**Summary:**

The manuscript "*Stratification-induced hypoxia shapes greenhouse gas dynamics in ponds*" by Esposito et al. describes the monitoring of GHG dynamics in 6 shallow ponds over the growing season. The authors used a combination of high frequency temperature sensors and profile measurements to assess stratification and mixing in each pond, combined with bi-weekly measurements of $CH_4$ and $CO_2$ concentrations, nutrient concentrations, and ebullitive emissions. The authors found that GHG concentrations and $CH_4$ ebullition were higher under periods of stratification and hypoxia, and attribute oxygen as a clear driver of GHG emissions.

Overall, this study presents a valuable dataset of pond GHG measurements and contributes to the sparse but growing literature linking stratification patterns to GHG dynamics in small, shallow systems. The manuscript would benefit from a clearer justification of how this work advances or differentiates itself from other recent studies addressing similar questions. Nevertheless, the topic is timely and addresses an important area of uncertainty in pond emissions. I believe there are several aspects of data interpretation and presentation that should be carefully considered and, if addressed, will substantially strengthen the manuscript.

**General comments:**

1. The definition and use of the term "hypoxia" needs to be clear and justified. The term hypoxia is central to this manuscript, appearing in the title, abstract, and throughout the manuscript, though an actual definition is not given until line 110 (as <0.5 mg/L DO). I am wondering if hypoxia is the correct term to use, as it seems as though the authors are referring to it as anoxia. While not a formal definition, many studies reference dissolved oxygen concentrations < 2 mg/L as hypoxia (e.g., Diaz & Rosenberg, 2008), while anoxia would refer to the complete absence of oxygen. What is the reason for the <0.5 mg/L cutoff? Is it because the DO sensor does not always completely register 0 mg/L? If so, perhaps it is best to define it as anoxia, as throughout the discussion it is seemingly referred to as the absence of oxygen.

I think this distinction is important because of the implications for methane oxidation. Oxidation can still occur (and may actually be highest) when oxygen concentrations are low (e.g. hypoxic), but not completely absent (anoxic). If the authors are essentially referring to hypoxia as the absence of oxygen, I would change to the term anoxia, or give this justification and their definition of hypoxia at the beginning of the manuscript.

2. While characterizing GHG dynamics in ponds is a timely and important topic, the specific knowledge gap addressed by this study needs to be articulated more clearly. The introduction provides a useful overview of recent work examining pond GHG emissions and mixing dynamics, but it does not sufficiently establish how this study advances beyond that existing literature or identify the specific mechanisms that remain poorly understood. Clearly defining the study's novel contribution and central knowledge gap would also help provide a stronger focus for the discussion and the remainder of the manuscript.

3. The interpretation of the results needs to be reconsidered or further justified, especially the argument that hypoxia is driving an increase in greenhouse gas production/ebullition. The authors make the claim that ebullition is higher in stratified conditions with hypoxia, and that the low oxygen availability is increasing ebullition rather than temperature (e.g., in the abstract: "indicating oxygen availability, rather than temperature, dominated GHG buildup and ebullition"). However, the data does not support this claim, and the authors do not provide mechanisms for why this may be the case. Methanogenesis fueling ebullition is occurring in sediments that are almost assuredly anoxic regardless of the oxygen concentrations in the overlaying water. The claim that increased ebullition is not due to temperature seems to stem from the fact that bottom water temperatures were not significantly different between hypoxic and oxic bottom waters, but not from a direct test of how temperature influences ebullition. In fact, Fig. S10 shows a significant positive correlation between ebullition and temperature that is stronger than the correlation between ebullition and oxygen, seemingly disputing the authors claim. In addition, the strong seasonality of ebullition is currently largely ignored. In Fig. S9, it seems all the ponds have clear seasonal patterns of ebullition, though the timing varies between ponds. While summer months with high ebullition rates also tend to have hypoxia, ebullition appears to increase before the onset of hypoxia and stay elevated after oxygen increases. These seasonal patterns and other drivers of ebullition must be acknowledged and further explored.

**Specific comments:**

<u>Abstract:</u>

Line 16: Freshwater is one word.

Line 16-17: This first sentence seems to suggest that because ponds are biogeochemical hotspots, they are then sources of $CH_4$ and $CO_2$, whereas this might not be the case for all ecosystems that are "biogeochemical hotspots".

Line 22: I would define hypoxia here (see above comment)

Line 27: I do not think the results support the claim that "oxygen availability, rather than temperature, dominated GHG buildup and ebullition" (see above comment).

Lines 27-31: These two sentences are repetitive, essentially making the same point.

<u>Introduction:</u>

Line 35: Here you say "small lakes", while later (line 39) you say "shallow lakes". The Richardson et al., 2022 paper cited here actually has different definitions for small and shallow lakes, so this distinction is important.

Line 56: The process of ebullition is not unique to shallow waterbodies, though this sentence seems to imply such. In addition, diffusive fluxes of gases still occurs across the thermocline in stratified systems, even if it is slowed down. The wording throughout this paragraph seems to imply that gas transfer is completely shut off.

Line 61: This is where I think the hypoxia-anoxia distinction is important. Hypoxia (low oxygen) does not inherently limit $CH_4$ oxidation, as if oxygen is present oxidation can still occur.

