Comment on essd-2022-346

Bridget Deemer (Referee)

Referee comment on "GRiMeDB: The global river database of methane concentrations and fluxes" by Emily H. Stanley et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-346-RC3, 2023

This data paper substantially updates and expands upon a previous fluvial methane database (MethDB) from a similar group of authors that was published in 2016 (Stanley et al. 2016). It will be an important resource for the aquatic biogeochemistry community in general. The paper identifies key gaps in the existing spatial representativeness of stream and river methane emission data (missing data from arid, high altitude, and arctic regions) and also highlights the lack of long time series methane flux data in fluvial ecosystems. The dataset is structured in a unique way that allows users to explore spatio-temporal variation. Instead of reporting mean emissions for a given system, the database provides within-system site and date specific emissions (and summary statistics across the day when relevant). The dataset will be a key resource for those interested in upscaling aquatic greenhouse gas emissions.

The paper contains helpful visualizations and tables that orient the reader to key aspects of the dataset. One useful addition the authors could consider adding is a database schema diagram that shows how unique IDs can be used to link the four.csv files (concentrations, fluxes, sites, and source). This type of diagram could provide a visual guide for a data user describing which connections are "one to one" versus "one to many". For example, on first look at the data files, I found myself a bit confused as to how I would link the flux data to the concentration data. Is there always a "one to one" link where each row of flux data has a matching row of data in the "concentration" data file?

We created a new figure (Figure 1) that provides an overview of the DB's structure and linkages between the tables to illustrate/explain when a row in the flux table did or did not have a matching row in the concentration table.

Another novel aspect of this database is the inclusion of both diffusive and ebullitive methane emissions and the comprehensive annotation of specific methodological approaches (where the previous MethDB database only reported diffusive emissions). I think this aspect of the dataset should be mentioned in the abstract.

Done

I also think it would be interesting to visualize and/or further describe differences in the emissions that are generated via different methods. There is already some discussion of the potential importance of ebullition in overall fluvial methane flux (lines 522-523 citing certain papers), but the authors do not report any statistics regarding the fraction of emissions that are ebullitive in their own dataset (for sites with independent estimates of both flux pathway). In my work with reservoir methane emissions, the potential predictors of emission became clearer when we stopped combining diffusive-only emission estimates with those that integrated across both flux pathways (Deemer et al. 2016). I think this division may be a bit harder to make in the river literature (since it may be harder to discern which estimates were truly diffusive-only), but I think some basic summary of the flux data by method would be helpful. In looking at the data, it looks like you only have 8 rows of data with total methane flux recorded. You do mention that 85% of the data is diffusive-only, but I'm surprised there are only 8 rows that have estimates of both diffusive and ebullitive emission. This could be explicitly called out.

Investigating the relationship between diffusive and ebullitive fluxes, including possible methodological effects is a great suggestion. However, we have elected not to pursue this suggestion in this paper because we are doing so in a separate effort that has a more in-depth consideration of CH_4 fluxes. However, we will note that given the reality that fluxes (total or diffusive) span 6 orders of magnitude, being able to see methods-driven differences at this spatial scale may be difficult.

With respect to having 8 rows (observations with both ebullition and diffusion data), this is slightly off. There are 64 observations in which ebullition and diffusion were measured simultaneously and independently. We suspect that the 8 observations noted in this comment are observations reporting ebullition and diffusion, but not total flux.

Thank you to the authors for this important contribution to the field.

We appreciate your generous words!

Line by Line Comments

Line 35- If you can fit it, I suggest including the ranges and/or standard deviations here
Ranges and SDs were added to the abstract.

Lines 88-95- I assume beaver ponds were not included as "marginal" fluvial systems, but you might explicitly mention this here. Also, what about river reaches upstream of weirs?
We added 'beaver ponds and immediately upstream of small dams' (weirs) to the list of excluded sites.

Line 131-132- Figure 2 doesn't really make this distinction regarding sites that were used in multiple studies. Consider either adding this or annotating it somehow directly in figure legend.
The first step in the process of entering sites into the Site Table for us was to determine if a site being considered was already entered (i.e., it had been studied before in a separate article). This is represented by the first box in Fig. 2 ("Site used previously?"). Comments were in fact entered regarding a site's use across >1 paper, including identifying the other sources using the site for their additional data collections. So we have not made any adjustments to this figure.

