

Anonymous Referee #2

This manuscript is badly organized with a lot of important information left out. There are a lot of issues with the experiment, and the results do not support the author's points at all. This manuscript should be rejected. The authors need to re-design and re-do all the experiments, and follow the instructions of Ocean Science journal to re-write the manuscript in the correction structure (Introduction, Experimental, Results and Discussion, Conclusion). **The manuscript has been restructured in line with the journal instructions and consistent with previously published manuscripts within the journal.**

1. The background info about the importance of pH in the ocean (line 1 – line 48) seems to be a bit tedious, which needs to be consolidated. **[1] The information has been consolidated.**
2. In the introduction section, it'd be very useful to list the accuracy requirements of each application (e.g. ocean acidification research, routine water monitoring, aquarium management, and aquaculture). **[2] Information about accuracy requirements has been added.**
3. The author should mention that the state-of-the-art lab method for seawater pH measurement is the indicator dye-based spectrophotometric method. A table of the accuracy and precision of each method (i.e. lab based spectrophotometric method, spectrophotometry-based sensors, potentiometry-based sensors, and the solid state sensor) would be very useful here. **Data pertaining to the accuracy and precision of each technique is readily available and so we feel a detailed comparison is not necessary in the current manuscript.**
4. The author should write about the chemistry behind the solid state pH sensor either in the 'Introduction' section, which is important for a technical note like this. **[3] The chemistry behind the sensor has been briefly mentioned. The scientific papers about both pH sensing and reference tracking chemistry previously published have been cited.**
5. Some contents of the 'Technology' section belong to the 'Introduction', and the rest should be moved to the 'Result and Discussion'. **The manuscript has been restructured.**
6. Line 91 – Line 92: Do you mean the 'peak height' changes for reference tracking electrode? **No, it is related to the actual peak potential of the reference electrode.**
7. The titles of Figure 2 need to be consistent with the figure caption ('(a) pH sensing electrode', '(b) reference tracking electrode'). **[4] Adapted.**
8. Figure 3(a): Need to add parameter name and unit for Y-axis. **Name and unit have been added.**

9. How is experiment done for Figure 3? What pH buffer solution was used? Was it a fresh water buffer or seawater buffer? How was the temperature controlled? What was the temperature? [5] The sensor was just deployed in synthetic seawater, and the resulting pH was calculated with the combination of both pH sensing and reference tracker electrodes. Temperature controlled information has been added.
10. Need to use a smaller Y-range in Figure 3b. I don't know how the experiments was done, but such pH variation is actually too big for most ocean acidification research. Smaller Y-range has been used for Figure 3B (newly Fig. 4B). The pH variation shows an accuracy of ± 0.05 in pH, as presented in the paper. However, this experiment was mainly performed to show the behaviour of the reference tracking electrode when the Ag/AgCl reference drifted, proving the ability of our system to self-calibrate. More controlled experiments were performed (Fig 4 – newly Fig 5) to study the accuracy of the system and the ability to use it as an ocean sensor.
11. Line 120: You need to use pH buffer solution for such experiment. Such as tris-buffer in synthetic seawater. How is the temperature controlled? [5] We do not use pH buffers solutions for calculating the pH as the drift is corrected by the reference tracking electrode. Thus, the system does not need to be calibrated.
12. Line 124: It is wrong to use a glass electrode as the gold-standard for such experiment, since the glass electrode itself could be drifting. You need to use the state-of-the-art lab-based spectrophotometric method for the validation. And the validation measurements need to be taken throughout the whole period. A full-ocean glass pH sensor was used as a standard because it is specifically for ocean monitoring applications, and no drift was assumed as it was freshly calibrated. The reason of not using it through the whole experiment is the actual drift; after leaving the glass pH electrode for that long it would very probably drift and not give any realistic value, at the contrary of being freshly calibrated.
13. Figure 4: Use time rather than measurements as the X-axis
14. If the temperature and the pH of the testing solution were stable, the result in Figure 4 shows: (1) there are really big and non-consistent differences between the two solid state sensors; and (2) for an individual sensor the readings were not stable at all. And the conclusion will be this solid-state sensor is really bad and not suitable for ocean acidification research. Both systems show pH readings between 8.30 and 8.35, with a standard deviation of ± 0.045 and ± 0.031 per sensor, and ± 0.015 and ± 0.012 between the 12 pH readings per measurement per sensor. Based on this numbers, the sensor would be suitable for ocean research as the accuracy required is ± 0.05 pH units.

