Associate Editor Decision: Reconsider after major revisions (30 Nov 2017) by Susan Natali
Comments to the Author:
Dear Dr. Siewert,
Thank you for your responses to the referees' comments. I agree with the reviewers' comments and your response to address issues related to the spatial resolution of the model. With this and other suggested changes, including editing for readability and inclusion of processing code, I feel that this will be a much improved and publishable manuscript.
Best regard,
Sue Natali

Response letter to the editor:

Dear Dr. Natali,

Thank you for your positive feedback to my manuscript. It has now been fully revised. Major changes include:
- A section testing the modelling approach at different spatial resolutions.
- Major revisions of the text addressing all reviewer comments as outlined below and editing for readability.
- link to the processing code at github as suggest by Referee #2

I hope that the manuscript will be judge publishable after these improvements.

I am looking forward to your feedback.
With best regards,
Matthias Siewert

Response letter to the reviewers:

Anonymous Referee #1
Received and published: 21 September 2017
This paper discusses a study that developed a high spatial resolution map of soil organic carbon for a sub-Arctic peatland in northern Sweden, using essentially Random Forest algorithms and a suite of environmental variables, including land cover,remotely-sensed vegetation indices, and digital elevation terrain modeling (DEM). The study is relatively straightforward, and demonstrates a reasonable approach for modeling/mapping soil carbon in high northern latitude systems. My only major issue had to do with clarification of the resolutions of the various input datasets, and the ultimate resolution provided by the model/map.

That and other minor points are listed here:

1)So, with regard to the resolution of the inputs and outputs, I found it slightly hard to follow, and I think it might help to put all of the resolutions on Figure 1 (right now only the orthophoto/DEM and the final map resolutions are on there). If I am understanding this correctly, the orthophoto is 1m and the DEM is 2m (this is actually slightly misleading in Figure 1, which has the orthophoto + DEM as 1m – but, I guess that the DEM was just “down-sampled” to 1m resolution. The SPOT data are either 10m or 20m, and the minimum size of a land cover classification was 130m 2 , so somewhat consistent with a SPOT pixel, although it’s unclear what the range of extents are for land cover regions.
The final map is then generated at the 2m resolution; why not 1m and utilize the more resolute orthophoto information?

Thank your for this very interesting comment that opens up a different perspective to this work. The resolution of the individual products is now mentioned in Figure 1 (Now Figure 3). A resolution of 2 m for the final model was originally chosen as a compromise between the available input variables, output quality, the benefit of higher resolution and processing time.

However, as this point has been mentioned by several reviewers, I ran the model at several spatial resolutions: 1m, 2m, 10m, 30m, 100m, 250 m and 1000 m for the Total SOC and at 1m for different depth intervals. This is in line with the suggestions made by reviewer 3 and 4. The outcome is discussed following the reviewers input with regard to the resolution of different input datasets.

2) Figure 6 – Does “Mean decrease in accuracy” indicate the accuracy reduction when that variable is removed from the analysis? If so, make that clear in the figure and caption.
Yes, that is the meaning of this measure. The figure caption has been reformulated to emphasis this:
Fig. 1. Variable importance for the prediction of total SOC measured as mean decrease in accuracy of the random forest model if the variable is excluded. The higher the value the more important is the variable.

3) Also, it’s interesting that the most important variable in the analysis was Land Cover, the variable at the coarsest resolution, followed by three SPOT variables. In fact, you don’t get a DEM variable until the 5th-most important (Elevation), and even then it’s unclear that the information is necessary at the 2m resolution (could be equally useful if aggregated to a coarser resolution). I know from first-hand experience that these systems can be highly variable in space over short distances with regard to SOC; however, it’s certainly interesting that most of the variability explained occurs at resolutions of tens of

meters, which puts into question the utility of a 2m resolution map. I think this is worthy of some additional discussion in the paper – particularly within the context of what is discussed on Page 13, Lines 1-8,where a fine resolution is necessary to capture the appropriate scale of variability in SOC.

