

**Author response to reviewer and associate editor comments on
“A cross-site comparison of ecosystem- and plot-scale methane
fluxes from wetlands and uplands” by Määttä et al.**

<https://doi.org/10.5194/egusphere-2025-5023>

**Please find our responses to each reviewer and associate editor comment below
in bold.**

Response to comments by associate editor

Dear Authors,

Thank you for your thoughtful responses to the reviewers' comments and suggestions.

Both reviewers acknowledged the significance of your cross-site comparison of two major methane flux methods and its implications for synthesizing a growing body of multi-scale methane flux data. While the overall evaluation is positive, the reviewers have identified several areas for improvement as well as a couple of important methodological concerns: the treatment of ebullition data and inadequate statistical analyses (e.g., non-normal data transformation).

With regard to your response to the ebullition data issue, I wondered if you could present the potential differences in chamber flux estimates with and without ebullition data. Although you can discuss the limitations of removing ebullition data without presenting additional tables or figures, this meta-analysis manuscript would benefit from an actual data presentation. This is merely a suggestion, as I am unaware of your data availability; feel free to decide whether to include this comparison.

Based on the referee reports, I recommend "Reconsider after major revisions." Upon receiving the revised manuscript, I may ask the original reviewers or new ones to assess whether the revisions have adequately addressed the reviewers' concerns.

When submitting the revised manuscript, please ensure that all changes are clearly shown in the track-changed version. Additionally, please include a point-by-point response to the reviewers' comments, specifying the line numbers of the revised text where applicable.

We look forward to receiving your revised manuscript.

Sincerely,

Ji-Hyung Park
Associate Editor, Biogeosciences

Dear Dr. Park,

We thank you for your careful evaluation and constructive guidance for our manuscript. We have substantially improved our manuscript according to your and the reviewers' comments.

In response to these comments, we have improved the manuscript for clarity and strengthened the justification for the statistical methods and interpretation of the results, and addressed the methodological concerns raised. The detailed point-by-point author response is shown below for each reviewer with line numbers from the revised manuscript.

Regarding your suggestion to analyze chamber CH₄ flux data with and without ebullition, we agree that this would be a valuable addition to our manuscript. However, the chamber CH₄ flux datasets were processed by the individual data providers and we do not have access to chamber CH₄ flux data with and without ebullition. Therefore, we were unable to conduct these additional analyses. As we have also described in our response to Reviewer 1 and further clarified in the revised manuscript, the main goal of this manuscript was to compare the CH₄ flux data commonly used by researchers when combining ecosystem- and plot-scale CH₄ flux data, where particularly the plot-scale data are commonly processed differently between data providers (e.g., Jentsch et al., 2025). We thus consider this chamber CH₄ flux data, where some sites excluded ebullition events while others did not, to reflect this common uncertainty when combining cross-scale CH₄ flux data across sites and to be appropriate for the aims of our manuscript. However, we have added a sentence describing this as a possible future research topic that might provide valuable insights into these cross-scale CH₄ flux differences (section 4.6, lines 788-792 in the revised manuscript):

“Given that our results indicated ebullition removal from some of the chamber FCH₄ data as one potential driver of Δ FCH₄, future studies could also conduct cross-scale FCH₄ comparisons based on chamber FCH₄ data with ebullition events both included and excluded from a variety of wetland types. Ebullition events are sometimes also removed from EC FCH₄ data following the standard data quality protocols and further standardization of EC-based ebullition measurements are needed.”

In addition to the reviewer comments received through Biogeosciences, our manuscript also underwent a peer review for USGS (required because of one of the coauthors' affiliation). We implemented these comments (related to site-specific differences in CH₄-relevant biogeochemistry and vegetation, as well as general language and clarifications of ecosystem-scale CH₄ flux increase in relation to plot-scale CH₄ flux, see e.g., Table 2 and lines 53-56 in the abstract) in the revised manuscript and highlighted them in green color in the tracked changes version.

We have also fixed some typos in the text and the figures, added clarifying sentences to improve the readability, as well as defined abbreviations in the figure

and table captions that were not previously described. We also added new acknowledgements that were missing from the original submission, and a link to the R scripts used for processing and analyzing the data (listed under “Code and data availability”).

In the revised manuscript with tracked changes, we have highlighted changes addressing Reviewer 1 comments with blue and Reviewer 2 comments with yellow color. Changes addressing USGS reviewer comments are in green. Tracked changes without highlighted colors were additional edits done by the authors.

We thank you again for the constructive feedback and for considering our manuscript for publication.

On behalf of all authors,
Tiia Määttä

Response to comments by Reviewer 1

<https://doi.org/10.5194/egusphere-2025-5023-RC1>

This manuscript presents a valuable cross-site comparison of methane fluxes measured by eddy covariance (EC) and chamber methods across 10 wetland and upland sites. The work addresses an important gap in understanding how these two common measurement approaches compare across multiple sites and temporal scales. The dataset is substantial, the statistical analyses are appropriate, and the findings have significant implications for combining multi-scale flux data in syntheses and modeling studies. The multi-site and temporal scales approach, practical recommendations and transparency about limitations are significant strengths.

We thank the reviewer for the careful and encouraging assessment of our manuscript. Their feedback has helped improve the manuscript considerably. Below, we address each point raised by the reviewer.

However, the manuscript would benefit from major revision to improve clarity, address methodological concerns, and better contextualize the findings.

Key issues to resolve:

1. Ebullition circular reasoning between methods and discussion
We have now clarified this in the manuscript and added multiple sentences to the main text (see details below).
2. Lack of wetland vs upland comparison despite title emphasis
We have changed the manuscript title to exclude mention of wetlands and uplands (see details below).
3. Figure and table readability and formatting (axis overlap, interpretive captions)
We have improved the figures and tables as suggested by the reviewer (see details below).
4. General language review for sentence structures, removing redundancies and making the text more concise and cohesive
We have improved the language of the manuscript as suggested by the reviewer (see details below).

We address these four issues as detailed in the sections below.

Section-specific comments and key issues:

Introduction

Line 90-91: “Thus, chambers provide a greater spatial site-level representation than EC sites and are needed to fill the missing data gaps.” – contradicting your statement about area representation in lines 80-81

We see the original wording may have been a bit misleading (sounding like chamber measurements would represent the sites better than EC, even though chambers cover a small area). We have now clarified this sentence to highlight that sites with chamber

measurements cover more global area than EC sites, which is why chamber data are needed to fill the gaps in EC data coverage:

“Thus, sites with chambers provide a greater global measurement coverage than EC sites and are needed to fill the missing data gaps.” (line 91 in the revised manuscript)

Line 99: “Plot and ecosystem-scale FCH₄...” : You mention in your line 78 that different ecosystem and hydrological subtypes within the EC footprint are defined as “plots” (see also line 80), but here it seems to me that by “plots” you mean chamber studies and chamber plots? To avoid confusion, consider re-wording/defining.

We see that the wording may have been a bit confusing, and we indeed mean chamber measurements by the “plot-scale” measurements (though they do represent different ecosystem and hydrological subtypes). We have moved the definition of the plot-scale from line 80 to the next sentence starting at line 80 and ending at line 82:

“At the ecosystem subtype scale (i.e., “plot scale”), chamber measurements represent fixed sampling points with well-defined spatial location but limited areal extent.” (lines 79-81 in the revised manuscript)

Line 113: “...than the true, mean ecosystem-scale FCH₄”: this insinuates that the EC ecosystem-scale measurement is the “true” FCH₄, but in reality, we do not know this for sure, as EC also comes with its own uncertainties.

