

On behalf of all authors, I would like to thank the three referees for their detailed reviews with relevant and constructive comments to improve the quality of the manuscript. Following are specific responses to each of the referees' comments.

*The referees' comments are in black and the author's responses in red*

#### Response to referee comment #1

General comments: The study provides a valuable dataset given its high spatial and temporal resolution and a variety of sampling methods, which is rare in the current literature on methane dynamics in inland waters. Therefore, the study has a good potential to offer original insights on the subject, with a unique perspective on spatial and temporal patterns and methodological biases. Data collection, curation, and interpretation are generally appropriate (although I have a limited expertise in the EC technique). However, the structure of manuscript, the presentation of the results, and the discussion around them can be vastly improved. I would like to acknowledge the hard work of the authors for producing this manuscript, and I am confident it will be suitable for publication after some modifications following external feedback. **Thank you for acknowledging the value of our measured dataset, and for your detailed review and comments that will help us to improve the quality of the manuscript.**

Goal definition: To help the reader follow the logical structure of the study, it would be helpful to define the aims of the research in a more specific manner. The gap that the authors are filling with their research is not clearly stated in the introduction. For instance, at page 2 line 25-26: "many questions about reservoir emission behaviour remain" is very vague. While the authors state that they "investigate biophysical drivers of CH<sub>4</sub>", they should be more clear about how their study differs from multiple other studies investigating CH<sub>4</sub> aquatic drivers, and how their unique dataset enables them to tackle more specific still unanswered questions on the subject. For instance, are the drivers similar at different temporal scales? How different method capture or miss those drivers and what are their biases/uncertainty when upscaling?

**We appreciate this helpful feedback which was echoed in the other reviewer's comments. We agree that the manuscript can be improved by better defining the study aims and focusing on key findings that contribute new information to the body of knowledge on aquatic CH<sub>4</sub> emissions.**

**Planned changes to improve the goal definition include:**

- 1. Changing the manuscript title to highlight the study findings: "Short-term emissions account for most of a two-fold inter-annual difference in methane emissions in a small eutrophic reservoir: insights from two years of monitoring with eddy covariance and spatial surveys"**
- 2. Defining the aims of the study more clearly in the abstract and introduction. We are re-focusing the manuscript around the questions:**
  - 2.1. What can we tell about the relevant importance of hot-spots vs. hot moments to sampling bias by comparing results from different methods?**

## 2.2. How important is interannual variability in one lake (in this case, the spring burst), and what causes it?

Main message: The manuscript provides a lot of scattered new information, however, the main conclusions are diluted and not clearly highlighted in the manuscript. Defining the study aim will help on that matter, but the authors also need to choose a few key results and conclusions and structure the manuscript to focus on them. The fact that the manuscript contains 3 tables and 12 figures (+10 supplemental figures) clearly reflects this issue!! Authors should select a few central figures and tables, and move the ones presenting secondary information to the supplemental document, but overall, the number of figures should be drastically reduced (main and supplementary). Accordingly, the structure of the discussion, the abstract, and the conclusion should be adapted to put the focus on the main findings.

Results presentation and discussion: The structure of the discussion is confusing. For instance, the first section named “Biophysical drivers” also outlines spatial and temporal trends, and the CH<sub>4</sub> drivers is also discussed in subsequent sections. Following previous comments, authors should find a more logical structure for discussing results. In general, the literature context for discussing the results can be improved, as the authors make little comparison with results from previous similar studies. Presentation of the results, especially in figures, should be streamlined as there is a lot of repetition.

The revised manuscript will focus on the results the three referees highlighted and the authors agree are the most important. These include the observed spring burst of FCH<sub>4</sub>, the role of sediment T, precipitation, and chl<sub>a</sub> in driving the spring burst, and the difference between methods in capturing drivers and in upscaling.