I agree with the reviewer in this point. Indeed, lower resolution input variables seem more important than higher resolution input variables. However, I believe that this is much an effect of the validation rather than true value in the spatial prediction using the model. Looking at the resulting maps of SOC in Figure 6, it is clearly visible that information from the DEM has a strong influence on the final map. This seems to be more relevant for fine scale and linear landscape features, while larger homogeneous areas are more influenced by lower resolution input data. The discussion was updated regarding this comment (see section 5.3, 2 paragraph).

4) Abstract, Line 10 – add “for SOC quantification” after “evaluated”
Changed

5) Abstract, Line 16 – change “surprising” to “surprisingly”
Changed

6) Abstract, Line 19 – add “s” to “scale”
Changed

7) Page 2, Line 2 – specify “Northern” high latitudes
Changed
8) Page 2, Line 8 – to what depth is the ∼1300 Pg SOC estimate?
It is now specified that this includes “soils to a depth of 3 meters and other unconsolidated deposits “. The reader can get more information on this under the specified reference.

9) Page 2, Line 10 and throughout – be consistent, either hyphenate “permafrost affected” or not – probably should hyphenate
Hyphenate is now used throughout

10) Page 2, Line 24 – remove “a” before “commonly”
Changed

11) Page 3, Line 1 – remove the hyphen from “higher-latitudes”
Changed

12) Page 3, Line 15 – I think LCC has not been spelled out yet in the paper
LCC has been spelled out on page 1 of the introduction.

13) Page 4, Line 19 – How long were the transects (i.e. what was the distance between sampling points)?
The following information was added: “between 50 to 300 m (Fig. 2)”. This should enable the reader to understand the sampling layout.

14) Page 4, Line 29 – “deeper soil horizons were sampled in 5-10 cm intervals” – what actually were the intervals, and what determined them?
To be more specific it was added “depending on horizon thickness”

15) Page 5, Line 4 – change “were” to “where”
Changed

16) Page 5, Line 6 – should the notation be “>2 mm,” if you are referring to the coarse fraction, or are you referring to the soil that is not the coarse fraction?
Changed to > 2mm.

17) Page 5, Line 13 – add “SOC” before “stored”
Changed

18) Page 8, Lines 8-10 – I understand the overestimation of SOC values due to the absence of sample point from bare ground surfaces, however, I just want to clarify the justification for using 0 as the quantity of SOC. First, I’m not sure I know what a “blockfield” is – maybe that’s just me, but I think a definition/description would be good. Also, one cause of bare ground in northern high latitudes is cryogenic disturbances (i.e. cryoturbation), and in many cases, these were once vegetated areas that can have quite a bit of SOC. Are these generally uncommon in your study area? In other words, are the dominant bare ground features these blockfields and stone beaches that I imagine have very little SOC?
The following has been added to clarify this:
“Originally, all models overestimated SOC contents for bare ground surfaces. These areas include exposed bedrock, blockfields (areas covered by shattered rock fragments with little or no fine substrate; Fig2b) and stone beaches along lake shores (alpine heat tundra with minimal soil development and cryogenic features form a separate class). “

19) Page 9, Line 4 – add “ed” to “collect”
Changed

20) Page 9, Line 10 – remove one “s” from “miss-“
Changed

21) Page 11, Line 3 – add “be” after the first “to” and remove the 2nd “to”
Changed

22) Page 11, Line 5 and throughout – Sphagnum should be capitalized and Italicized
Changed

23) Page 12, Line 26 – don’t capitalize “Geographically”
Changed

24) Page, 12, Lines 28-29 – I’m not sure that I understand the statement that “very strong environmental gradients” would “suggest low spatial autocorrelation.” I would think that strong environmental gradients would lead to high spatial autocorrelation.
Rephrased to “These sharp transitions in SOC storage between different land covers suggest low spatial autocorrelation at local scale, i.e. little relationship in SOC values between points far apart”

25) Page 18, Line 2 – change “adoptions” to “adaptations” – I think that’s what you are meaning to say?

Changed
Thanks, that's clearly what I meant.

26) Page 18, Line 3 – need to reword "release them into the carbon cycle" – even if a carbon pool is stable for a long period of time, it's still in the carbon cycle.
The sentence was reworded to avoid this construct:
Rapid future permafrost degradation in peatlands may lead to erosion of organic sediments. This would transfer presently stored carbon into lakes and potentially into the atmosphere.

