

Comment on **essd-2022-8**

Anonymous Referee #1

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-RC1>, 2022

Major comments

The authors present a very large dataset of soil samples from different land cover classes in the permafrost region. They used the data to calculate carbon and nitrogen storage estimates for the northern permafrost region with an upscaling approach. This is a very valuable study and especially the nitrogen storage estimates as these have not been the focus of many studies in permafrost regions. However, there are some things that could be improved to this study which I listed below, and I recommend major revisions of the manuscript before publication.

I agree that it is important to distinguish between Yedoma and non-Yedoma sediments. However, I don't understand why this distinction goes only so far to say that there is Yedoma tundra or forest. Isn't it also important to distinguish between the broadleaf and needle leaf forests within the Yedoma region, and between the shrub and graminoid / forb tundra? Also, I would leave out water bodies and snow/ice out of this table, as you have no samples from these land cover classes, and only mention them in the text instead, just to make the table clearer. Thus, I would propose to introduce Yedoma as a separate level (or tier) of the land cover class system and distinguish between Yedoma or non-Yedoma sediments, then between the four classes (forest, tundra, wetland and barren) and then the corresponding subclasses.

I think the methods chapter is very long. You could consider to move part of it to the

supplements and have a more concise description in the manuscript itself.

I am completely missing the description (results) and interpretation (discussion) of the spatial distribution of C and N storage, as well as from the other soil parameters (C/N ratio, $\delta^{13}\text{C}$, BD, volumetric fractions, texture). Please incorporate this!

Minor comments

Abstract

L34: please rephrase "within the soil area" (for example: "in soils in the northern...") as you report the C and N storage estimates for a volume, not an area.

L38: the sentence should not start with "of which"

L40: "but show different spatial patterns" -> this is the only time in the paper that you say anything about the spatial distribution

L41-43: this is not the right place to cite these datasets

Introduction

L47: temperatures can't warm, please rephrase to "warming of the soil" or "increasing soil temperatures"

L50: isn't the accelerating you mention part of the feedback? Please rephrase

L54: introduce the abbreviation OM in line 48 instead

L60: I am missing what you are upscaling, maybe you can add the word data or estimates

L69: here you refer to data as singular whereas in L51-52 you are referring to data as plural

L72: I think you mean aims here (what you hope to achieve)

L76-78: I would not introduce another aim here, you can leave out this sentence

Methods

The subchapters about the sampling, lab analyses and calculations all fall under the main chapter "Dataset structure" which I think is not so fitting. Consider to rename the chapters such as: 2. Methods, 2.1 Dataset structure, 2.1.1 Class definitions of soil pedons to land cover types, 2.2 Soil sampling, 2.3 Laboratory analysis, etc.

Also, you could combine chapter 2.4 and 2.5 or make a clearer distinction between the chapters.

A few times you mention "at most sites", "for some locations", "occasionally", "normally", "when possibly", etc. This is not very helpful if it is unclear why you only carried out certain procedures on a subset of the sites and what happened for the other sites. Please explain

L101-102: use small letters for the land cover types

L131-132: you can remove this here

L134-136: this is a bit vague. How many samples were taken in these 100-200 m intervals?

L150: can you indicate here how many soil pedons exceeded 1 m or reached 3 m?

L153: rephrase "measuring the block volume in the field" to "and the block volume was measured in the field"

L153-154: in L145 you mentioned you took 3 replicates samples of the organic layer but here you say that you took replicates sometimes. Later (L237) you say you only used the first of the three replicates. Why is that? And even later (L287-288), you mention the replicates were only considered for pedons reaching the full depth. Do you here refer to other replicates than the organic layer replicates?

L158-160: what is the relevance of this information?

L160: permafrost-free should be non-permafrost

L168-L171: this can be left out

L172-172: what do you mean with "following recovery"?

