We thank the reviewers for taking the time to evaluate our manuscript and for the positive and constructive comments. We have carefully considered all the suggestions and detailed responses are provided below. Reviewers' comments are written in black, our responses in blue. Line numbers refer to the submitted manuscript.

As suggested, we mainly focused our changes on:

- The term "hypoxia" will be replaced with "anoxia". This is supported by recent sources (Nkwalale et al., 2026) that justify the term "anoxia" for freshwater oxygen concentrations ≤ 0.5 mg/L. The threshold for the oxic conditions in our paper (> 0.5 mg/L) is unchanged, so results are unchanged
- Methane ebullition fluxes will be corrected with the average of a 25% underestimate of emissions of static chambers as reported in Davidson et al., 2024. We will present this in fig. 6 and 7 and discuss the implication in the Discussion section.
- We re-analysed CH_4 ebullition and bottom-water CH_4 and CO_2 with mixed-effects models including bottom-water temperature and anoxia together, with pond as random effect and a CAR1 structure for temporal autocorrelation. Anoxia significantly increased dissolved CH_4 and CO_2 in the bottom water at a given temperature (both $p < 0.001$). For ebullition, the two drivers acted at different scales: within each pond, ebullition followed the seasonal changes in temperature ($p < 0.001$), whereas between ponds, those with more frequent anoxia had significantly higher ebullition ($p = 0.003$). The abstract, results and discussion will be revised accordingly.
- All statistical analysis and models will be fully described and reported with all the necessary parameters.
- All the figures, table and text will be carefully checked and the missing references will be added
- Stratification is now defined as a surface–bottom temperature gradient ≥ 1 °C m^{-1} (temperature difference divided by the distance between the measurements), calculated from the manual profiles at each sampling. The gradient (°C m^{-1}) is also used as a measure of stratification strength and shown in a new figure in the text, together with the high-frequency logger data.

Review of *Stratification-induced hypoxia shapes greenhouse gas dynamics in ponds* by

Esposito et al.

General comments

This study presents a valuable dataset characterizing the development and frequency of thermal stratification and associated hypoxia, as well as their effects on CH_4 concentrations and fluxes over an entire growing season in six small, shallow ponds. Although the main findings generally support previous observations from lakes and ponds, the study provides a novel contribution through its detailed, depth-resolved measurements of water

temperature and dissolved oxygen, combined with measurements of CH_4 and CO_2 concentrations in both surface and bottom waters. This approach allows the authors to assess how short-term transitions between stratification and mixing, and the resulting alternation between oxic and hypoxic conditions, affect methane concentrations and fluxes. The study therefore offers valuable insight into greenhouse gas dynamics in small ponds. Overall, I found the manuscript valuable and relevant. However, several aspects would benefit from further clarification or more detailed consideration. In particular, I consider it important to address the uncertainty associated with the CH_4 ebullition estimates and its potential implications for the reported total CH_4 fluxes and the relative contribution of ebullition. In addition, the description of the statistical analyses should be expanded to provide sufficient information about the model structures and how repeated measurements and temporal autocorrelation were addressed. Finally, reporting the actual diffusive CH_4 fluxes would provide valuable context and facilitate comparisons with other studies (see the specific comments below for further details).

We thank the reviewer for the positive comments and constructive suggestions.

As mentioned above, due to the choice of using static chambers to measure CH_4 ebullition, the ebullitive fluxes are underestimated. Initially in the paper we decided to present this underestimate without correction and acknowledge the underestimation. However, to address the reviewer's concern we have now applied a 25% correction based on the mean underestimate described by Davidson et al., 2024 where they compared the fluxes measured with static chambers vs fluxes measured by automatic flushing chambers. The corrected fluxes (accounting for a 25 % underestimation, i.e. measured fluxes divided by 0.75) will be shown next to the measured fluxes in Fig. 6 and Fig. 7. We will show the total budget with and without the underestimation – changing therefore the contribution of ebullition to the total CH_4 budget. Despite the change in absolute values (in absolute terms, high fluxes increase more than low fluxes), the correction is proportional – all fluxes are divided by the same factor (1 – 0.25) – so the statistical relationships with temperature and anoxia, which are tested on log-transformed fluxes, do not change. We will then discuss the implication of the new results in the Discussion section.

