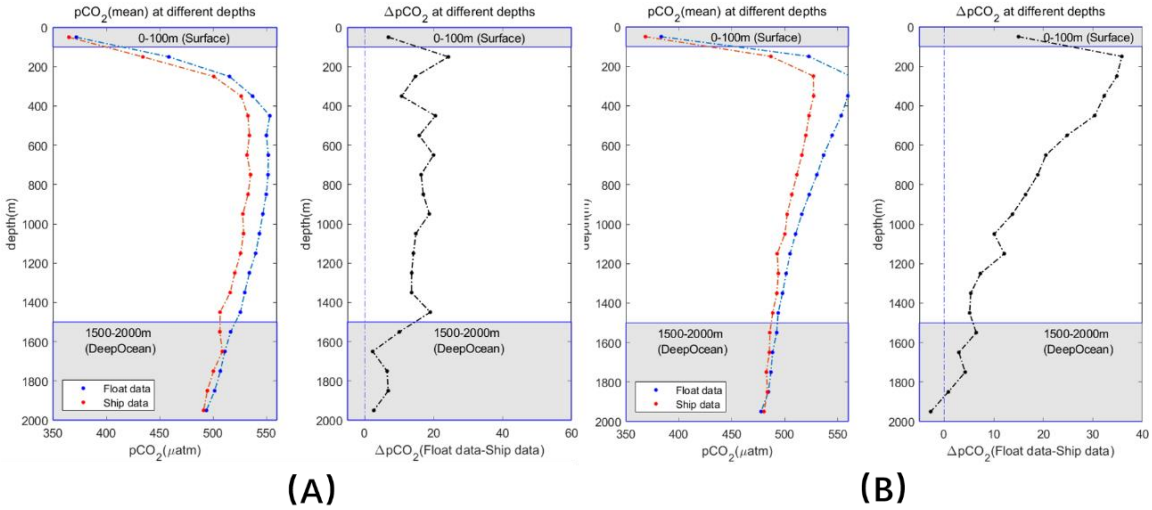

Figure (8): pCO_2(mean) and difference at different depths in different regions. (A) float data and ship data located in 50°S-60°S region; (B) float data and ship data located in south of 60°S region.

The solubilities of gases dissolved in seawater are mainly controlled by temperature. Oxygen and carbon dioxide gas concentrations therefore tend to be higher in colder waters. To counteract the effect of a potential sampling bias in temperature (due to a greater proportion of ship data coming from further south where waters are colder; Figure 6) potentially leading to a bias in CO_2 and O_2 gas concentrations), we calculated the O_2 saturation anomaly ($\Delta[O_2]$) ($\Delta[O_2] = \left(\frac{[O_2]_{observed}}{[O_2]_{\text{saturation}}} - 1\right) \times 100\%$) and the CO_2 saturation anomaly ($\Delta[CO_2]$) ($\Delta[CO_2] = \left(\frac{[CO_2]_{observed}}{[CO_2]_{\text{saturation}}} - 1\right) \times 100\%$) and show the results below.

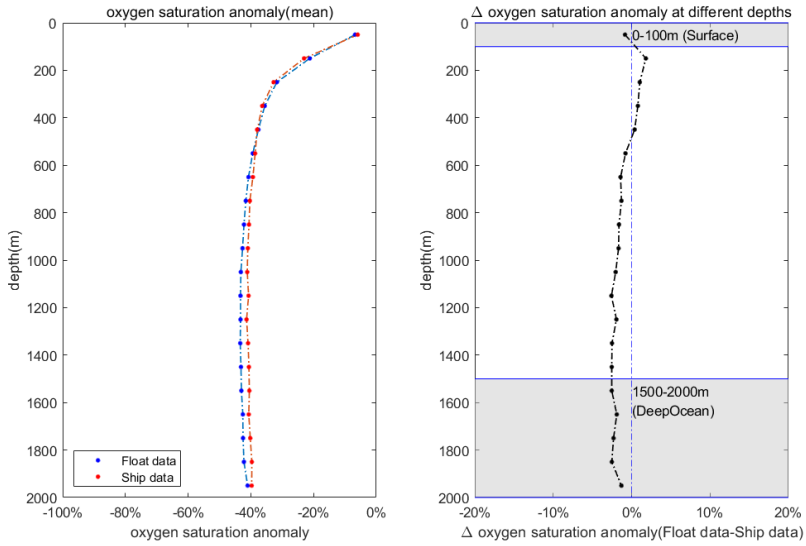

Figure (9): Oxygen saturation anomaly(mean) at different depths. The left panel shows the average float and ship oxygen saturation anomaly at different depths; the right panel shows the difference between the two (float oxygen saturation anomaly minus ship oxygen saturation anomaly).