Agreed that EC CH₄ flux estimates can have positive or negative bias and therefore not represent the ‘true’ mean. We have removed the word “true” from the sentence:

“However, chambers capture a small portion of the landscape, are often placed in high-emitting hotspots, do not sample over tall vegetation patches, and may incorporate sampling location biases (Bansal et al., 2023b), leading to higher observed fluxes at the individual sampled plots, than the mean ecosystem-scale FCH₄ as measured by EC (but see Voigt et al., 2023).” (lines 110-113 in the revised manuscript)

In the introduction, you could also discuss the different chamber systems used in GHG studies, i.e. manual sampling with gas chromatography analysis, or in-situ measurements with portable gas analyzers. This distinction affects comparability and should also be discussed in the limitations section.

Agreed. We have added sentences discussing this in the introduction and discussion (sections 4.5 and 4.6):

Introduction: “Chamber and EC FCH₄ measurements also contain different uncertainties due to varying methods for measuring gas concentration in the chamber measurement techniques (e.g., gas chromatography vs high-precision CH₄ analyzers) and different EC and chamber instrument makes and models (Christiansen et al., 2015; Peltola et al., 2014; Pihlatie et al., 2013).” (lines 98-100 in the revised manuscript)

Section 4.5:

“The specific EC CH₄ analyzers can also differ in signal noise (Peltola et al. 2014).” (line 726-727 in the revised manuscript)

“It may be valuable to compare chamber and EC FCH₄ using both linear and exponential fits for chamber FCH₄ (from both high-precision CH₄ analyzers and gas chromatography) to better understand ΔFCH₄ trends across sites.” (lines 758-760 in the revised manuscript)

Section 4.6:

“Thus, in order to produce results that would be better generalizable to other sites and regions (e.g., tropics), future studies could include more sites from a variety of climates, dominant vegetation types, and chamber measurement systems (i.e., automated and manual, gas chromatography and high-precision CH₄ analyzers) (n>3 sites per group to allow statistical inference).” (lines 763-766 in the revised manuscript)

“In addition, our cross-scale FCH₄ comparisons may contain large uncertainties due to differences in chamber FCH₄ outlier removal (Table C2), design and the gas analyzer used (Table C1) (Jentsch et al., 2025; Levy et al., 2011; Pihlatie et al., 2013; Pumpanen et al., 2004).” (lines 785-787 in the revised manuscript)

New reference added:

Pumpanen, J., Kolari, P., Ilvesniemi, H., Minkkinen, K., Vesala, T., Niinistö, S., Lohila, A., Larmola, T., Morero, M., Pihlatie, M., and Janssens, I.: Comparison of different chamber techniques for measuring soil CO₂ efflux, *Agric. For. Meteorol.*, 123, 159–176, <https://doi.org/10.1016/j.agrformet.2003.12.001>, 2004.

Methods:

Key issue: Inconsistencies in ebullition treatment

Most sites removed ebullition events from chamber data (Text A1). However, ebullition is mentioned repeatedly as potentially explaining EC > chamber fluxes, and the authors suggest EC captures ebullition better than chambers (lines 477-482, 491-494, 497-499, 569-570). This is circular reasoning: chambers are processed to remove ebullition, then the authors express surprise that EC is higher. Recommendation: Clarify whether ebullition removal from chamber data is appropriate for this comparison or acknowledge it as a limitation. Discuss implications: if ebullition is real and important, should it be removed from chambers when comparing to EC?

This is a very important point, and we agree. The goal was indeed to compare the ecosystem- and plot-scale CH₄ fluxes in the form they are often used in syntheses

(i.e., gap-gilled EC CH₄ fluxes and chamber CH₄ fluxes filtered differently by different data providers), to highlight that these two data streams should be combined carefully if used with non-standardized chamber CH₄ flux data especially. We have added a clarification of this goal to the end of the introduction:

“To achieve this, we utilized FCH₄ data commonly used by the FCH₄ community, i.e., gap-filled EC data and chamber data quality-controlled in different ways by data providers (ebullition events were removed in some datasets as is common for chamber FCH₄ data; Jentsch et al., 2025) (see 2.2.1 and Table C2).” (lines 134-136 in the revised manuscript)

We also now highlight the decision of including chamber data with ebullition events filtered out in the methods section: “The decision to utilize chamber FCH₄ data with differing ebullition removal protocols across data providers was intended to reflect the non-standardized and site-specific data processing typical of chamber FCH₄ data (Jentsch et al., 2025), which may contribute to differences between bulk ecosystem- and plot-scale FCH₄ estimates.”

We also now discuss the influence of ebullition on plot-scale CH₄ fluxes and cross-scale differences and its implications in more detail in the discussion:

End of last paragraph in section 4.1:

“Altogether, the mismatch in EC footprint and chamber measurement coverage, as well as chamber CH₄ ebullition removal, could be important Δ FCH₄ drivers, as FCH₄ can vary strongly between surface cover types and within them even during the same growing season (Voigt et al., 2023). This highlights the need to account for EC and chamber footprint representativeness as well as chamber data quality control when combining plot- and ecosystem-scale FCH₄ data, particularly at high-FCH₄ sites and periods (Fig. 4).”

(lines 595-599 in the revised manuscript)

First paragraph of section 4.2:

“As ebullition events are often removed from chamber FCH₄ data, these results highlight that the large variation in chamber FCH₄ data processing protocols between researchers could also increase Δ FCH₄ and thus uncertainty in multi-site syntheses combining cross-scale FCH₄ data, at least in the sites included in this study (e.g., Jentsch et al., 2025; Levy et al., 2011).”

(lines 607-610 in the revised manuscript)

Third paragraph of section 4.5:

“As previously discussed (see 4.1 and 4.2), plot-scale FCH₄ could have been generally underestimated due to the removal of ebullition events from some of the chamber FCH₄ data (Table C2), calling for standardization of chamber-based ebullition measurements and data processing (Jentsch et al., 2025).”

(lines 744-747 in the revised manuscript)

Section 4.6:

“Given that our results indicated ebullition removal from some of the chamber FCH₄ data as one potential driver of Δ FCH₄, future studies could also conduct cross-scale FCH₄ comparisons based on chamber FCH₄ data with ebullition events both included and excluded from a variety of wetland types. Ebullition events are sometimes also removed from EC FCH₄ data following the standard data quality protocols and further standardization of EC-based ebullition measurements are needed.”

(lines 788-792 in the revised manuscript)

Conclusions:

“We attribute the higher ecosystem-scale FCH₄ than plot-scale FCH₄ mainly to the combination of selective chamber placement, ebullition removal from chamber FCH₄ data, and the spatio-temporal dynamics of the EC footprint which may have captured CH₄ emission events that were not detected by chambers.”

(lines 804-806 in the revised manuscript)

“● Standardized protocols for chamber FCH₄ data quality control, especially related to ebullition removal (see Jentzsch et al., 2025 for recent recommendations for chamber FCH₄ data processing), and accounting for these differences when combining chamber and EC FCH₄ data”

(lines 818-820 in the revised manuscript)

Table 1: I wonder if Table 1 could be better presented in the final publication? Perhaps landscape orientation would make the column titles and words not be cut off.

Agreed- we have changed the table position to landscape in the revised version.