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. 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 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 have been made to address the reviewers comments: Some detail was added to the individual methods. However, a lot of literature is available on these methods and the interested reader is pointed on several occasions to recent key literature. The manuscript was edited throughout which hopefully improved the readability. The redundant sections of the manuscript were shortened or deleted. A link to a public github repository was added to the supplement material. The repository contains code relevant to reproducability of the article.
An analysis to investigate model performance at different resolutions of 1m, 2m, 10m, 30m, 100m, 250m and 1000m was added. A full section addressing the discrepancy to the NCSCD and the effect of a reduced spatial resolution in the model was added to the discussion section. This includes a discussion on how to bridge both scales and which satellite data could be used to improve circumpolar estimates. This is in line with the suggestions made by the other reviewers.

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.

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:
"The mapping approach will be discussed with regard to SOC estimation in permafrost regions at local to circumpolar scale."

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.
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. A link to a github repository was added in the supplement.

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.

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.

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 was revised as suggested to include a discussion on scales, how these can be bridged and how circumpolar SOC stock estimates could be improved.

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 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 added estimates at spatial resolutions of 1m, 2 m, 10 m, 30m, 100m, 250m and 1 km. To keep the article focused this was implemented using the RF model, as it showed overall the best results. 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 was added to the results in Table one and the predicted maps are shown in the supplement. A brief discussion point was 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. I will keep this notion in mind for future project.

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:
"In Abisko, the distribution of SOC is defined by the occurrence of peatlands (Fig. 5 and Fig. 8), to the extent that two separate populations of soil pedons can be identified (Table 1). This strong non-linearity may be the reason, why some models perform better. In the future, it should be tested if in such a case separate models for different populations of soil pedons can improve the prediction."

Anonymous Referee #4
Received and published: 3 October 2017
Author present comparison of four digital soil mapping techniques in predicting high-resolution (2x2m) SOC stocks of sub-Arctic peatland terrain. Study reports that Random forest performed better in comparison to other three techniques used and land cover types derived from a high resolution remote sensing data was the most important predictor of SOC stock variability. Author also report that most of the SOC of study area is relatively new carbon (∼ 2000 years old). Author report interesting findings and the outcome should be of interest to a wide readership of Biogeosciences. However, the current manuscript can be improved in multiple different ways as suggested below:

Thank you for your review.

- The sentence structure at multiple places is awkward so a careful editing is required.
My apologies. The manuscript was revised throughout with a focus on readability.

- Its not clear to me how 2x2m spatial resolution for SOC stock was defined? Author seem to have a variety of environmental datasets with spatial resolution ranging from 1 m to 20 m.
The spatial resolution of 2x2m was chosen as a compromise between the available input variables, output quality, the benefit of higher resolution and processing time. However, as several reviewers have highlighted interest in the exploration of different resolutions, I changed the changed the standard resolution to 1x1 m and added an anlysis on estimates for resolutions at 1m, 2m, 10m, 30m, 100m,250m and 1000m.

- I don’t agree with the term internal validation used in this manuscript. Using model training dataset as a model validation is not correct. It provides an incorrect metric of map accuracy. For validation, you have to either use the split sample in the beginning (like you did for 20% data) or it has to be take one out approach (cross validation; using remaining samples to predict at the data point by taking out that data point from the model calibration data).
The internal validation was completely removed. The manuscript metrics are now based on cross validation. All values have been updated.

- Its not clear to me how land cover data was treated in different models used, were all the land cover types were equally important predictors of SOC? or it was only a subset of all the land cover types? Please provide results.
The land cover types were treated as equally important predictors.

- I will like to see a section on uncertainty in this manuscript. Either calculate the uncertainty or provide a discussion of potential sources of uncertainty involved this study.
A discussion of sources of error is provided on page 13 L 18-33 (original manuscript). The section was update to point out uncertainty and was given a separate heading to make it easier accessible for the reader.