Planned changes to the results and discussion include:

3. Expanding the results section describing the warm-season and annual budgets to compare budgets from different methods
4. Clarification of how and why we use FCH<sub>4</sub> results from the different monitoring methods to interpret different FCH<sub>4</sub> phenomenon
5. Deemphasizing and reducing the discussion of FCH<sub>4</sub> diurnal patterns and intra-reservoir spatial patterns
6. Moving five figures (current Figures 4, 5, 6, 7, and 12) from the main manuscript to the supplement
7. Restructuring the discussion to directly address upscaling implications based on the study results. Instead of breaking the discussion into sections that still overlap (4.1: Biophysical drivers, 4.2: Temporal patterns, 4.3: Comparison with other systems and methods), we plan to structure the discussion to answer our guiding questions:
  - 7.1. Comparison with other systems and methods
  - 7.2. Implications for upscaling

Specific Comments:

- Line numbering should be continuous, not restarting on each new page. Thank you for pointing this out; we will use continuous line numbering in our revised manuscript.
- Page 2 line 17: “in space in time” replace by and OK
- Page 9 line 3: “elevated are positively” replace by and OK
- Page 9 line 11-12: “The period...smaller median” this sentence could be simplified as follows: ...if 1) the difference between daytime vs nighttime FCH4 median was >50 %. OK
- Section 3.1, the title of this section could be replaced by “Temporal patterns in FCH4” since it does not only focus on seasonal trends. We agree and will make this change in the revision
- Page 9 line lines 22-26: the two sentences are repetitive and can be combined into one. OK
- Page 10 line 2 “in contrast...” and line 11-13 “This difference...”, page 11 line 2-3 “Much of this behaviour...” statements like these belong in the discussion section. This is helpful for guiding the process of streamlining the results and discussion in the revision.
- Page 10 line 20-25: Was there any investigation done concerning the CH4 drivers on a day to day scale? It seems like an important component if looking at drivers at different temporal scales. We did not directly investigate FCH4 drivers on a day-to-day scale. This scale is inherently part of the ANN gap filling model. We emphasized seasonal, interannual, and diurnal time scales because of the potential impact of biased upscaled estimates. We would expect day-to-day variability to be more stochastic.
- The first paragraph of section 3.3 belongs to the method section. The second paragraph of this section could be moved to section 3.1 as it relates to the temporal measurements and drivers of CH4. Also, main drivers of CH4 derived from the ANN analysis should be mentioned in this result section rather than just referring to the figure. We agree and will make these changes as part of streamlining the results and discussion in the revision.
- Section 3.4 should be restructured to present the overall budgets from different methods and comparing them before discussing the differences between years which was already discussed in section 3.1. We plan to rearrange the results section to present the overall budgets first, and expand this section to compare budgets from different methods
- The first paragraph of section 4.1 mostly contains information that belong in the method and results sections. We agree and will make this change in the revision.
- Page 13 line 11-12 “Our analysis...” authors should be careful with this statement as they have not performed an analysis that specifically support that statement. The cited figures are only visual aids but do not include any statistical testing of this hypothesis. We will rephrase this to clarify that while supported by observations, the connection from precipitation to algal biomass to FCH4 is not unequivocal.
- Section 4.2.1: here the authors should include a wider range of literature studies linking CH4 to Chla at global spatial scales, in several temporal studies, and discussing its known link to pelagic oxic methane production. This section discussing drivers of the 2018 spring burst will undergo substantial revision. We will add discussion related to the following recent studies demonstrating links between chla and FCH4: Zhang et al., 2021; Bartosiewicz et al., 2021; McClure et al., 2020. Will also add discussion the potential importance of pelagic oxic methane production, citing Hartmann et al 2020.
-

- Section 4.2.2: When talking about diurnal CH<sub>4</sub> drivers, authors mention that nondiurnal factors may contribute to the variability in CH<sub>4</sub>. While these other factors may influence CH<sub>4</sub> on different temporal scales, by definition, they do not affect its diurnal variability. Thus, I do not see the point in mentioning them when talking about diurnal variability, and the authors should hypothesize another explanation for this. **We would argue that this analysis is an important contribution toward understanding the role of diurnal patterns in emissions in lentic systems, and whether FCH<sub>4</sub> magnitudes tend to be higher or lower during the day. Thus, the lack of diurnal pattern and potential reasons behind that is just as important a result as observations of strong diurnal patterns. We plan to condense the discussion of FCH<sub>4</sub> diurnal patterns. We plan to touch on these findings in brief as part of the implications of our findings on upscaling.**
- Authors do not discuss the limitations and potential biases of the EC method compared to other techniques, and do not discuss the reasons behind a more elevated flux when using this method. This should be addressed. **We plan to address these items in more depth by expanding the comparison of methods in the results.**