Line 215: I'd like more information about seasonality of measurements and when they were conducted. I see this in supplementary material (Figure B1), but I think this is important information for the main text. Perhaps integrate to Table 1?

We have added monthly coverage ranges to each site in Table 1 where we also refer the reader to Figure B1 for more details in the caption: “Month coverage shows the range of months covered across years per site (see details in Fig. B1).”

Results:

Line 310: What do you mean by large CV? Please quantify.

We referred to the CVs reported in Table 2, but did not specify that in the sentence. We have now included the CV values in the sentence together with a reference to Table 2:

“However, the coefficient of variation (CV, %) for Δ FCH₄ was large, particularly in daily (674%) and weekly (467%) aggregations (Table 2).”

(lines 393-394 in the revised manuscript)

Lines 313-315: I think it is good practice to put units behind all numbers, even if you are repeating them/listing them.

The lack of units here was an accident. We have now added the units after these values: “...where median ΔFCH_4 ranged between $0.28 \text{ nmol m}^{-2} \text{ s}^{-1}$ (annual) and $1.23 \text{ nmol m}^{-2} \text{ s}^{-1}$ (half-hourly)...” (line 398 in the revised manuscript)

Figure 3:

Caption reads like interpreting results rather than figure description. I would say that figure captions should describe content, not provide conclusions.

Add R^2 and RMSE in addition to Spearman's ρ – they are more relevant for linear regression.

The labels on both axes could be more sparse (they nearly overlap and are hard to read) and x axis labels could be written at an angle for better readability.

Try to keep only one Obs. count legend.

We clarify that this figure is not showing linear regression but Spearman correlation coefficients due to the non-normal data. Linear regressions in this case would have required data transformations, and even then, proper modeling would have required linear mixed models, which would have made model visualization unnecessarily complicated for this figure. Thus, we propose not to add linear regressions and their corresponding R^2 values to this figure. However, we have added normalized RMSE values to the figure, as suggested, and explain our decision in the methods section:

“As the data were non-normally distributed and did not meet the assumptions of linear regression, we also used Spearman correlations together with normalized root mean square error (using the standard deviation of pooled EC and chamber FCH_4 as the denominator at each temporal scale) to assess the direction and strength of the relationship between EC FCH_4 and chamber FCH_4 , manual and automated chamber FCH_4 , as well as FCH_4 magnitude (row-wise mean of EC and chamber FCH_4) and absolute ΔFCH_4 .”

(lines 317-321 in the revised manuscript)

We have changed the labels to be more sparse and x-axis at an angle for better readability. We have also combined the number of observations legends for a and b plots.

We have revised the figure caption. We agree that we did not include a direct descriptive statement in the caption, and that was added. Nonetheless, we feel that there is some benefit in providing a succinct summary of the discussion in the caption, interpreting the figure, i.e., answering: why are we showing this figure? What is the take home message we want the reader to internalize? We did move the interpretive statements to the end of the caption, and clarified the descriptive ones at the start.

The revised caption is:

“Results of correlation test (Spearman rank correlation coefficient, ρ , its significance level, p , and the normalized root mean square error, NRMSE) between plot-scale (chamber) methane (CH_4) flux (FCH_4) and ecosystem-scale (eddy covariance; EC) FCH_4 at half hourly (a), hourly (b), daily (c), weekly (d), monthly (e), and annual scales (f). For visualization, the plot axes (a-f) were transformed with inverse hyperbolic sine to spread out points in the low FCH_4 range and retain negative values (see untransformed plots in Fig. B3). Spearman ρ was calculated with untransformed data. NRMSE was calculated by dividing RMSE by the standard deviation of untransformed ecosystem- and plot-scale FCH_4 at each temporal aggregation. In a) and b) the points for half-hourly ($n=74482$) and hourly ($n=40072$) aggregations are shown in hexagonal density clouds with \log_{10} -transformed color range to highlight trends in high point density areas (colors represent number of observations per hexagon). Agreement between chamber and EC FCH_4 improves from finer to coarser temporal aggregations (a-f), as indicated by ρ . The high observation densities in a) and b) reveal site-specific trends in the discrepancy between ecosystem and plot scales (e.g., at $x=0$ and $y=5$). For daily (c), weekly (d), monthly (e), and annual (f) aggregations, sample sizes were $n = 1879, 349, 121,$ and $22,$ respectively. The dashed line represents 1:1 line.”
(lines 437-448 in the revised manuscript)

Figure 4:

Similar comment about the caption as Figure 3.

Add R^2 and RMSE in addition to Spearman's ρ – they are more relevant for linear regression.

Consider adding a regression line onto Figure 4, to show the deviation from $y=0$ (the blue) line.

As in our response to Reviewer 1 comment on Figure 3, linear regressions would have required additional data transformations and linear mixed modeling where model visualization would have become unnecessarily complicated for the purpose of this figure. Thus, we did not add linear regressions and R^2 values to this figure. However, as for Figure 3, we made this decision clearer in the methods section and added normalized RMSE (based on ΔFCH_4 and standard deviation of EC and chamber FCH_4 at each temporal scale) to each temporal aggregation plot. While we understand how adding linear regression lines would show deviation from $y=0$ more clearly, for the aforementioned reasons, we did not add them in this figure.

As with figure 3, we revised the caption to remove the interpretive sentences and added a descriptive sentence to the beginning of the caption:

“The relationship between methane (CH_4) flux (FCH_4) magnitude (FCH_{4_mean}) and absolute difference between ecosystem-scale (eddy covariance; EC) and plot-scale FCH_4 (ΔFCH_4) from half-hourly (a) to annual (f) scales, represented by Spearman correlation coefficient, (ρ), its significance, (p), and normalized root mean square error of ΔFCH_4 (NRMSE). FCH_{4_mean} is the row-wise mean of EC FCH_4 and chamber

FCH₄. In a) and b) half-hourly and hourly points are shown in hexagonal density clouds with a log-transformed color range to highlight trends in high point density areas (colors represent number of observations per hexagon). Plots c-f show daily, weekly, monthly and annual aggregations, respectively. The blue dashed line represents $\Delta\text{FCH}_4=0$ meaning complete agreement between ecosystem and plot-scale FCH₄. Higher Spearman correlation coefficient ($\alpha=0.05$) represents stronger deviation from $\Delta\text{FCH}_4=0$. NRMSE was calculated by dividing RMSE (of ΔFCH_4) by the standard deviation of ecosystem- and plot-scale FCH₄ at each temporal aggregation. For visualization, outliers were removed from daily (n=3), weekly (n=10), monthly (n=8) and annual (n=1) plots but the Spearman correlations and NRMSE are based on original data. See plots with outliers in Fig. B17 and a figure showing how high CH₄ emissions from ecosystem and plot scales contribute to annual CH₄ emissions per site in Fig. B2.”
(lines 472-483 in the revised manuscript)

Figure B6-B7: without reading the captions, there is no way to understand what these figures are illustrating.

As we think that including these figures in the Appendices is necessary (the results are referred to in the main text), we improved the readability of these figures by adding clear labels to the heatmap axes (“Hourly bins (from 0-1 to 23-24)”) and added descriptive titles (e.g., “Pairwise comparisons (Conover-Iman) in ΔFCH_4 between hours, Half-hourly aggregation”).

Table 3: Formatting needs improvement - currently difficult to follow. Try to make more compact.

Agreed- we made the table more compact with less space between rows.

