

We are grateful to the reviewer, Thank you so much for this thorough, constructive, and in-depth review. Thank you for pointing out concrete needs for clarifications and improvements. Below we provide some preliminary response to the main comments and all numbered comments. We will also provide detailed response to technical corrections at a later stage. (Reviewer's comments are underlined for clarity).

Specific comments:

- Are the studied lakes representative for other lakes or waterbodies? Please clearly outline limitations of this study in terms of impact and application in a broader sense.

Thank you for raising this. Admittedly, the significance of our study and representativeness was somewhat neglected. Addressing this will definitely improve the manuscript. Short answer is “Yes”, Svartkulp is particularly representative of Northern Hemisphere lakes, typically found in granitic bedrock regions in North-East America and Scandinavia. It is a typical low-productivity, heterotrophic, slightly acidic to neutral, moderately humic lake. Similar lakes are found in Southern Norway (de Wit et al., 2023), large parts of Sweden (Valina et al. 2014), and Finland, Atlantic Canada (Houle et al., 2022), Ontario, Québec and North-East USA (Skjelkvåle and de Wit 2011; Weyhenmeyer et al., 2019). Note that most of the lakes in Norway are more acidic than Svartkulp (de Wit et al., 2023), and have lower alkalinity and would therefore be more sensitive to acidification caused by HgCl₂ fixation. Even if Svartkulp is among the best buffered lakes in Norway, our findings are also relevant to more acidic, lower ionic strength lakes found in Norway and large parts of Northern Canada.

Lundebyvannet, is also representative of a large group of these Northern lakes, however, it is quite a productive lake with high photosynthetic activity, which is more of an end-member case (e.g., worst-case scenario for Norway, related to CO₂ flux overestimation with HgCl₂ fixation).

- Is it feasible to assume unfixed samples to represent “real” concentrations/fluxes, as control? Could you discuss this further and eventually consider renaming “control” to “unfixed” for the first experiment?

You agree with the suggestion to rename “control” samples to “unfixed” samples. In fact, we should reorganize the discussion and start with a discussion on how representative these unfixed samples are to real conditions.

For CO₂ and O₂, processing unfixed samples kept dark and cold is considered “real concentrations”. In particular, a few studies have used unfixed samples (only preserved dark at 4°C) up to 48h after sampling to determine CO₂ or DIC concentrations (e.g., Sobek et al. 2003, Kokic et al., 2015). So, the DIC samples in Lundebyvannet and the CO₂ and O₂ concentrations at t=0 should be fine. We will refer to these studies in the revised manuscript.

However, for CH₄, Fig. 1 shows that already within 24h, the CH₄ concentration in the unfixed sample is below atmospheric saturation while it is consistently much higher in all three sets of fixed samples. So, unfixed samples do not represent real CH₄ concentrations.

- The Methods section lacks necessary detail. I would suggest to restructure the section to make the experimental setup and respective study lakes clearer. E.g. in the study area section Lake Lundebyvannet should also be introduced, and ideally both lakes should be presented with the same level of detail relevant to the respective experiments. More importantly, I'm missing information on sampling procedures and their feasibility. Since this

is a technical note, I believe the Methods should be sound. I added several comments in this regard below.

Excellent point, yes we will add all necessary details, better describe Lundebyvannet. It is right that the sampling procedure needs to be clearer for others to be able to reproduce it and to evaluate their feasibility in other field contexts. Thank you for your help and all the specific line comments related to this, this is very helpful.

- I think the results of this study could be put into clear recommendations for future studies, and this could be part of the abstract and expressed more clearly in the discussion.

Another nice suggestion. We will add a section on recommendations in the Discussion and represent this section in the abstract. While some recommendations are formulated L. 535 onwards, they deserve a stronger position in the manuscript.

- The Introduction could benefit from adding some information about the other preservatives studied. The application of HgCl₂ is broadly introduced (could be shortened), but the description of the substitutes dealt with in this study falls short. Are there any other studies where CuCl₂ or AgNO₃ were used to determine dissolved gas concentrations? Are there differences expected between the application of those two?

