

Ocean Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/os-2021-126-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on os-2021-126

Phil Bresnahan (Referee)

Referee comment on "Technical Note: Calibration-Free pH Sensing of Ocean and Estuarine Waters" by Monica Miranda Mugica et al., Ocean Sci. Discuss.,
<https://doi.org/10.5194/os-2021-126-RC1>, 2022

Review of Technical Note: Calibration-Free pH Sensing of Ocean and Estuarine Waters

High-Level Summary:

The authors present a novel technology for the measurement of estuarine and ocean pH and report on its performance in laboratory and field settings. The sensor is stated to have the advantage of being calibration free and capable of drying out without causing problems. Furthermore, the sensor's performance can be renewed through a simple servicing routine with sandpaper and sensor heads can be replaced without difficulty. Such a sensor could be a valuable addition to the existing lineup of potentiometric and colorimetric technologies.

Reviewer Comments:

In summary, I believe that the authors make claims about their sensor's accuracy that need much more careful quantification/justification. There needs to be more detail about how the voltammetric measurement is converted to a pH value and how those pH values are to be trusted. Accuracy needs to be not just mentioned as something to be taken for granted but clearly defined, calculated, and written out in this article. How reference values are collected also needs a more thorough description. Figures, especially those that show intermittent performance of the instruments, need much more explanation. The article as a whole could use a different structure that allowed deeper explanation of methods used followed by a careful explanation of results, rather than a sprinkling of methods and results which is repeated for multiple scenarios. Below are more specific suggestions for these high-level comments followed by line-specific editorial remarks.

As one of the key points of the article is to describe the sensor's accuracy in the absence of manual calibrations, I believe much more detail needs to be added to justify the performance claims. Both additional text and analysis (in the form of tables of statistics and alterations to figures) are necessary in order to present a more compelling case for this technology.

The authors describe accuracy in an only semi-quantitative fashion, but it is mostly left to the reader to decide whether the solid-state sensor agrees with the (potentially faulty—see below) validation data based on visual interpretation of a number of figures. What figures of merit are used to quantify sensor accuracy? The authors state ± 0.05 early in the "technical details" section, but do not rigorously quantify this value elsewhere in the article nor cite a prior reference where such work was conducted. For example, I see no mention of RMSE or bias or anything that might represent a comparison of a trusted reference value to the sensor values.

Furthermore, the chosen reference value may be problematic in its own right. The reference value is glass electrode pH (collected/reported only sporadically) and the authors describe (in the Introduction and elsewhere) and depict (in especially Fig. 9) numerous shortcomings of that technology, making it harder to justify their accuracy claims. I am in agreement with the authors that glass electrode pH can present challenges, and it is not clear how they have overcome these challenges in order to use glass electrode data in the demonstration of the accuracy of their sensor.

pH scale (total? free? seawater?) is never mentioned, which leaves me with some concern about the comparison among sensors. How is the glass electrode—the sole provider of reference/validation data—calibrated? There are a number of other ways of validating ocean pH data (e.g., spectrophotometric, estimation via total alkalinity and dissolved inorganic carbon, cited in, for example, Dickson 2007 Guide to Best Practices for Ocean CO₂ Measurements) that would leave the reader with more confidence in the sensor's performance.

How validation measurements are made with the glass electrode could be better explained. Is there a chance for the water temperature to change from in situ conditions? Presumably so, if samples are collected in bottles (e.g., Line 225). How is pH adjusted back to in situ conditions for comparison with the sensors?

Figures deserve much more thorough explanation. For instance, Fig. 7—a key figure as it is the longest time-series and only experiment with multiple days, multiple sensors, and multiple reference points—appears to show sensors starting and stopping sporadically but this is not addressed in the text, which suggests problems with operation that are not described here. Furthermore, there are very clear disagreements among sensors in multiple figures, but it is hard to estimate the degree of disagreement using only visual aids and, therefore, additional metrics like RMSE should be used.

Several figures use overextended y-axis limits (e.g., Figs 3, 4, 8) which obscures important features (e.g., trends, offsets vs. reference samples, offsets between sensors, etc.). Axis limits are indeed somewhat subjective but in many cases here should be pulled in.

