CC1: 'Comment on egusphere-2024-2853'

Christopher Sabine - 25 Nov 2024

General Comments:

This is a well written and well documented manuscript, but I am concerned about the accuracy of the measurements and the robustness of the conclusions. Based on the final corrected pH plots in figure 3, it looks like the pH changes along the two transects were very minimal. The authors calculate that the largest influence on pH over a 24-hour cycle is temperature, with an occasional hint of a biological signal. This really isn't a surprise. These variations are occurring in both time and space, so it is impossible to quantitatively separate the effects. The authors conclude, "…although the processes responsible for these pH variations are well-understood, high-resolution data highlight the challenge of identifying a dominant factor at the fine-scale due to their complex interplay." This makes me question what is the point of this manuscript? Perhaps the authors can do a better job of explaining what is new and novel about this work.

We appreciate the reviewer's time and thoughtful feedback on our manuscript. While we acknowledge that some of our findings—such as the role of temperature in short-term pH variations—are well-established, our study highlights the challenge of disentangling the multiple drivers of pH variability at fine spatial and temporal scales. The complexity of interactions between temperature, salinity, and biological activity makes it difficult to isolate a single dominant factor, even with high-resolution observations.

Rather than resolving these dynamics, the novelty of our work lies in demonstrating how high-frequency measurements capture fine-scale variability and expose the limitations of traditional approaches in attributing pH changes to individual drivers. Although our results suggest that while large-scale, lower-resolution datasets are likely sufficient for capturing ocean-basin scale CO_2 variability, fine-scale observations remain essential for identifying regional differences, constraining equilibration timescales, and refining predictive models of ocean acidification and air-sea CO_2 exchange. To clarify this, we have refined the manuscript's focus to emphasize both the challenges revealed by high-resolution data and the unique capabilities of our experimental setup.

Specific Comments:

Lines 256-257: How do the authors define unreasonable drift patterns? This approach sounds very subjective.

To clarify, the flagged data points corresponded to periods of high seas, during which the underway system intermittently received air, as reported in the cruise logs. This resulted in abrupt and erratic pH fluctuations inconsistent with the expected behavior of both the optode and surface ocean pH dynamics. Thus, the identification of unreliable data was not solely based on subjective judgment but was supported by independent observations of system disturbances during the cruise. To reflect this clarification, we have revised the sentence as follows:

"To ensure the reliability of the dataset, an initial screening was conducted to identify and flag unreliable continuous pH data points, primarily attributable to optode stabilization issues. Specifically, data points recorded during documented periods of particularly rough weather, when air intrusion into the underway system was reported, were flagged as unreliable due to the resulting abrupt and erratic drift patterns inconsistent with expected optode and surface ocean pH behavior." [L256 – 259]

Lines 259-263: The authors are correcting the measured pH values to pH calculated from TA and DIC, but how do they know the calculated values are correct? Which constants were used for the calculations? Is there no direct measurement of pH to validate the corrections? What about the two-point calibrations described in the methods? Were these calibrations not helpful? How frequently were they done?

While there were no direct spectrophotometric pH measurements available for validation, we collected duplicate TA and DIC samples twice daily (approximately every 12 hours) when conditions allowed. These discrete samples provided a means to assess the drift of the optode over time. Regarding the two-point calibration and the additional third-point correction using a Certified Reference Material (CRM), this approach was indeed beneficial in initially calibrating the system. However, because the system was operating continuously, recalibration during the measurement period was not feasible. This limitation necessitated the use of discrete underway samples to evaluate and correct for potential drift in the continuous pH measurements. To clarify these points, we have revised the manuscript as follows:

