

**Author response to reviewer 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 comment below in bold.**

**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 99: “Plot and ecosystem-scale FCH<sub>4</sub>...” : 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.”**

Line 113: “...than the true, mean ecosystem-scale FCH<sub>4</sub>”: this insinuates that the EC ecosystem-scale measurement is the “true” FCH<sub>4</sub>, but in reality, we do not know this for sure, as EC also comes with its own uncertainties.

**Agreed that EC CH<sub>4</sub> flux estimates can have positive or negative bias and therefore not represent the ‘true’ mean. We have removed the word “true” from the sentence: “...often leading to higher observed fluxes at the individual sampled plots, than the mean ecosystem-scale FCH<sub>4</sub> as measured by EC (but see Voigt et al., 2023).”**

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 CH<sub>4</sub> flux 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 CH<sub>4</sub> analyzers) and different EC and chamber instrument makes and models (Christiansen et al.,**

2015; Peltola et al., 2014; Pihlatie et al., 2013).”

Section 4.5: “The specific EC CH<sub>4</sub> analyzers can also differ in signal-to-noise ratios (Peltola et al. 2014).” and “It may be valuable to compare chamber and EC FCH<sub>4</sub> using both linear and exponential fits for chamber FCH<sub>4</sub> (from both high-precision CH<sub>4</sub> analyzers and gas chromatography) to better understand ΔFCH<sub>4</sub> trends across sites.”

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 CH<sub>4</sub> analyzers).” and “In addition, our cross-scale FCH<sub>4</sub> comparisons may contain large uncertainties due to differences in chamber FCH<sub>4</sub> outlier removal (Supplementary Methods A1), design and the gas analyzer used (Table C1) (Jentsch et al., 2025; Levy et al., 2011; Pihlatie et al., 2013; Pumpanen et al., 2004).”

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 CO<sub>2</sub> 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 CH<sub>4</sub> fluxes in the form they are often used in syntheses (i.e., gap-gilled EC CH<sub>4</sub> fluxes and chamber CH<sub>4</sub> fluxes filtered differently by different data providers), to highlight that these two data streams should be combined carefully if used with non-standardized chamber CH<sub>4</sub> flux data especially. We have added a clarification of this goal to the end of the introduction:

“To achieve this, we utilized FCH<sub>4</sub> data commonly used by the FCH<sub>4</sub> 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 FCH<sub>4</sub> data; Jentsch et al., 2025) (see 2.2.1 and Table C2).”

We also now highlight the decision of including chamber data with ebullition events filtered out in the methods section: “The decision to utilize chamber FCH<sub>4</sub> data with differing ebullition removal protocols across data providers was intended to reflect the non-standardized and site-specific data processing typical of chamber FCH<sub>4</sub> data (Jentsch et al., 2025), which may contribute to differences between bulk ecosystem- and plot-scale FCH<sub>4</sub> estimates.”

We also now discuss the influence of ebullition on plot-scale CH<sub>4</sub> 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 CH<sub>4</sub> ebullition removal, could be important  $\Delta$ FCH<sub>4</sub> drivers, as FCH<sub>4</sub> 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 footprint representativeness as well as chamber measurement location, frequency, and data quality control when combining plot and ecosystem-scale FCH<sub>4</sub> data, particularly at high-emitting sites and periods (Fig. 4).”

First paragraph of section 4.2:

“As ebullition events are often removed from chamber FCH<sub>4</sub> data, these results highlight that the large variation in chamber FCH<sub>4</sub> data processing protocols between researchers could also increase  $\Delta$ FCH<sub>4</sub> and thus uncertainty in multi-site syntheses combining cross-scale FCH<sub>4</sub> data, at least in the sites included in this study (e.g., Jentsch et al., 2025; Levy et al., 2011).”

Third paragraph of section 4.5:

“As previously discussed (see 4.1 and 4.2), plot-scale FCH<sub>4</sub> could have been underestimated due to the removal of ebullition events from some of the chamber FCH<sub>4</sub> data included in this study (Table C2), calling for standardization of chamber-based ebullition measurements and data processing (Jentsch et al., 2025).”

Section 4.6:

“Given that our results indicated ebullition removal from some of the chamber FCH<sub>4</sub> data as one potential driver of  $\Delta$ FCH<sub>4</sub>, future studies could also conduct cross-scale FCH<sub>4</sub> comparisons based on chamber FCH<sub>4</sub> data with ebullition events both included and excluded from a variety of wetland types. Ebullition events are sometimes also removed from EC FCH<sub>4</sub> data following the standard data quality protocols and further standardization of EC-based ebullition measurements are needed.”