## Response to referee comment #2

General Comments: I agree with reviewer #1 on the high potential of this well conducted study on CH<sub>4</sub> emissions from a temperate eutrophic reservoir which includes 2 years of continuous monitoring of total CH<sub>4</sub> emissions by eddy covariance (EC) and gap-filling with ANN and ebullition with automated bubble traps at shallow and deep sites and six extensive field surveys during which diffusion (floating chambers) and ebullition (manual bubble traps) were measured at more than 10 sites. The interpretation on the spatial and temporal variability of CH<sub>4</sub> emissions can be done on the basis of meteorology (Rainfall, temp, atmospheric pressure), energy balance (H, LE), hydrodynamics (Brunt-vaiala Freq, temp profiles), hydrology (water inputs, water levels) and biogeochemistry (O<sub>2</sub>, Chloa). **Thank you for acknowledging the quality of this study, and for your detailed and constructive comments.**

Major comments: My first major comment is about the result section which does not depict the whole dataset. Indeed, only CH<sub>4</sub> fluxes are described but not correctly (see below). Information on meteorology and hydrology would be very welcomed. Description of the energy balance, thermal stratification and its spatial variability, vertical biogeochemical stratification (O<sub>2</sub>, CH<sub>4</sub>...) and their spatial variability and chlorophyll a data and its spatial variability are required.

**This comment speaks to the tension between focus and thoroughness in a manuscript. We provide key information on meteorology and hydrology in results section 3.1, which are depicted in Figure 3 (temperature, LE and H, precip, stream inflow, water level, the Brunt-Vaisala frequency, and water temperature profile). Information on vertical stratification of pCH<sub>4</sub> is provided in the supplement (Figure S4). Estimating the energy balance over open water is challenging because of the high degree of uncertainty in the storage term. In contrast to terrestrial systems, the energy balance would have limited utility in diagnosing the quality of**

the EC measurements in our study. Similarly, while there are some limited data we could add about the spatial variability in dissolved nutrients and chlorophyll a, it would need to directly contribute to the main findings of the manuscript.

For CH<sub>4</sub> emissions, I would recommend to separately describe ebullition (funnels, bubble traps), diffusion (floating chambers) and total emissions from EC. We plan to expand the results section describing the warm-season and annual budgets to compare budgets from different methods.

As a matter of fact, I wonder whether the gap-filling is not already a kind of interpretation as the gap-filling is based on the covariation of the fluxes with other variables when EC data are available. Therefore, it has to be decided by the authors to keep it in the result section or move it to the discussion. Independently of where the gap-filled fluxes are described (results or interpretation), it would be very informative for the reader to have information on the validated fluxes (“real data”) and on the EC fluxes after gap filling for comparison. It is true that the gap-filled EC flux dataset is dependent on driver variables. For this reason, we only use the directly measured/non gap-filled EC data in the diurnal analysis, and the ecoQ10 and 2DKS analysis relating FCH<sub>4</sub> to sediment T. We realize this is not clearly explained in the manuscript and will clarify this point in the revision. For interpreting overall patterns in FCH<sub>4</sub>, and CH<sub>4</sub> budgets, it is better to use the gap-filled dataset, as it mitigates any bias due to data coverage.

The second major comment is related to the absence of information regarding the calculation of total emissions from the reservoir. A critical discussion on the comparison of the different type of measurements is required in order to determine the adequate methodology to combine them for a robust estimation of total emissions. We currently ignore whether the emission factor given in the manuscript is an average of all measurements, whether it is only based on EC... Did the author take into account the bathymetry for the extrapolation of ebullition from the reservoir since ebullition at deep sites is lower than at shallow sites? We agree with this comment and plan to address this in the revision by expanding the results section that describes the budget from different methods, and adding a discussion of our assumptions in estimating total reservoir emissions.