Lines 69-71: These two questions need to be better differentiated, as they seem like essentially the same question (how to temperature dynamics and oxygen dynamics influence GHGs).

Methods:

Lines 79-82: These are results and should be in the results section. These results on nutrient concentrations are actually coming before you state the methods on how you collected this data.

Lines 96-99: It should be stated the exact depths that temperature sensors were placed in each pond. Does bottom mean directly on top of the sediment? How were sensors kept from getting buried in the sediment? Etc. Perhaps a supplemental table would be helpful.

Line 101: It would be helpful if you split the supplemental text into sections (e.g. "supplemental text S1) so that you could refer to specific supplemental text. I know currently there are only two supplemental text paragraphs, but this would still be useful.

Line 103: This definition of stratification is not clear to me, do you mean that the maximum temperature gradient had to be above 1°C per meter? Or that you only look at the surface temperature and the temp at 1 meter to see if there was a >1°C difference?

Additionally, because stratification is central to the main questions of the manuscript, I think there should be measure of the strength of stratification. Perhaps the density gradient from top to bottom, or just the temperature difference from top to bottom. I think this could be a figure in the main manuscript, as Fig. S5 shows strong temporal and diel patterns in stratification that is central to the interpretation of the results.

Line 106: I do not think that stratification always means hypoxia, and the data in Fig. S4 also shows this is not the case.

Line 109: I am confused by this logic, if you are using a cutoff for stratification of 1°C per meter, shouldn't the cutoff be less than 1°C if the pond is less than a meter deep?

Line 111: How were samples collected from the entire water column?

Line 118: What depths were sampled for surface and bottom water? How were bottom samples retrieved to not disturb the sediment? How were samples handled to not introduce air bubbles? More details are needed here.

Additionally, how was all sampling conducted by boat? Or on foot?

Line 141: This reference (Bastviken et al., 2015) is not listed in the references at the end of the manuscript.

Line 151-153: This is an extremely important point, and this must be taken into account when interpreting these results throughout the rest of the manuscript. This is a very large range of uncertainty,

and this needs to be stated when presenting the results. In the cited paper, Davidson et al., 2024, they state that the "static-chamber method cannot be said to accurately quantify CH4 emissions", and a similar statement should be made here. The ebullition results are currently presented as though they are meant to quantify emissions, despite this limitation. Even in comparison among these ponds, I think there needs to be caution, as ponds with higher ebullition would have higher amounts of $CH_4$ diffuse back into the water, potentially muting the actual patterns of ebullition.

Results:

Line 178: This does not seem to reference the correct figure, Fig. S3 shows oxygen concentrations. As stated above in the methods comments, I think a figure showing the stratification strength over time in each pond would be very beneficial, and should go in the main manuscript.

Line 179: Based on Fig. S5, I am not sure that it is true that RH1, RH4, and RH6 were stably stratified for that period. These ponds are very shallow, and the stratification and mixing dynamics are likely not as straightforward as presented. In RH4 and RH6, there seems to be times (likely at night) where bottom and surface temperatures are almost the same in late July and early august. RH2 actually seems more stably stratified compared to 4 and 6.

Line 183: Here and elsewhere, it is stated that relationships are "weak but significant". Here the p values is <0.001, so what exactly is meant by "weak"? This seems like a strong and convincing relationship to me. Perhaps remove these qualifiers and just state the p value.

Lines 183-186: Here and throughout the results, please state the actual results. The effect size, p-values, means, etc. should be stated here.

Line 191: But fig. S6 shows that bottom $CO_2$ and $CH_4$ was significantly correlated to bottom temperature? For $CH_4$ this p value was < 0.001.

Line 212: Here you say anoxic, which makes me think you are referring to anoxia when using "hypoxia"?

Line 273: As stated above, how are you accounting for temporal patterns and seasonal variation when looking at ebullition under hypoxic and oxic conditions?

Line 276-277: Based on fig. S9, I think this pattern is present in several ponds and not just RH3 (e.g. RH2, RH1, RH6)

Lines 294-295: If ebullition is underestimated, it would greatly influence these results. With the way ebullition was measured, I am not sure if this type of analysis is appropriate.

Discussion:

Line 320: What does non-correlated mean here?

Line 327: I do not think this supports the hypothesis that low oxygen is driving GHG production. For $CH_4$, production is often largely occurring in sediments that remain anoxic regardless of oxygen in the overlaying water. While anoxia may lead to more labile substrates reaching the sediment, this was not

tested or discussed. Rather, it seems more likely that anoxia is limited $CH_4$ oxidation, but this is not discussed here.

Additionally, here and throughout the discussion, the differences in $CH_4$ and $CO_2$ need to be more carefully interpreted. For $CO_2$, aerobic respiration is simultaneously consuming oxygen while producing $CO_2$. Therefore, high $CO_2$ concentrations may be correlated with low oxygen concentrations, but low oxygen is not "driving" $CO_2$ production.

Line 379: There can still be diffusion of gases across stratified layers, this should be acknowledged.

Line 392: Given the uncertainty of the ebullitive fluxes, I do not think it is appropriate to use these as an accurate estimate of emissions.

Line 399: I again think the wording needs to be more careful here. Hypoxia does not necessarily cause the pond to become a source of GHGs, rather, increased respiration may shift the metabolic balance, and create hypoxia as a result of oxygen consumption.