Another major issue that needs to be addressed is that there needs to be substantially more description and justification of the method and data analysis. There is only a short and high-level treatment of the principle of operation and depiction of the voltammetric response, but no illustration/statistics of peak potential vs. pH over a range of physical characteristics (e.g., ionic strength, temperature, pressure). Relatedly, I assume peak potential is a/the key metric, but how is this quantified? There is no clear/flat baseline for the peaks. Please expand on your methodology. L169–170 says “The pH output is calculated using an inbuilt algorithm which combines a knowledge of the peak potential of both the pH sensing and reference tracking elements and the temperature to produce the pH.” I understand if the authors wish to keep certain trade secrets; however, this is too vague a description for a peer-reviewed article.

A frequently mentioned claim is that the technology does not require calibration. Where is the equation that results some physical measurement to pH and does not have any calibration coefficients/offsets? An equation relating peak voltage to pH (or at least the form of such an equation) would help the reader understand how the technique works.

I suggest you choose a single name for the technology and use it throughout. You occasionally use terms like calibration-free and solid-state and voltammetric. All of these terms may be appropriate, but if you choose one single name that you use consistently and also add other details as necessary, your presentation will be clearer.

As a final high-level comment, the organization strikes me as unusual. Why are there no discrete “methods” and “results” sections? You include a large amount of experimental data in the “technology” section; I suggest moving much of this down to the latter section and much of the information regarding how experiments were conducted up to an earlier section. Otherwise it is difficult to know how data were gathered. If Ocean Science accepts less common section headers, I have no problem with that approach, but nonetheless think that more detail around methods is essential.

Additional Details:

Abstract: I recommend including performance metrics here. Line 10 says sensor performance is demonstrated in lab and field—what is level of performance in terms of accuracy, precision, stability, etc.?

Fig. 4 and description appear to show the performance that you describe in the Methods,

but also do so for 12-sample averages. Averaging is acceptable, but should be described when defining your version of accuracy.

L9: salinity units might be better chosen. PSU? g/kg?

L10: "additional" salinity? Clarify if you mean that no salinity measurement is needed at all. E.g., "without the need for auxiliary salinity measurement"

L13: I'd replace "efficiency" with "ability" or similar. Similar later in the article.

L20–29: I recommend removing this paragraph. It does not describe pH. In a journal like Ocean Science, it seems more appropriate to describe things like ocean acidification, the effects of eutrophication on pH, and other more common oceanographic issues. Including mention of pH of effluent in a list of phenomena related to pH might be sufficient, but this paragraph is superfluous.

L37: given your mention of the decrease in concentration of CO₃ in L40, I recommend not mentioning "release of carbonate ions" here as it is misleading/imprecise.

L39: change to something like "average global ocean surface pH is expected to..." and add a citation.

L41–44: if you are describing changes in buffering capacity, a citation will be necessary, but I do not believe this sentence is critical here.

L45–49: It is fine to say that it is worthwhile to measure pH near vents, but this is not a type of pollution. Please edit.

L50: I recommend citing a review article on pH sensors. E.g., Byrne's 2009 OceanObs paper, Martz's 2015 Technology for OA research. Could even mention Sorensen's original 1909 paper which describes potentiometry and colorimetry, and note that these two methods have dominated for the past 100 years.

L57: there are many ISFET pH references, including Martz 2010, Johnson papers on Deep Sea Durafet, Shitashima 2002. Please include others here.

L60: appropriate to cite original spec pH papers, including Clayton and Byrne as well as Seidel 2008. DeGrandpre 1999 and 2014 provide good overviews of the theory and application as well.

L68: is it perhaps more accurate to say that it does not require any recalibration? It is hard to understand what the authors mean by referring to this sensor as calibration-free while neglecting to include any theory or empirical equations. See high-level "Reviewer Comments" above for more detail.

L74: it seems critical to justify the accuracy claim that is made here with thoroughly analyzed and visualized data in the Results section, and I do not believe that this has been achieved. It would be appropriate to describe in Methods the accuracy of a sensor that was previously used/published, but not that of a novel sensor.

