General comments to the authors

This manuscript introduces a Floating Membrane Equilibrator (FME) designed to resolve greenhouse gas concentration gradients in calm, standing waters at fine vertical resolution. The description of the instrument and its construction is clear and will enable reproducibility, thereby supporting broader ecological measurements and improving flux estimates. The pool experiments convincingly reveal distinct vertical CO2 gradients within a 20 cm water column and demonstrate the rapid response of the FME when the water column is homogenized.

My principal concern relates to the comparison between the FME and conventional CO2 probes under field conditions. In the current setup, the instruments were not deployed at the same depth. Given the strong vertical gradients shown in the pool experiment, this difference in measurement depth makes direct performance comparisons problematic and may lead readers to overinterpret the relative performance of the FME versus conventional sensors. The observed discrepancies are likely driven by sampling different layers with distinct dynamics rather than by sensor response characteristics alone.

To address this and avoid potential misunderstanding, I suggest the following:

Option A (preferred): Conduct an additional field experiment in which both devices are co-located at the same depth. Even a shorter deployment would provide a robust basis for the claimed device-to-device comparison. The existing field results could then be reframed to illustrate how deeper measurements miss surface-layer dynamics and may yield biased flux estimates.

Option B: Moderate the comparative claims throughout the manuscript to explicitly state that the devices were not co-located and that observed differences likely arise from sampling different layers with distinct diffusivities and adaptation times, rather than from sensor performance per se.

I strongly encourage Option A, as Option B would weaken the manuscript's comparative narrative and the case for the FME's advantages. Overall, I find the FME to be a valuable contribution that can facilitate measurements of dissolved greenhouse gases and improve global flux estimates. With a clarified or strengthened comparison, the paper will be significantly more compelling.

General Response:

We thank the reviewer for reading, providing helpful comments and mentioning concerns, which will help us improve our manuscript. We are pleased to see that the reviewer acknowledged the benefits of our floating membrane equilibrator to the research field of standing water CO_2 measurements.

The principal concern of the reviewer relates to FME deployments under field conditions and the comparability of our FME to conventional sensors under such deployments. We agree that, in parts, the manuscript does not state the benefit of the FME over conventional probes clearly enough. The difference between the FME and conventional probes is that conventional probes cannot be installed that close to the water surface. We know from previous studies about increased CO_2 concentrations at the water surface, but due to the size and measuring principle of conventional probes, we do not have CO_2 measurements as close to the water surface as with the FME to compare to.

We agree that an additional experiment will help to improve the manuscript. Therefore, we will perform a new pool experiment with the FME and an AMT probe installed at the same depth. We will measure in water at equilibrium with the air to avoid drifts or rapid changes. We will further take gas samples to be analyzed in a gas chromatograph. The data will be presented in a revised version of the manuscript.

Details

Abstract

Line 14: Please clarify that the reported bias primarily applies to calm, standing waters—the conditions under which the FME offers clear advantages.

Response: We added “of standing waters” and removed “, potentially introducing bias” from that part to avoid confusion. The next sentence clarifies that calm conditions cause vertical gradients and potential measurement biases.

Line 22: See general comment for direct device comparison.

Response: We will add data for a direct comparison of measurement accuracy of the two devices.

Line 24. To streamline the narrative, I suggest to relocate the two sentences reporting the pool-experiment results to the section where that experiment is described.

Response: We agree that this was confusing and restructured that part according to the reviewers comment.

Introduction

Line 31: Clarify ‘lakes and reservoirs are sources’—do you mean sources to the atmosphere?

Response: We added “to the atmosphere” to the sentence.

Line 40: Consider merging the two sentences for concision and flow.

Response: We merged and modified the two sentences for better flow.

Line 82: Ensure consistent reporting of vertical resolution—use a single unit (either cm or m) throughout the manuscript.

Response: We agree with the reviewer that varying units can be confusing. However, we think in some parts of the manuscript the reading flow is better when using decimal-free measurements with “cm”, while sometimes, on larger range measurements, using “m” gives better perspective. When revising the manuscript, we will review the units and make sure the units used are appropriate.

Materials and methods

Line 88: Readers could benefit from a list of materials and an up-to-date cost estimate to assess the effort and expense of replicating the device.

Response: This is a good suggestion. We originally decided not to include such a list because of varying prices in different countries, but will include the parts-list in the revised manuscript.

Line 90: The ‘also’ is not needed here.

Response: We deleted it.

Figure 1: If possible, increase number font size in the image.