- The manuscript will benefit if authors can provide reasoning to the observed results. For e.g., why the environmental predictors changed with depths, why certain environmental controllers were significant predictor at certain depth and not other.
The revised version contains insights regarding the influence of different predictors with depth (section 5.3). However, there is of course limited space for such detail analysis.

-How the multicollinearity and non-linear relationships were handled?
Multicollinearity was tested using a cross-table of the predicting variables. In the revised version, highly correlated predictive variables are excluded. Non-linear relationships can be handled by the chosen models. See the updated methods and the discussion in section 5.1.

- Fig. 5 need to be replaced, please remove pseudo sampling points from the plots, provide the number of samples used for model validation. Provide separate plots for 4 mapping techniques using validation samples only. Add R2, RMSE, and CCC values in each plots.

I see the need to replace Figure 5. However, if the figure is replaced according to the suggestion of the reviewer (excluding training and pseudo sampling points) it would mean that it will only be based on 10 validation points per model (as the original pedon dataset is rather small). Using the full dataset (excluding pseudo sampling points) will provide much more information to the reader than just ten points. The R2, RMSE and CCC are now derived from cross-validation (one out approach) as suggested by the reviewer earlier on.

- Table 1: Please remove metrices calculated using model calibration datasets, and after adding these values in plots suggested earlier, you will not need this table. In results section, please describe what readers should learn from these map accuracy measures.

Table 1 has been removed. The information will be added in Fig. 5. Detail was added in the result section to describe what the reader should learn from the measures in terms of accuracy and precision.

High-resolution digital mapping of soil organic carbon in permafrost terrain using machine-learning: A case study in a sub-Arctic peatland environment

Matthias B. Siewert [1,2]

[1]Department of Physical Geography, Stockholm University, Stockholm, SE-106 91, Sweden

[2]Department of Ecology and Environmental Science, Umeå University, Umeå, SE-901 87, Sweden

Correspondence to: Matthias B. Siewert (matthias.siewert@umu.se)

Abstract

. Soil organic carbon (SOC) stored in northern peatlands and permafrost-~~-~~affected soils are key components in the global carbon cycle. ~~I~~This article quantif~~y~~ies SOC stocks in a sub-arctic mountainous peatland environment in the discontinuous permafrost zone in Abisko, northern Sweden. Four machine-learning techniques are evaluated for SOC quantification: multiple linear regression, artificial neural networks, support vector machine and random forest. The random forest ~~approach~~model performed best and was used to predict SOC for several depth increments at a spatial resolution of 1 m (1~~2~~ × 1~~2~~ m). A high-resolution (1 ~~× 1~~m) land cover classification generated for this study is the most relevant predictive variable. The landscape mean SOC storage (0–150 cm) is estimated to ~~7.9~~8.3 ± 8.0 kg C m^{-2} and the SOC stored in the top meter (0–100 cm) to 7.~~0~~7 ± 6.~~3~~2 kg C m^{-2}. The predictive modeling highlights the relative importance of wetland areas and in particular peat plateaus for the landscape SOC storage. The total SOC was also predicted at reduced spatial resolutions of 2 m, 10 m, 30 m, 100 m, 250 m and 1000 m and shows a significant drop in land cover class detail and a tendency to underestimate the SOC at resolutions >30 m. This is associated with the occurrence of ~~large number of surprising A~~many small scale wetlands ~~areas are mapped~~ forming ~~very~~ local hot-spots of SOC storage that are omitted at coarse resolutions. ~~The results show that robust SOC predictions are possible with the available methods and very high-resolution remote sensing data.~~ Sharp transitions in SOC storage~~Strong environmental gradients~~ associated with land cover and permafrost distribution are the most challenging methodological aspect. However, in this study, at local, regional and circum-Arctic scales the main factor limiting robust~~, high-resolution~~ SOC mapping efforts is the scarcity of soil pedon data from across the entire environmental space. For the Abis~~i~~ko region, past SOC and permafrost dynamics indicate that most of the SOC is barely 2000 years old and very dynamic~~ in wetland areas with permafrost related landforms~~. Future research needs to investigate the geomorphic response of permafrost degradation and the fate of SOC across all landscape compartments in post-permafrost landscapes.