The statistical tests will also be better described, to provide sufficient information about how we addressed the problems repeated measurements and we will now also assess temporal autocorrelation using the autocorrelation function (ACF) of the normalised model residuals within each pond and show the ACF plots in the Supplement.

Results from the diffusive fluxes will be reported in a table and further discussed in the Discussion section.

Specific comments

Lines 62–63: Could the authors briefly explain the mechanism through which shifts in primary producers may enhance greenhouse gas emissions?

Yes, we will add some lines explaining how the shift between a plant dominated pond to a phytoplankton dominated pond might alter GHG emissions. Higher submerged plant abundance can potentially reduce CH_4 emissions, due to the structural role of plants, but also due to the CH_4 oxidation activity associated with them.

However, dense vegetation can also increase CH_4 emissions, and the effect may depend on plant identity. Conversely, phytoplankton dominance can increase the supply of labile organic matter to the sediment and reduce oxygenation near the bottom, both of which favour CH_4 production and accumulation. We will support the new statements with references as DelSontro et al., 2016, Davidson et al., 2018, Rabaey et al., 2026, Holgerson et al., 2026.

Lines 123–125: Did this calculation account for the background concentrations of CH_4 and CO_2 in the air used for the headspace method described in lines 118–120?

Yes, they did. We will make sure to report it carefully in the Methods section

Lines 126–137: Except for the relative contributions of ebullition and diffusion, I could not find any results on diffusive fluxes in the Results or Discussion sections. I understand that the diffusive fluxes were calculated from dissolved gas concentrations, whose temporal patterns are described in detail in the Results. Nevertheless, reporting the actual diffusive flux values would provide useful additional context and facilitate comparisons with other studies. I therefore suggest including, at a minimum, the mean and range of the diffusive fluxes.

That is a good point. The diffusive fluxes from each pond will be reported in a table in the text

Lines 151–153: The sampling design likely leads to some underestimation of CH_4 ebullition because of CH_4 exchange between the chamber headspace and the pond water. Given that the magnitude of this uncertainty was quantified in the cited study, could the authors clarify why it was not taken into account when reporting the results? Is the uncertainty range reported in that study also applicable to the ponds investigated here? If so, I suggest presenting values corrected using this uncertainty range alongside the original, uncorrected estimates. Such a correction could affect both the estimated total CH_4 fluxes and the relative contribution of ebullition to the total flux, which appears comparatively low relative to values reported for other ponds when based on the current conservative estimates. The potential effects of the selected methodology on the results, including a comparison of corrected and uncorrected CH_4 ebullition estimates, should also be considered in the relevant part of the Discussion.

Yes, thank you for the suggestion. We will now make a new fig. 6 and fig. 7 with the comparison of the conservative ebullition fluxes and the new ebullition fluxes with the 25% of underestimation due to the use of static chambers. In fig. 7 we will account for the underestimation in the total budget, increasing, therefore, the possible contribution of ebullition to the total CH_4 budget.

Lines 163–173: The description of most of the statistical analyses is rather general and would benefit from additional detail. More specifically, for the analysis described in lines 168–172, which aspect of hypoxia was tested: its occurrence, its duration, or the dissolved oxygen concentration? Please also specify which other variables were included as fixed effects and which, if any, were included as random effects. Was the final model assessed for temporal autocorrelation? This could influence the results because repeated measurements were conducted in the same ponds. Similarly, for the analysis described in lines 172–173, please specify the type and structure of the linear model used. Did the model account for repeated measurements within ponds and potential temporal autocorrelation? These observations cannot be considered fully independent.