Response: We will rework the figure and increase the font size.

Line 96: The ‘of’ is not needed here.
Response: We deleted it.

Line 98: Can you provide any diffusivity values through the silicon tube, in order to gauge, if that might influence system response times?
Response: We cannot provide diffusivity values. Further, this might be a product-specific value. While designing and testing the FME, we used silicone tubes from different suppliers and have not noticed differences. However, we always used 4 mm inner diameter and 6 mm outer diameter tubes. We believe that the thickness of the silicone tube is the driving parameter for good performance.

Line 100: Please indicate whether the two analyzers were the same model. To avoid focusing too much on instrument-related offsets, I suggest rephrasing with stronger emphasis towards the aim of using a single analyzer: e.g.: ‘Because using two analyzers can introduce systematic differences—even between identical models—we used a single analyzer and alternated measurements between the two FMEs via a switching manifold.’
Response: Thank you for that comment. We will change the text according to your suggestion.

Line 109: Please clarify the gas-flow design: does your system use closed-loop circulation with a pump (as in Hari 2008; Heiskanen 2014; Provenzale 2018), or an open-path setup with continuous inflow? If the latter, specify which gas was used and how potential dynamics in inflow gas concentration was handled.
Response: Thank you for the comment. We used a closed loop design, powered by the pump of the connected CO_2-analyzer. We will add this information to the revised manuscript.

Line 109: Please clarify whether the concentration measured by the analyzer is (a) the equilibrated headspace gas concentration or (b) the dissolved concentration in water. If it is (a), add how you convert to (b)
Response: The analyzer measures the equilibrated gas phase concentration. We will add additional information in the revised manuscript.

Line 110: Specify if the zero calibration performed with N2.
Response: The zero calibration is performed by eliminating CO_2 using soda-lime in a specific calibration loop of the analyzer. When doing the calibration, the gas-loop changes from in/out to an internal loop that includes the soda lime.

Line 117: Please state whether temperature was stable. If not, describe how you corrected concentrations for temperature variations.
Response: The temperature was stable during the experiments.

Line 138: I assume this is an NDIR sensor, with a gaseous headspace equilibrating with the water phase through a semipermeable membrane; please briefly explain the probe’s working and equilibration principles.
Response: Yes, the probe is based on an NDIR sensor. In brief, the sensor chamber contains a CO_2-sensitive optical element that is separated from the surrounding water by a gas-permeable silicone membrane. Gaseous CO_2 diffuses through the membrane and equilibrates with the water phase. The

equilibrated CO_2 concentration is then detected by the internal NDIR sensor based on the Single-Beam Dual Wavelength principle.

Line 139: Even though the CO2 probe's vertical resolution is constrained by the membrane diameter, co-locating measurements at identical depths is essential for a direct, device-to-device comparison (presented throughout the manuscript). According to my general comments, please align depth levels (or justify deviations and moderate the comparative claims throughout the manuscript).

Response: As mentioned in the general response, we will add a figure showing a comparison of FME data and conventional probe data at the same depth during the field deployment.

Line 142: Rather than pressure-induced errors, I expect the dominant issue to be likely mixing of gas volumes equilibrated at different concentrations, especially in the two-depth setup.

Response: We agree that this is an error source too. However, the fluctuation also appears when closing the valve and opening it again, using the same FME. In the revised manuscript, we will add the suggested error source by the reviewer, because we agree its valid and should be discussed.

Results

Figure 3A: The FME represented by black dots seems to be in equilibrium at the start of the shown measurement interval. Was the FME represented by red dots deployed later, or does this reflect loop-volume mixing/equilibration (similar to 3B?)? You might consider clarifying this in the text.

Response: What can be seen is the gas volume mixing to the new depth. We will describe how "old gas" in the tubes influences measurements. The graph also shows nicely how this artifact vanishes quickly.

Line 159: Please specify which "two tests" you mean? Was the experiment replicated or other variables changed?

Response: We mean the repetition of "same depth" experiments, shown in Fig. 4 A and B.

161: The 'see' is not needed here.

Response: We removed it.

Figure 4: Correct the 'blue dots' in the figure caption.

Response: We changed it to "black".

Figure 4A: If the rise in the black trace reflects loop-volume mixing until the gas is fully equilibrated, state this explicitly in the text and indicate the timescale.

Response: We will add this to the revised manuscript.

Line 163: Please clarify how the sensor response time was computed. In particular, explain how the final (100%) value was defined or assumed.