15. The 'Experimental' section should be right behind the 'Introduction' Section. Structure of the paper has been adapted, and the 'Experimental' section is now after the 'Introduction'.
16. Some contents of the 'Demonstration of deployments' should be in the 'Experimental' section, and the rest should be part of 'Result and Discussion' section. Structure of the paper has been adapted. 'Results and Discussion' has been named instead of 'Demonstration of deployments'.
17. Section 3.1 and 4.1: Again, it is wrong to use a glass electrode as the gold-standard for such experiment, since the glass electrode itself could be drifting. You need to use the state-of-the-art lab-based spectrophotometric method for the validation. A freshly calibrated full-ocean glass pH sensor can be trusted for a pH measurement point (however, not for a deployment, as demonstrated in figure 9).
18. Figure 5 is unnecessary. The aim of this figure is to show the sensor's output from the program, and explain the health number of the sensor, which tells the end user when maintenance is needed.
19. Texts in Figure 6 are very blurry. It has been changed.
20. Section 4.1: What is the point to change sensors during the long-term experiment? It'd be better to use all three sensors for the entire experiment period so that both accuracy and the reproducibility of different sensors can be evaluated. The idea of changing the sensors was to compare a third one between the initial two at a different point or lifetime of the experiment. This indicates that sensor 1 maintains its calibration during the course of the experiment compared to the freshly prepared sensors.
21. Figure 7: It would be more useful to plot the sensor values against the pH measured by standard method, do a linear regression and give the statistics (e.g. slope, intersect, r square, and RMSE). The aim of Figure 7 is to show the response of the sensor when injecting CO₂ into the system several times, while the seawater recovers the pH value in between additions. The response of the sensor against a constant variation in pH levels is shown.
22. Section 4.2: It is wrong to use a glass electrode as the gold-standard for such experiment, since the glass electrode itself could be drifting. You need to use the state-of-the-art lab-based spectrophotometric method for the validation. A freshly calibrated full-ocean glass pH sensor can be trusted for a point pH measurement validation.
23. Figure 8: only one validation point is far from enough. Validation measurements need to be taken throughout the entire period. After the validation point was taken, the glass pH sensor was deployed into the seawater, with the idea of validating all the test. However, the glass pH data was unstable (as shown in Figure 9). Access to the field location facilities meant the experiment could not be repeated.

24. Figure 9: This figure doesn't mean anything. First of all, there is no validation and we don't know if the solid state sensor or the glass electrode are correct. Secondly, 'more stable' doesn't mean it is correct. In contrast, it could mean it is not sensitive enough. As mentioned in comment 14, a non-stable pH reading would mean unsuitable pH probe for ocean, which is what the glass pH electrode shows. The pH value calculated with our sensor matched the validation of the calibrated glass pH electrode. Knowing our sensor is measuring the right pH value, and because of the expected response of the sensor against the tidal pH variation originated by the inflow of the fresh SW into the bay.
25. Section 5: The experiments don't support the author's claims at all. Based on the data presented in this manuscript, it seems that this sensor is not suitable for ocean acidification research. We believe the sensor has proved to achieve the accuracy presented (0.05 in pH) and to be sensitive enough to follow the tidal pH variations, as well as a good reproducibility between different sensors. The accuracy of the sensor has been successfully validated using a freshly calibrated full-ocean pH probe.