We had a hard time trying to find studies where CuCl₂ and AgNO₃ have been used as preservatives for freshwater samples for the determination of dissolved gases. In fact, these preservatives are often only briefly mentioned in the methods which makes it difficult to identify relevant studies. We can nevertheless re-equilibrate the introduction between HgCl₂ and the two other alternatives by describing the expected impacts on water chemistry of the fixation with AgNO₂ and CuCl₂.

Please note, the numbers at the beginning of each comment denote the line numbers.

1. 26-27: can you be more specific about time periods (3w, 3m)?

Yes, initial increase to from 24h to 3 weeks, and then decrease from 3 weeks to 3 months.

2. 29: are low ionic strength / high DOC lakes representative?

Yes, as describe above, we will rephrase this (see our response to your first comment above)

3. 30: are these estimations valid for other lakes?

These are probably among the worst case scenarios regarding over-estimation of CO₂ efflux, we will clarify this in the abstract and discussion.

4. 31: I think explicitly adding recommendations here would be useful.

Yes, will do (see also our response to your 4th comment above)

5. 59: better in regards to what?

“yielded concentrations after preservation and storage closer to real concentrations”, will clarify this.

6. 67: what is the impact of higher H⁺ concentrations?

Shift in the equilibrium between CO₂ and HCO₃ with consequences on CO₂ and DIC estimations.

7. 70: could you elaborate why this leads to an overestimation of CO2 concentration?

Yes, we will elaborate here. If the pH is lower than thought, the measured CO2 signal by the GC represents a higher proportion of DIC, compared with a situation where the pH equals the expected pH.

8. 82-92: it would help the reader if it was made clearer here that two different experiments were conducted and two different lakes were sampled for that, for example by 1)... 2)...

As suggested, we will introduce the distinction here.

9. 87: This assumes that unfixed samples are the control, or "real results". Is that feasible?

We will rephrase here also consistent with our response to your second comment above.

10. 99: Did you collect the water from the surface? Did you use anything other than the bottles to avoid bubbling or degassing? Were samples temperature controlled (or otherwise controlled) between sampling and analysis?

We will describe this better.

11. 100: slowly poured - a bit vague? How could you guarantee no degassing?

We will describe in details the procedure, e.g., "The sample bottles were almost horizontal when starting to pour and slowly tilted as the water level was increasing to avoid any turbulence, bubbling and the like".

12. 103: Are the results you got from the water samples representative for lakes in the region in terms of magnitude? Do the numbers represent means of the sub-samples or was each sub-sample used for determination of one of the parameters?

Yes, see also our response to your 1st comment.

13. 105: As far as I understand, the concentration of platinum does not necessarily describe the color characteristics of water?

We will come back to this, I believe these are standard units, but color is not a major parameter here, DOC, Alkalinity, and nutrients are more important.

14. 106: how did you measure the temperature? is this an important information if the water was transported to the lab? Or did you preserve this temperature during transport?

This will be clarified, 18.5C was in the lake, we also monitored the temperature in the lab.

15. 111: technically 3 treatments and one control. Which of the scenarios would presumably result in the most "real" concentration?

This depends on which gas, CH4 is closer to real concentrations in the fixed samples, while CO2 and O2 in the unfixed samples at t=0 is the closer to real concentrations as also described in our response to your 2nd comment.

16. 112: why did you choose these time steps? could you elaborate if these times are representative?

These timesteps were selected and thought to be representative for average storage times, 24h as "direct measurements following a field trip", and 3 months as a long term storage scenario.

17. 116: Is the preparation of the solutions part of the experiment? do the yielded concentrations have an uncertainty? or would a derivation not have an impact on the outcome?

We used standard chemistry lab equipment with high accuracy scale to weigh the salt and volumetric flasks. Reference to the suppliers of the preservative salts will also be included in the description.