In the original analysis, the tested metric was bottom-water oxygen status at each sampling (categorical: oxic > 0.5 mg L^{-1}, anoxic ≤ 0.5 mg L^{-1}), with oxygen status as fixed effect and pond as random intercept to account for repeated measurements within ponds. We have now extended this analysis. Durbin–Watson tests on the model residuals showed temporal autocorrelation in two ponds (RH1 and RH2). We therefore added a continuous-time first-order autoregressive correlation structure (CAR1) within ponds, which improved model fit and removed the residual autocorrelation (ACF plots will be shown in the Supplement). Bottom-water temperature was added as a fixed effect, and anoxia was split into a between-pond component (proportion of anoxic samples per pond) and a within-pond component (anoxia at the time of sampling; see also our response to Reviewer 2). For the analysis in L172–173, the correlations will be replaced by the same model structure. All model structures will be described in the Methods section.

Lines 279–281: Three different terms appear to be used for CH_4 ebullition within these three lines, including bubble flux and CH_4 bubble emissions. This terminology may be confusing because, in some studies, bubble flux refers only to the volume of released gas, irrespective of its CH_4 content. To avoid ambiguity and ensure consistency, I recommend using the term CH_4 ebullition throughout the manuscript.

Thanks for spotting it. We will use CH_4 ebullition consistently throughout the manuscript.

Lines 293–295: Please also report the ranges (minimum–maximum) for the relative contributions of diffusive fluxes and ebullition. Could the comparatively low contribution of ebullition observed in these ponds be, at least partly, an artefact of the selected methodology, which may substantially underestimate CH_4 ebullition?

Yes, we will report the diffusive fluxes of each pond in a table and we will consider that the static chambers are underestimating the CH_4 ebullition of on average 25%, so we will show the new ebullition results next to the results that were calculated without accounting for the underestimation.

Lines 291–300: Given that ebullition was generally higher in ponds experiencing more frequent and prolonged hypoxic events, was this pattern also reflected in a greater relative contribution of ebullition to the total CH_4 flux? Alternatively, were diffusive fluxes also higher in these ponds, resulting in a similar relative contribution of ebullition across ponds?

We will also calculate the relative contribution of ebullition to total CH_4 emissions in each pond, for both measured and corrected ebullition, and test whether it is higher in ponds with frequent anoxia, or whether diffusive fluxes also increased in these ponds.

Lines 326–328 and 334–335: Can the available data clearly distinguish whether CH_4 accumulation in the bottom waters of these ponds was driven primarily by enhanced CH_4 production rather than by reduced CH_4 oxidation? One could argue that the sediment layers where methanogenesis occurs remain anoxic immediately below the sediment–water interface, irrespective of whether hypoxia develops in the overlying bottom water. This interpretation may also be supported by the observations from pond RH3, where ebullition increased before hypoxic conditions developed in the bottom water (lines 276–277). What evidence indicates that the observed CH_4 accumulation resulted mainly from enhanced methanogenesis rather than reduced methanotrophic activity? This distinction should be clarified or the interpretation presented more cautiously.

Yes, that is true. We have no strong evidence proof that bottom-water anoxia is enhancing CH_4 production over suppressing CH_4 oxidation, we cannot separate the role of the two processes. However, other studies have found sediments overlain by anoxic water released two to four times more dissolved CH_4 per unit area than oxic sediment in the epilimnion (Bastviken et al. 2008). We will need to explicitly mention this uncertainty over which process is driving increased concentrations in the bottom was water in the discussion.

Line 348: Please specify the duration of the hypoxic events described here as "short."

We will specify in the text the number of bottom-water anoxic measurements that we define as "short".

Technical corrections

Lines 80–86: Table 1 does not include the TN ratio, although this variable is mentioned in both the main text and the table caption.

Yes, that was a mistake. We will remove the TN ratio phrasing, as we are directly showing TN and TP concentrations.

Line 178: Figure S3 does not appear to present the manually measured temperature profiles. Please check and correct the figure reference

Thanks for spotting it. We will revise that.

Figures 3 and 4: These figures are not boxplots, as stated in their captions. Please revise the figure descriptions accordingly.

Thanks for spotting it. We will revise that.

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.

We thank the reviewer for the positive evaluation and the constructive comments, which we address below.

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.