Conclusions:

“We attribute the higher ecosystem-scale FCH<sub>4</sub> than plot-scale FCH<sub>4</sub> mainly to the combination of selective chamber placement, ebullition removal from chamber FCH<sub>4</sub> data, and...”

**Conclusions- recommendations:**

**“Standardized protocols for chamber FCH<sub>4</sub> data quality control, especially related to ebullition removal (see Jentsch et al., 2025 for recent recommendations for chamber FCH<sub>4</sub> data processing), and accounting for these differences in quality control when combining chamber and EC FCH<sub>4</sub> data”**

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 will change 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  $\Delta$ FCH<sub>4</sub> was large, particularly in daily (674%) and weekly (467%) aggregations (Table 2).”**

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  $\Delta$ FCH<sub>4</sub> ranged between 0.28 nmol m<sup>-2</sup> s<sup>-1</sup> (annual) and 1.23 nmol m<sup>-2</sup> s<sup>-1</sup> (half-hourly)...”**

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<sup>2</sup> and RMSE in addition to Spearman's  $\rho$  – 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-normal and violated linear regression assumptions, 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  $\Delta FCH_4$ .” 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,  $\rho$ , its significance level,  $p$ , and the normalized root mean square error, NRMSE) between plot-scale (chamber)  $CH_4$  flux ( $FCH_4$ ) and ecosystem-scale (EC)  $FCH_4$  from half hourly (a) to 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  $\rho$  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  $\rho$ . 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  $n = 1879, 349, 121, \text{ and } 22$ , respectively. The dashed line represents 1:1 line.”

Figure 4:

Similar comment about the caption as Figure 3.

Add  $R^2$  and RMSE in addition to Spearman's  $\rho$  – 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 will not add linear regressions and  $R^2$  values to this figure. However, as for Figure 3, we have made this decision clearer in the methods section and have added normalized RMSE (based on  $\Delta FCH_4$  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 will not add them in this figure.

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

“The relationship between  $CH_4$  flux ( $FCH_4$ ) magnitude ( $FCH_{4\_mean}$ ) and absolute difference between ecosystem-scale (EC) and plot-scale  $FCH_4$  ( $\Delta FCH_4$ ) from half-hourly (a) to annual (f) scales, represented by Spearman correlation coefficient, ( $\rho$ ), its significance, ( $p$ ), and normalized root mean square error of  $\Delta FCH_4$  (NRMSE).  $FCH_{4\_mean}$  is the row-wise mean of EC  $FCH_4$  and chamber  $FCH_4$ . 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  $\Delta FCH_4=0$  meaning complete agreement between ecosystem and plot-scale  $FCH_4$ . Higher Spearman correlation coefficient ( $\alpha=0.05$ ) represents stronger deviation from  $\Delta FCH_4=0$ . NRMSE was calculated by dividing RMSE (of  $\Delta FCH_4$ ) by the standard deviation of ecosystem- and plot-scale  $FCH_4$  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  $CH_4$  emissions from ecosystem and plot scales contribute to annual  $CH_4$  emissions per site in Fig. B2.”

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 manuscript Appendices is necessary (the results are referred to in the main text), we have 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  $\Delta FCH_4$  between hours, Half-hourly aggregation”).

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

**Agreed- we have made the table now 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  $\Delta FCH_4$  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  $FCH_4$  differences between these ecosystem types. Thus, we have now changed 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 have now rearranged this paragraph as follows: "Across all temporal aggregations, ecosystem-scale (EC)  $FCH_4$  was higher than at the plot scale (chamber). Supporting these results, higher EC  $FCH_4$  than chamber  $FCH_4$  have been observed in an arctic peatland with area-weighted chamber  $FCH_4$  (Budishchev et al., 2014), a managed peat meadow with upscaled chamber  $FCH_4$  (Schrier-Uijl et al., 2010), a peatland with down-scaled EC  $FCH_4$  (Forbrich et al., 2011), a temperate forest with spatial chamber  $FCH_4$  averages (Wang et al., 2013), and a temperate salt marsh with spatio-temporal chamber and EC  $FCH_4$  averages (Hill and Vargas, 2022b). Other studies at individual sites have observed higher chamber  $FCH_4$ ..."**