"$pH_{TA, DIC}$ was calculated from TA and DIC discrete samples using the carbonate system equilibrium constants from Lueker et al. (2000). These calculated values were then aligned with the continuous pH dataset to determine the offset between $pH_{TA, DIC}$ subsamples and the continuous optode-based pH measurements (Fig. 3). A Piecewise Cubic Hermite Interpolating Polynomial (PCHIP; Fritsch & Carlson, 1980) was fitted to the offset to provide a continuous correction across the pH timeseries. While continuous recalibration of the optode was not possible, a two-point calibration, supplemented by a third-point correction using a CRM, was performed prior to deployment (see Section 2.3). Additionally, discrete underway TA and DIC samples, collected twice daily when possible, were used to assess and correct for drift in the optode measurements." [L262-272]

Figure 3: Some of the pH changes are very large and the final pH curves look nothing like the raw data. If all the calculated pH values are used to adjust the underway pH measurements, then what confidence do we have that the values in between the calibrations are correct? Are there any independent pH measurements that were not used for calibration that we can use to assess accuracy?

While direct spectrophotometric pH validation was not available, we collected duplicate TA and DIC samples twice daily to assess optode drift.

According to the manufacturer (PyroScience PK8T specification sheet), the pH sensor has a specified drift of <0.005 pH units per day at 25°C. During our deployments, in situ seawater temperatures ranged from 13.4 to 22.0°C in the North Atlantic (SO279) and from 19.8 to 27.7°C in the South Pacific (SO289). While parts of the North Atlantic deployment occurred at temperatures below 25°C — potentially reducing drift — much of the South Pacific cruise took place at or near the reference temperature, where full drift rates are expected.

Even under the conservative assumption that drift occurred at the maximum specified rate, cumulative drift over the 35-day North Atlantic deployment could reach 0.175 pH units, and 0.28 pH units over the 56-day South Pacific deployment. These estimates are consistent with the magnitude of deviation observed in the raw underway data (Figure 3, pink line), especially during the later stages of each cruise.

Moreover, even if an intermediate recalibration had been feasible mid-deployment, drift before and after that point would still lead to offset and inconsistency across the time series. Recalibration alone, especially using standard buffers, does not fully address environmental effects such as pressure, light path shifts, or biofouling — nor does it guarantee agreement with carbonate system equilibrium.

To ensure alignment with in situ carbonate chemistry, we instead corrected the underway pH data by referencing $\mathrm{pH}_{\mathrm{total(TA/DIC)}}$ values from discrete subsamples (black circles in Figure 3), derived from total alkalinity and DIC measurements. This allowed us to apply a bootstrapped drift correction (blue line and uncertainty shading) that is traceable to chemical standards and consistent with best practices for autonomous ocean pH measurements. The corrected dataset thus provides an accurate and internally consistent record suitable for biogeochemical interpretation.

To further enhance confidence in our corrected underway pH dataset, we have now included independent validation using high-resolution surface measurements of both pH and pCO_2:

- Independent pH validation: A factory-calibrated spectrophotometric pH sensor (Sunburst Sensors SAMI), sampling every 15 minutes, was used as an independent check of the optode-corrected pH values. The comparison yielded an RMSD of 0.0329 pH units (Figure S1), reflecting modest scatter consistent with inherent sensor measurement noise. Importantly, the mean error of the SAMI pH measurements was approximately 0.02 pH units, while the mean error for our optode-corrected pH was smaller, at approximately 0.008 pH units. This indicates that our corrected optode data provide improved accuracy relative to independent spectrophotometric measurements, clearly demonstrating the effectiveness of our drift correction procedure.

- Independent fCO_2 validation: Continuous pCO_2 measurements from a commercial HydroC sensor (4H-JENA Engineering GmbH) were adjusted to in situ conditions and compared to fCO_2 calculated from our corrected optode pH and discrete alkalinity measurements. While the resulting RMSD was relatively large (30.76 µatm, Figure S2), this variability largely reflects uncertainties inherent in alkalinity estimations and carbonate system calculations. To contextualize, the mean error of the directly measured fCO_2 was approximately 5 µatm relative to discrete carbonate system samples, whereas our calculated fCO_2 had a mean error of approximately 12.7 µatm. This level of accuracy remains acceptable for autonomous carbonate chemistry measurements and further confirms that our optode-corrected pH provides a reliable improvement over raw sensor measurements, suitable for robust biogeochemical analyses.These additional validations using independently measured carbonate system parameters underscore the accuracy and consistency of our corrected pH dataset, strengthening the dataset's suitability for robust biogeochemical interpretation. We have clarified these points in the revised manuscript and included these detailed comparisons (Figures S1 and S2) as supplementary materials.

