

Earth Syst. Sci. Data Discuss., referee comment RC2  
<https://doi.org/10.5194/essd-2022-8-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on **essd-2022-8**

Kristen Manies (Referee)

---

Referee comment on "A high spatial resolution soil carbon and nitrogen dataset for the northern permafrost region based on circumpolar land cover upscaling" by Juri Palmtag et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-8-RC2>, 2022

---

This manuscript describes the data collection methods for over 6500 soil samples and uses these data to provide C and N stock estimates for the circumpolar permafrost region. These data are very useful and important. The Introduction and Results are presented well. However, I do think that the Discussion section is missing a paragraph discussing the caveats of the data. This information is only briefly mentioned in the Conclusions (line 415) and needs to be discussed more in depth in the Discussion portion of the paper. As stated in line 415, their data are concentrated in non-North American locations, such that a more complete picture could be obtained by combining their data with other datasets. In addition, the dataset only contains one high alpine site, so this ecosystem type is underrepresented. I don't expect this manuscript to do analysis beyond what is presented here, but I do think it's important to be clear about the limitations of their data and what next steps (i.e., combining with other datasets) could be taken to expand our understanding on this important information.

In addition, there were many times when I was reading the methods that I had unanswered questions regarding specific methodology and/or how their methods might impact data quality. I would like to see many, if not all, of the following questions answered, such that others who want to use the data truly understand how it was collected. Areas where I had questions include (line #: question):

119: Do you mean that the following field descriptions were classified as wetlands? Also, why is "mineral" a wetland?

124: I don't understand what your reason is (as no reason was stated in the previous sentence).

127: How did you define the Yedoma region? I don't think that this classification is something you can determine with site photos.

135: How many soil descriptions per site usually?

145: Does the top organic layer mean all organic soil? Or does it mean organic soil to a certain depth? Or organic soil to a certain (estimated) bulk density?

146: If a steel pipe was not used, how was permafrost soil sampled?

[It would probably be helpful for the reader if the paragraphs from line 145 and 152 were combined. There would be less duplication and some of my questions raised reading the 1<sup>st</sup> paragraph were answered when reading the 2<sup>nd</sup> paragraph.]

156: Was the active layer never deeper than 50 cm? If so, how was the deeper active layer sampled? There must have been areas where the organics were deeper than 50 cm, especially in the wetlands. How were these soils sampled?

156: The way this sentence is currently written it sounds like only in the "few cases from natural exposures" were the horizontal sampling rings used. But, according to Figure 2, this is the method used for the entire active layer. This sentence needs to be rewritten to clarify this point as well as include the information requested below.

157: I'm also worried about the sampling that happened at fixed depth intervals. How frequent were these intervals (every 5, 10, 20 cm)? Could you have missed changes in soil horizons (and thus bulk density and C concentration, affecting your C stock values) by sampling this way?

158: I don't understand what "emphasis" means here. Or what was done when there was a lot of spatial variability within a soil pit. I think additional descriptions (i.e., depths were measured every 10 cm on the photograph and then averaged) or an example is needed here.

For clarity to the reader, maybe describe the normal way you measured the active layer. Then give the details about the special cases (natural exposures, spatial distribution).

170: Please clarify that the length of this pipe was measured each time it was used, so that the bulk density measurements are accurate.

185: How were you able to do hand manipulations on the frozen sections? Did you let part of the sample thaw and then test for soil texture? Also, if you are taking these subsamples out for texture analysis, did you return that subsample to the bag so that the weights remained accurate?

213: It'd be nice to have a few more details about how the determination of the presence of inorganic C was done. For example, were they chosen by eye? Or if the sample fell a certain percentage off the 1:1 line between LOI550 – LOI950?

217: Since you are using LOI data to predict C for some samples, I'd like to know a) the percentage of samples for which this predication was used (i.e., no C data available) and b) how good the fit of this relationship for these data are.

237: I am confused why you took three organic samples when only one was used for C stock measurements. I'm assuming that you only used OL1 because it matched with the rest of the soil profile. Maybe clarify that the other two samples were taken to quantify the variability of this layer (if this is the correct reason) and are available as data for others but aren't considered in these results.

Additional, but minor suggestions, are as follows (preceded by the line number):

37: Are you missing a verb here? "to be 380 Pg"?

38: I found this sentence a very confusing.

41: What is the difference between the 2 datasets? It would be nice to have this detail, so readers know which link to use.

66: misspelling "volumetric water content for organic soil"

68: The word "cover" before "stones and boulders" initially made me think you were looking at those data as a percent cover. Consider removing this word (maybe need to say percentage of stones and boulders?).

105: I found these two sentences confusing and think would be more understandable if they were a) a part of section 2.1 and rewritten a bit. For example, "All sites were classified with Tier 1 descriptions using field descriptions and, where possible, assigned a more detailed (Tier 2) description."

120: It might be clearer if you say something like "Where the permafrost status within the top 2 m of a site was known, a Tier 2 status was assigned."

165: Please revise – you say earlier in this sentence that these soils weren't always frozen.

191: I was confused why this information was presented, especially since I didn't see this information discussed in the results or presented in the datasets.

Figure 2 is very nice.

214: If you place the information that these regressions were done for each study area in this sentence readers won't be left wondering (as I was) until they read on.

218: Aren't C:N ratios usually based on percentages of these elements, not weights?

220: Simpler to say "more decomposition"?

222: This ratio? I'm confused what "this" is referring to.

270: I'm confused by the words "indicated by permafrost area". I don't understand what this phrase is clarifying for the previous statement "but not the actual area underlain by permafrost"?

271: Simplify to "This dataset"?

273: I suggest you move this paragraph to right after the paragraph on line 260 that

introduces the ESA database.

283: I think the first Tier mentioned should really be "Tier I".

286: Better to put the equation here?

346: I don't see Yedoma tundra (yellow line) on the Figure 5 panels for the silt+ clay nor sand panels. I also see a lot of variability in the Non-permafrost wetlands for many data types, so you may want to also mention this class.

355: The scientific communities don't have high spatial resolution, the dataset does.

358: I think this sentence is a better topic sentence, with the sentences following this sentence explaining how it's better than what previously existed.

374: Despite? Or because of different upscaling approaches? I find this and the following sentences to be confusing/too wordy. I think you need to focus on the points: although your values are a bit lower than their estimates, they're within each other errors. You used different upscaling approaches, which could be the cause of some of these differences. Your upscaling approach was chosen because...

382: I found this wording confusing. Maybe "estimate of 66 Pg (+/- 35 Pg) by Harden et al."?

384: If you're going to have this paragraph on C:N ratios in here I think you need to discuss your results more (i.e. how they vary with land type, etc.). Right now it's just saying what you already said at line 218. There are other data you don't discuss. Maybe the focus of this paragraph should be about the other data available in this dataset and what their uses could be? Otherwise, I'd delete this paragraph.

390: I think "although" fits better at the beginning of the sentence as it's currently written.

393 & 395: This what? Make sure to follow the word "this" with a noun so readers don't get confused about the subject you are discussing.

395: I don't think you need to say "in this study" here.

396: I'm not sure "throughout" is the appropriate word here since you're only discussing wetland classes. I suggest deleting it.

397: It might be clearer to say "exchange the ESA wetland areal coverage for the values in Hugelius". Also, you give us your updated estimates, but please remind us how those relate to the other estimates and what those values are. (Otherwise I have to reread the paper to find them.)

401: I think your argument needs to be that you present a more complete dataset in regard to variables used to parametrize models. Because other data sets have similar data, maybe just not to the completeness you do.