1 Introduction

Northern h~~H~~igh-latitudes are among the regions most affected by increasing temperatures and climate change (IPCC, 2013). Large amounts of soil organic carbon (SOC) and the abundance of wetlands as a substantial source of methane (CH_4), are factors that make th~~ese~~is region~~s~~ a key component in the global carbon (C) cycle (McGuire et al., 2009). Frozen conditions, cold temperatures and water-logging are characteristics of wetlands, peatlands and permafrost-~~-~~affected soils that reduce decomposition rates of SOC (Davidson and Janssens, 2006; Ping et al., 2015). This has led to the accumulation of large stocks of SOC in high-latitude ecosystems (Tarnocai et al., 2009). SOC stocks in the circumpolar permafrost region are ~~currently~~ estimated to ~1300 Pg, including soils to a depth of 3 meters and other unconsolidated deposits (Hugelius et al., 2014) . ~~and~~This ~~represent~~corresponds to around half of the global SOC stocks (Köchy et al., 2015). ~~Circumpolar mapping efforts of this SOC provide important input data for Earth System Models, while Hhigh-resolution mapping efforts are necessary to map SOC to understand the substanstial local scale spatial and vertical variability of SOC in permafrost-affected soils (Siewert et al., 2015, 2016).~~ A significant proportion of this SOC is stored in northern wetland and peatland areas (Gorham, 1991). However, warming temperatures, environmental changes caused by warming of soils and consequent permafrost degradation are projected to lead to a gradual and prolonged release of greenhouse gases in the future (Schuur et al., 2015).

~~This article investigates SOC storage and longterm SOC dynamics in the Abisko region, sub-Arctic Sweden, where numerous ecosystem dynamics related to climate warming have been documented (Callaghan et al., 2013). DramaticSignificant changes in surface structure and vegetation in a peat mire have been reported between 1970 and 2000 (Malmer et al., 2005). During this period, degradation of permafrost and vegetation changes can be associated with increases in landscape scale CH4 emissions (Christensen et al., 2004). An analysis of present day C fluxes indicates that losses from soil and over the hydrosphere currently offset C accumulation in peatlands and above ground biomass (Lundin et al., 2016). To improve our understanding of permafrost C dynamics, particularly over longer timescales, high-resolution maps of landscape distribution and partitioning of SOC, including the vertical partitioning, and an integration into numerical models is necessary (Mishra et al., 2013; Schuur et al., 2015). Combined with a better temporal framework of past C dynamics, this will improve the projection of future global temperatures.~~

Circumpolar mapping efforts of SOC provide important input data for Earth System Models. At the same time, high-resolution mapping efforts are necessary to understand the substantial local scale spatial and vertical variability of SOC in permafrost-affected soils (Siewert et al., 2015, 2016). Thematic maps are ~~a~~ commonly used to ~~upscale~~map SOC from point measurements to landscape scale in permafrost environments (Hugelius, 2012). This method ~~has been~~is used in combination with soil maps to estimate SOC storage in the circumpolar permafrost region using the Northern Circumpolar Soil Carbon Database (NCSCD) (Hugelius et al., 2014; Tarnocai et al., 2009). Land cover maps ~~have~~are also~~been~~ used at local to regional scales to estimate SOC values in numerous circumpolar environments (Fuchs et al., 2015; Hugelius et al., 2010, 2011; Hugelius and Kuhry, 2009; Palmtag et al., 2015; Siewert et al., 2015; Zubrzycki et al., 2013). While soil maps may better reflect soil properties and soil forming processes, a land cover classification (LCC) has the advantage that it can be readily generated from remote sensing data using the spatial resolution of the respective sensor. However, thematic mapping also represents a strong generalization, as equal soil properties are assumed for all areas covered by the same mapping class. Furthermore, for land cover maps there is an implicit assumption that land cover alone reflects below-ground soil properties (Hugelius, 2012).

