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

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 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 circum-

polar 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 no 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 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. NC-SCD). 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 of related SOC stocks would be really useful, especially if the manuscript is edited to improve clarity.

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.

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.

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

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.

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

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

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

P5 L4: Typo.

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

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.

P7 L28: What are 'visual sound results'?

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

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

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

P10 L8: 'Underestimated opposed' is confusing wording.

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

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

P13 L24: Type 'led' not 'let'

P14 L13: How generalizable are these results then?

P14 L17: 'Incrementally'

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?

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