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.

First of all, thank you very much for your valuable comments. We fully agree, that it is important to distinguish Yedoma in the way we did with the other classes (forest, tundra, wetland and barren) but unfortunately, we have too few sites. As shown in table 5 and 7, we only have 8 pedons to 100 cm in Yedoma tundra and only 1 pedon in Yedoma forest. However, the idea was to create this subdivision and to point out the importance. We also introduced the tier levels, which can be extended beyond tier 2. As mentioned in several places, there are other areas (Canada, Alaska, Tibet, high alpine areas) which are
underrepresented. Since we provide all the parameters and coordinates, this dataset can be easily combined and extended with these important but underrepresented areas.

As suggested, we removed the water bodies and snow/ice classes to make it clearer for the reader.

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, δ13C, BD, volumetric fractions, texture). Please incorporate this!

The methods chapter is now shortened. Additional information and interpretation about the spatial distribution of C and N is added in both sections, in the result and the discussion section. The core objective is to provide and describe the dataset and secondly, to quality test same dataset to quantify the carbon and nitrogen pools within the northern circumpolar permafrost region. As described in the methods section, there are many more soil parameters available but are beyond the focus of this “Earth System Science Data” paper.

**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.

Changed as suggested

L38: the sentence should not start with “of which”

I removed this sentence as it is not relevant for the abstract.

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

Thank you. I removed the “different spatial distribution pattern” part from the abstract as we are not actually comparing the pattern here.

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

I removed the references and added following as suggested by ESSD: “Dataset references and DOIs are presented in the “Data access” section in the end.”

*Introduction*

L47: temperatures can’t warm, please rephrase to “warming of the soil” or “increasing soil temperatures”

Thank you for the comment, changed as suggested.

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

I rephrased the sentence to “The release of these greenhouse gases to the atmosphere would in turn generate further climate change, resulting in a positive feedback on global warming (Turetsky et al., 2020).”
L54: introduce the abbreviation OM in line 48 instead

Introduced the abbreviation earlier as suggested

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

Added estimates as suggested.

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

Changed to plural here as well, thank you.

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

Changed to aims

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

I removed 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.

Changes made as suggested: Methods chapter restructured and section 2.4 and 2.5 combined.

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

Agree and thankful for the good comment. Methods section restructured and partly rewritten. All the above issues are now hopefully solved.

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

Changes as suggested.

L131-132: you can remove this here

Removed as suggested

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

Additional information added in the text “582 out of 651”

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

The number of pedons which extended below 100 cm, is 313. This number is added. In
addition, the n for pedons reaching the depth of 100, 200 and 300 is shown in Table 4.

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

Changed as suggested.

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?

Thank you for the comment and pointing out the confusion. As explanation, in some cases, 3 replicates were collected for the organic layer due to substantial variation, but these were not used in the SOC/TN calculations. Sections where OL replicates were mentioned are now removed as they are not part of the calculations and therefore nor relevant in this manuscript.

Comment to the line 287-288: Yes, replicates here refers to sampled pedons. A pedon was only considered suitable if the full depth was reached. Since the OL replicates are now removed from the text, hope this is now more clear.

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

Agree and removed as not relevant here.

L160: permafrost-free should be non-permafrost

Changed to non-permafrost

L168-L171: this can be left out

Removed from the manuscript

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

The whole part was rewritten to: “Samples were split lengthwise into two halves: one half was analyzed to determine sediment characteristics, volumetric ice content, and gravimetric water content.”

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

Moved up

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

As suggested, moved this part to the Laboratory section and renamed the section.

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

Agree with the reviewer that this part is not giving any necessary information and is therefore removed from the manuscript.
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?

Added additional comment that the bulk density was obviously determined from the weight and the volume. Also, comment to the drying procedure. All samples, now with a n of 3684 out of 5230, which were organic rich or fine grained were dried again at 105°C to exclude the possibility of remaining water. Remaining samples which were not dried again, were sand or course grain samples and showed in tests no remaining water.