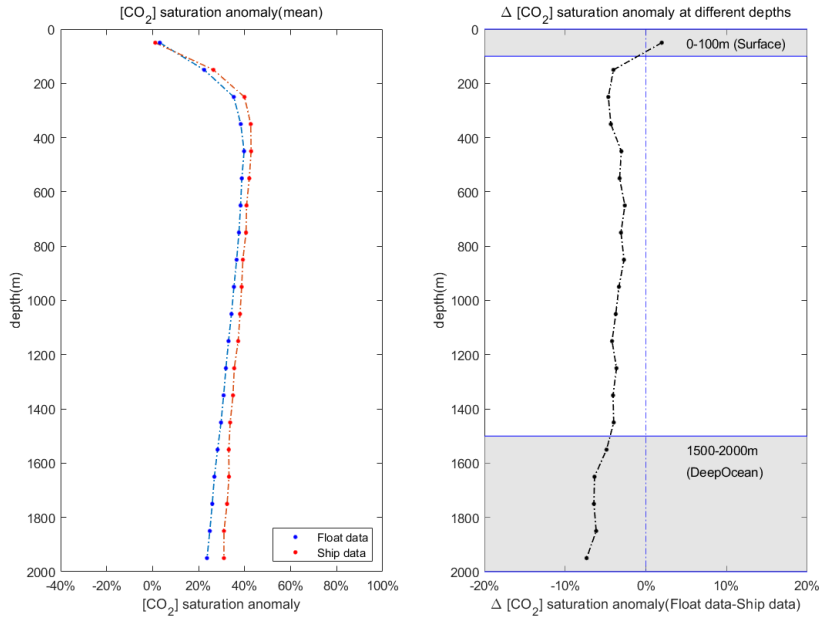

Figure (10): $[CO_2]$ saturation anomaly(mean) at different depths. The left panel shows the average float and ship oxygen saturation anomaly at different depths; the right panel shows the difference between the two (float $[CO_2]$ saturation anomaly minus ship $[CO_2]$ saturation anomaly).

Based on the saturation anomaly results (which correct for temperature differences), the float $\Delta[O_2]$ is in rough overall agreement with ship $\Delta[O_2]$ (Figure.9). The float surface $\Delta[CO_2]$ is however approximately 2% higher than ship surface $\Delta[CO_2]$, which converts (at pCO_2 of 400 μatm) to a pCO_2

difference of around 8 μatm. After correcting for various effects and possible biases, the calculated discrepancy in pCO_2 is thus close to the results in our manuscript. The large-scale patterns across the Southern Ocean are similar between ship and float data for nitrate, salinity, temperature and oxygen (after removal of latitude/solubility effects) (Figure.2-5,7,9). The fact that the large-scale patterns do not agree for pCO_2 even after correcting for sampling biases (Figure.8,10) is therefore a point of interest.

We corrected for accumulation of anthropogenic CO_2 over time in surface waters by using the same method as Wu et al. (2019) (their section 2.1), which in turn built on methods described by Takahashi et al. (2009) (their section 2.4). Moving the results of this analysis into the results section, as recommended by the reviewer, will make this clearer and we are happy to do this in the amended version.

The uncertainty in each independent float pCO_2 data value does not affect our finding that float pCO_2 is systematically high; assuming a normal distribution in individual float uncertainties, as our results are based on a significantly large number of data points, and the standard error of the mean (the standard deviation of mean values) decreases as a function of $\left(1/\sqrt{N}\right)$ where *N* is the number of data points, then the effect of individual point estimate uncertainties becomes negligible. Williams et al. (2017) estimated the uncertainty of an individual float pCO_2 value to be around ±11 μatm when float pCO_2 is 400 μatm. In the figure below we show the probability density function of average float pCO_2 minus ship pCO_2 from 1000 Monte Carlo iterations. This figure was generated by the following procedure: (1) assume ship average pCO_2 to be 400 μatm, (2) generate 30,000 independent float pCO_2 values, each equal to $400 + G(0,11)$, where $G(\mu, \sigma)$ is a random number from a normal (Gaussian) distribution with mean of μ and standard deviation of σ, (3) calculate the average float pCO_2 and then the difference between ship and float average values, (4) repeat 1000 times to obtain 1000 differences, (5) plot the frequency distribution of the differences. The effect of uncertainty in each single point of float pCO_2 data on the difference in the final float mean is minor (Figure.11). This procedure assumes that errors are random and independent. It does not hold for systematic biases, but that of course is what we are investigating in our study.