Discussion

Key issue: The title of the manuscript emphasizes "wetlands and uplands" but there is minimal comparison between these ecosystem types in the results and discussion. Lines 592-595 briefly mention this, but given the title's prominence, this deserves substantially more attention. How do the ΔFCH_4 patterns differ between the two ecosystem types? Do different processes drive differences in these contrasting systems? If data are insufficient for robust comparison, consider revising the title.

This is a very good point. Originally, we wanted to highlight that the sites included both wetlands and uplands, but we now see this may be misleading in the title. Unfortunately, the number of upland (n=2) and wetland (n=8) sites did not allow for robust statistical comparisons in cross-scale FCH₄ differences between these ecosystem types. Thus, we hchanged the title to remove mention of wetlands and uplands: “A cross-site comparison of ecosystem- and plot-scale methane fluxes across multiple sites”.

Line 450: Avoid starting immediately with “contrary to our hypothesis.” Start by summarizing key findings, then contrast with hypothesis.

We rearranged this paragraph (and added a more general sentence of the main goals of the study in the beginning, as suggested by USGS peer reviewer) as follows:

“As a first step to reconcile the discrepancies in FCH₄ data obtained from ecosystem-scale EC and plot-scale chamber measurements used increasingly in combination in various syntheses and FCH₄ modeling, we explored the cross-scale differences across ten sites and six temporal aggregations. Across all temporal scales, ecosystem-scale (EC) FCH₄ was higher than at the plot scale (chamber). Supporting these results, higher EC FCH₄ than chamber FCH₄ have been observed in an arctic peatland with area-weighted chamber FCH₄ (Budishchev et al., 2014), a managed peat meadow with upscaled chamber FCH₄ (Schrier-Uijl et al., 2010), a peatland with down-scaled EC FCH₄ (Forbrich et al., 2011), a temperate forest with spatial chamber FCH₄ averages (Wang et al., 2013), and a temperate salt marsh with spatio-temporal chamber and EC FCH₄ averages (Hill and Vargas, 2022b). Other studies at individual sites have observed higher chamber FCH₄...”

(lines 544-551 in the revised manuscript)

Lines 457-460: long sentence, consider splitting into two.

Fixed (please see the previous response).

Lines 463-474: This section could use some citations to back up your claims and relate to previous findings.

Agreed. We added new references to some of the arguments that are not based on our results, referred to Table 1 for our results and clarified the text:

“... The improved agreement is likely a result of the data aggregation, which reduces the influence of inter-daily FCH₄ variability and inflates correlation coefficients (e.g., Clark and Avery, 1976; Pollet et al., 2015). In addition, mean Δ FCH₄ in weekly, monthly and annual aggregations was negative (Table 1), indicating higher plot-scale than ecosystem-scale FCH₄, and the CV for the weekly aggregation in particular was large (467%) (Table 1). Our results suggest that high CH₄ emissions and FCH₄ variability in plot-scale measurements are associated with higher Δ FCH₄, particularly at time scales longer than daily (Table 2 and Fig. B2); suggesting that combining plot- and ecosystem-scale bulk FCH₄ at heterogeneous sites is particularly problematic at coarse temporal scales. However, footprint-aware comparisons between upscaled chamber or downscaled EC FCH₄ could show better agreement between ecosystem and plot scales (e.g., Schrier-Uijl et al., 2010) (see 4.6). Nonetheless, this highlights the practice of selective chamber placement on high-emitting locations and time periods within the study sites (Hill and Vargas 2022b; Vargas and Le 2023)...”

(lines 560-570 in the revised manuscript)

New references:

Clark, W. A. and Avery, K. L.: The effects of data aggregation in statistical analysis, *Geogr. Anal.*, 8, 428–438, <https://doi.org/10.1111/j.1538-4632.1976.tb00549.x>, 1976.

Pollet, T. V., Stulp, G., Henzi, S. P., and Barrett, L.: Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data, *Am. J. Primatol.*, 77, 727–740, <https://doi.org/10.1002/ajp.22405>, 2015.

Line 475 and onwards: however, you could also discuss here that chamber artifacts can increase chamber fluxes: e.g. disturbance of soil/water surface and thus causing a pulse (especially for manual sampling).

We have now added a sentence discussing this in this paragraph:

“Ebullition can also be triggered by chamber placement onto water or waterlogged soil surface, or by soil disturbance around the chamber (e.g., Jentzsch et al., 2025), but as ebullition events were removed from some of the chamber FCH₄ data (see 2.2.1 and Table C2), this was unlikely to contribute to the general ΔFCH₄ trends.”
(lines 577-580 in the revised manuscript)

Lines 481-485: long sentence, consider splitting into two.

We have now split this sentence into two and improved its readability:

“FCH₄ hot spots and hot moments can also vary in space and time, which manual chamber FCH₄ measurements (n=6 sites) may not capture due to sporadic daytime measurements in weekly or monthly intervals (Anthony and Silver, 2021, 2023; Vargas and Le, 2023). This may result in uncertainties in spatio-temporal FCH₄ and ΔFCH₄ variation across temporal scales (Anthony and Silver, 2021, 2023; Vargas and Le, 2023).”
(lines 583-586 in the revised manuscript)

Lines 485-486: Yes, but Meijide et al found the opposite trend – that chamber fluxes exceeded EC fluxes. Can you make direct comparisons with your results and how would your interpretation change?

This was included in the original manuscript version but removed to shorten the text for the final manuscript draft. However, as Meijide et al. studied only one site and we included multiple sites, we think this might just show that, whichever scale (ecosystem or plot) has higher FCH₄, the difference between them is amplified in high-emitting sites and periods, and could be influenced by the chamber placement within the sites (as discussed later in the text). However, the reasons behind these trends are still unclear, and more research is definitely needed to look into the dynamics behind this.

We added this discussion briefly to this section and the last sentence:

“...This highlights the need to account for EC and chamber footprint representativeness as well as chamber data quality control when combining plot- and ecosystem-scale FCH₄ data, particularly at high-FCH₄ sites and periods (Fig. 4).”
(lines 597-599 in the revised manuscript)

Related to this, we also added a new sentence clarifying that this needs to be studied more in the future in the first paragraph of section 4.4 where it is discussed in a bit more detail:

“However, further investigations into the exact mechanisms behind these trends are needed.”
(lines 701-702 in the revised manuscript)

Lines 486-489: So what may be the implications of this poor detection of differences at low fluxes? How may it compromise/influence your results?

We think this may mainly influence the interpretation of Fig. 4 (increasing Δ FCH₄ with increasing FCH₄ magnitude), as well as Δ FCH₄ in low-emitting sites (e.g., some upland land cover classes at US-Ho1). We have now added a sentence starting from line 489 (in the original manuscript) discussing this:

“Thus, the Δ FCH₄ trends in low FCH₄ (Fig. 4) and possibly low-emitting uplands (e.g., part of US-Ho1) should be interpreted with caution.”
(lines 594-595 in the revised manuscript)

Lines 489-492: again, consider splitting into two sentences

We split this sentence into two and emphasized the point of increasing Δ FCH₄ with increasing FCH₄ magnitude (see previous response to Lines 485-486 of original manuscript) in the end of the sentence:

“Altogether, the mismatch in EC footprint and chamber measurement coverage, as well as chamber CH₄ ebullition removal, could be important Δ FCH₄ drivers, as FCH₄ can vary strongly between surface cover types and within them even during the same growing season (Voigt et al., 2023). This highlights the need to account for EC and chamber footprint representativeness as well as chamber data quality control when combining plot- and ecosystem-scale FCH₄ data, particularly at high-FCH₄ sites and periods (Fig. 4).”
(lines 595-599 in the revised manuscript)

Line 496: consider if statistics such as p-values are relevant for discussion or if they should be kept only in results.