An alternative to thematic mapping is the use of ~~P~~predictive modeling methods~~of SOC values~~. These can yield well resolved pixel based estimates of SOC~~and provide a potential improvement over thematic mapping~~. ~~Quantitative methods in soil science are largely based on the works of Jenny (1941, 1980) and have significantly developed since. Digital soil mapping is the successor of these concepts using modern methods.~~ A comprehensive summary of these methods, commonly called digital soil mapping, ~~has been~~is published by McBratney et al. (2003) and many examples are available ~~for example in~~, e.g. in Boettinger et al. (2010). However, ~~at higher-latitudes~~ in sub-Arctic and Arctic permafrost environments the adoption of ~~these~~ predictive modelling methods ~~has been slow and only few studies apply predictive modeling methods~~ to ~~upscale~~map soil properties has been limited. Some examples include ~~in sub-Arctic and Arctic permafrost environments~~ Bartsch et al., 2016; Baughman et al., 2015; Ding et al., 2016; Mishra and Riley, 2012, 2014 and Pastick et al., 2014. Th~~is~~e limited adoption has several reasons including: the limited availability of environmental input data, the limited amount of soil pedon data (Mishra et al., 2013) and the large local scale variability of permafrost-affected soils (Siewert et al., 2016). To cope with these limitations new mapping methods for permafrost environments are necessary to better constrain SOC stocks in the northern circumpolar permafrost region (Mishra et al., 2013). Such new methods include the use of machine-learning in digital soil mapping (Hastie et al., 2009; Li et al., 2011). Machine-learning in soil science covers a set of data-mining techniques that can recognize patterns in data-sets and learn from these to predict quantitative soil variables. Many algorithms are available and robust prediction results are possible (Hastie et al., 2009; Li and Heap, 2008; Li et al., 2011). Testing these methods at different spatial resolutions can eventually improve local and circumpolar scale estimates of SOC.

Numerous ecosystem dynamics related to climate warming have been documented in sub-Arctic Sweden (Callaghan et al., 2013). These include the degradation of permafrost ~~(Åkerman and Johansson, 2008; Johansson et al., 2011)~~(Åkerman and Johansson, 2008; Johansson et al., 2011), significant changes in surface structure in peat mires and changes in vegetation (Malmer et al., 2005). These changes can be associated with increases in landscape scale CH_4 emissions (Christensen et al., 2004). Analysis of present day C fluxes indicate that losses from soil and over the hydrosphere currently offset C accumulation in peatlands and above ground biomass making these ecosystems a C source (Lundin et al., 2016). To improve our understanding of these long-term permafrost region C dynamics, high-resolution maps of landscape distribution and partitioning of SOC are necessary. This should include the vertical partitioning of SOC and provide data that can be integrated into numerical models (Mishra et al., 2013; Schuur et al., 2015). Combined with a better temporal framework of past C dynamics, this will improve the projection of future climatic changes.

~~This study aims to compare a variety of machine-learning techniques for the prediction of SOC in permafrost and peatland environments and to estimate carbon storage under different land cover types~~This study aims to compare four different machine-learning techniques for the prediction of SOC in a high-latitude permafrost and peatland environment. ~~Combined with radiocarbon dates to estimate SOC accumulation, this will advance our knowledge on SOC distribution and longterm C dynamics in high-latitude permafrost environments.~~ The mapping approach will be discussed with regard to its suitability to estimate SOC stocks in permafrost environments at local to circumpolar scale using different spatial resolutions. ~~Machine-learning in soil science covers a set of data-mining techniques that can recognize patterns in large data-sets and learn from these to predict quantitative soil variables. Many algorithms are available and robust prediction results are possible (Hastie et al., 2009; Li and Heap, 2008; Li et al., 2011). The workflow of this study is outlined in Fig. 3. A 1 × 1 m very high-resolution LCC is generated. Four prediction models are compared: a multiple linear regression (MLR) model, an artificial neural network (ANN), a support vector machine (SVM) and the random forests (RF). The best performing model is used to demonstrate a high-resolution (2 × 2 m) spatial regression modeling approach of SOC pedon data to landscape scale. The LCC is used for stratified extraction of the SOC per class.~~The results will~~As an outcome, I~~ provide high-resolution SOC storage data for a key sub-Arctic research site ~~on ecosystem adaption to climate change~~ located in Abisko, northern Sweden. Th~~e~~~~is includes~~ study will give insights on the spatial and vertical partitioning of the SOC under different land covers and its association with different environmental variables. The temporal evolution of the SOC stocks over the Holocene ~~is~~will be interpreted from eight radio-carbon dates and the future development of SOC stocks and potential C release in high-latitude environments ~~is~~will be ~~discussed~~addressed.

