Zhang review

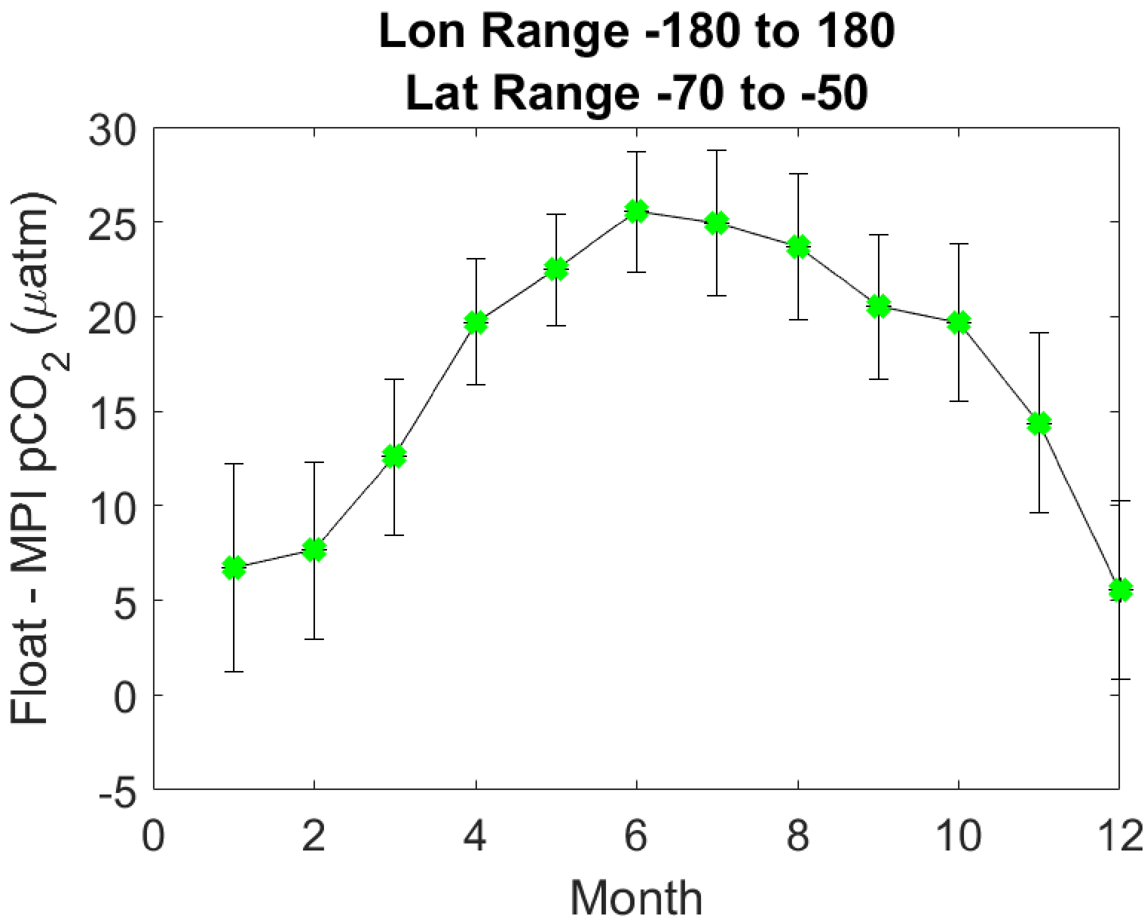
This paper is focused on assessing possible biases in pCO2 values computed from profiling float pH data and estimates of total alkalinity derived from various empirical models fitted to GLODAPv2, shipboard data.This is the third paper from the lead authors on this topic (Wu et al., 2022; 2023). It reiterates some earlier analyses, adds a few new examples, but doesn't add much fundamental new information.

