

We thank the reviewer and appreciate the very thorough and insightful comments which have helped improve the quality of the manuscript. Please find our responses to the reviewers' suggestions below. Our author responses are denoted "AR" with blue text.

Reviewer #3

This manuscript describes a genuinely valuable dataset with 2375 chamber-based CH₄ flux measurements from five Swedish lakes spanning 68°N to 57°N, collected over 2016–2022 with a deliberately standardized design. Methodologically consistent multi-lake, multi-year flux programs are rare, and the value of this one will only grow as the time series lengthens. The dataset is openly archived and clearly falls within ESSD's scope.

My reservations concern the balance and the rigor of the two halves of the paper. As a data paper, the dataset documentation is thinner than ESSD normally expects — there is no variable/file description, no QA/QC section, and no analytical performance metrics, with readers instead referred to an external documentation file. As an analysis paper, the temperature-extrapolation exercise (Sect. 2.4, 3.3) carries more weight in the abstract and conclusions than its statistical treatment can support, and several of its assumptions are either untested or, in a few places, internally inconsistent. I would recommend the authors either strengthen the extrapolation substantially or present it more explicitly as an illustrative example rather than a headline result.

AR: Thanks for valuable comments. A detailed protocol covering sampling, sample analysis, and data packaging is linked from the revised manuscript. This is the same document providing meta-information to the public database while we use this link, rather than add it as a supplementary file to the ESSD paper to avoid double-publication. We have in the revised version provided a doi to the same document to create a direct and persistent link. We further provide the link to the SITES data portal where all data is openly published. The data are published as a collection of five separate files, one per lake. Each data file includes a well-developed header describing all variables including quality flagging. We further provide a supplementary file to the paper with the spreadsheet used to generate the published data used in the paper. Thereby the manuscript can add depth to the meta-description of the approach and data and provide examples of investigations that become possible with the generated data. We have tried to clarify this and that the extrapolation is to be seen as an example in the abstract and elsewhere.

Below I have suggestions and comments. None of these issues are fatal, but several need to be addressed before publication.

The diffusive/ebullitive partitioning rests on an unvalidated threshold (L163–169). The assignment of chambers to "diffusion-dominated" using the minimum apparent k and a 2-fold cutoff is the single most consequential methodological choice in the paper — it determines the reported 62% (ALM) vs 19–36% (other lakes) diffusive fractions and, downstream, every F_{diff} result. Yet the threshold is inherited from earlier work without testing here, and the qualifier "provided that k was in a realistic range given the wind speed" is left undefined: what range, based on which wind data, and how many chamber-deployments were rejected on this basis? Please also state whether the minimum k is identified per lake-deployment (across all 12 chambers) or per depth-zone. Given the very

different wind exposure of ERK versus the small forest lakes, a sensitivity analysis using alternative cutoffs (e.g. 1.5×, 2×, 3×) would considerably strengthen confidence in the partitioned data product.

AR: It is correct that the diffusive/ebullitive flux partitioning depends on the k600-ratio threshold choice as pointed out in the review comment. We have added a sensitivity analysis in which we vary the k600-ratio threshold from 1.5 to 4 (New figure in appendix) explaining the results and the choice of the threshold of 2 in the figure legend. Briefly, the data from the studied lakes did not point towards a clear threshold choice, which is why we took guidance from previous data supporting a 2-fold spatial variability in mean k-600 over a 24 to 48-hour period. In general, the ebullition fraction of total flux varied < 0.25 over the threshold range tested and this variability was smaller than the variability among measurement days at every specific threshold choice. The revised manuscript text clearly acknowledges that this threshold choice leads to uncertainty regarding the relative proportion of ebullition and diffusive flux in the new figure legend and also clarify that the total flux is rather insensitive to the threshold choice.