Response: We will add this information to the final manuscript. In short: we assumed the 100 % to be the steady state of each experiment. We then calculated the time to reach 90 % of the difference between the starting and ending concentration.

Line 164: The referenced table appears to be in the Appendix; please update the citation accordingly.
Response: Thank you for highlighting that. We will change the reference.

Figure 5: The sharp decline likely reflects increased turbulence and mixing that raised gas-transfer velocity leading to overall lower concentrations in the water phase. This might be an interesting fact for readers.
Response: We will add this to the discussion.

Figure 5: Please explain the initial rise in concentration, as in Figures 3A and 4A. Clarify whether it reflects loop-volume mixing/equilibration, delayed deployment/start-up, switching between FMEs, or a step change in ambient/headspace composition. All subsequent observations reflect the fact that the sensors were not co-located in depth.
Response: Thank you for the comment. We will add this to the discussion.

Line 180: The direct comparison of response times is likely confounded by differing depth-specific concentrations. If the FME (near-surface) and the probe (25 cm) are equilibrating to different target levels, the observed "response time" may reflect different driving gradients rather than instrument dynamics.
Response: We agree. The objective of that experiment was to show that measurements at the water surface (using the FME) reveal dynamics that are missed when measuring with conventional methods (slower probe at lower depth). We will rephrase this part in the manuscript to make it more clear.

Line 184: The scientific contribution would greatly increase with an additional experiment demonstrating that the FME measures the same as the probe, but faster, thereby enabling small-scale diurnal processes to be uncovered at finer spatial scales.
Response: We agree that an additional experiment will help this. As mentioned in our general response, we agree on performing another experiment (described above).

Line 186: Briefly mention the process behind this increase: that it is likely due to higher turbulence at the water surface and thus a higher CO_2 uptake from the water phase.
Response: We will add details about this. It is likely the combination of higher turbulence and the CO_2 carried to the water by the water droplets.

187: Line 187: Even if the FSME might react faster to concentration variations, it is likely that the probe's concentration was delayed due to the slower diffusivity of CO_2 in water, rather than sensor response times. A follow-up experiment could provide validation for your claim.
Response: We agree again, but this is not the scope of the study. Our study is meant to introduce a method for a higher resolution of CO_2 measurements at the surface. The take-home message here is that we would miss the event with conventional measurements, while these observations clearly show CO_2 dynamics that are relevant for water-atmosphere flux calculations.

Figure 6: In my version, the columns representing precipitation are not blue; please correct if necessary.
Response: We will check, but in our version, they are blue.

Figure 6: I assume the FME measures the CO_2 gas phase that is in equilibrium with the water phase, whereas the probe (per the Methods) measures dissolved concentration directly; please specify this, and, if the devices measure different quantities, how you convert between them.

Response: The probe and the gas analyzer connected to the FME use the same principle and derive the same type of measurements. Both use an NDIR to measure CO_2 in a gas phase. The AMT probe uses a membrane to equilibrate CO_2 from the water to the probe's gas volume, the FME uses the silicone tubes for the same matter. There are no units converted.

Commented [PA1]: Matthias fragen

Commented [PA2R1]: ppmv in der AMTR sonde

Figure 6: Readability would improve if you mark, with dotted vertical lines in the respective colors, the times at which the probe and FME are assumed to have reached in-situ concentration.

Response: We think that this would add confusion and too much information to the figure.

Figure 6: Please explain why some FME measurements are missing (technical errors?), and why the FME pattern is not reflected by the probe—neither with a temporal nor a magnitude shift; clarify what happened there.

Response: The gaps come from the FME that was installed in the AMT probes depth, which we will add for the revised manuscript. The point of the plot was to compare surface dynamics to underlying water dynamics. Since we were only able to switch to the deeper depth three times, we decided to leave that data and focus on the earlier described differences. We will add this for a later version, since it also introduces the direct comparison between AMT and FME.

Figure 6: There are darker areas (barely visible in my version); please intensify the color and explain in the caption what they represent.

Response: The darker areas highlight nighttime. We will add this information to the caption and will make sure it has better visibility.

Discussion:

Line 211: The manuscript clearly articulates how water-phase turbulence influences equilibration times, and I agree that a key strength of the device is its avoidance of artificial turbulence and preservation of natural heterogeneity. To improve clarity and interpretability, I recommend adding a concise description in the Methods of how T90 is computed, as this would ground the subsequent discussion. In addition, you could elaborate on device-specific determinants of equilibration time: for instance, can diffusivity through the silicone tubing be treated as negligible, implying that equilibration is primarily governed by the concentration gradient? It would also help to state explicitly whether the system operates with a closed or open gas loop, and to discuss how this design choice affects T90. Finally, when addressing device-specific factors, consider including guidance on how equilibration might be optimized—e.g., whether increasing gas flow can reduce T90.