#### Minor comments

- Throughout the manuscript: Does “Static pressure” depict atmospheric pressure or the sum of atmospheric and hydrostatic pressure? The sum of atmospheric and hydrostatic pressure. We will specify this where we introduce static pressure as a driver of the ANN in the methods.
- Did the author explore the role of hydrostatic pressure (water level and their variations) on CH<sub>4</sub> emissions? Yes, as noted above, hydrostatic pressure was included as a component of static pressure.
- Did the authors attempt to decipher diffusive fluxes and ebullition from the EC dataset (at least when they have concomitant surface concentrations and or chamber measurements with EC measurements)? We used the results from the inverted funnel

and chamber measurements to characterize the relative importance of these two main emission pathways (Figure 7) and found that ebullition typically accounted for > 75% of total emissions. Deciphering between the two pathways in the EC dataset based on these measurements has limited value given the high level of spatial variability. There are a few studies that use wavelet analysis to partition CH<sub>4</sub> fluxes into diffusive and ebullitive is an emerging technique (see Iwata et al., 2018; Taoka et al., 2020), but it is outside the scope of this study to apply their novel method.

- As the manuscript require substantial rewriting/reorganization in order to properly present the dataset and better focus on key results in the discussion no detail comments are provided. Thank you for serving as a referee. We hope you will provide comments on the revised manuscript.

Response to referee comment #3:

### General Comments

This paper deals with methane emissions in a small temperate eutrophic lake. Emissions were assessed from a variety of measurement techniques (floating chambers, submerged funnels and eddy covariance) together with some environmental parameters (sediment temperature, atmospheric pressure, heat fluxes, met data...) and a neural network (ANN) approach. The paper discusses the links between CH<sub>4</sub> fluxes and the biophysical parameters, as well as it provides an analysis of the temporal and spatial variability of those emissions. The subject is of great interest since methane emissions from reservoirs are still poorly studied and constrained at the global scale. There are very few eddy covariance-based studies with long series (2 years) as presented here. As stated before by reviewers #1 and 2, There is no doubt that the data base gathered here is worth publication in the Biogeosciences journal. Some rearrangements would be welcome before publication. Thank you for acknowledging the value of our study, and for your helpful comments.

One of the most striking results presented here is the difference between 2017 and 2018 seasonality and cumulated emissions. Unfortunately, though well argued, there are no direct measurements of nutrients and carbon (TOC, DOC, POC, quality of OM) to support these assumptions. We must disagree with this comment. We used direct measurements of the chl<sub>a</sub> concentration (e.g., Figure 11 and discussion in Section 4.2.1), which is a strong indicator for algal biomass and a widely used proxy for reservoir productivity. We do plan to revise the discussion around the spring burst away from speculating about the potential role of autochthonous C vs. allochthonous C.

Discussion on the diurnal patterns is also a bit disappointing since the results are not unequivocal. As stated above in response to RC1, our finding of dynamic diurnal patterns is an important contribution toward understanding the role of diurnal patterns in emissions in lentic

systems, and whether FCH<sub>4</sub> magnitudes tend to be higher or lower during the day. Thus, the lack of diurnal pattern and potential reasons behind that is just as important a result as observations of strong diurnal patterns. We plan to condense the discussion of FCH<sub>4</sub> diurnal patterns. We plan to touch on these findings in brief as part of the implications of our findings on upscaling.

Authors should focus the paper on the main findings which can be supported by the data provided in the paper, and subsequently, present figures might be a little bit too numerous in that perspective of a more focused paper. This is a recurring theme in the RCs, and as stated above we plan to focus the paper in the revision and move five figures (current Figures 4, 5, 6, 7, and 12) to the supplement.

The end of the abstract is mentioning "...there is a trade-off in intensive measurement of one water body versus short-term and/or spatially limited measurements in many water bodies", and also "The insights from multi-year, continuous, spatially extensive studies like this one can be used to inform both the study design and emission upscaling from spatially or temporally limited results". These statements are indeed interesting and I wish the paper would give clearer insights and develop more on this matter in the discussion and conclusion. We appreciate that you highlighted this section of the abstract. As stated above, we plan to directly address the difference between methods and the implications for upscaling in the revision.

Rearrangements suggested by Rev 1 and 2 would improve the paper a lot since results and discussion are all mixed together at the moment. I am particularly sensitive to the place devoted to ANN gap-filling and on the way it impacts final emission numbers.