We removed the repeated p-values from the discussion and instead refer the reader to Table 3 for details.

Lines 495-499: This is a good thought about PA, however, you bring out that your results show decreasing ΔFCH_4 with higher PA at the weekly scale, which is not something relevant for ebullition events (very short term). I wonder if you could elaborate on this point.

We included discussion of decreases in PA as a possible trigger for ebullition in the weekly scale because a recent FCH_4 synthesis suggests that PA and FCH_4 are inversely related and possibly due to ebullition. This synthesis also observed similar trends at a similar temporal scale (2.7-21.3 days), which represents synoptic weather variations (based on wavelet transform time scale decomposition in Knox et al., 2021; <https://doi.org/10.1111/gcb.15661>). However, it is still possible that the effect of PA on ebullition is more relevant at finer temporal scales (i.e., diel and daily). We have now added a clarification of the multiday scale in this paragraph:

“PA is a strong predictor of daily and multiday (ca. 3-21 days) FCH_4 (Knox et al., 2021), and ΔFCH_4 decreased with higher PA (Table 3).”
(lines 602-603 in the revised manuscript)

Lines 501-502: but consider that low u^* values are typically removed from EC data if using standard protocols – how would this influence your results’ interpretation?

We changed the sentence starting from line 501 (in the original manuscript) to discuss this important topic:

“While EC FCH_4 can be underestimated in low u^* , leading to decreased ΔFCH_4 , EC FCH_4 under low u^* were filtered out by the FLUXNET- CH_4 team, so low u^* was unlikely to influence the observed ΔFCH_4 trends (Aubinet, 2008; Baldocchi, 2003; Knox et al., 2019; Delwiche et al., 2021).”
(lines 611-613 in the revised manuscript)

Line 504: change ‘may be’ to ‘was’. I think overall the word ‘may’ is used excessively in the discussion. Try not to use ‘may’ if you are referring to your own results: your results are certain, not ‘maybe’; the implications of your results and how they relate to previous findings, that is where the word ‘may’ comes in.

We changed this sentence as follows, and removed the excessive “may” words related to our results in the rest of the manuscript:

“The importance of plant activity is further supported by the marginally-significant TS (Table 3), a possible proxy for increased plant activity in the peak growing season months in the northern hemisphere (July and August; Table 3).”
(lines 617-619 in the revised manuscript)

Line 511: ‘As expected’ replace with ‘As hypothesized’, or similar

Changed to “As hypothesized,..” (line 625 in the revised manuscript)

Lines 517-519: You talk about soil CO₂ respiration hotspots and hot moments and then bring those same conclusions to explain CH₄ hotspots and moments. I think CO₂ respiration hotspots do not directly translate to CH₄ dynamics. You could try to find studies describing CH₄ hotspots/hot moments with chamber measurements; or better integrate the CO₂ ideas.

It is true that CO₂ dynamics do not directly translate to CH₄, but as other multi-site EC-chamber CH₄ flux comparison papers are currently lacking, we wanted to explore how similar comparisons have been affected by flux hot spots and moments. In this context, CO₂ has been studied the most, and similar results have been found for CO₂. Thus, we deem it reasonable to mention here. However, we agree that it is important to mention the differences in CH₄ and CO₂ hot spots and moments. Thus, we changed the sentence to better integrate the CO₂ thoughts to CH₄:

“While CH₄ cycling is driven by different controls than CO₂, chambers capturing CH₄ emission hot spots and hot moments may have similarly led to the large ΔFCH₄ CVs and negative mean ΔFCH₄, particularly in the daily and weekly aggregations in both median and mean-based temporal aggregations (Table 2 and Table C4).“
(lines 630-633 in the revised manuscript)

Line 521: By overcome do you mean exceed, i.e. EC FCH₄ is higher? Overcome is a strange verb to use here.

“Exceeded” is indeed a better word for this context. We changed the sentence accordingly:

“Nevertheless, despite the possible importance of chamber CH₄ emission hot spots and moments in driving ΔFCH₄, cumulative plot-scale FCH₄ is increasingly exceeded by higher ecosystem-scale FCH₄ at coarser temporal scales, but with site-specific trends (Table C5, Fig. B16).”
(lines 634-637 in the revised manuscript)

Line 530-531: Good point about using representative chamber patches, but discuss the issue of how to best determine these?

We now give some examples for determining representative chamber measurement locations and times as the last sentence in this paragraph:

“Therefore, using representative chamber patches and measurement times to upscale chamber FCH₄ to the EC footprint could potentially decrease ΔFCH₄ (Schrier-Uijl et al., 2010; Vargas and Le, 2023). This could be achieved for example by utilizing statistical optimization for temporal sampling (Vargas and Le, 2023) and matching the chamber, EC and site spatial heterogeneity by surveying the vegetation, hydrological and edaphic properties of the study site, EC footprint, and the surrounding area/region that the footprint represents (e.g., Chu et al., 2021; Schrier-Uijl et al., 2010, Riutta et al., 2007) (see also 4.6).”
(lines 652-657 in the revised manuscript)

Line 535: add ‘the significant *effect* of uWD...’

Fixed. (line 662 in the revised manuscript)

Lines 542-544: Rewrite the final sentence of this paragraph; it is confusing. What is ‘significant April’?

We have now clarified this sentence and added another one to strengthen the message:

“The significant effect of April in the monthly model (Table 3) was likely influenced by site-specificity, as only three out of ten sites had observations in that month (Fig. B1, Table 1). Thus, more sites with year-round FCH₄ observations are needed to confirm the significance of, and the possible ΔFCH₄ drivers in, April.”

(lines 669-672 in the revised manuscript)

Lines 546-548: Why would chambers not capture the release of stored CH₄ below the ice and snow cover? Freezing and thawing dynamics also occur inside chambers. I also struggle to see very specific “higher ecosystem-scale FCH₄ at CN-Hgu and US-Ho1 in cooler months” in Figure B14. But maybe this is an issue with figure readability.

The idea here was that the EC footprint may have covered areas where CH₄ was released as a result of thaw, similar to the plant-mediated CH₄ transport discussed earlier in the text. Since this is unclear, we improved the sentence to include the EC footprint as the possible reason, not necessarily the chambers:

“The higher ecosystem-scale FCH₄ at CN-Hgu in cooler months (February-April, Fig. B15) may have resulted from spring snowmelt releasing stored CH₄ below the ice and snow cover (Hargreaves et al., 2001; Morin et al., 2017; Rinne et al., 2007; Zhang et al., 2012) which could have been captured by the EC footprint but not by the smaller chamber footprints, especially since chamber placement in frozen conditions tends to be located further from ice cracks and fissures.”

(lines 674-678 in the revised manuscript)

We improved Fig. B14 for readability by removing the FCH_{4,median} points and adding transparency to the ΔFCH₄ points (now hollow instead of filled triangles) which overlap with the EC and chamber points. We also created a new figure to the Appendix (Fig. B15 in the revised manuscript) to highlight the higher February-April CH₄ fluxes occurring at CN-Hgu, which are indeed not clearly visible in the current Fig. B14. The higher FCH₄ at US-Ho1 in cooler months was accidentally left to the text from an older version, and is now thus excluded from this sentence- we thank the reviewer for pointing this out. This new figure replaces the US-Ho1 figure (previously Fig. B15) showing the outliers in the daily-aggregated US-Ho1 data, which was not as essential for understanding the main results of the study.