We thank the reviewer for this comment, and will follow the suggestions of using the ≤ 0.5 mg/L threshold to define anoxic rather than hypoxic conditions. Following Nkwalale et al., 2026, who define anoxic levels as oxygen ≤ 0.5 mg/L, we will replace "hypoxia/hypoxic" with "anoxia/anoxic" throughout the manuscript, including the title. We will state in the Methods that anoxia is operationally defined as ≤ 0.5 mg L^{-1} and report the accuracy of the oxygen sensors at low concentrations (± 0.1 mg/L). As the threshold is unchanged, all analyses and results remain the same.

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.

Thank you for finding this gap. We will highlight in the introduction and in the discussion section the novelty of this study:

- shallow waters can stratify and become anoxic several times during the warm season;
- high frequency measurements are necessary to better quantify GHG emissions from shallow waters, as they are very variable also due to the storage of CH_4 and CO_2 in the bottom waters during periods of anoxia related to thermal stratification.

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.

We thank the reviewer for this comment. We have re-examined the data and re-run the analyses to better separate the effects of anoxia and temperature. Methane ebullition was analysed with a linear mixed-effects model on log-transformed ebullition including bottom-water temperature and anoxia, with pond as a random intercept and a continuous-time first-order autoregressive structure (CAR1) for temporal autocorrelation. To address the seasonality and the observation that ebullition sometimes increased before the onset of anoxia, anoxia was split into a between-pond component (proportion of anoxic samplings in each pond) and a within-pond component (anoxia at the time of sampling), following van de Pol and Wright (2009).

The results confirm that anoxia is a key factor, but show that it acts at different scales.

- Within ponds, ebullition followed bottom-water temperature over the season (+24 % per °C; $p < 0.001$), and anoxia at the time of sampling had no significant additional effect ($p = 0.26$).
- Ponds with frequent anoxia had substantially higher ebullition at a given temperature (+32 % per 10 % increase in the proportion of anoxic samplings; $p = 0.003$), also when excluding the most frequently anoxic pond (RH6; $p = 0.044$). These ponds were not warmer than the others.
- Dissolved CH_4 in the bottom water was 6.8 times higher and CO_2 2.0 times higher under anoxic than under oxic conditions at the same temperature (both $p < 0.001$). Models using bottom-water oxygen concentration or the surface–bottom temperature difference instead of anoxia fitted the ebullition data less well.

We will therefore revise the abstract, results and discussion. Rather than stating that oxygen dominated over temperature, we will report that anoxia strongly increased the accumulation of CH_4 and CO_2 in the bottom water and that ponds experiencing frequent anoxia had higher ebullition, while seasonal changes in temperature best predicted ebullition within ponds. We will discuss whether seasonal temperature acts as a direct control or mainly as a predictor that integrates other seasonal changes. We will also discuss the possible mechanisms: prolonged anoxia may reduce CH_4 oxidation at the sediment–water interface and in the bottom water,

and the supply of labile organic matter from the water column may enhance CH_4 production in the surface sediments (e.g. Zhou et al., 2019).

Specific comments:

Abstract:

Line 16: Freshwater is one word.

Whilst it is true that fresh water is one word as an adjective, it can be two words as a noun. E.g. freshwater ecology is the study of the ecology of fresh waters. We believe the text is correct as stands but will check it to be sure.

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"

We will revise the text, removing the 'as such'.

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

We will, as mentioned above, defining anoxia where oxygen concentrations are ≤ 0.5 mg/L, according to recent literature.

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).

We agree that this sentence needs revising. Based on the new analyses as mentioned in the comments above, we will replace it with: "Anoxia strongly increased CH_4 and CO_2 accumulation in the bottom water, while CH_4 ebullition increased with temperature within ponds and was higher in ponds experiencing more frequent anoxia."

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

Only one sentence will be kept.

Introduction:

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

We will use the terms consistently and clarify them in the text.

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.

We will rephrase the text to make clear that CH_4 ebullition is more important in shallow waters, but does not happen only in these environments. In addition, stratification slows but does not stop diffusive exchange.

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.

