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.

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.

Line 22: See general comment for direct device comparison.

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.

Introduction:

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

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

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

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.

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

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

Line 96: The ‘of’ is not needed here.

Line 98: Can you provide any diffusivity values through the silicon tube, in order to gauge, if that might influence system response times?

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.’

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.

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)

Line 110: Specify if the zero calibration performed with N2.

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

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.

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).

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.

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.

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

161: The ‘see’ is not needed here.

Figure 4: Correct the ‘blue dots’ in the figure caption.

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.

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

Line 164: The referenced table appears to be in the Appendix; please update the citation accordingly.

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.

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.

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.

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.

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.

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.

Figure 6: In my version, the columns representing precipitation are not blue; please correct if necessary.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

Line 235: Here arguably the vertical spatial resolution, where your device clearly outperforms common methods also plays a role.

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.

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

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.

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

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?

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

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