L77: "comms" should be "communications"

L85-86: this reads as though the sensor was previously published by Batchelor-McAuley and Lu and Compton. I imagine that is not the case, though, correct? If so, I recommend rewording such that you are stating that your sensor is based on a previously published principle of operation.

L87: "combination of both" is redundant

L88: is the reference electrode the pH inactive molecule? This seems logical but isn't made explicit. It should be. It becomes clearer by L90, but I think it should be included in the first description of the active vs. inactive molecules.

L90: you frequently describe this as a calibration free sensor, but then mention recalibration. I appreciate the distinction between user calibration and autonomous, but suggest you clarify here.

L95: there ought to be a figure of pH vs. potential and relevant statistics describing the goodness of fit or predictability of pH from potential (and perhaps other metrics).

Fig. 3: I appreciate the value of this figure but it deserves more explanation. What is thought to be the cause of the shift in reference voltage?

Additionally, it could be argued that there is a meaningful trend in pH with respect to time here. What is the slope of Fig. 3b? Is it statistically different from 0?

L124: "perfect agreement" needs quantification and probably isn't an appropriate word choice. What are the values retrieved with the glass electrode? How/when was it calibrated? You've previously suggested that glass electrodes may not be trustworthy, so more justification/description is merited here.

Fig. 4: this is the first time I've noticed that you are describing sensor performance after averaging multiple samples. This is fine, but should have been mentioned previously if it is commonly done. If it is not common to average samples like this, then what is the difference in accuracy between averaged vs. individual samples?

Furthermore, how fast is the sensor sampling at this time? Is it 12 samples @ 15 second measurement frequency? Describing time would be helpful here.

Glass electrode pH shown in Fig. 4 does not merit the aforementioned characterization of "perfect agreement."

L137: does this mean that the electrodes are cycled in their use to extend overall lifetime of the transducer head? If so, please state this.

L144: please explain "assesses the voltammetric profile of each electrode." This sounds rather subjective. More detail would help.

L169-170: this should be mentioned before any pH data are shown. More detail is essential. It is fine if you cannot give away a proprietary algorithm, but can you describe anything else? How do you know the form of the equation? Is it purely theoretical? Empirical?

L171: add "Figure"

L192: how was glass electrode calibrated? How frequently? Immediately before use each time?

Section 3.1: this material seems quite relevant to the figures shown previously under

Section 2. Is it not? I think it should be moved up much higher in the article.

L196: interesting but seemingly irrelevant

Fig. 6: increase resolution. Cannot be interpreted at this blurriness.

L207: add "Figure"

L207: how was CO₂ added?

Section 4.1. Some of this text might fit better in the figure caption (e.g., "open circles represent..")

Fig. 7: it is difficult to verify accuracy claims based on this figure. The open circles which represent validation samples are very large. What is the RMSE of sensor vs. validation sample?

It would be beneficial to show some depiction of sensor vs. reference anomaly here so that we can more readily assess the offsets. An additional minor detail: "measured pH" is not a good label for your reference pH as all dots represent some form of measured pH.

L223: add "Figure"

L227: "semi-diurnal" rather than "bi-daily"

L229: I am not familiar with the "tidal coefficient" concept. Why not use tidal range in physical values (e.g., meters, dbar, etc.)?

Fig. 8: there is no need to extend the y-axis much beyond 7.6–7.9. Perhaps 7.5–8.0 could be justified for its use of round(er) numbers. Extending the axis as far as shown here seems to hide the signal and potentially obscure any offset with the reference sample.

L241–244: I do not believe these claims are justified without some other way of validating pH data. From everything else presented so far, I find it just as easy to believe that the glass electrode is correct and not the solid-state electrode as vice versa. Even if there is indeed a problem with the glass electrode, this does not mean that the solid-state electrode is accurate. If anything, this figure only suggests that using the glass electrode for validation is unsatisfactory.

This is also another instance of insufficient introduction of the methods. All other uses of the glass electrode are discrete measurements, but now we have continuous glass electrode pH. Is it the same sensor? Its own datalogger? How were the instruments deployed?

L255: this is the second use of “intra” (within) that I believe should be “inter” (between). I believe you are describing the reproducibility between sensors, correct? See also L211.