We will now change the text to "anoxia" instead of "hypoxia", according to recent literature. We will rephrase the text saying that CH_4 oxidation is suppressed under anoxic conditions, not under low-oxygen conditions.

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).

We will clarify the questions so that there is not overlap between the two.

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.

We will move the data to the Results section.

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.

The temperature sensors were not placed directly on top of the sediment. Each logger was attached to an anchor rope approximately 25 cm above the anchor, and the rope was secured to a float at the water surface. This setup was used with the intention of keeping the sensors above the sediment and reducing the risk of burial. As the anchor bags may have settled into the soft sediment to an unknown extent, there is some uncertainty around their height above the sediment, which may be less than 25cm, however, we are confident they lay above the sediment.

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.

We will do that.

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?

We thank the reviewer for pointing this out. The original definition was applied as a ≥ 1 °C difference between surface and bottom water regardless of the distance between the two measurements. We have now recalculated stratification for each sampling from the manual profiles (Aqua TROLL, accuracy ±0.1 °C). For each profile, we calculated the vertical temperature gradient as the surface–bottom temperature difference divided by the

distance between the surface (0.2 m) and bottom measurements. A similar approach was used by Holgerson et al. (2022) to classify mixing in ponds and shallow lakes, although they expressed the top–bottom gradient as water density rather than temperature. The distance between the measurements was close to 1 m in most profiles (mean 1.3 m, range 0.75–1.9 m). A pond was considered stratified when the gradient was ≥ 1 °C m^{-1} (Wetzel, 2001). This takes the different depths of the ponds into account. This changed the classification of 15 of the 90 samplings, mostly spring samplings and samplings in RH1, the deepest pond, where a surface–bottom difference slightly above 1 °C corresponds to a gradient below 1 °C m^{-1}. With this definition, 64 of the 90 samplings were stratified. The revised classification is used in all figures and analyses. The Methods will then state:

"Thermal stratification was assessed at each sampling from the vertical profiles as the temperature gradient between the surface (0.2 m) and the deepest measurement, i.e. the temperature difference divided by the vertical distance between the two depths (mean 1.3 m, range 0.75–1.9 m), similar to the top–bottom density gradient used by Holgerson et al. (2022). Ponds were considered stratified when the gradient was ≥ 1 °C m^{-1} (Wetzel, 2001). The gradient (°C m^{-1}) was also used as a measure of stratification strength. The high-frequency temperature loggers were used to describe stratification and mixing dynamics between samplings."

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.

We agree and thank the reviewer for this suggestion. The temperature gradient (°C m^{-1}) now provides a continuous measure of stratification strength at each sampling. The mean gradient per pond ranged from 1.8 °C m^{-1} (RH1, RH3) to 2.3 °C m^{-1} (RH2, RH6), with maxima of 5.0–8.2 °C m^{-1}. A new figure in the text will show, for each pond:

- the daily mean and diel range of the gradient from the high-frequency loggers (28 June–4 November 2021);
- the gradient measured in the profiles at each sampling, with oxic and anoxic samplings marked;
- The probability of anoxia increased with stratification strength (see replies to L106).

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

We agree. Most anoxic samplings (28 of 34) occurred under stratification and the temperature gradient was higher at anoxic than at oxic samplings (mean 2.6 vs 1.7 °C m^{-1}). Stratification strength alone did not explain which ponds became anoxic: RH2 was as strongly stratified as the other ponds (mean 2.3 °C m^{-1}) but was anoxic at only 2 of 15 samplings, compared with 8 of 15 in RH1 and 12 of 15 in RH6. Conversely, RH6 remained anoxic also when mixed, likely due to its dense submerged macrophyte beds. We will state this clearly and

discuss that, while stratification favours anoxia, the tendency of each pond to become anoxic likely depends on other factors such as sediment oxygen demand and vegetation

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?

The reviewer is correct. A fixed 1 °C difference is not equivalent to 1 °C m^{-1} when the distance between the surface and bottom measurements differs from 1 m. In our ponds this distance was close to 1 m in most profiles, but it was longer in the deepest pond (up to 1.9 m in RH1). We now use a vertical gradient (temperature difference divided by the distance between the measurements, analogous to the top–bottom density gradient of Holgerson et al., 2022), so the threshold scales with the depth of each pond. This changed the classification of 15 of the 90 samplings, mostly spring samplings and samplings in RH1. The new classification is applied throughout the manuscript.