Lines 550-554: the chamber measurement system issue seems to not belong together in the same paragraph with seasonality issues. Consider re-arranging. Would chamber system issues better be placed in limitations? Perhaps you could also integrate the use of different

chamber flux analysis methods (GC vs portable analyzers) and create a new paragraph.

We placed the chamber system changes in this part of the manuscript (about monthly and annual ΔFCH_4 trends), as the changes were probably most influential at the monthly and annual scales due to the number of chambers per land cover class changing between years at US-Ho1 and US-Uaf, which may have affected monthly and annually-aggregated ΔFCH_4 values (the temporal scale discussed in this section). Therefore, we decided to keep it in this section. However, we have added a sentence about the chamber flux analysis methods in sections 4.6 (limitations) and 4.5 (last paragraph with detailed discussion of CH_4 flux analysis methods), as suggested (see the revised sentences in our earlier response to Reviewer comments on the introduction).

Line 556: I'd suggest removing 'As we expected' and just starting the sentence with 'Diel analyses revealed...'

Fixed. (line 691 in the revised manuscript)

Line 559: What do you mean by 'higher diurnal EC FCH_4 than chamber FCH_4 '? Do you mean higher diurnal variation? Or higher daytime flux overall?

We mean higher daytime flux overall. We have improved the sentence to make this clearer:

"...higher daytime EC FCH_4 than chamber FCH_4 ..." (line 694 in the revised manuscript)

Line 573: this sentence needs a citation.

The references for this sentence were listed after the second sentence, which was related to this one. We clarified this as follows:

"Related to VPD, pressurized plant-mediated CH_4 transport typically peaks in the late morning to afternoon, as temperature and humidity gradients between cooler belowground tissues and warmer, drier aboveground air enhance internal-external pressure differences that drive gas flow through aerenchyma (van den Berg et al., 2020; Knox et al., 2021; Morin et al., 2014; Vroom et al., 2022; Whiting and Chanton, 1996). However, very high VPD can induce stomatal closure, thereby reducing CH_4 transport (Grossiord et al., 2020)."

(lines 707-711 in the revised manuscript)

Lines 599-601: This sentence is confusing, consider re-writing.

We have improved the sentence as follows:

"While manual chambers allow researchers to capture higher spatial FCH_4 variation than automated chambers (e.g., Vargas and Le, 2023), the use of spatial medians for chamber FCH_4 may have reduced manual chamber FCH_4 variation so that the resulting median FCH_4 was similar to the FCH_4 measured by automated chambers."

(lines 738-741 in the revised manuscript)

Limitations:

Could the specific instrument for chamber measurement (e.g. Licor, Picarro etc) also play a role in method discrepancies?

We added a brief (to decrease the amount of text) mention of these possible differences in the discussion (section 4.6):

“In addition, our cross-scale FCH₄ comparisons may contain large uncertainties due to differences in chamber FCH₄ outlier removal (Table C2), design and the gas analyzer used (Table C1) (Jentsch et al., 2025; Levy et al., 2011; Pihlatie et al., 2013; Pumpanen et al., 2004).”

(lines 785-787 in the revised manuscript)

For balance, we also added a sentence about possible differences in EC CH₄ analyzers in section 4.5:

“The specific EC CH₄ analyzers can also differ in signal noise (Peltola et al. 2014).”

(lines 727-728 in the revised manuscript)

Acknowledge that lack of winter measurements at most sites is a major limitation and may bias annual comparisons.

We added two sentences about this in the discussion (section 4.6):

“In addition, year-round FCH₄ observations were lacking, which introduced uncertainty, particularly into the annual ΔFCH₄ trends. While challenging to measure, nongrowing season FCH₄ can be significant (Treat et al., 2018). Thus, future syntheses could include nongrowing season FCH₄ observations to improve annual ΔFCH₄ estimates and investigate the possible effects of ice thaw and snowmelt on ΔFCH₄.”

(lines 766-769 in the revised manuscript)

New reference added: “Treat, C. C., Bloom, A. A., and Marushchak, M. E.: Nongrowing season methane emissions – a significant component of annual emissions across northern ecosystems, *Glob. Change Biol.*, 24, 3331–3343, <https://doi.org/10.1111/gcb.14137>, 2018.”

Supplementary materials

Try to keep figures and figure captions on one page.

While we completely understand why this would be important in a supplementary material document, Biogeosciences publishes Appendices directly in association with the main text, so the figure and table captions will be automatically and clearly associated with them (see e.g., <https://bg.copernicus.org/articles/23/831/2026/#section6>).

Just suggestions: Confusion in Supplementary Material/Appendices naming: Text A1-A3 and figures B1-B19 and Tables C1-C14. They make sense once you make it all the way down to the supplementary material and realize that the letters refer to Appendix A, Appendix B and

Appendix C, but when they first pop up in-text, it is confusing. Also, instead of Text A1-A3, you could do Supplementary Methods 1 etc. maybe, as “Text” is ambiguous.

We understand the confusion. However, it is our understanding that Biogeosciences requires items in the Appendices to be referred to in the text (more information in: <https://www.biogeosciences.net/submission.html#manuscriptcomposition>), and we therefore kept the naming as is. We agree on the “Text” reference and changed it to “Supplementary Methods”.

Text A1 regarding QC/QA of chamber measurements on each site: I’d like to see a more uniform description of these. Currently, some sites mention R2 filtering, some not. Some say ebullition was not significant and thus it was removed, some say ebullition was not significant and thus it was *not* removed. Perhaps you could combine Text A1 into a table with different QC/QA criteria for better comparison (i.e. R2 threshold, dark/transparent chamber, GC vs PGA instrumentation, pressure vent and fan, H₂O correction, ebullition removal...). Similar to Table C1.

Large differences between chamber CH₄ flux QA/QC processing between data providers are very common, and standardized methods for the chamber flux community are still being developed (see e.g., Jentsch et al. 2025; <https://doi.org/10.5194/essd-17-2331-2025>). This large variety of different QA/QC protocols is reflected in this study as well. Combining these different QA/QC methods into a table is a good idea, and we added it to the Appendices (as Table C2) together with chamber placement rationale for each site, as recommended by Reviewer 2 (<https://doi.org/10.5194/egusphere-2025-5023-RC2>).

Response to comments by Reviewer 2

<https://doi.org/10.5194/egusphere-2025-5023-RC2>

Määttä and co-authors conducted a metaanalysis comparing chamber and eddy covariance based methane (CH₄) flux measurements in peatlands. They then explore parameters that can explain the difference observed between these two measurement methods.

This is an important and timely contribution that falls well within the Biogeosciences scope. The authors have carefully compiled a unique dataset which holds the potential for important contributions. Unfortunately, I find that the statistical approach is not very mature (see comments below), and I'm therefore not sure if all the reported conclusions are sufficiently supported. The manuscript is clearly written and reads well.

We thank the reviewer for the thorough review and thoughtful insights into the methodology of this manuscript. We address each point raised by the reviewer below.