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 have 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  $FCH_4$  variability and inflates correlation coefficients (e.g., Clark and Avery, 1976; Pollet et al., 2015). In addition, mean  $\Delta FCH_4$  in weekly, monthly**

and annual aggregations was negative (Table 1), indicating higher plot-scale than ecosystem-scale  $FCH_4$ , and the CV for the weekly aggregation in particular was large (467%) (Table 1). Our results suggest that high  $CH_4$  emissions and  $FCH_4$  variability in plot-scale measurements are associated with higher  $\Delta FCH_4$ , particularly at time scales longer than daily (Table 2 and Fig. B2); suggesting that combining plot- and ecosystem-scale bulk  $FCH_4$  at heterogeneous sites is particularly problematic at coarse temporal scales. However, footprint-aware comparisons between upscaled chamber or downscaled EC  $FCH_4$  could show better agreement between ecosystem and plot scales (e.g., Schrier-Uijl et al., 2010) (see 4.6). Nonetheless, this highlights the importance of selective chamber placement on high-emitting locations and time periods within the study sites (Hill and Vargas 2022b; Vargas and Le 2023)..."

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 manual chamber placement onto water or waterlogged soil surface, increasing plot-scale  $FCH_4$  (e.g., Jentzsch et al., 2025), but as ebullition events were removed from some of the chamber  $FCH_4$  datasets by data providers (see 2.2.1 and Table C2), this was unlikely to contribute to the general  $\Delta FCH_4$  trends observed here."

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

We have now split this sentence into two and improved its readability: " $FCH_4$  hot spots and hot moments can also vary in space and time, which manual chamber  $FCH_4$  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 large uncertainties in spatio-temporal  $FCH_4$ , and possibly  $\Delta FCH_4$ , variation across temporal scales (Anthony and Silver, 2021, 2023; Vargas and Le, 2023)."

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  $FCH_4$ , 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 have now added this discussion briefly to this section and the last sentence: “...This highlights the need to account for EC footprint representativeness as well as chamber measurement location and frequency when combining plot and ecosystem-scale  $FCH_4$  data particularly at high  $FCH_4$  sites and periods (Fig. 4).”

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 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  $\Delta FCH_4$  with increasing  $FCH_4$  magnitude), as well as  $\Delta FCH_4$  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  $\Delta FCH_4$  trends in low  $FCH_4$  (Fig. 4) and possibly low-emitting uplands (e.g., part of US-Ho1) should be interpreted with caution.”

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

We have now split this sentence into two and emphasized the point of increasing  $\Delta FCH_4$  with increasing  $FCH_4$  magnitude (see previous response to Lines 485-486) in the end of the sentence: “Altogether, the mismatch in EC footprint and chamber measurement coverage could be an important  $\Delta FCH_4$  driver, as  $FCH_4$  between surface cover types and within them can vary strongly even within the same growing season (Voigt et al., 2023). This highlights the need to account for EC footprint representativeness as well as chamber measurement location and frequency when combining plot and ecosystem-scale  $FCH_4$  data particularly at high- $FCH_4$  sites and periods (Fig. 4).”

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

We have now 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  $\Delta FCH_4$  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 fine 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  $\Delta FCH_4$  decreased with higher PA (Table 3).”**

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 draft) to discuss this important topic: “While EC  $FCH_4$  can be underestimated in low  $u^*$ , leading to decreased  $\Delta FCH_4$ , EC  $FCH_4$  under low  $u^*$  were filtered out by the FLUXNET- $CH_4$  team, so low  $u^*$  was unlikely to influence the observed  $\Delta FCH_4$  trends (Aubinet, 2008; Baldocchi, 2003; Knox et al., 2019; Delwiche et al., 2021).”**

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 have now changed this sentence as follows, and removed the excessive “may” words related to our results in the rest of the manuscript.**

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

**Changed to “As hypothesized,..”**

Lines 517-519: You talk about soil  $CO_2$  respiration hotspots and hot moments and then bring those same conclusions to explain  $CH_4$  hotspots and moments. I think  $CO_2$  respiration hotspots do not directly translate to  $CH_4$  dynamics. You could try to find studies describing  $CH_4$  hotspots/hot moments with chamber measurements; or better integrate the  $CO_2$  ideas.

**It is true that  $CO_2$  dynamics do not directly translate to  $CH_4$ , but as other multi-site EC-chamber  $CH_4$  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,  $CO_2$  has been studied the most, and similar results have been found for  $CO_2$ . Thus, we deem it reasonable to mention here. However, we agree that it is important to mention the differences in  $CH_4$  and  $CO_2$  hot spots and moments. Thus, we have**

changed the sentence to better integrate the CO<sub>2</sub> thoughts to CH<sub>4</sub>:

**“While CH<sub>4</sub> cycling is driven by different controls than CO<sub>2</sub>, chambers capturing CH<sub>4</sub> emission hot spots and hot moments may have similarly led to the large  $\Delta FCH_4$  CVs and negative mean  $\Delta FCH_4$ , particularly in the daily and weekly aggregations in both median and mean-based temporal aggregations (Table 2 and Table C3).”**

Line 521: By overcome do you mean exceed, i.e. EC FCH<sub>4</sub> is higher? Overcome is a strange verb to use here.

