Review of manuscript essd-2022-346
"GRiMeDB: The global river database of methane concentrations and fluxes" by Stanley et al.

GriMeDB is a valuable compilation of published and unpublished global data of methane (CH_4) concentrations and fluxes in flowing waters (rivers and streams), and the add on of values for CO_2 and N_2O. This valuable data set can support improving global and regional carbon and nitrogen budgets.

It is well known that rivers are generally emitters of relevant greenhouse gases to the atmosphere (i.e., CO_2, CH_4 and N_2O), and an updated global compilation in fluvial systems was very much needed, especially as the authors point out, with the increasing density of measurements over the last years in response to improved technologies. Such compilation was necessary in order to better quantify the role and contribution of fluvial systems to greenhouse gas emissions, as well as to better identify current geographic gaps.
The authors not only compiled these data sets into one single source, but also carefully curated the data in order to standardize the data base (e.g., use of same units in concentrations, daily aggregated values). Additionally, they provide interesting information as a result of statistical analysis showing trends in the data available regarding spatial coverage and temporal trends, over- and under-sampled areas and values with respect to channel types. These results certainly bring new opportunities and recommendations for future studies.
The GriMeDB data base includes detailed information on the methods used for quantification of the three methane emission pathways (diffusion, ebullition and plant-mediated transport), pointing out the large dominance of diffusive fluxes, followed by total fluxes (sum of diffusion and ebullition). Hence, they also provide information regarding gaps in current methodologies.
The data is easy to access and handle, organized in tables with clear description in headers for any user. The manuscript is well written and contains clear flow charts that explain to the reader the protocol followed for the preparation and curation of the data sets. The graphic work is also clear and contains summarized information of the global data base analysis. I have only few recommendations/comments and minor comments in the hope that these can be useful to the authors.
I can only thank the authors for the great effort put in this work. I am sure that the great effort made by the authors to compile this data set will be greatly appreciated by the community, and therefore I recommend the publication of this work in ESSD.

General comments:

- The authors mentioned that the diffusive fluxes have large uncertainties due to the various methods used to calculate the gas transfer velocity *k* in the different studies. This is a well-known issue, and k is the most uncertain parameter, and what brings the largest uncertainty to any diffusive gas calculation. I wonder if the authors are able to provide in this context:

1) An estimate of uncertainty in the flux calculation to estimate the error of the comparison between fluxes?

2) This might be out of the scope of this study, but having an overview of all the studies and summarizing the >25 different references for *k* model sources, can the authors provide an own view/recommendation of the "best model", or at least "most commonly used" model, for estimation of *k* in fluvial systems? This can benefit greatly to future studies to at least make use of a common model, allowing for a reduction in the uncertainty between studies for comparison purposes. This might vary in channel type though, but a distinction between e.g., hydraulic models, wind speed-based models, might hint at a best approach.

- Can the authors assess the uncertainty reduction that the data compiled in GriMeDB provides when compared to the estimate contribution of fluvial methane emissions or wetland streams (WS category) in the current global methane budget (Saunois et al., 2020)? In fact, the dominant bottom-up inventory of inland waters in the global methane budget is from wetlands, and including rivers and streams to these budgets might contribute to reduce the difference between top-down and bottom-up approaches.

A comparison to current budgets e.g., global methane budget should be assessed in order to compare the previous knowledge and added value that GriMeDB gives in reducing the uncertainty of carbon fluvial emissions.

- The authors do not mention in the manuscript if GriMeDB has the potential to be a living data base with a call to scientist to inform and/or deposit new (self-curated) data and to keep this data set growing.

The current GriMeDB contains I believe the majority of studies published with CH4 data in fluvial systems, but this list is not exhaustive, especially with missing recent published works (some in spotted under sampled regions), such as:

Canning et al., 2021, https://doi.org/10.5194/bg-18-3961-2021
Castro-Morales et al., 2022, https://doi.org/10.5194/bg-19-5059-2022
Patel et al., 2022, https://doi.org/10.1016/j.watres.2022.119380
Wesley et al., 2022 (still work on discussion, unpublished),
https://doi.org/10.5194/egusphere-2022-549
Zhao et al., 2022, https://doi.org/10.1016/j.envpol.2021.118769
Zhu et al., 2022, https://doi.org/10.1038/s41467-022-31559-y

Maybe some of these studies, do not meet the requirements for GriMeDB, but it is just an example of current works that could be potentially added to the data base.

Additionally, and as part of my comment, can the authors provide some sort of protocol for sampling and curation of future or current available data set that are not yet part of GriMeDB and can help the authors to include and expand GriMeDB with community contributions. An example of such living data base is SOCAT (https://www.socat.info/), which provides clear protocols for scientists that wish to contribute with their data sets to this growing data base. This might not be on the scope of the authors at this stage, but given that they host the method it would be interesting if a call to the community can be done on the interest of keeping GriMeDB growing with a certain quality control.

Even if there is no current perspective by the authors to include more data sets to the current data base, an added short protocol as part of this manuscript would be extremely useful. Such protocol can include, e.g., parameters to measure and report (e.g., concentration in certain units), method used to determine *k*, measurement or information of other parameters (e.g., channel type following Strahler scale, channel slope), in a way that it can facilitate future additions of new data sets to GriMeDB.

Already the authors provide certain general recommendations (e.g., determine and report detection limits and include samples falling below these limits, include information on habitat conditions, studies expanding temporal dimensions are encouraged, routine CH_4 measurements with as part of water quality monitoring programs are encouraged), that can also be included and summarized more clearly in a protocol, with potential of including recommendations for e.g., *k* parameterization to be used. If the authors cannot provide a protocol at this stage, I encourage them to add a section 4.3. in the discussion section with "Recommendations for future studies".

Minor comments:

P4,
L4 – How the authors assessed the quality control in unpublished data sets?
L84 – EU Zenodo is missing

P7,
L141 - it is better at this stage to mention that R package was used and not until section 2.5
L159 - It is missing the code NORM in Table 1, it is only mentioned in the caption of Fig. 14

P10,
L222 - it is necessary to add explicitly the units of concentrations and fluxes in each column, where it corresponds, in Tables A3 and A4, otherwise it is only possible to visualize the units by accessing the headers at the tables directly.

P21, L371 - there is a dot instead of a comma after the parenthesis

P22,
L396 – I believe the words "between CH_4 physical site attributes" need to be removed, so the sentence can only read "As with relationships between CH_4 concentration of flux and water chemistry parameters … "
L398 – Did the authors try to calculate correlations of Figs. 12 and 13 per latitude bands? They can be biased due to density of observations but at least some meaningful correlations might be seen between the selected parameters and methane.
L401 – refer here to Fig. 13a
L403 – refer here to Fig. 13b

P23, L408 – define here IMP as (impounded reaches), as all the other listed channel types were defined in this paragraph, the same for TH (thermogenic CH_4 inputs) in L415.

P25, Fig. 14 – These site-averaged concentrations need to be normalized to sample size so variations can be reduced due to the varying sizes and a better comparison between channel types can be done.

P29, L551 – “compared **TO** our previous efforts”

P30, L610 – additionally data assimilation models will strongly benefit from the GriMeDB database

P31, L628 – “… the expansion of GHG data **FOR** world streams and rivers …”