Major comments:

1. In the introduction I'm missing a discussion of the representativeness of chamber and especially EC measurements for peatland surface-averaged FCH₄. I'm thinking about ecosystem models here, which delineate peatlands, and then ascribe an average flux to the peatland surface, which is compared to empirical measurements. EC measures the average flux in the footprint, which is different from the delineated peatland. Is there a bias introduced by placing the tower in the center of a peatland, potentially under-sampling lower-flux locations at the edge of the site, and does this introduce a systematic error to model/data comparison?

We added the following sentence in the introduction:

"For example, placing chambers over CH₄-emitting hollows in a peatland could bias ecosystem FCH₄ estimates, as the lower FCH₄ in other peatland microtopographic forms and margins may not be captured (e.g., Bubier 1993; 1995; Juselius-Rajamäki et al., 2025; Waddington and Roulet, 2000)."
(lines 113-116 in the revised manuscript)

Since this paper is not specifically about peatlands, we do not focus too much on them in the introduction. We do agree with the bias regarding position in the ecosystem, and we also added more details of this in the discussion.

New references added:

Bubier, J., Costello, A., Moore, T. R., Roulet, N. T., and Savage, K.: Microtopography and methane flux in boreal peatlands, northern Ontario, Canada, *Can. J. Bot.*, 71, 1056–1063, <https://doi.org/10.1139/b93-122>, 1993.

Bubier, J. L.: The relationship of vegetation to methane emission and

hydrochemical gradients in northern peatlands, *J. Ecol.*, 83, 403–420, <https://doi.org/10.2307/2261594>, 1995.

Juselius-Rajamäki, T., Piilo, S., Salminen-Paatero, S., Tuomaala, E., Virtanen, T., Korhola, A., Autio, A., Marttila, H., Ala-Aho, P., Lohila, A., and Väiliranta, M.: External and internal drivers behind the formation, vegetation succession, and carbon balance of a subarctic fen margin, *Biogeosciences*, 22, 3047–3071, <https://doi.org/10.5194/bg-22-3047-2025>, 2025.

Waddington, J. M. and Roulet, N. T.: Carbon balance of a boreal patterned peatland, *Glob. Change Biol.*, 6, 87–97, <https://doi.org/10.1046/j.1365-2486.2000.00283.x>, 2000.

2. ΔFCH_4 , as studied in this manuscript, is the combined product of (i) true difference (plot vs. ecosystem flux have different drivers), (ii) measurement uncertainty and artefacts in each method, and (iii) human behaviour (e.g. where do researchers place chambers). I feel like (iii) could be better discussed in the manuscript, and information on chamber placement strategies/goals could be provided in the appendix. Also should also be discussed that (iii) is prone to change over time, possibly in response to this manuscript. It is therefore possible that the manuscript, in exploring predictors of ΔFCH_4 , itself changes these predictions.

This is a great point, and we have added more discussion about chamber placement in the study sites in the discussion (e.g., in the last paragraph of section 4.3 where chamber system changes are discussed: “This further highlights the influence of selective site-specific chamber and EC tower placement and the development of methods for plot selection over time on ΔFCH_4 .” lines 688-689 in the revised manuscript).

We also added the rationale of chamber placement within each site in the Appendix (we combined this with the chamber CH_4 flux QA/QC information, as suggested by Reviewer 1; <https://doi.org/10.5194/egusphere-2025-5023-RC1> in Table C2)- we thank the reviewer for this useful recommendation.

3. The authors deal with a challenging dataset: ΔFCH_4 is clearly non-normal and crosses 0, as are both chamber and EC flux data, and the data analysis requires mixed effects models to account for repeated measurements at each site. The authors address non-normality through data transformation (power transformation). However, such transformation fundamentally changes model structure: log-transformation turns additive models into multiplicative models, and more complex transformations leave models without clear interpretability (like, random intercepts in MEMs mean something very different after this

transformation but I cannot intuit what that means in the real world). Why would e.g. a one-degree temperature difference have distinct effects on low, high, or negative fluxes that follow a certain (but not intuitive) mathematical function? I highly encourage the authors to explore statistical methods that do not rely on such complex transformations, which leave it almost impossible to draw meaningful conclusions from the results.

The dataset is indeed very challenging to analyze due to different sample sizes across sites and non-normal data. It is true that interpreting the slopes based on Yeo-Johnson-power-transformed values is difficult, but we made the decision of using this transformation for three main reasons: 1) the strongly non-normal data that, if untransformed, would have violated the normality assumptions of linear mixed effects models, 2) the FCH₄ data included zero values, and 3) the main goal was to investigate how the direction of ΔFCH_4 (i.e., whether absolute ΔFCH_4 increases or decreases, indicating trends toward higher ecosystem- or plot-scale FCH₄, respectively) changes with the chosen predictors, instead of the precise effect sizes (e.g., change in ΔFCH_4 per 1 °C). Trying to define precise effect sizes would be very challenging due to the unbalanced dataset, as mentioned by the Reviewer. Thus, we kept the Yeo-Johnson-transformation for ΔFCH_4 in the models. However, as the rationale for this decision seems unclear, we clarified this decision and the aim of looking mainly at the directionality of the ΔFCH_4 response in the methods (section 2.4.2):

“We built linear mixed models to estimate the predictors of ΔFCH_4 . The aim was to explore how the predictors influence the direction of ΔFCH_4 (i.e., more positive or negative ΔFCH_4 or, in other words, increase ecosystem-scale FCH₄ in relation to plot-scale FCH₄ or vice versa) at the ten sites. To meet the assumptions of linear mixed modeling and to improve residual diagnostics (normality and homoscedasticity of residuals) for model inference, we applied Yeo-Johnson power transformation (Yeo and Johnson, 2000) to absolute ΔFCH_4 values using the function `yeojohnson` from `bestNormalize` (Peterson, 2021). This transformation can be applied to zero values, and it improved our residual diagnostics, which were important for model inference. Acknowledging the difficulty to interpret the precise effect sizes after this transformation, we used this model only to investigate the directionality of ΔFCH_4 . All models were built with the function `lme` from `nlme` (Pinheiro et al., 2000, 2023)”.

(lines 327-334 in the revised manuscript)

We also shifted the focus more from the effect size magnitude to its direction in the discussion.

4. ΔFCH_4 vs. mean FCH₄: analysis seems to run into some artifacts, especially at half-hourly and hourly level. I highly recommend the authors show Fig B3 instead

of Fig 4. B3 shows how this apparent correlation of ΔFCH_4 vs. mean FCH_4 is driven by sites where either only chambers or only EC showed high CH_4 emissions (see also fluxes in B8 and B9 where chambers show no CH_4 emissions at all. Unless there's some conversion error here, this seems to be driven by chamber placement on non-emitting sites in the tower footprint?). B8-B12 show very distinct relationships between chamber and EC fluxes, and I wonder if it would be better to analyse these relationships for each site first and then compare the outcomes rather than pool the datasets and then try to identify overall drivers.

As Fig. B3 represents the general relationships between EC and chamber FCH_4 with untransformed data (to provide an untransformed version of Fig. 3) while Fig. 4 shows the relationship between mean FCH_4 and absolute ΔFCH_4 , we don't think replacing Fig. 4 with Fig. B3 would be necessary here. Based on the figure description, we assume the Reviewer means Fig. B2 instead. We think showing Fig. 4 is important here, as similar trends have been observed, for example, in a rice paddy (Meijide et al., 2011; <https://doi.org/10.5194/bg-8-3809-2011>) and for EC FCH_4 in a synthesis by Knox et al. (2019) (<https://doi.org/10.1175/BAMS-D-18-0268.1>). This figure highlights the importance of deviations from perfect agreement between ecosystem- and plot-scales with increasing FCH_4 magnitude at different temporal scales, which Fig. B2 does not show (it is based on the unaggregated dataset). However, we agree that Fig. B2 provides important additional site-specific information of high CH_4 emission contribution to ΔFCH_4 , and added more references to this figure in the results and in the caption of Fig. 4.