Minor comments:

- Page 4, line13: How was used time-lapse camera in this study? The time lapse camera was used to identify periods of ice-cover. We will add this information to this section of the methods.
- Page 4, line 27: there were no  $u^*$  filtering at EC-S1? If so, you should argue why We did not use  $u^*$  filtering at EC-S1 due to insufficient temporal coverage to determine the  $u^*$  threshold. We will clarify this in the revision.
- Page 5, line 33: more details are needed on the way Akaike information criterion (AIC) was used to determine fitting rate of change in the chambers. See below
- Page 6, line 10-11: vertical profile were done manually, detail procedure( how long for each level) See below
- Page 6, line 30: give more details about:" a probability design that has been shown to reduce uncertainty relative..." See below

These three comments highlight the tension in finding a balance between including adequate details in the methods and streamlining the manuscript. We can expand these sections somewhat (for example, clarifying the connection between the spatially balanced probabilistic survey and the survey sites located near the swimming beach),

but we do provide the relevant references to publications with more details on these methods.

- Page 9, line 26: you should give the information that “both quantitative analyses of the relationship between FCH4 and SedT yielded statistically significant results” before implying a link between those two parameters in lines 22-24 **OK**
- Page 11, line 3: I understand that the sandy substrate mention here was brought for recreation use (beach). Is there any point to measure fluxes at the very specific place? **Yes, the probabilistic GRTS design is a hybrid between a random and gridded design. Their inclusion in the survey sites reflects our effort to characterize reservoir-wide emissions.**
- Page 11, lines 23-24: comment on absolute and relative importance of each factor **The variable importance factors are ranked in terms of their % importance. I’m not clear on the distinction between relative and absolute importance in this context.**
- Page 11, lines 28, 29 and 30: table 3 instead of table 2 **OK**
- Page 13, line 4-5: any assessment of the mentioned transfer? **The transfer in question is the transfer of heat to the deeper sediment and nutrient transfer to the deeper site, in their impact on the phase shift in FCH4 and sedT at the shallow and deep sites. The heat transfer is well documented by direct measurements. The nutrient transfer is more speculative and the reference to this will be removed.**
- Page 13, line 21: any nutrients data to support the suggestion mentioned here? **As mentioned above, the chl<sub>a</sub> measurements are a strong indicator of algal biomass.**
- Page 13, line 26-27: any measurement of residence time and output/input of C to support this? **This section of the discussion on the role of autoOC and alloOC will be reduced in the revision.**
- Page 14, line 2: is this consistent with kinetic found by Grasset et al, 2018? **This section of the discussion on the role of autoOC and alloOC will be reduced in the revision.**
- Page 14, line 28: pattern and patterning instead of patter and pattering **OK**
- Page 15, line 32: detail input parameter of the model used **OK**
- Page 15, line 33: Del Sontro et al 2018 ref missing or is this Del Sontro et al 2016? **2018, will add the reference in the revision.**

## References:

Bartosiewicz, Maciej , Roxane Maranger, Anna Przytulska, Isabelle Laurion, Effects of phytoplankton blooms on fluxes and emissions of greenhouse gases in a eutrophic lake, *Water Research*, 10.1016/j.watres.2021.116985, 196, (116985), (2021).

Grasset, C., Mendonca, R., Villamor Saucedo, G., Bastviken, D., Roland, F., & Sobek, S. (2018). Large but variable methane production in anoxic freshwater sediment upon addition of allochthonous and autochthonous organic matter: Methanogenic potential of different OC types. *Limnology and Oceanography*, 63(4), 1488–1501. <https://doi.org/10.1002/lno.10786>

Hartmann, J. F., M. Gunthel, T. Klitzsch, G. Kirillin, H. P. Grossart, F. Keppler, and M. Isenbeck-Schroter. 2020. High Spatiotemporal Dynamics of Methane Production and Emission in Oxic Surface Water. *Environmental Science & Technology* **54**:1451-1463.

McClure, R.P., M. E. Lofton, S. Chen, K. M. Krueger, J. C. Little, C. C. Carey, The Magnitude and Drivers of Methane Ebullition and Diffusion Vary on a Longitudinal Gradient in a Small Freshwater Reservoir, *Journal of Geophysical Research: Biogeosciences*, 10.1029/2019JG005205, 125, 3, (2020).

Zhang, Lei, Cheng Liu, Kai He, Qiushi Shen, Jicheng Zhong, Dramatic temporal variations in methane levels in black bloom prone areas of a shallow eutrophic lake, *Science of The Total Environment*, 10.1016/j.scitotenv.2020.144868, 767, (144868), (2021).