The Monte Carlo treatment of uncertainty is not defensible as implemented (L213–221, L334–337, Table A3). The confidence intervals in Table A3 appear to be intervals on the means of 1000 draws, not the 2.5–97.5 percentiles of the simulated flux distribution. For ALM 2016 the SD is 0.0012 while the CI spans 0.0032–0.0033. Readers need the predictive interval; as presented, the CIs give a misleading impression of precision. Consequently, the among-year ANOVA and the significance letters in Fig. 6 are not interpretable. With $N = 1000$ arbitrarily chosen simulation draws, p-values can be driven to any desired level by increasing N . These are simulated replicates, not independent observations.

AR: In response to this comment, we have removed the CIs and the ANOVA test. Figure 6 is now solely based on box-plots, and we use pairwise Mann-Whitney U along with Holm-Bonferroni correction to evaluate differences among years in a way that is not dependent on confidence intervals. We also clarify that the evaluation of differences among years are based on the simulated uncertainty distributions.

The “open water period” definition produces implausible integration windows (L216–218, Table A3). Defining the open-water season as all days with modelled T_w above the lowest observed T_w gives ERS 266–361 days per year — essentially the entire year in 2020. The threshold may include winter/ice-covered periods unless checked against ice records, and the air-temperature regression (Eq. 3) has no mechanism to represent ice cover. Fluxes are then extrapolated across ~300 days from roughly a dozen summer-weighted campaigns. Please either impose an ice-phenology constraint (do SITES stations have ice records?) or restrict integration to the observed sampling window, and report how sensitive the annual totals are to this choice.

AR: We have ice-in and ice-out records for some years in the studied lakes and available data suggest a high between-year variability. We have revised the analysis in the following way: For the ice-time records available, we set a minimum surface water temperature threshold yielding an open water period corresponding to actual data (on an average 3-8 days shorter as there are some days with partial ice cover). The thresholds were 6 °C in ALM (being subarctic and alpine; colder but also wind exposed), 3 °C in ERK (being a large wind-exposed lake with large waves which delays ice-in and speed up ice-out), 5 °C the small, forest surrounded lakes STO and FER, and 2 °C in ERS

having highest average temperatures and being located in a high precipitation area. As a result, the related numbers in the manuscript tables have been revised.

This was an important revision as it revealed the importance of considering not only direct temperature effects but also the indirect temperature effects on the length of the open water period. While results in ALM, STO, ERK, and ERS remained largely unchanged with regards to differences among years and showing 2018 as an outlier in STO, ERK and ERS, the results for FER changed by better consideration of the length of the open water period. It turned out that despite the higher mean summer temperature 2018, the total open water period in FER was considerably shorter than the other years. Hence, total accumulated CH₄ emissions over 2018 did not stand out from the other years (higher flux per day during the warm 2018 but overall fewer open-water days; Fig.6).

Annual fluxes are reported for years in which some lakes were not sampled. Table 2 gives 4 sampling years for ALM, 3 for FER, 5 for ERK — but Table A3 and Fig. 6 present 2016–2022 estimates for every lake. For FER in particular, four of the seven annual estimates appear to rest on no observations from that lake in those years, and Fig. 2 suggests ERK was not sampled early in the period either. Since the 2018 warm-summer signal is a headline finding, it matters whether 2018 was actually observed at each lake. Please state the sampled years explicitly per lake and mark purely extrapolated years in Fig. 6.

AR: We are now explicitly pointing out that field-collected data does not exist for all years in all lakes and that the extrapolation examples extend beyond extrapolation within-year to also extrapolation among years. It is also mentioned that the underlaying assumption is that the relationships derived from three or more years are comprehensive enough to support extrapolation among years during a seven-year period. We also point out that the years with and without data are visible in the left panels of Fig. 2. In comparison with previous extrapolations of lake CH₄ emissions our data support seems relatively abundant even if some years are missing, given the general scarcity of multi-year lake emission data. However, it is of course key that results seem reliable, and we hope the above measures adequately address this comment and make it possible for readers to evaluate our results themselves in the light of the data availability.

Internal inconsistency in the ALM temperature parameters (Table A2). For all three ALM depth-zones, F_{tot20} and F_{diff20} are identical (0.10), with $\theta_{diff} > \theta_{tot}$ in every case. Taken literally, this parameterization implies that ebullition falls to zero at 20°C and becomes negative above it, which is mechanistically backwards and inconsistent with the reported 38% ebullition contribution in ALM. Please check whether this reflects rounding, a lower bound imposed during fitting (the recurrence of exactly 0.10 across eight entries is striking), or a genuine fitting problem and constrain the fits so that $F_{tot} \geq F_{diff}$ across the applied temperature range.