L175-176: I think this sentence should be moved up before the lengthwise splitting

L188-191: please move this to the chapter "Laboratory analysis" and rename this subchapter "Soil sampling"

L191-193: I think these sentences are not necessary. If they are, please put them into context

L202-208: this is unclear to me. How did you determine the bulk density with only the weight before and after? Why did you not dry all samples at 60-70 °C and 105 °C so that you can use the samples dried at a low temperature for further analyses and the weight difference from samples dried at a high temperature for the calculations? Was the correction really necessary or in other words, was the weight difference very different for the samples dried at low and high temperatures? If yes, how can you assume that the weight difference is correct for those samples where you did not dry the subsamples at the higher temperature?

L211: rephrase "every second sample" to "half of the samples"

L215: introduce the abbreviation organic C % here

L216-217: why in most cases? What was the alternative?

L221: write 13 in $\delta^{13}\text{C}$ in superscript

L242: explain "from laboratory results" better

L256-258: in line 231 you mention different intervals. Why did you average the values with a 1 cm resolution if you use 1 value per depth interval for the actual calculations?

L264: what do you mean with "majority statistics"?

L266-268: I don't understand what you mean with this sentence

Results

I find the pedon grouping confusing. You binned the data into intervals of 0-30 cm, 0-50 cm, 0-100 cm, 100-200 cm, 200-300 cm and 0-300 cm. This is not consistent as the 0-100 cm interval contains the 0-30 and 0-50 cm, but then you separate 100-200 and 200-300 cm. Why? It would be clearer to have intervals from 0-30, 30-50, 50-100, 100-200 and 200-300 cm and then have the "summary intervals" 0-100 and 0-300 cm. This way, the amount of pedons in Table 4 would also add up and it would be clearer how many pedons cover what intervals.

L299 and 323: use a different word instead of "bulk"

L302: the graph also shows the distribution of the depth 0-100 cm; please describe the results of the spatial distribution of the C storage (and the same for N in the next subchapter).

L321-322: can you back this statement statistically? I can't really confirm this in Fig 5.

L342: rename this subchapter and describe the data visualized in Fig. 5

L347-348: what is the relevance of this information?

Discussion

I would recommend to restructure the discussion to better follow the two study aims. It feels like the first paragraph can be left out as it repeats parts from introduction and methods.

L376: rephrase and clarify "within each other's error estimates"

L377: "in comparison" does not fit to the sentence; I suggests to move this sentence as more of an outlook

L384-388: this paragraph is very general and is quite similar to the text in the methods. Instead, really discuss the actual data.

L387: reformat d13C; with "locates the areas... vulnerable to permafrost degradation" don't you mean the organic matter vulnerable to decomposition? Or can you please explain how you can define vulnerable areas with the $\delta^{13}\text{C}$ and C/N values?

L397-400: this feels a bit awkward, as you chose the surface areas from the land cover map for a reason and now you say your area is wrong?

Conclusion

L409-411: first start to mention the actual SOC estimates and then you can mention that this is lower than previous studies (although not significantly?). The part about the wetlands is not needed here I think.

References

The notation of the DOI in the references is not consistent: mostly it is written as <https://doi.org/10...> (which is the correct way), but sometimes it is written as doi.org/10..., [doi:10...](https://doi.org/10...) or [DOI:10...](https://doi.org/10...)

Hugelius 2012: move year to the end

Kracht and Gleixner (2020): DOI is missing

Figures and tables

Figure 1: source of map should be Natural Earth Data

Figure 1, 3 and 4: add a space between the degree sign and the direction

Figure 3 and 4: add labels to figure panels (a) and (b)

Figure 5: write parameters in the caption

Table 1: add degree sign and direction for the longitude and latitude

Table 4 and 6: add unit of depth

Supplements

Please add more information to the caption of Figure S1 to explain what information is recorded for every sample.

Formatting

Please make sure to check the journal's guidelines on figure content and mathematical notation and equations (e.g., spaces between number and unit, units written exponentially)