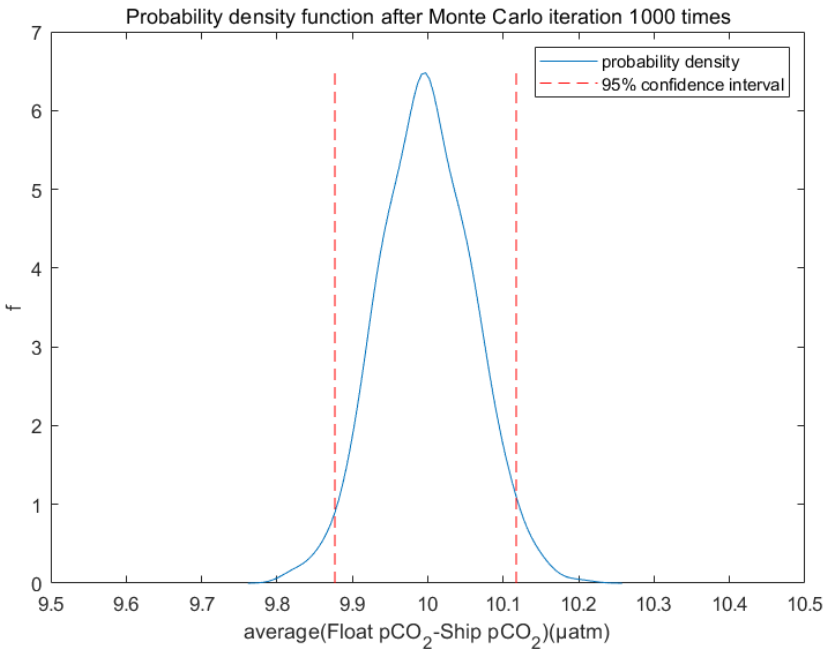

Figure (11): Assessment of the impact of uncertainty in individual float pCO_2 data on the uncertainty in the overall value of (float pCO_2 – ship pCO_2), based on Monte Carlo calculations.

We thank the reviewer for raising this point and will add our response to it to our manuscript. We hope that this, together with the other additions, will be seen to have improved the discussion of the merits of this method and additionally will have addressed the reviewer's concerns that our method is not suitable.

3. **Structure:** Result sections 3.1, 3.2 and 3.3 should be merged as the subsections merely contain different plots. The discussion section comprises the presentation of additional analyses, thereby resembling more of a result section.

   Response: we agree with this suggestion. We will restructure the results and discussion sections according to this comment in the next version of the manuscript.

4. **Choice of visualization:** The content in Figure 2, 3, S1 as well as 4, 5, S2 could be merged (remove scatterplots, add error bars to line plots).

   Response: we appreciate this suggestion, and we will take it into account and make this change in the next vision of the manuscript.

5. Authors doubt/question data quality without further arguments (l.200-206). After adjusting the means, they did not go into "float pH data quality issues". I would have appreciated a discussion on why the quality is perceived as poor and how it could be improved etc.

   Response: we thank the reviewer for this constructive suggestion. We highlight the quality of float pCO_2 data (estimated from pH data) because this is the most likely explanation for the finding in our results that float pCO_2 exhibits an overall bias in all our analyses, even after considering (and, where relevant, correcting for) various possible sampling biases. Another reason is that no such bias is seen when comparing O_2 data corrected for temperature, whereas a significant bias is seen in pCO_2 data corrected both for increasing temperature and increasing atmospheric CO_2 (figure 10). We hope that the additions we will make to the manuscript (described above) provide the further arguments the reviewer would like to see.

   A float pCO_2 bias explains well the large difference between the fluxes calculated by the floats and the fluxes calculated by the other observing platforms. We look forward with anticipation to improved calibration of the float pH data and estimated pCO_2, but it is not within the scope of this study to suggest how it should be done. Instead, we present new evidence that the float pCO_2 is anomalously high through novel methods, bringing new information to an important field of research. Subsequent work will hopefully ascertain the reasons for this and therefore the solutions. The best process for processing float pH data (and from it float pCO_2 values) remains open to discussion; we expect that our findings will eventually contribute to higher accuracy of float pCO_2 data.