AR: We very much appreciate this note that forced us to double check all numbers and repeat all calculations which improved the manuscript. We found a code typo behind these numbers that led to errors in a few tables and figures and in the z-score comparisons among depths. These figures and tables have been corrected and updated in the revised manuscript. The main results were not affected by this apart from some patterns across depth zones becoming clearer.

Dataset documentation is insufficient for an ESSD paper (Sect. 2.2, 2.3, 5). The paper should be self-contained enough that a user can work with the archive without reconstructing the protocol from a separate file. Currently missing: a table of file structure, variable names, units and flags; GC calibration standards, analytical precision and detection limits; blank handling; the effect of vial storage duration on measured concentrations; how failed or lost chambers were treated. On the last point, ERS reports 887 fluxes from 82 campaigns (~10.8 per campaign against a nominal 12), implying routine losses that are nowhere described. A short QA/QC subsection plus a variable table would address most of this.

AR: A detailed protocol covering sampling, sample analysis, and data processing is linked from the revised manuscript. This document provides the information requested and also constitutes the metainformation to the public SITES data portal, while we keep it separate to avoid double-publication. Each of the data files publicly available via the SITES data portal further contain a quality flagging described in the data header and which explains why measurement from individual chambers and sampling occasions in some cases are missing. We now better describe this in the paper.

No code availability statement. Processing spanned Python (curve_fit), a supplemental spreadsheet, and GraphPad Prism. Given that the flux calculations, the k-threshold partitioning and the Monte Carlo routine all involve non-obvious choices, depositing the processing code alongside the data would substantially raise the reuse value of the dataset and is consistent with ESSD practice.

AR: We have tried to clarify the choices in the manuscript text and also provide the core Python code used with meta-information via links to time-stable repositories in the revised manuscript.

Minor comments

L20, L33–34 (Abstract) and L343–344: The claim that “increases of whole season mean water temperature between 0.5 and 3.2°C corresponded to increases in CH₄ fluxes ranging from 16 to 74%” pairs two quantities from Table A3 that are computed over different year pairs. $\Delta_{\text{max}} T_w$ is the range in mean T_w (e.g. ERK 2018 vs 2020), while $\Delta_{\text{max}} F_{\text{tot}}$ is the ratio of maximum to minimum flux (ERK 2018 vs 2017). The two should not be presented as a cause-and-effect pairing.

AR: Agreed and this text is now revised to highlight what the comment points out. We also changed the approach from comparing min and max years to comparing estimated temperature and flux anomalies within the studied time period.

Table 1: TN is used in Fig. A3 but is absent from Table 1. Please add it. Mean annual air temperature and typical ice-cover duration per site would also be useful.

AR: This information has now been added to Table 1.

Table A3 caption: Related, the caption defines $\Delta_{\text{max}} F_{\text{tot}}$ as “the change in emission from the year with the lowest T_w to the year with the highest T_w expressed as percent,” but the tabulated values (121, 138, 174, 120, 116) are max/min flux ratios $\times 100$. Either relabel as a ratio or recompute as a percentage change. Also: “lower and upper bond” $\rightarrow$ “bound.”

AR: We have fundamentally changed this Table structure, and this comparison is no longer made. Instead, we show flux and temperature anomalies versus the mean of the studied period.

L96: “measurements of to seven years” → “of up to seven years.”

AR: Thanks, now fixed.

L109–110 vs Table 1: Asa is given as 57.18°N in the text and 57.17°N in the table.

AR: 57.17°N is correct and this has been corrected.

L134–139: With only two sampling points ($t = 0$ and 24–48 h), the non-linear flux model cannot be verified against the observed accumulation curve. Please note this limitation, and clarify whether fluxes are normalized to the actual deployment duration, which appears to vary by a factor of two.