18. 133: Did the fact that pH was not measured affect your study? Or was that one reason to use the PHREEQC model?

Admittedly, the experimental design was imperfect and pH measurements following fixation should have been included. It would have yielded valuable information on the impact of preservation addition on e.g., pH. So yes, part of the reason why we used the PHREEQC model was because we didn't have pH data. However the PHREEQC model proved to be much more useful than just estimating the pH upon preservative addition, it was also predicting if the precipitation of some mineral phases was plausible.

19. 137: Is the sampling strategy outlined different than the one for the first experiment? How did you achieve sampling water from different depths? It would also be nice if both lakes were described with the same level of detail.

Agreed, both lakes need to be described with the same level of details. This will be corrected. Water was collected from different depths using a standard water sampler (information on this will be included). The sampling procedures will be described in more details here. They included direct sampling on the lake in this case.

20. 145: Could you clarify the purpose of DIC analysis in this study?

DIC analyses were performed to obtain an independent estimation of CO₂ and DIC concentration, compared to the GC analysis. This will be added here.

21. 147: Add name of TOC analyzer and/or merge with other sections below to avoid repetition. You state that samples were not fixed – why not?

Samples were not fixed to avoid any interaction with the sample. However, the samples were analyzed within 2h. Note that this is in line with previous studies.

22. 151: Did you compare the pH data with that measured with the pH-meter as mentioned above, or why measure twice?

We apologize, this is an error. pH was not measured in the laboratory, we only used the pH data from the *in situ* HOBO sensor. This sentence will be corrected.

23. 160: The temperature was recorded during shaking – do you mean the water temperature? What was the purpose?

Yes, the water temperature (sorry for the lack of clarity) was recorded during shaking to ensure it is stable throughout the equilibrium process. The temperature was stable, and the value was used to calculate back the dissolved gas concentrations.

24. 171: Do you mean ambient air was used for calibration? Did you know the concentrations of the ambient air?

Yes, ambient air was used for O₂ and N₂ calibration. Ambient air is measured regularly and is stable through time.

25. 175 section: I think it would help to directly add formulas in a section in the appendix for better understanding and reproducibility.

Good suggestion, thank you.

26. 187: Could you explain the purpose of DIC analysis here or in earlier sections. Are the CO₂ concentrations calculated in addition to the concentrations measured by GC for comparison? I think it may not always be clear where you used measured or calculated CO₂ concentrations.

This will be introduced earlier, see comment #20-

27. 244: Is it important to mention what files were input and output files? For someone who doesn't know the program, this info seems meaningless.

We will clarify this.

28. 255: Is this analysis done in retrospect to make up for not measuring sample pH directly after storage (among other things)?

Yes partly, this will be clarified.

29. 320: What temperature did you use to determine the Schmidt number?

I believe we used the ambient water temperature, but I need to confirm this and add this information here, agreed.

30. 328 f: It is nice to have different temporal resolutions, but what is the purpose of that for this study? Would your measurements not reflect instantaneous fluxes (could maybe be considered as daily fluxes) rather than weekly?

The main idea to show fluxes with these three temporal aggregations is to highlight the magnitude of the mis-estimation of the fluxes when HgCl₂ is used as preservatives for the water samples. The error magnitude can be much larger over shorter timescales (see. E.g., Fig. 4).

31. 340: A rather general comment to this study: what is the assumption about the development of gas concentration in between times 0, 3w, 3m? E.g., what would the concentration after 2w or 2m supposedly look like? Did you examine that?

We haven't looked at other time points, this is difficult to predict.

32. 362: Would preservation with AgNO₃ then be preferable rather than with HgCl₂ due to its toxicity? Can you draw conclusions regarding CH₄ from your results?

Yes definitely. These recommendations will be summarized into a dedicated section in the discussion later on.