L211: rephrase “every second sample” to “half of the samples”

Done

L215: introduce the abbreviation organic C % here

Done

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

Changed to: “A third or fourth order polynomial regression model was used...”

L221: write 13 in δ13C in superscript

Changed to superscript throughout the manuscript

L242: explain “from laboratory results” better

Changed to: “were calculated based on the laboratory analysis for all the individual samples." How it was calculated is described below.

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?

Thank you for pointing out the confusion. Our data is on 1 cm resolution. But the depth increments are used for comparison with other publications. Text now changed in the manuscript to: “SOC content for each pedon was calculated by summing up individual samples on 1 cm resolution until the maximum sampling depth was reached. The pedons were assigned to a specific land cover class and the SOC content averaged for different depth intervals (30 cm, 50 cm, 100 cm, 100-200 cm, 200-300 cm, and 0-300 cm).”

L264: what do you mean with “majority statistics”?

Majority statistics referred to the fact that the land cover type that occurred most frequently during the selected years was chosen.

To make it clear, the wording “majority statistics” was removed and sentence changed to “by identifying prevailing land cover classes within this period”.

L266-268: I don’t understand what you mean with this sentence
Thank you for the comment and fully agree, that this sentence is rather confusing and not of relevance here. Sentence removed.

**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.

The intervals in both tables are now regrouped following your suggestion.

L299 and 323: use a different word instead of “bulk”

Exchanged to most in both cases.

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).

A few more additional statements added in the result and the discussion section.

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

This statement is removed from L321-322, but added additional information in the section 3.3 “The data shows clear differences occurring in the more variable top meter in comparison to the rather stable second and third meter. With an exception in Non-Permafrost wetlands where the TN is more variable below 100 cm depth, which results from only 2 stratigraphy different available pedons where TN data is available (Table 7).”

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

Chapter title renamed to “Soil stratigraphies“ and additional description to several other classes added.

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

We removed this sentence from the manuscript.

**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.

Discussion partly restructured, rewritten and first paragraph greatly shortened.

L376: rephrase and clarify “within each other’s error estimates

Sentence rephrased to: “Although our values are a bit lower than their estimates, they are within each other errors.”

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

Agree and this part is now in the last section of the discussion.

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

Following the comment from reviewer 2, this paragraph is now removed. The focus to test the database is on SOC and TN data. Several other parameters are only mentioned as part of the data, but not discussed.

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 δ13C and C/N values?

Section removed, see above.

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?

Thank you for the understandable comment. We choose the ESA product because it offers a high resolution land cover product on a circumpolar scale. Even it is a great product, it has its limitations and unfortunately a relatively low accuracy in the very important for SOC and TN natural and semi-natural aquatic vegetation class. We conclude, that this is partly the reason for our lower estimate. But we also pointing out several other underrepresented areas which additionally increase the uncertainties. At the same time, this is a point based database which can be simply extended with additional sampling sites or to different land cover products.

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.

Restructured and the wetland part removed.

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... or DOI:10...

DOI’s are now updated using the correct way.

Hugelius 2012: move year to the end

Moved to the end.

Kracht and Gleixner (2020): DOI is missing

DOI added

Figures and tables

Figure 1: source of map should be Natural Earth Data
Source text “Made with Natural Earth” is the suggested way from their homepage.

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

Unfortunately, this is a presetting in Esri’s ArcMap program and I can not change that.

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

Labels “a” and “b” added to both figures

Figure 5: write parameters in the caption

All the parameters added to the caption as also added labels for the figure panels.

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

Degree sign and direction added as suggested

Table 4 and 6: add unit of depth

Units of depth added in both tables.

Supplements

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

Additional information in the caption added as suggested.

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), powered by

Thank you for the comment. Went through the document and corrected the typos such as spaces, etc.

We thank reviewer #1 for the constructive comments, which helped to improve our manuscript and hope we addressed all the questions raised by the reviewer.