   We appreciate the reviewer's many minor, detailed comments and will attend to these in the revised version of the manuscript.

References:

FAY, A. R., LOVENDUSKI, N. S., MCKINLEY, G. A., MUNRO, D. R., SWEENEY, C., GRAY, A. R., LANDSCHÜTZER, P., STEPHENS, B. B., TAKAHASHI, T. & WILLIAMS, N. 2018. Utilizing the Drake Passage Time-series to understand variability and change in subpolar Southern Ocean pCO 2. *Biogeosciences,* 15**,** 3841-3855.

GRAY, A. R., JOHNSON, K. S., BUSHINSKY, S. M., RISER, S. C., RUSSELL, J. L., TALLEY, L. D., WANNINKHOF, R., WILLIAMS, N. L. & SARMIENTO, J. L. 2018. Autonomous biogeochemical floats detect significant carbon dioxide outgassing in the high‐latitude Southern Ocean. *Geophysical Research Letters,* 45**,** 9049-9057.

JIN, Y., KEELING, R. F., STEPHENS, B. B., LONG, M. C., PATRA, P. K., RÖDENBECK, C., MORGAN, E. J., KORT, E. A. & SWEENEY, C. 2024. Improved atmospheric constraints on Southern Ocean CO2 exchange. *Proceedings of the National Academy of Sciences,* 121**,** e2309333121.

JOHNSON, K. S., PLANT, J. N., COLETTI, L. J., JANNASCH, H. W., SAKAMOTO, C. M., RISER, S. C., SWIFT, D. D., WILLIAMS, N. L., BOSS, E. & HAËNTJENS, N. 2017. Biogeochemical sensor performance in the SOCCOM profiling float array. *Journal of Geophysical Research: Oceans,* 122**,** 6416-6436.

LONG, M. C., STEPHENS, B. B., MCKAIN, K., SWEENEY, C., KEELING, R. F., KORT, E. A., MORGAN, E. J., BENT, J. D., CHANDRA, N. & CHEVALLIER, F. 2021. Strong Southern Ocean carbon uptake evident in airborne observations. *Science,* 374**,** 1275-1280.

MACKAY, N. & WATSON, A. 2021. Winter air‐sea CO2 fluxes constructed from summer observations of the polar southern ocean suggest weak outgassing. *Journal of Geophysical Research: Oceans,* 126**,** e2020JC016600.

SUTTON, A. J., WILLIAMS, N. L. & TILBROOK, B. 2021. Constraining Southern Ocean CO2 flux uncertainty using uncrewed surface vehicle observations. *Geophysical Research Letters,* 48**,** e2020GL091748.

TAKAHASHI, T., SUTHERLAND, S. C., WANNINKHOF, R., SWEENEY, C., FEELY, R. A., CHIPMAN, D. W., HALES, B., FRIEDERICH, G., CHAVEZ, F. & SABINE, C. 2009. Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans. *Deep Sea Research Part II: Topical Studies in Oceanography,* 56**,** 554-577.

WILLIAMS, N., JURANEK, L., FEELY, R., JOHNSON, K., SARMIENTO, J. L., TALLEY, L., DICKSON, A., GRAY, A., WANNINKHOF, R. & RUSSELL, J. 2017. Calculating surface ocean pCO2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles,* 31**,** 591-604.

WIMART-ROUSSEAU, C., STEINHOFF, T., KLEIN, B., BITTIG, H. & KÖRTZINGER, A. 2023. Technical note: Enhancement of float-pH data quality control methods: A study case in the Subpolar Northwestern Atlantic region. *Biogeosciences Discuss.,* 2023**,** 1-26.

WU, Y., HAIN, M. P., HUMPHREYS, M. P., HARTMAN, S. & TYRRELL, T. 2019. What drives the latitudinal gradient in open-ocean surface dissolved inorganic carbon concentration? *Biogeosciences,* 16**,** 2661-2681.

WU, Y. & QI, D. 2022. Inconsistency between ship-and Argo float-based p CO 2 at the intense upwelling region of the Drake Passage, Southern Ocean. *Frontiers in Marine Science,* 9.