Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-323-AC3, 2017

Biogeosciences Discussions
Open Access
EGU

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 #3

Received and published: 3 October 2017

General comments:

The study involves the evaluation of different methods for detailed mapping of SOC in permafrost regions. It targets a relevant topic and the methodological approach is

Printer-friendly version

Discussion paper

sound. The manuscript is well written and thoroughly deals with all sections. Some improvements could be made though. The different machine learning methods were utilized a diverse set of input parameters, including individual parameters (e g spectral bands), derived parameters form single data sources (e g NDVI, TWI) and integrated parameters (landcover/LCC). The single best predictor was LCC which is not surprising since the LCC integrates several remote sensing sources and also involves manual processing. These diverse types of parameters make it difficult to conclude which raw data sources are most important for SOC mapping. A brief discussion about the importance of different sources could be added to the discussion. Further it would be very interesting to see the performance of LCC alone for mapping as a single predictor. This could be achieved by providing the performance of LCC alone in Table 1. The study focusses on high-resolution mapping (e g 2x2 meters) which is good, but in addition it would be of interest to see how the different methods perform at coarser scales. Unbiased estimate at the 100x100 meter scale or 1x1 km scale is of great importance for global SOC mapping initiatives. A summary of landscape estimates for all the different methods (including LCC) could be added to the results. The SOC distribution in the Abisko area is strongly dependent on the occurrence of peatland areas. In Fig. 4 it can be seen clearly that the modelling mainly separate peatland areas from minerogenic soils. This is not discussed in relation to method performance and implications of the findings.

Thank you for your review. As several reviewers have suggested to investigate modeling at different spatial scales, I will add to the article estimates at spatial resolutions of 1m, 2 m, 10 m, 30m, 100m and 1 km. This has now been tested and will provide meaningful results using the random forest model. As this model seems to work best at all resolutions, I think there is little point to investigate the other models for all scales other than an initial test at 1x1 m. Furthermore, I considered and tested to model the SOC using only the LCC, but the results where not very promising and don't seem to add to the manuscript in a coherent

way. A summary of landscape estimates for different resolutions will be added to the results. A brief discussion point will be added regarding the differentiation of peatland soils and minerogenic soils in the model. Indeed, very different controls for these two major SOC populations can be imagined.

Detailed comments:

P1 L21: Abisko is misspelled.

Changed

P2 L15: Describe more specifically which “dramatic changes is peat mires...” that you refer to.

Replaced by “ Significant changes in surface structure and vegetation in a peat mire...”

P5 L6: I believe it should be “>2mm” instead of “<2 mm”.

Thanks, changed

P5 L6: How was the coarse fraction volume determined?

Added: “determined by sieving of the sample“

P7 L29: Change “visual” to “visually”.

Changed

P9 L24 (also P12 L17): Explain why the external validation was so much superior for RF compared to the other methods. What is the implication of this?

It is hard to explain why exactly one machine-learning method would perform better than others. I don't see any straightforward answer to this question from the literature. RF is generally known to be a versatile algorithm, while other algorithms can perform better in certain situations, but also require very detailed fine tuning. RF seems to be an overall reasonable recommendation.

P13 L24: Change "let" to "led".

changed

P13 L35: Clarify that LCC is an integrated parameter combining many other data sources.

This is now clarified further down in the paragraph

P13 L35-: In this section please discuss the inference of your results based on the fact that the distribution of SOC in the Abisko landscape is so strongly dependent on the distribution peatland.

A section will be added to address this.

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

Printer-friendly version

Discussion paper