

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 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 have added the following sentence in the introduction: "For example, placing chambers (or EC tower) in the middle of a peatland covered by CH₄-emitting hollows could bias ecosystem FCH₄ estimates, as FCH₄ varies between peatland microtopographic forms and margins, and the FCH₄ variation can further differ across peatland types (e.g., Bubier 1993; 1995; Juselius-Rajamäki et al., 2025; Waddington and Roulet, 2000)."

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 have 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 will add a more detailed discussion about chamber placement strategies 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 placement and the development of methods for plot selection over time on ΔFCH_4 .”).

We have 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)- thank you 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 FCH_4 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 FCH_4 , 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 will keep the Yeo-Johnson-transformation for ΔFCH_4 in the models. However, as the rationale for this decision seems unclear, we have now 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 FCH_4 in relation to plot-scale FCH_4 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)”.

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

4. ΔFCH_4 vs. mean FCH_4 : 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 have now 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 will now use CN-Hgu (Fig. B9) as an example of chamber placement influence on ΔFCH_4 patterns in the discussion. Related to this, we have also now added results of high negative plot-scale FCH_4 (i.e., CH_4 uptake) increasing absolute ΔFCH_4 to the last paragraph of section 3.1:

“Sites also differed in whether the trends in negative FCH_4 came from higher plot or ecosystem-scale FCH_4 : 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 $FCH_{4\dots}$.”

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 will now add discussion on useful ways to redo this work by ecosystem type once there is enough replication within each ecosystem type.

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 will 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 are 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. We therefore will highlight more 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 have now 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. 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. While we have discussed the site-specificity in the discussion, we will now add additional sentences about the models' inability to predict ΔFCH_4 at other sites. 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 will keep the multi-site models in the analyses. We thank the reviewer for this suggestion.