Response: We will add information on the calculation of T90 response times and that we used a closed loop principle. Since we only used one analyzer, we can only hypothesize about how to improve the response time. We agree that higher airflow could improve the performance and will add it to the revised manuscript.

Line 216: Here you would have the opportunity to highlight another key strength of the device—its capacity to capture gradual concentration changes (e.g., diurnal variability). Additionally, to illustrate how rapid changes are resolved, expand your explanation of the event on the afternoon of September 26 in Figure 6, as changes there are much more striking than after the first rain event. Use it as a case study to substantiate the device's performance.
Response: Thank you very much for this suggestion. We will rework this part of the discussion.

Line 219: Here, I suspect the observed difference is less a device-specific effect and more a consequence of slower dynamics in the water column at 25 cm depth. While you convincingly demonstrate enhanced vertical resolution, a complementary measurement with both devices deployed at the same depth would more robustly substantiate this point.
Response: We agree. As mentioned above, we will add data for the AMT and FME at the same depth.

Line 225: Do you mean spatial resolution as the ability to resolve a specific depth (i.e., vertical resolution)? If so, please state this explicitly—otherwise readers may interpret "spatial resolution" as referring to larger horizontal or areal scales.
Response: Yes, we refer to vertical resolution.

Line 229: Again, the probe might have missed this due to its slower response times, but it might also be an effect of slower dynamics in the deeper water. You finally address this issue in line 234, but direct device comparison is therefore difficult.
Response: As mentioned above, we agree and will add data on it.

Line 232: Yes, the rainfall had an effect on the concertation, likely due to higher diffusive CO2 uptake by the water column, but the concentration change is both faster and larger in the afternoon of September 2026. I miss the explanation for that event.
Response: The explanation for the sharp increase of that day is the switch from autotrophic dominated to heterotrophic dominated processes when the sun has set. We will add this to the discussion.

Line 235: Here arguably the vertical spatial resolution, where your device clearly outperforms common methods also plays a role.
Response: We agree and rephrased the sentence.

Line 238: Several studies (e.g., Attermeyer et al.) have documented day–night differences in CO_2 concentrations and fluxes. Please position your approach relative to these methods and clarify the added value of your device.
Response: We will discuss this in a revised manuscript. However, the sentence described was probably phrased poorly. We are not stating that diurnal changes are difficult to measure, there are many references, including our own, that show that already. We were trying to highlight that the conventional method was very slow in this experiment and would have missed a substantial part of the nightly increase. This could be due to neuston community activity or slower increase in the "deeper" water, but this was not the point. We will rephrase this part to avoid misunderstandings.

Line 239: The probe here failed completely to reflect any day-night pattern. Why?

Response: We disagree. In Figure 6 one can see the nightly increase in the CO2 probes data, as well as a decrease during the day, and subsequently an increase in the second night. The change in the probes depth was either low or slow.

Line 244: Here you could move the sentence about the FME towards the end, thereby first address comparison and difficulties and then end nicely with the advantages of the FME.

Response: We agree and will change the order.

Line 254: Making it unpracticable for long-term deployment? Or with a rather high frequency maintenance, if biofilm grew after two days already.

Response: That is correct. We mentioned this in the next sentence (Line 255) in the original manuscript. However, we would like to highlight that we did not see an influence on the measurements during that time, but wanted to warn about possible influences due to fouling.

Line 255: Does this imply the device is only truly usable under calm conditions? If wave-induced motion causes vertical displacement, wouldn't that compromise the FME's key strength of high vertical resolution?

Response: We understand the reviewers concern. The effect described happened in conditions of strong currents on the side of a boat (as can be seen in the video supplement). The FME does work as intended in calm to low winds, which are the situations with unresolved CO_2 gradients.

Line 263: Since you invoke "optimization," either provide concrete guidance on how to optimize or remove the term here.

Response: We removed "With further optimization" here.

Appendix Table A1: Please explain the origin of the differing initial mixing ratios, especially the elevated laboratory values, and define your equilibrium criterion.

Response: We will add this information to the legend. The differing mixing ratios are mostly lake specific dynamics. The starting conditions were air mixing rations.

Commented [PA3]: Add why values were elevated