Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-569-AC1, 2018

Biogeosciences
Discussions
Open Access

Interactive comment on "Spatial variation and linkages of soil and vegetation in the Siberian Arctic tundra – coupling field observations with remote sensing data" *by* Juha Mikola et al.

Juha Mikola et al.

juha.mikola@helsinki.fi

Received and published: 16 March 2018

Thank you for your very helpful comments. We have revised our manuscript accordingly. Our responses to your comments are listed below one by one. The page and line numbers refer to the revised manuscript, which we provide as a supplement to this response.

......

Referee #1 comments

The presented manuscript analyses spatial variation of plant and soil properties and

their relations to each other for a field site in the Siberian Arctic tundra. Furthermore, it is tested to what extent remote sensing data can be utilised to capture variation in these properties and, consequently, to extrapolate vegetation and soil effects on ecosystem carbon fluxes to the large scale. The study highlights difficulties in predicting soil properties from NDVI, since they are not linked to vascular plant LAI, but moss biomass, which cannot be captured well by remote sensing. Instead, a classification of vegetation and soil properties according to land cover types is recommended to capture their spatial variation. The manuscript is well written and of good scientific quality. However, some details of the methodology and the results should be explained more clearly in order to make the manuscript easier to understand for readers who are not familiar with remote sensing techniques. I recommend to publish it with minor revisions, as outlined below.

General comments (1) The abstract is too long in my opinion, the middle part contains many details and figures which are not essential for the message of the study. I therefore recommend to shorten it by around 50%. REPLY#1 – We shortened the abstract.

(2) A schematic of the individual working steps described in the methods section and how they relate to each other may be helpful to understand the approach (e.g. how are data from field sampling combined with remote sensing). REPLY#2 – We added a diagram of the main working steps (Figure 2).

(3) As far as I understood Sect. 2.5, the 9 land cover types (LCTs) were determined through personal judgement in the field. Subsequently, a statistical classification approach (random forest) was applied to construct a model which predicts these LCTs based on remote sensing data. The authors report that 109 features of the remote sensing images were used in this model to predict the LCTs. However, no information is provided on what exactly these features are and why such a large number of input variables is needed. Please explain (a) what properties these 109 features represent (in general terms). (b) how you tested the model for overfitting, i.e. couldn't you have

produced a similar prediction of LCTs with less input variables? (I am aware that the number of original features was reduced from 262 to 109, but this still seems a lot to me) (c) why the features lead to a superior recognition of LCTs than the NDVI-based information. REPLY#3 – (a) Previous studies have shown that inclusion of multiple features improves classification accuracies. For this reason, we used several features, calculated from different datasets, to capture topographic and spectral variation. We earlier explained in general terms what these features were, but now we also included more justification for why we used so many (P7, L3-5). (b) Random forest is insensitive to overfitting and handles well multidimensional data. Therefore, the high number of features is not a problem. However, it has earlier been shown that feature selection improves the performance of random forest. We included this reasoning in the manuscript (P7, L16-18). (c) The features included NDVI, but also captured other aspects. This ultimately leads to better recognition of LCTs than with NDVI alone.

(4) Figures 3 to 6 show how vegetation and soil properties depend on LCT, while Figs. 11 and 12 show relations between moss biomass/vascular LAI and NDVI, and soil properties and NDVI, respectively. While around half of the relations in Figs. 11 and 12 are relatively weak (R squared < 0.3), the others do suggest that NDVI provides information about vegetation and soil properties. Moreover, the LCT classification does show some weak relations to soil and vegetation properties, too, and the external accuracy of the statistical model which predicts LCT is only 49%. Hence, it is not obvious to me why the derivation of vegetation and soil properties via LCTs is better than the NDVI-based derivation. It would be nice if you could provide a more quantitative prediction of soil and vegetation properties based on LCTs. Early growing season NDVI, for instance, explains 23% of moss biomass. Would it be possible to come up with a comparable figure for the LCT approach, e.g. given the 50% accuracy in predicting the LCT, and the standard deviation in moss biomass within an LCT, how much of the variance is explained? REPLY#4 – This is a good point and we therefore calculated measures of uncertainty for the predictions of the LCT approach as well (P7, L36 – P8, L3; P10, L23-29; Table 3, see also our REPLY#18). These show that the uncertainty

in capturing and predicting moss biomass cannot be avoided in LCT map either (P12, L3-8; P15, L23-27). We revised the discussion and conclusions based on these new results.