33. 364, Fig. 1: Concentrations of all gases (except CO₂) show largest ranges for AgNO₃ addition (largest bars) after 3w. Is there an explanation for that? Add to caption: what do the boxplots show, presumably 25th and 75th percentiles and the median?

Unfortunately, no we have no explanation for the relatively larger range for the AgNO₃ fixed samples after 3w. We will add the description of the boxplots in the caption, yes.

34. 375 f: Are you arguing that this process is slowed down in freshwater?

Not necessarily slowed down because of freshwaters, there are many parameters playing a potential role here, e.g., temperature, substrate concentration, etc. Rees et al. (2021) performed their incubations at ambient temperatures which is the most likely explanation for the difference seen with our observations. Our samples were stored at 4C for 3 months.

35. 380-381: Is this assumption reflected in your results by the decrease of N₂ concentration? Is there also an explanation for the N₂O consumption following production?

The interpretation of small changes in the N₂ data should be avoided since none of the groups are significantly different from each other.

36. 385: Did you perform a statistical test here too? Is it worthwhile mentioning that the concentrations seem to have the opposite response over time than N₂O?

These changes do not have the same magnitude, μM for N₂, nM for N₂O. The expected changes in N₂ from the process mentioned in comment #35 is not detectable.

37. 422: The opposite of what you state in the text is shown in Tab. 3. Is there a mistake in the labels?

Yes, there is a mistake in the labels in Table 3, these are not consistent with Fig. 4. Thank you for spotting that.

38. 426 f: Wouldn't we expect to see a shift in pH then in Fig. 2 (top panel)? It appears as if the fixed and unfixed samples have the same pH?

The pH plotted here in the *in situ* pH which is the same for both sample sets. pH was determined in the lab only on DIC samples, not on GC samples. This will be clarified in the caption.

39. 435, Tab. 3: What was the reason to show fluxes calculated following Cole and Caraco and not the other wind-based models here?

This was to avoid overloading the table, the values are different but the relative differences between fixed and unfixed samples is bound to the original concentration data. We believe there is no point in showing three models showing the same differences.

40. 466: Why was this cut-off of 20 μM chosen?

This 20 μM cut-off was chosen as the maximum likely error from e.g., pH error of 0.05. This will be clarified in the caption.

41. 503: Which of those shown in Tab. 3 and Fig. 4 were obtained from DIC analyses?

We apologize, there is an inversion in the "Preservatives" labels in Table 3, lower values (1st and 4th rows) are from "unfixed" samples analyzed for DIC while the highest values were those from samples fixed with HgCl₂ (2nd and 5th rows). Same in Fig. 4, blue lines and crosses (lower values) are from "unfixed" samples analyzed for DIC.

42. 507 f: This estimate is only valid for the tested lakes. What would be the implication for other lakes? Is this also valid for sea water samples? Do you have recommendations or a protocol that should be followed? And what about greenhouse gases other than CO₂?

Excellent questions, thank you. We will answer those questions in a specific section in the discussion. The overestimation of CO₂ concentration of up to 300% is likely among the worst-case scenarios since it is restricted to lakes which present both a slightly acidic to near neutral pH and intensive photosynthetic activity. However, such lakes with similar pH and productivity are relatively abundant in the Northern Hemisphere (see our response to the first comment). Note that this overestimation is not valid for sea water samples where the fixation with HgCl₂ only caused a insignificant acidification because of the high buffering capacity of seawater (see Chou et al. 2016). For the other greenhouse gases, AgNO₃ seems to be as good as, or slightly better than HgCl₂ and CuCl₂ for preservation. All of these findings will be further discussed and presented as recommendations in a specific section in the discussion.