AR: We have clarified the trade-offs with this strategy in the revised manuscript. We also have clarified that fluxes were calculated and normalized to the actual deployment period.

L193–194: “Hydrological” → “Hydrological.”

AR: Fixed.

L196–200: The text states that regressions were developed “between the observed surface water CH₄ concentrations and three MESAN temperatures,” but Eq. 3 predicts T_w . This appears to be an error — please correct to “observed surface water temperatures.”

AR. Indeed and fixed. Thanks again for careful reading and detecting this typo.

L232–233 and L259–261: Several of the quoted fold-ranges don’t match Table 2. F_{tot} varies ~2100-fold in ERK (not “100- to 1000-fold”); F_{diff} varies ~30-fold in STO (below the stated “50- to 200-fold”). Please recheck.

AR: We thank the reviewer for spotting this. We realize that in some cases, min-max values were used and, in some cases, means. We have re-checked and updated this for consistency.

L252: “becomes important is both measurement design and data processing” → “in both.”

AR: Fixed.

L262–266 (Fig. 3): The F_{tot} regression has a negative intercept ($-0.24 \text{ mmol m}^{-2} \text{ d}^{-1}$), which is physically impossible. Given the strong heteroscedasticity evident in the panel, consider a log-log fit or a zero-intercept model, and note that the reported R^2 depends on this choice.

AR: These equations are not used for any flux estimations. They are simply reported to illustrate empirical relationships between observed fluxes and surface water concentrations which is interesting because of increasing use in literature of only water concentrations to estimate fluxes. Therefore, it is important that physical impossibilities in these regressions, such as those pointed out by the reviewer, are clearly visible and hopefully sends the message that it not reliable to estimate total flux from surface water concentrations. Accordingly, we prefer to not make any adaption, nor adopt log-log relationships which would look much stronger than they are. However,

we have clarified in the legend that these relationships were not used to estimate the fluxes reported.

L286: “Natichimuthu” → “Natchimuthu.”

AR: Fixed.

L293 vs L346: “Bastviken and Johnson 2025” vs “Bastviken and Johnsson 2025” — inconsistent.

AR: Fixed.

L318–323 (Fig. 5 caption): The caption states that significant differences were found “except for F20 for diffusive flux and for Caq”, but the figure shows $p = 0.015$ and $p = 0.010$ for precisely those two panels, with “ns” on the other three. The caption reverses the result and contradicts the text at L313–314.

AR: We forgot a “no” before “significant”. This has now been fixed.

L316–317: “provides support for a lateral transport of dissolved CH₄ from shores towards central parts of lakes” overstates what a near-shore-to-offshore concentration gradient can show; the same gradient is equally consistent with local shallow-sediment sources plus bubble dissolution. Please soften.

AR: We agree with the Reviewer that there could be multiple mechanistic reasons behind the observed pattern of higher concentrations and fluxes near the shore including those stated in the review comment. We here simply aimed at (1) describe the empirical observations of higher fluxes and concentrations near compared to further away from the shore, and (2) mention that the fact that a concentration gradient was observed means that there is empirical support for a gradient driven transport of dissolved CH₄ from the shore towards central parts of the lakes (simply following Fick’s Law). Perhaps our previous wording was ambiguous, and we have tried to clarify it in the revised manuscript.

L365–367: “annual whole lake F_{tot}” is misleading given that ice-cover, ice-out and overturn fluxes are excluded (L218–219). Please use “open-water season” consistently, and note that the omission of ice-out likely makes these lower bounds on true annual emission.

AR: Thanks. We have adopted these suggestions.

L409: “conduced” → “conducted.”

AR: Fixed.

L559 (Table A2 caption): “F20 and θ in Eq 1” should refer to Eq. 2.

AR: Fixed.

Figs. A2–A5 captions: “blown crosses” → “brown crosses”

AR: Fixed. Thanks!

Title/scope: The program is described as a Greenhouse Gas Flux Program but only CH₄ is presented. If CO₂ data exist in the archive or are forthcoming, a sentence in Sect. 5 would help users.

AR: We have clarified the focus on CH₄ data on Section 2.