**“Exceeded” is indeed a better word for this context. We have now changed the sentence accordingly: “Nevertheless, despite the possible importance of chamber CH<sub>4</sub> emission hot spots and moments in driving  $\Delta FCH_4$ , cumulative plot-scale FCH<sub>4</sub> seems to be increasingly exceeded by higher ecosystem-scale FCH<sub>4</sub> at coarser temporal scales, but with site-specific trends (Table C4, Fig. B16).”**

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:**

**“Determining representative patches and time periods for chamber FCH<sub>4</sub> measurements 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).”**

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

**Fixed.**

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 EC and chamber CH<sub>4</sub> flux observations are needed to confirm the significance of, and the possible  $\Delta FCH_4$  drivers in, April.”**

Lines 546-548: Why would chambers not capture the release of stored CH<sub>4</sub> below the ice and snow cover? Freezing and thawing dynamics also occur inside chambers. I also struggle to see very specific “higher ecosystem-scale FCH<sub>4</sub> 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 CH<sub>4</sub> was released as a result of thaw, similar to the plant-mediated CH<sub>4</sub> transport discussed earlier in the text. Since this is unclear, we have now improved the sentence to include the EC footprint as the possible reason, not necessarily the chambers: “...which might 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”

We have improved Fig. B14 for readability by removing the FCH<sub>4,median</sub> points and adding transparency to the ΔFCH<sub>4</sub> points (now hollow instead of filled triangles) which overlap with the EC and chamber points. We have also created a new figure to the Appendix (Fig. B15 in the revised manuscript) to highlight the higher February-April CH<sub>4</sub> fluxes occurring at CN-Hgu, which are indeed not clearly visible in the current Fig. B14. The higher FCH<sub>4</sub> at US-Ho1 in cooler months was accidentally left to the text from an older version, and is now thus excluded from this sentence- thank you 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<sub>4</sub> 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<sub>4</sub> values (the temporal scale discussed in this section). Therefore, we decided to keep it still 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<sub>4</sub> 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 559: What do you mean by 'higher diurnal EC FCH<sub>4</sub> than chamber FCH<sub>4</sub>'? 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<sub>4</sub> than chamber FCH<sub>4</sub>...”**

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 have now clarified this as follows: “Related to VPD,**

pressurized plant-mediated CH<sub>4</sub> 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<sub>4</sub> transport (Grossiord et al., 2020).”

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

**We have improved the sentence as follows: “While manual chambers often capture higher spatial FCH<sub>4</sub> variation than automated chambers (e.g., Vargas and Le, 2023), the use of spatial medians for chamber FCH<sub>4</sub> may have reduced the manual chamber FCH<sub>4</sub> variation so that the resulting median FCH<sub>4</sub> was similar to the FCH<sub>4</sub> measured by automated chambers.”**

Limitations:

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

**We have 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 FCH<sub>4</sub> comparisons may contain large uncertainties due to differences in chamber FCH<sub>4</sub> outlier removal (Supplementary Methods A1), design and gas analyzer (Table C1) (Jentsch et al., 2025; Levy et al., 2011; Pihlatie et al., 2013; Pumpanen et al., 2004).”**

**For balance, we also added a sentence about possible differences in EC CH<sub>4</sub> analyzers in section 4.5: “The specific EC CH<sub>4</sub> analyzers can also differ in signal noise (Peltola et al. 2014).”**

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

**We have added 2 sentences about this in the discussion (section 4.6):**

**“In addition, year-round CH<sub>4</sub> flux observations were lacking, which introduced uncertainty, particularly into the annual ΔFCH<sub>4</sub> trends. While challenging to measure particularly with chambers, nongrowing season CH<sub>4</sub> fluxes can be significant (Treat et al., 2018). Thus, future syntheses should also try to include nongrowing season CH<sub>4</sub> flux observations to further improve annual ΔFCH<sub>4</sub> estimates and investigate the possible effects of ice thaw and snowmelt on ΔFCH<sub>4</sub>.”**

**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 will therefore keep the naming as is. We agree on the “Text” reference and have now 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 CH<sub>4</sub> 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 have 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>).**