References:

- Chou W.C., Gong G.C., Yang C.Y. & Chuang K.Y. (2016) A comparison between field and laboratory pH measurements for seawater on the East China Sea shelf. *Limnology and Oceanography-Methods*, 14, 315-322. <https://doi.org/10.1002/lom3.10091>
- Clayer, F., Thrane, J.-E., Brandt, U., Dörsch, P., & de Wit, H. A. (2021). Boreal Headwater Catchment as Hot Spot of Carbon Processing From Headwater to Fjord. *Journal of Geophysical Research: Biogeosciences*, 126(12), e2021JG006359. <https://doi.org/10.1029/2021JG006359>
- de Wit, H. A., Garmo, Ø. A., Jackson-Blake, L. A., Clayer, F., Vogt, R. D., Austnes, K., Kaste, Ø., Gundersen, C. B., Guerrero, J. L., & Hindar, A. (2023). Changing Water Chemistry in One Thousand Norwegian Lakes During Three Decades of Cleaner Air and Climate Change. *Global Biogeochemical Cycles*, 37(2), e2022GB007509. <https://doi.org/10.1029/2022GB007509>
- Houle, D., Augustin, F., & Couture, S. (2022). Rapid improvement of lake acid–base status in Atlantic Canada following steep decline in precipitation acidity. *Canadian Journal of Fisheries and Aquatic Sciences*, 79(12), 2126–2137. <https://doi.org/10.1139/cjfas-2021-0349>
- Kokic, J., Wallin, M. B., Chmiel, H. E., Denfeld, B. A., & Sobek, S. (2015). Carbon dioxide evasion from headwater systems strongly contributes to the total export of carbon from a small boreal lake catchment. *Journal of Geophysical Research: Biogeosciences*, 120(1), 13–28. <https://doi.org/10.1002/2014JG002706>
- Rees, A. P., Brown, I. J., Jayakumar, A., Lessin, G., Somerfield, P. J., & Ward, B. B. (2021). Biological nitrous oxide consumption in oxygenated waters of the high latitude Atlantic Ocean. *Communications Earth & Environment*, 2(1), Article 1. <https://doi.org/10.1038/s43247-021-00104-y>
- Skjelkvåle, B. L., & de Wit, H. A. (2011). Trends in precipitation chemistry, surface water chemistry and aquatic biota in acidified areas in Europe and North America from 1990 to 2008 (ICP Waters report 106/2011). In 126. Norsk institutt for vannforskning. <https://niva.brage.unit.no/niva-xmlui/handle/11250/215591>
- Sobek, S., Algesten, G., Bergström, A.-K., Jansson, M., & Tranvik, L. J. (2003). The catchment and climate regulation of pCO₂ in boreal lakes. *Global Change Biology*, 9(4), 630–641. <https://doi.org/10.1046/j.1365-2486.2003.00619.x>
- Valiente, N., Eiler, A., Allesson, L., Andersen, T., Clayer, F., Crapart, C., Dörsch, P., Fontaine, L., Heuschele, J., Vogt, R., Wei, J., de Wit, H. A., & Hessen, D. O. (2022). Catchment properties as predictors of greenhouse gas concentrations across a gradient of boreal lakes. *10(880619)*. <https://doi.org/10.3389/fenvs.2022.880619>
- Valinia, S., Englund, G., Moldan, F., Futter, M. N., Köhler, S. J., Bishop, K., & Fölster, J. (2014). Assessing anthropogenic impact on boreal lakes with historical fish species distribution data and hydrogeochemical modeling. *Global Change Biology*, 20(9), 2752–2764. <https://doi.org/10.1111/gcb.12527>
- Weyhenmeyer, G. A., Hartmann, J., Hessen, D. O., Kopáček, J., Hejzlar, J., Jacquet, S., Hamilton, S. K., Verburg, P., Leach, T. H., Schmid, M., Flaim, G., Nöges, T., Nöges, P., Wentzky, V. C., Rogora, M., Rusak, J. A., Kosten, S., Paterson, A. M., Teubner, K., ... Zechmeister, T. (2019). Widespread diminishing anthropogenic effects on calcium in freshwaters. *Scientific Reports*, 9(1), Article 1. <https://doi.org/10.1038/s41598-019-46838-w>