

This manuscript by Hugelius et al. is a valuable contribution to the existing permafrost soil database. However, the quality of the data need to be addressed more clearly (please see 'Data quality' below). In addition, the paper suffers from a lack of final editing and could be much improved by adding more structure to the discussion, a more consistent use of abbreviations and elimination of extremely long sentences.

### **Data quality**

- It is difficult to understand the quality of the incorporated datasets, because I do not understand one of the requirements and it is hard to gauge from the text how many pedons were incomplete.
  - In P77L19 the 2nd criteria states: 'at least >150 cm'. Does this mean  $\geq 150$  cm?
  - P79L10-13 Please state the number or percentage of pedons for which data was not available to the full 200 or 300 cm.
  - P79L24 Please add number of profiles affected by lack of data on C in mineral soil underlying peat (n= ?). 'In some cases' is too casual
  - P80L4-6 Here it is unclear if the regional calculations include pedons that were gapfilled and, or extrapolated or only includes pedons for which measured data exists.
  - P83L27/28 'The Mollisols are also significant [...] but are not represented in the current database'. It would be good to explain why the Mollisols were excluded, since they seem to be an important C reservoir.

### **Structure and Style**

- The results and discussion section could be improved by adding more structure. At the moment it is one huge block of text. I recommend adding sub-headers such as 'Organic soils' (P82L19-P83L3), 'Permafrost' (P83L4-P84L11), 'Bedrock and ground ice' (P81L13-24). Also, I strongly recommend moving the section about organic vs. mineral soils (P84L25-P85L18) to after the 'Organic soils' (P82L19-P83L3). That would give a nice transition from organic to mineral soils to bedrock and improve the flow of the discussion.
- The manuscript contains a lot of VERY long sentences that make it hard to read. It would be great if one of the authors could read through the manuscript to reduce the occurrence of megasentence.
  - P76L25-28. I suggest chopping the sentence into 2 (end after 'SOC estimate'); comma missing before 'because'
  - P81L25-P82L2, P82L5-L9
  - P83L5-11
  - P84L7-11. This sentence contains 3 'because' – this is not necessary. It is also awkward to start a sentence with 'Therefore, because' I suggest 'For upscaling in the NCSCDv2, all permafrost-free mineral soil orders were

- aggregated into one class with circumpolar coverage, because they have an uneven geographical distribution [...], are similar with regards to 0-300 [...] and several soil orders are not represented in the current database.'
- The text uses a lot of abbreviations, which makes it hard to read. This can be improved:
    - I find it confusing that NCSCD is also NC-SCD (e.g. P75L24/25, P76L16/17), when there is a line-break. Since there are so many abbreviations in the text, I would keep this one together at all times.
    - P77L16/17 Is it really necessary to introduce another abbreviation (OC), since the text already uses 'C' and 'SOC', maybe here it could be 'SOC', too?
    - P80L21 'SOC storage' – is this SOCC?
  - Section 2.1.1. The conversions from LOI to %OC do not belong under the header 'gap filling'. I suggest moving this C-section up to after P78L16 'we refer to Hugelius et al. (2013)', where the laboratory methods are discussed. Then, start a new paragraph 'In many places [...]' And, unless there are many different equation for the mineral soil LOI-OC conversion, I would like to see the regression function in this manuscript as well and not just the citation Hugelius at el. 2011.
  - Table 2 is not referred to in the main text.

### **Minor comments**

- There is a major formatting change starting section 2.1.2, when suddenly new paragraphs shown after an insert, while before there are none. Please adjust the formatting to be consistent throughout the entire manuscript.
- P75L5: 'is a digital Geographical Information systems (GIS)'; change to 'System'
- P76L15-17, This statement is missing words? What was not included in the first version? 'Based on limited field data (46 pedons), and not included in the first version of the spatially distributed NCSCD [...]'
- P77L9 Capitalize 'systems'
- P77L19 start new paragraph after ([...] see Sect. 2.1 below)'
- P79L14 remove d from 'exposed' (should read 'to expose')
- P79L18 change 'carbon-free' to C-free
- Please consistently use 'sub-orders' or 'sub orders' or 'suborders'. Please check the entire manuscripts, e.g. P80L17, P81L3, P81L27
- P82L19-22
  - It is not clear to me why there was a concern that the Siberian Lowland Peats would be different from other organic soils (=peatlands?) in the Russian Permafrost zone. While the authors mention that the West Siberian Lowland peats formed in the early Holocene, it is unclear when the other peatlands

formed and, or why their age is likely to affect C storage. I take from the rest of the paragraph that the differences are due to density. It would be great if the authors could add more explanation.

- I suggest cutting this sentence in two, ending the first one after '(Smith et al. 2004)'. That way, one would also avoid this awkward construction 'and therefore there' (P82L20/21).
- P83L28/29 'Hugelius (2012) found that in a regional study of SOC storage in periglacial terrain, [...]' The 'that' is out of place, I suggest changing this sentence to 'In a regional study of SOC storage in periglacial terrain, Hugelius (2012) found that [...]'
- P84L12 & L14/15 It is not necessary to repeat 'In many parts of the northern circumpolar permafrost region' within 4 lines of text