Regarding the comments on Figs. B8 and B9, the Reviewer's observations are accurate. At site CN-Hgu (Fig. B9), the chambers were placed in more CH_4 -consuming patches, whereas the EC footprint likely captured CH_4 -emitting patches, resulting in these patterns at the multi-site scale as well (Fig. B8). Because we used medians, the resulting central tendency for the plot-scale FCH_4 is "dragged" lower for chambers in Fig. B8 (all sites combined) even though some sites sometimes reached higher plot-scale CH_4 emissions than ecosystem scale (but this is shown as high IQRs in Fig. B8). It is good that this was pointed out though, as we now used CN-Hgu (Fig. B9) as an example of chamber placement influence on ΔFCH_4 patterns in the discussion, e.g.:

"In addition, the lower plot- than ecosystem-scale diel FCH_4 at CN-Hgu (e.g., Fig. B9) likely reflected the selective chamber placement at CH_4 -consuming areas whereas the EC footprint captured CH_4 emission events more often (Table C2)."

(lines 715-717 in the revised manuscript)

Related to this, we also added results of high negative plot-scale FCH_4 (i.e.,

CH₄ uptake) increasing absolute ΔFCH_4 to the last paragraph of section 3.1:

“Sites also differed in whether the trends in negative FCH₄ came from higher plot or ecosystem-scale FCH₄: for example, at US-Uaf and CN-Hgu, 100% and 91% of ΔFCH_4 observations at $FCH_{4_mean} < 0$, respectively, consisted of higher plot-scale FCH₄ while ca. 66% of hourly and half-hourly observations ($FCH_{4_mean} < 0$) in US-Ho1 came from higher ecosystem-scale FCH₄.”

(lines 466-469 in the revised manuscript)

It is true that many of the observed trends show site-specificity, and we do discuss this in the manuscript. We did consider site-specific analyses (incl. modeling) but found it to be difficult to glean insights from. This is because it would have resulted in ca. 60 linear mixed models in total, each model having different variance structures and temporal autocorrelation structures. If we had enough replication for each ecosystem type, this could have been a better way forward as we could generate ecosystem-specific hypotheses. We added mention of redoing this work by ecosystem type once there is enough replication within each ecosystem type:

“Thus, in order to produce results that would be better generalizable to other sites and regions (e.g., tropics), future studies could include more sites from a variety of climates, ecosystem types, dominant vegetation types, and chamber measurement systems (i.e., automated and manual, gas chromatography and high-precision CH₄ analyzers) (n>3 sites per group to allow statistical inference).”

(lines 763-766 in the revised manuscript)

“Given that our results indicated ebullition removal from some of the chamber FCH₄ data as one potential driver of ΔFCH_4 , future studies could also conduct cross-scale FCH₄ comparisons based on chamber FCH₄ data with ebullition events both included and excluded from a variety of wetland types.”

(lines 788-791 in the revised manuscript)

5. Similarly, the dataset is highly unbalanced, including hundreds of entries from some sites but only a few from other sites. This means that any relationships identified in the dataset will be highly driven by a few sites, while relationships at other sites are barely taken into account. Some of this can be addressed by allowing for random slopes rather than just random intercepts in MEMs and/or by downsampling over-represented datasets.

We also highlight this in the discussion and now emphasize it more in discussion and limitations. While we use site as a random intercept nested

with temporal variables (e.g., month) to account for the different temporal measurement coverages between sites, we decided not to use random slopes in order to reduce the complexity of the models and to avoid model non-convergence.

6. Generalization: I'm not quite sure what aim of the statistical analysis is, and specifically, what 'general population' do the authors want to make inference about? (i.e., ultimately, what kind of prediction do they want to support?). My guess is that they will respond 'explain the observed data' but statistics cannot explain data. Is this exercise meant to provide a means for comparing (or converting between) EC- and chamber-based flux estimates? Is it such a comparison/conversion meant to be valid beyond to sites outside this dataset (i.e., should this help us compare the chamber fluxes from one site to the EC fluxes at another site)? In that case we're talking about predicting chamber/EC flux the 'general population' the authors are trying to make conclusions about is the population of (potential) flux stations, i.e., if I conduct chamber and EC measurements at a new site, what will ΔFCH_4 based on various predictors? If that is the case, there's a lot of pseudoreplication going on throughout the study, as different measurements within a site should not be considered as independent. The mixed effect models used somewhat alleviate this in some analyses but not everywhere.

This is an important point. As we were limited to the data provided from ten sites as a response to the community call for data and as the data are very unbalanced, we were not able to generalize these results to other sites with this dataset (also see our response to Reviewer 2 comment 7). We attempted to account for the pseudoreplication issue by including temporal autocorrelation structures in the models, and the spatially-aggregated medians of chamber CH_4 fluxes should remove pseudoreplication within sites. Throughout the manuscript, we therefore now highlight that these analyses were meant to find potential predictors across these sites, but generalization to other sites is not yet warranted based on the small number of sites and unbalanced sampling efforts between sites. Our high-level goal was to provide some guidance to efforts that combine chamber and EC data and we find that, at least at our 10 sites, doing this kind of combination is feasible at annual scales but may result in errors at finer temporal scales.

7. Validation: the authors could test the robustness of their results e.g. with leave-one-site-out analyses. This would provide some additional robustness to the results.

Good idea. We added leave-one-site-out analyses to the models (in Table 3 and Tables C12-C15 in the Appendix) and added them to the methods section 2.4.2:

“To test how well the models generalize to other sites, we validated the models with leave-one-site-out cross validation and evaluated model performance with R^2 , mean absolute error (MAE) and root mean square error (RMSE) between observed and predicted values. To allow for predictions to new sites with the training data, the fixed effect VEG had to be removed from the models, as some of the VEG classes (tree and ericaceous shrub) were represented only by a single site and the effect of these classes cannot be estimated when they are withheld in the test data. Similarly, due to uneven temporal coverage across sites, observations (e.g., date or year-month) included in the test data but not present in the training data were excluded from evaluation.”

(lines 426-432 in the revised manuscript)

The leave-one-site-out analyses ($R^2 < -1$, and MAE and $RMSE > 1$) confirmed that the models do not predict well for new sites, highlighting the site-specificity of the potential ΔFCH_4 predictors. We added a sentence highlighting this in the results text:

“In addition, the high conditional R^2 and high negative LOOCV R^2 , high MAE and RMSE showed that the ΔFCH_4 predictors are specific to the sites included in this study (Table 3).”

(lines 608-609 in the revised manuscript)

While we have discussed the site-specificity in the discussion, we now added an additional sentence about the models' inability to predict ΔFCH_4 at other sites in the discussion:

“However, given the limited sample size in the models ($n=9$ sites) and the low model performance based on leave-one-site-out analyses (Table 3), these results are influenced by site selection and generalizations to other sites are not possible.”

(lines 857-859 in the revised manuscript)

However, as discussed above, as our main goal was to explore the potential predictors at the ten sites included in this study instead of generalizing to other sites, we kept the multi-site models in the analyses. We thank the reviewer for this suggestion.