(5) I am missing a few words on the outcome of the decomposition experiment in the discussion section. What is the implication for ecosystem carbon fluxes? REPLY#5 - We now better illustrate the implication of the results of the tea bag trial for ecosystem carbon fluxes (P13, L15-28).

Specific comments p3,l1 Could you please expand the sentence by one or two examples stating which soil properties affect ecosystem carbon exchange and how they do that? REPLY#6 – We added temperature as an example of important soil properties (P2, L34-35).

p4,l25 Please add a few words on why this sampling point pattern was chosen (increasing distances between points with larger distance from EC tower). In particular, explain how this pattern is suitable to capture soil and vegetation properties at different spatial scales (from smaller to larger distances) for the study area. REPLY#7 – We added our reasoning for the chosen study plot pattern and contemplate its effects on capturing variation in soil and vegetation properties (P4, L18-21).

p5,l4 Please explain shortly why these soil properties were measured? How do they relate to carbon exchange fluxes? REPLY#8 – We added reasons for measuring the chosen soil properties (P5, L1-3).

p10,l30 & p11,l12 & p14,l1 Shouldn't moss biomass relate to topography via wetness? At least Sphagnum should show a link to low elevation. Figure 8 seems to show a good correlation between topography and dry/wet areas, which correlate well with vegetation type. Therefore, the explanation regarding microtopography seems not very satisfactory to me. Please explain this in more detail and maybe show a map of the topography of the study area for comparison. REPLY#9 – We revised and elaborated the discussion about the reasons for the lack of link between topography and moss mass (P15, L4-11).

p11,l15f If remote sensing reflectance cannot capture well moss biomass and associated soil properties, how can remote sensing be successfully used to classify LCTs, which also largely depend on vegetation properties? This means, how can the random forest classification distinguish between the 7 LCTs which differ mostly in vegetation properties? Please provide more details on the 262/109 features (see also general comment 3 & 4). REPLY#10 – The RF classification was specifically trained to classify LCTs, i.e. we used training data of LCT occurrence when constructing the classification (P7, L14-16). In addition, in the RF classification, we used eight spectral bands and three spectral indices of two satellite images taken at different phases of the growing season as well as several topographic features. In the regressions, we used one spectral index and topographical features only. Nevertheless, although the larger set of features in LCT classification helped in capturing variation in vegetation, the overall classification accuracy was 49%, thus suggesting that there is uncertainty in the LCT classification as well (P15, L30-31). The 262 features consisted of 15 features calculated from spectral bands and 2 features calculated from spectral index and topographical layers. The 109 features is a subset of these 262 features and included spectral, topographic and textural features. These details are now better explained in the text (P7, L11-13 and 20-21).

Comments on style p4,l13 "..the soil is in continuous.." - the "in" seems to be misplaced here. REPLY#11 – Right, “in” was deleted.

p4,l20 What do you mean by the word "manuscript" in the cited studies? At least Tuovinen et al do not appear in the bibliography. Could you please correct that and use "submitted" instead? REPLY#12 – We removed citations to unpublished papers.

p6,l12 Please explain the abbreviation "GCP", e.g. putting it in brackets in line 10. REPLY#13 – Done.

p9,l29 It is not clear what "user and producer accuracies" are. REPLY#14 – We added

explanations for the two accuracies (P10, L10-11).

p13,l10 Please provide a reference to the figure/table which illustrate this finding at the end of the sentence. This should be done also for the rest of the discussion. REPLY#15 – To help the reader, we added references to Figures and Tables throughout discussion.

Please also note the supplement to this comment:
https://www.biogeosciences-discuss.net/bg-2017-569/bg-2017-569-AC1-supplement.pdf

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-569, 2018.