

## ***Interactive comment on “High-resolution digital mapping of soil organic carbon in permafrost terrain using machine-learning: A case study in a sub-Arctic peatland environment” by Matthias B. Siewert***

**MB Siewert**

matthias.siewert@umu.se

Received and published: 8 November 2017

Anonymous Referee #2

Received and published: 26 September 2017

Siewert presents a study that maps soil organic carbon (SOC) stocks at high spatial resolution (~2m) for a sub-Arctic study site in Sweden. Four machine-learning algorithms are compared to assess which is best for predicting SOC. The Random Forests method creates the most accurate predictions. The results revealed that

C1

vegetation/land-cover type explained most variability in SOC, and thus the spatial distribution of SOC is controlled largely by landcover. On average, landscape scale estimates of SOC are in line with other high-resolution estimates generated at the landscape scale, and these are generally substantially lower than the best available circumpolar estimates generated using thematic maps. Overall the research is good quality and helps advance understanding of spatial variability in high-latitude SOC dynamics. Revisions are required before the manuscript can be considered further for publication. In general I find the science presented in this study to be sound. However some of the methods could benefit from additional detail. The writing could also be improved to enhance the clarity of the paper. There are quite a few wordy, run-on sentences that are hard to decipher. In other places there are generalities that do not actually convey much information. As a result of these things some very important key points are easy to miss, and this makes the paper seem less important than it actually is. Substantial editing will greatly improve the manuscript. I suspect that it should be possible to reduce the length of the paper quite a lot without losing any of the current content. As I mention above, and in specific comments below, aspects of the methods would benefit from additional detail. In particular, the details of several machine learning approaches are unclear. I realize that you use many different data sources, software tools, and analytical approaches, and so there are many details. However, it is becoming more common to publish processing scripts and data (where feasible) with your papers (using a repository such as GitHub, etc. . .). I myself am working to do this, and I encourage others to do the same. This has many benefits, and few downsides. With regards to the content of the article, one area that I believe should be improved is the discussion of your results in comparison to circumpolar SOC estimates (i.e. NCSCD). The discrepancy you report is large and seems important, but this is not the first case. Can you discuss potential approaches to bridge these two scales? Would Landsat or MODIS data be appropriate? Since land cover is an important determinant of SOC, it seems as though this could be feasible. Some discussion of how to extend remote sensing methods of SOC prediction to regional and circumpolar scales, and implications for estimates

C2

of related SOC stocks would be really useful, especially if the manuscript is edited to improve clarity.

**Thank you for this detailed review. The following changes will be made to address the reviewers comments: More detail will be added to the individual methods. However, I don't think that this should mean longer descriptions. There is a lot of literature available on these methods and the interested reader is pointed on several occasions to recent key literature. The writing will be revised throughout the manuscript. The manuscript will also be shortened to emphasis the most relevant results. This will in particular affect the last part of the manuscript that deals with the temporal evolution of the SOC storage. I will consider for future publications to structure my code and workflow in a way that it makes sense to publish the processing scripts.**

**I will add a full discussion section regarding different spatial scales in order to bridge local scale measurements to circumpolar scale. This is in line with the suggestions made by the other reviewers. This would include modeling of the SOC at a scales of 1m, 2m, 10 m, 30 m , 100m and 1000m. Corresponding to available remote sensing data (including Landsat and MODIS) and resolutions used by different model approaches. This will be discussed in context of improvements over the NCSCD at circumpolar level.**

Specific Comments:

P2 L14: This seems like an odd place to state the purpose of the articles, especially when it is re-stated in more detail later in the introduction. The introduction should begin with broad context and then gradually narrow to the scope of the present study, whereas this seems to bounce back and forth a bit.

C3

**The introduction has been restructured and shortened to provide a clearer overview to the topic.**

P2 L34-37: Could you elaborate on the evolution of quantitative soils methods, or get rid of this passage. It seems strange to say that methods have changed without at least a brief description of how.

**The specific passage has been deleted.**

P3 L1-4: Six studies seems like more than a few.

**Thanks. The wording has been changed.**

P3 L10-12: Will this really advance knowledge of SOC in all permafrost environments? Perhaps just this particular one, with potential for improved understanding in others.

**The wording was changed to adopt the perspective of the reviewer:**

**“this will improve our understanding of the SOC distribution and long-term C dynamics in high-latitude ecosystems.”**

P3 L14-22: This reads more like methods. It would be better to include this as methods.

**The paragraph has been moved to the methods section.**

P3 L33-34: Probably only need to note the 2002-2011 period just once.

**Changed**

P4 L4-13: This paragraph would fit better with the climatological information, before the detailed soils description.

C4

**The paragraph has been moved.**

P5 L4: Typo.

**Corrected**

P5 L33-35: This is ambiguous and not necessarily reproducible. Ideally you should publish your scripts/code with the paper.

**Thank you for your encouragement. I will consider to publish my scripts in the future.**

P7 L11: Did you use the caret package to fit the model as well, or was this just for cross-validation? The methods are a little vague here.

**Yes, caret was used to fit the model.**

P7 L28: What are 'visual sound results'?

**"Changed to visually meaningful results"**

P7 L28-30: This is a run-on sentence.

**Changed.**

P8 L2: It would be helpful to specify the number of points (i.e. how many is 20%).

**Changed.**

P10 L6-7: This sentence is discussion and doesn't belong in the results.

**The sentence was deleted.**

C5

P10 L8: 'Underestimated opposed' is confusing wording.

**Thank you, the entire paragraph has been edited to improve language.**

P10 L21-27: There is a lot of discussion in here.

**All sentences that discuss the results will be deleted or moved to the discussion section.**

P12 L24: In which environments to other algorithms perform better, and why might this be?

**At this point the general conclusion in the literature is only that no algorithm serves all landscapes. This most likely relates to statistical properties and underlying assumptions of each algorithm and how it can cope with the input data. A sentence was added to underline this.**

**"This indicates that different machine learning algorithms might suit different landscapes and that several algorithms should be compared (Forkuor et al., 2017)."**

P13 L24: Type 'led' not 'let'

**Thanks, Corrected**

P14 L13: How generalizable are these results then?

**Of course there is a limit to what geographical extent a set of input points can be generalized. It is reasonable to assume that a similar environment will feature a similar pattern of SOC distribution, but higher or lower SOC mean values depending on climate.**

C6

P14 L17: 'Incrementally'

**Changed**

P15 L15-20: This seems important – can you expand to discuss how these scales might be bridged? Does this mean all areas underestimated? What does this mean for circumpolar SOC stocks?

**Thank you for your interest. The article will be revised as suggested to include a discussion on scales, how these can be bridged and how circumpolar SOC stock estimates could be improved.**

---

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2017-323>, 2017.