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

We will add some text to carefully explain the procedure.

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.

Yes, we will add more details in the text about the sampling depths and procedure.

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

By boat. We will state it carefully in the text.

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

Thanks for spotting the gap – we will add it to the reference list.

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 CH_4 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.

We agree. We will present corrected and measured ebullition together (see comments above) and state in the Results and Discussion that our estimates are conservative. We will also discuss that CH_4 re-dissolution in the chambers may reduce differences between ponds, especially where ebullition is highest.

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.

We will correct the figure reference and include a new figure in the text showing, for each pond, the surface–bottom temperature gradient (°C m^{-1}) over time: the daily mean and diel range from the high-frequency loggers (28 June–4 November 2021), the gradient measured in the profiles at each sampling with oxic and anoxic samplings marked, and the 1 °C m^{-1} threshold.

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.

We thank the reviewer for this careful observation, which is confirmed by the logger data. Using the same threshold as for the profiles (1 °C m^{-1}), between 20 July and 10 August the water column mixed at least partly (gradient < 1 °C m^{-1} for part of the day), mostly at night, on 19 of 22 days in RH5, 15 in RH3 and 13 in RH4, but only on 5 days in RH6, 1 day in RH1 and no days in RH2. As the reviewer noted, RH2 was the most stably stratified pond. We will revise the description of the stratification patterns accordingly.

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.

Yes, we will remove the qualifier and only use the statistical values.

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

All the statistical results will be reported.

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.

We agree that the original text was not clear about the relationship with temperature and CO_2 and CH_4. We have now tested bottom-water temperature and anoxia together in linear mixed-effects models with pond as a random effect and a CAR1 autocorrelation structure (as for ebullition; see comments above). Bottom-water CH_4 increased with temperature (+32 % per °C; $p < 0.001$), but at a given temperature it was 6.8 times higher under anoxic than oxic conditions ($p < 0.001$). Similarly, bottom-water CO_2 increased with temperature (+10 % per °C; $p = 0.003$) and was twice as high under anoxic conditions ($p < 0.001$). Both temperature and anoxia therefore contributed to the accumulation of dissolved gases in the bottom water, with anoxia having the larger effect. We will revise the text accordingly and report the results of these models in a new table. Figs. S6 and S8 will be kept to show the raw data, with captions referring to the model results.

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

The terminology is now corrected.

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

The ebullition model now includes bottom-water temperature, which captures the seasonal variation, and a CAR1 structure for temporal autocorrelation within ponds; anoxia is split into within- and between-pond effects (see comments above).

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)

We will revise the text to better describe the patterns.

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.

We will show the comparison using both measured and corrected ebullition (see comments above) and discuss its limitations.

Discussion:

Line 320: What does non-correlated mean here?

We will rephrase the text.

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.

As addressed above, the role of the overlying water on CH_4 emissions is well established. As mentioned above, the oxygen status of the overlying water can affect CH_4 dynamics: in lakes, sediment overlain by anoxic water released two to four times more dissolved CH_4 per unit area than sediment under oxic water (Bastviken et al., 2008), and CH_4 accumulated in the bottom water while it remained anoxic (Encinas Fernández et al., 2014) However, the reviewer is correct that we have not sufficiently discussed the role of CH_4 oxidation, which can be substantial. We will discuss the possible relative importance of production versus oxidation of CH_4 in the revised manuscript.

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.

Agreed we will rephrase our interpretation of CO_2 to reflect that both result from respiration.

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

Thanks. We will add this.

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.

We do not present them as quantitatively accurate but rather as the least worst emissions estimate, we also attempt a correction of the degree of underestimation.

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.

This is a good point, the change in metabolism causes oxygen depletion which alters the balance of GHG concentration (figure 5) and will revise this sentence to reflect that increased respiration can both deplete oxygen and increase GHG concentrations.