There is no real doubt that there are biases in the profiling float pCO2 estimates, relative to highly calibrated observations made from ships. The biases, when correctly assessed, generally appear to be within the error limits set in the original Williams et al. paper. That is 2.7% (standard uncertainty) of the pCO2 value. The critical issue, though, is whether the autonomous observations are useful, despite the biases. The paper doesn't contribute any real understanding in this regard. It seems more oriented to providing a quite negative perspective. It does this by first reporting uncorrected comparisons, such as those in Figure 3, and stating the bias in float pCO2 is 37 µatm (line 201). Only after that do the authors ultimately apply a set of corrections for age of the ship data, etc., and then, pages later, find the bias from this comparison is 9 µatm (line 332, Figure 9). Nor do they make it easy for the reader to appreciate that the shipboard pCO2 values that they are using as a target also likely have significant biases. The shipboard values are not direct measurements. They are derived from dissolved inorganic carbon and alkalinity with their own, attendant uncertainty.

Perhaps the most robust bias assessment in the manuscript is the direct comparison to SOCAT pCO2 measurements in the Drake Passage, where they find a bias (float minus SOCAT) of 8.3 µatm. A value of 8.3 µatm is about what we would estimate (5 to 7 µatm with the most recent SOCAT data from the Southern Ocean) and that value is still within the error limits of 11 µatm at near surface pCO2 values reported by Williams et al. That is nowhere near the values they state on line 201 (37 µatm) and various other places in the manuscript. The SOCAT comparison is not particularly a new result since values from a similar approach (but more rigorous cross over requirements) are reported in Bushinsky and Cerovecki (2023). It is also much less comprehensive than the Bushinsky and Cerovecki result, which considered all SOCAT crossover data.

Finally, despite the very negative assessment of the paper, the pCO2 values reported in the paper are proving quite useful. Examples would be the papers by Prend et al. (2022; doi: 10.1029/2021GB007226), Chen, Haumann et al. (2022; doi: 0.1029/2021GB007156), or a new paper Carranza et al. (provisionally accepted, Extratropical storms induce carbon outgassing over the Southern Ocean).

Another way to look at the value of the float pCO2 estimates is a comparison of float pCO2 with the machine learning products trained on SOCAT data (Johnson et al., presentation at the 2024 Ocean Sciences Meeting). The machine learning products are the values actually used to compute air-sea CO2 flux. The following figure shows the monthly mean differences in all float surface pCO2 estimates from 70S to 50S (the authors area of focus) with the MPI SOM-FFN values in this region and in each unique month. During Southern Ocean summer, when SOCAT data are available, the float minus MPI offset is about what one would expect from the float minus SOCAT comparisons. The mean is about 7 µatm. But in winter, the differences increase dramatically to as high as 25 µatm. There are really only two explanations for this. One is that the float pCO2 bias increases in winter. There is no apparent explanation for that. The second is that machine learning methods, with little to no data to constrain them in winter, also have a bias. In this case, the machine learning value is underestimating the pCO2 increase observed by floats during winter.



A second example would be the recent paper by Hauck et al. (2023; doi: 10.1029/2023GB007848). In their Figure 5, the air-sea CO2 flux during Southern Ocean summer that is computed from a machine learning product that merges ship and float data is no different than a flux computed with a product based on ship only data. But in Southern Ocean winter (Hauck Figure 4), the ship only product has a much lower outgassing (pCO2 too low?) in the 50 to 70 S region when compared to the ship plus float product.

I'd argue these results demonstrate the value of float pCO2 estimates. That is not a message one would get from this paper.

In summary, a significant paper of this type does would do one of three things. It might identify an unknown bias. It might report a known bias and propose a solution (e.g. the various papers on air-oxygen recalibration of profiling float oxygen sensors). Or it might assess whether the data are useful despite the bias This paper does none of these. The general assessment of bias reported here has been discussed in more than 10 other papers, including 2 by these authors, and is well known. The mean bias of those assessments is generally consistent with the error estimates in Williams et al. It is also very unclear from this paper what the authors think the bias is; 37 µatm (line 201), 8.3 µatm (line 346), etc. Finally, it clearly doesn't appreciate that there is significant value in the pCO_2 estimates derived from float profiles. These observations lead me to conclude the paper contributes little to our understanding of pCO_2 or air-sea CO_2 flux.

Ken Johnson
MBARI