Line 175- Consider changing the wording in this title (and/or in the text directly below it) to "Concentrations Table and Fluxes Table" to make extra clear that they are two separate tables.
Done

Line 183- Delete duplicate use of the word "both"
Done

Lines 221-228- So, is it true that in some cases the same concentration data might be applied to many rows of flux estimates (one to many)?
This relates to the earlier general comment about how the tables in GRiMe are linked, and hopefully the new Fig. 1 helps to clarify this situation. To respond to this specific question, no- it is not true that one concentration observation would be applied to many rows of flux estimates. If a site-date combination has concentration data with supporting water chemistry and also flux data, there will be 1 row in the Concentration Table and 1 row in the Flux Table for this site-date combination. If the site-date combination has flux and chemistry but no concentration, there will also be 1 row in the Flux Table and 1 row in the concentration table. In this case, the Concentration Table entry has the supporting water chemistry data, but does not contain CH_4 concentration data (since it doesn't exist). If a site-date combination has only flux data (no concentration, no supporting water chemistry), then there will be 1 row in the Flux Table and no corresponding row in the Concentration Table.

Lines 287-289- This pattern is also true for lake and reservoir methane data—65% of the lake/reservoir methane emission estimates in a recent dataset were collected since 2015

(Rosentreter et al. 2021, Deemer and Holgerson 2021)
Not surprising! Perhaps data increases reflect the growing use of portable GHG analyzers.

Lines 292-294- Wow! I can't believe how short the longest flux record is—much shorter than the lake literature.
Yes- it's very surprising. We hope drawing attention to this brevity will inspire continued collection of records to generate longer time series.

Line 335- Do you mean 4% of the global river surface? I don't think you mean land surface from looking at the map (more than 4% seems to be shaded darker tones of orange, but some of this is in surface-water poor areas like the Sahara).
Yes- thanks for catching this omission. Wording has been fixed.

Line 408- Include the definition for "IMP" like you do for the other site types.
Done. Also done for "TH" later in the paragraph.

Figure 12- The relationship between flux and total N & P looks stronger than for DOC or dissolved oxygen, but this isn't called out where you discuss drivers (lines 571-590) . You might consider citing some of the wetland, lake and reservoir literature that has linked methane emission to productivity/chlorophyll a.
Text was added to this paragraph in the discussion about the relationships between TN, TP, and CH_4 emissions along with noting papers that have found similar positive relationships between CH_4 and eutrophication/nutrient enrichment.

Line 480- You could discuss insights on spatial/temporal resolution from the lake literature here. Wik et al. 2016 Geophysical Research Letters showed that spatial and temporal under-representation generally led to underestimates of emission in lakes.
We are hesitant to suggest that limited temporal sampling is likely to lead to underestimated emissions in streams and rivers at this point, given the very different controls on gas concentrations and fluxes between lentic and lotic systems.

Line 509- Remove either "few" or "several"
Done

Line 522- It isn't clear if this 30-90% range comes from your entire database, or just from the few papers cited here. Or maybe there are only three papers that quantify both pathways together? In the lake and reservoir literature, the fraction of emissions that are ebullitive can range dramatically (undetectable to almost all of the emission), with ebullitive emission contributing a median of 78% of methane emissions in reservoirs and 54% in lakes- Deemer and Holgerson 2021).
The wording was edited to clarify that the 30-90% of total emissions was reported by the papers cited here.

Line 584- Rosentreter et al. 2021 also used latitude to upscale stream and river emissions.
Citation added

Line 599- I thought Burns et al. 2018 reported rather high methane emissions from glacial systems?
Although Burns et al. argue for very high emissions, outflow concentrations are actually below the global average.

Line 628- Add the word "from" between "data" and "world"
Done

Tables A3 and A4- I suggest explicitly clarifying that "new" units are the relevant units for the data you report.
These table describe the 'new units' as the current common units for all concentrations