

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2021-141-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-141**

Anonymous Referee #2

Referee comment on "BAWLD-CH₄: a comprehensive dataset of methane fluxes from boreal and arctic ecosystems" by McKenzie A. Kuhn et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-141-RC2>, 2021

This paper describes data that combines terrestrial and aquatic CH₄ flux measurements and associated supporting information for northern boreal-arctic biomes. As the authors clearly state, methane-producing habitats are particularly abundant in this part of the world and current estimates of boreal and Arctic emissions are highly uncertain. This uncertainty is not unique to latitudes above 50°N, but (not mentioned by the authors) it is arguably particularly important in this case given predictions and early evidence of increasing rates of emissions associated with rapid climate change.

A key 2-part feature of this paper is (a) its construction in tandem with a land cover data set and (b) the argument supported by preliminary analyses that land cover classes can be identified/assigned based on their CH₄ emission behaviors. The case for this argument is compelling for the boreal and Arctic region. But even if a researcher is skeptical about this argument, this dataset includes both flux data and the methane-relevant land cover data, thus providing the raw material for hypothesis testing as well as for cross-system comparisons and upscaling studies. The combination of terrestrial and aquatic fluxes and support data in the BAWLD-CH₄ dataset is a significant expansion of prior and separate terrestrial and aquatic data compilations, and when paired with the land cover data (BAWLD spatial data) it creates a 'one-stop shopping' resource for researchers studying high latitude GHG dynamics. I appreciate the clear delineation of what data are/are not included and potential limitations of the dataset. The limits of the current state of the data serve as useful raw material for ending the discussion with a section on directions for future research. There is much to like about this paper; it provides the best-available dataset for those interested in high latitude methane dynamics, which is a topic of substantial and rapidly increasing scientific interest, the paper is very well written, and it is thorough in the presentation and explanation of dataset assembly.

Most of my specific comments emphasize aquatic data issues due to my familiarity with these ecosystems. However, I have two general comments. First, following a preliminary reading of this manuscript, I mistakenly thought that this paper was discussing two complementary data sets: BAWLD and BAWLD-CH₄. I made this assumption because it makes sense that these datasets would be presented together given their construction, and also because a fair amount of text in this paper is dedicated to describing aspects of BAWLD (the land cover classes). Presumably the land cover categories are described in

the Arctic Data Center (ADC) data publication, so could this part of the paper be streamlined to put the emphasis back on the CH₄ flux data?

Second, given that a strength of this dataset is being able to link it to the BAWLD spatial dataset, it was surprising and disappointing to see several rows with missing or highly aggregated information on site location, including data records from papers written by co-authors on this effort. Often, many sites are lumped together and assigned the same very coarse-level lat/lon (e.g., 64, -148) despite including different types of ecosystems. This is disappointing, as it restricts future analyses that could have been done if and when new and better spatial data become available. I understand that nothing can be done about some of these records, but this is not the case for all records with vague/absent data.

Aquatic storage flux (lines 392-395)- Can you provide any more detail about determining storage fluxes? I think of this flux as being estimated by quantifying the mass of CH₄ under ice and then assuming that this mass (or some fraction of this mass) is lost to the atmosphere following ice out—because of mixing. Thus, I am not clear on how storage flux is being differentiated from spring mixing flux.

Section 3.3 figure citations- do you mean Fig. 10 (not fig 2.10) in this section? And re: fig. 10, while the 2 different blues are easily distinguishable in Fig. 9, they are less so in this figure.

Lines 248-249- I am also unaware of flux measurements taken from glacier surfaces, but there are now a handful of papers quantifying fluxes at glacial outflows or termini (of course including the Anthony et al. 2012 paper). It may be helpful to add in a sentence or two as to how these habitats were handled or why they were excluded given the increasing interest in glacial retreat and GHGs.

Line 376- There are several k models, beyond Cole and Caraco; is this the dominant model used in data sources? (why was it singled out?)

Line 402- E.LOCATION refers to the location for ebullition only, and this should be specified here. But more importantly, it seems odd to have this category (1) only for ebullition and not also diffusion and (2) given that there is only one row corresponding to a single observation that reports an ebullition flux measured from the lake edge. In short, this field does not really provide much useful information, so is it worth including?

Line 410/Table 2- Very minor! "Dataset" is used in the Table, "Database" is in the csv

Line 500/Fig 8- There may be a better way to report that shared letters denote no

significant differences among categories within each comparison set in this figure.

Line 616-617- small suggestion: add in 'spatial' before differences to remind readers of this focus.

Line 646- depth may also not be a significant predictor because of unreliability of depths that were reported. In the absence of thorough surveys, depths are often estimated or measured at the center of the lake or where fluxes were measured, and thus may not be a reliable measure of lake mean or maximum depth.

BAWLD_CH4_Aquatic.csv and BAWLD_CH4_Terrestrial.csv contain data drawn from papers that are not cited in the body of the manuscript, and thus are not cited in the reference section of the manuscript. I was also. Unable to find a list of citations for data sources for these files on the ADC dataset web page. It seems appropriate to include this information somewhere- perhaps as a supplement (and apologies if this information was overlooked).

Very minor grammatical details-

Line 74- 'like' means similar to, whereas I think you are saying that marshes and fens are actual examples of graminoid-dominated wetlands (long way of saying- use 'such as' instead of 'like')

Lines 77-80- this is a run-on sentence. Can it be divided into two at the point of the second 'which'?