We have clarified this in the revised manuscript:

"The manufacturer specifies a drift of <0.005 pH units per day at 25 °C, though drift may vary slightly with temperature. During our deployments, seawater temperatures ranged from 13.4°C to 22.0 °C in the North Atlantic and from 19.8°C to 27.7 °C in the South Pacific. While the cooler North Atlantic conditions may have reduced drift rates slightly, temperatures during the South Pacific cruise were close to or above 25 °C for extended periods, and full manufacturer-specified drift is likely to have occurred. Even if drift remained at the nominal rate of 0.005 pH units per day, this would amount to a cumulative offset of up to 0.175 pH units over ~35 days (SO279) and 0.28 pH units over ~56 days (SO289), in line with the deviation observed in our raw data (Fig. 3). Notably, even with a recalibration during the cruise, drift before and after that point would still introduce offsets. Thus, over timescales of days and longer, the accuracy of the measurement is dependent on the correction to the TA and DIC samples.

To further assess the robustness of our drift correction, we compared our optode-corrected pH and derived fCO_2 values with independent continuous measurements from autonomous sensors (Figures S1 and S2). The corrected optode pH exhibited a lower mean deviation (~0.008) from discrete pH (TA/DIC) measurements than the independent spectrophotometric SAMI sensor (~0.02 pH units), indicating that our correction effectively minimized drift-related inaccuracies. The overall scatter between the optode-corrected and SAMI-measured pH was modest (RMSD = 0.0329), suggesting reasonable consistency despite inherent sensor measurement noise. Similarly, calculated fCO_2 from corrected optode pH showed a larger mean deviation (~12.7 µatm) compared to directly measured fCO_2 (~5 µatm). The relatively higher RMSD (30.76 µatm) reflects variability in calculated fCO_2 arising from uncertainties in alkalinity estimates and carbonate system calculations rather than fundamental flaws in the optode pH correction itself. Thus, these validations collectively confirm that our corrected pH data represent a reliable improvement over raw optode measurements, suitable for robust biogeochemical analysis." [L317-357]

Line 265-279: It seems a bit circular to use TA to adjust the pH values, then turn around and use TA together with the pH to calculate the other parameters. How do the Lee et al. TA values compare to the measured values? Are these uncertainties smaller than if the authors simply took all the measured TA and DIC values to calculate the other carbon parameters?

The empirical TA estimates from Lee et al. (2006) were used for continuous carbonate system calculations because they provide a complete, high-resolution dataset, whereas measured TA was available only from discrete samples. As shown in Fig. S1 in the supplementary information, the Lee et al. (2006) equations align well with measured TA values, with deviations generally smaller than the uncertainty of carbonate system calculations. Using all measured TA and DIC values instead would result in a significantly lower temporal resolution. We have clarified this in the revised manuscript:

"While direct TA measurements were collected twice daily, their limited temporal resolution made them unsuitable for continuous carbonate system calculations. The empirical TA equations from Lee et al. (2006) provided a high-resolution dataset that allowed for more comprehensive system reconstructions. A comparison of measured and estimated TA values (Fig. S1 in the Supplementary Information) shows good agreement, with deviations generally within the uncertainty of carbonate system calculations." [L293-298]

Line 382: This is not a traditional use of the term water mass. One does not normally think of water masses as surface features because external forcing (warming/cooling, precipitation/evaporation, etc.) can make water properties quite variable, compared to traditional subsurface water masses that have stable properties that can be defined and tracked as they move into the ocean interior. I understand what the authors are trying to say, but I think a different term for waters with different properties needs to be used.

We acknowledge that surface waters are more dynamically influenced by external forcing and do not exhibit the stable properties characteristic of subsurface water masses. To improve accuracy, we have replaced "water mass" with "hydrographic variability" and "waters with distinct temperature-salinity properties" to better reflect the transient nature of surface conditions. This revision ensures clarity while maintaining the intended distinction between recent and longer-term influences on pH.

Line 392: This section is focused on the influence of T and S on pH, which seem to have signals that are less than 0.01. I am wondering how robust these signals are if the raw pH values had to be corrected by ~0.4 units (fig 3) and the uncertainties in the final values are around +/-0.01 (fig 4 and 5). How do the authors know these are signals and not just noise that they are interpreting?

While the raw pH values underwent a significant correction (~0.4 units), the consistency of observed pH trends across multiple cycles and their alignment with expected pH changes from temperature and salinity variations suggest that these signals are not random noise. Noise would present as random scatter, whereas the observed pH trends align with expected physical and chemical processes. To further assess robustness, we note that the observed diurnal variations in pH align with well-established temperature-driven changes in carbonate chemistry. Additionally, the uncertainties in the final pH values (~±0.01) are similar to or smaller than the observed trends in many cases, reinforcing the validity of the signals. If these variations were purely noise, we would expect random scatter rather than structured diurnal patterns that correspond with temperature fluctuations.

To clarify this, we have revised the manuscript to explicitly discuss the uncertainty and its implications for interpreting fine-scale pH variability. We have also included an additional analysis in the Supplementary Information to better illustrate the signal-to-noise ratio in our dataset.

Line 414: I do not understand this statement about the waters having lower thermal inertia. What do the authors mean? Lower than what? Water has a low thermal inertia compared to the air, but the sentence is trying to explain why there is a diurnal temperature signal in the surface water. The deeper waters do not have a higher heat capacity, they are just removed from the forcing.

Our intention was to highlight that surface waters experience greater temperature fluctuations due to their direct exposure to atmospheric forcing, while deeper waters remain more thermally stable due to their isolation from these rapid changes. We have revised the sentence to remove the reference to "thermal inertia" and clarify that deeper waters are insulated from direct atmospheric influence rather than having a higher heat capacity. However, as per the other reviewer's suggestion, we have removed this discussion of rapid thermodynamic responses, as the timescales we analyze (hours to days) do not align with split-second carbonate system dynamics. Instead, we have refined the text to focus on the more relevant diurnal temperature variations and their impact on pH.

Line 635-637: This sentence seems to convey the essence of this manuscript: However, although the processes responsible for these pH variations are well-understood, high-resolution data highlight the challenge of identifying a dominant factor at the fine-scale due to their complex interplay. What is the take home message that you are trying to convey? It sounds like there isn't much point in doing these high-resolution measurements.

While the fundamental drivers of pH variability are well-established, our high-resolution measurements reveal the difficulty of attributing pH changes to a single dominant factor at fine spatial and temporal scales. Rather than diminishing the value of high-resolution observations, this finding underscores their importance in exposing the complexity of surface ocean carbonate chemistry—something that lower-resolution datasets may overlook.

Our results suggest that broader ocean-basin scale analyses based on lower-resolution data are indeed likely sufficient for global CO_2 cycle assessments, at least in the regions where our measurements were taken. However, this was not known for certain before carrying out these measurements. Also, even though fine-scale variability was relatively small, we could still detect diel patterns in pH which were consistent with hydrographic (temperature, salinity) and biological (chlorophyll) variability. In other parts of the ocean, and at other times, the fine scale variations in pH may be greater, and ignoring them may indeed have an impact on basin-scale analyses. Given that we were able to detect and attribute these very subtle patterns, a similar measurement system should be more than capable of doing them same where the fine-scale variability is greater. We have clarified this point in the revised manuscript to ensure that the significance of our findings is properly conveyed.