~~Figure 1 near here~~

2 Study area

The study area is a sub-Arctic mountain environment in the Abisko region near Stordalen ~~along the shores of lake Torneträsk in~~, northernmost Sweden (Fig. 1 and Fig. 2). Environmental monitoring and research has been conducted for more than a century in the region and a particular interest has been the main peatland complex called Stordalen mire (Callaghan et al., 2013; Jonasson et al., 2012). The mapping extent covers two major peatland complexes, Stordalen and Storflaket, ~~east of the Abisko Scientific Research Station,~~ the surrounding birch forest and the adjacent alpine tundra zone. The altitude ranges from 342 m a.s.l. ~~corresponding to the lake level of Torneträsk~~ to 932 m a.s.l.~~ in the mountain zone~~. The total mapping area is 65 km².

A mean annual air temperature of 0.5°C~~ (2002–2011)~~ and a mean annual precipitation of 332 mm have been measured for the period 2002–2011 in Abisko~~. These values are approximate indicators as topography and rain-shadow effects have strong local influence in the region~~ (Callaghan et al., 2013). The study area is located in the zone of discontinuous permafrost (Brown et al., 1997). The onset of late Holocene permafrost aggradation in Stordalen was around 2650 cal BP with a first phase that lasted until 2100 cal BP and a second phase after ca. 700 cal BP (Kokfelt et al., 2010). Today, the occurrence of permafrost at lower elevations is confined to peat mires due to insulation effects of the peat. Here it can be several meters thick below elevated permafrost peat plateaus (palsa) (Johansson et al., 2011). At higher elevation, permafrost was modeled to occur above 850 m a.s.l. on north-east and east-facing slopes and above 1000 to 1100 m a.s.l. on west and south facing slopes west of Abisko Station (Ridefelt et al., 2008). Widespread permafrost degradation has occurred at least since the 1980s. This was associated with increased active layer thicknesses, decrease in permafrost thickness and complete disappearance of permafrost in some areas~~ (Åkerman and Johansson, 2008) and with a decrease in permafrost thickness from 15 m in 1980 to 9 m 2009 in one borehole~~ (Åkerman and Johansson, 2008; Johansson et al., 2011).

Wetland soils in the study area are of organic nature (Histosols). Soils in the surrounding forest have mostly characteristics of Podzols or micro-Podzols with a bleached horizon below the organic surface layer. ~~The a~~Alpine soils are often limited to shallow surface organic layers over rock (Leptosols) or very weakly developed soils in unconsolidated slope or moraine material (Regosols). ~~However,~~ Soils classify as Cryosols if permafrost occurs within 1 m at higher elevation or when cryoturbation occurs and permafrost ~~occurs~~can be detected within 2 m~~, then soils classify as Cryosols~~ (FAO, 2015).

~~The study area is located in the zone of discontinuous permafrost (Brown et al., 1997). The onset of late Holocene permafrost aggradation in Stordalen was around 2650 cal BP with a first phase that lasted until 2100 cal BP and a second phase after ca. 700 cal BP (Kokfelt et al., 2010). Today, the occurrence of permafrost at lower elevations is confined to peat mires due to insulation effects of the peat. Here it can be several meters thick below elevated permafrost peat plateaus (palsa) (Johansson et al., 2011). At higher elevation, permafrost was modeled to occur above 850 m a.s.l. on north-east and east-facing slopes and above 1000 to 1100 m a.s.l. on west and south facing slopes west of Abisko Station (Ridefelt et al., 2008). However, widespread permafrost degradation has occurred at least since the 1980s. This was associated with increased active layer thicknesses and complete disappearance of permafrost in some parts of Stordalen mire (Åkerman and Johansson, 2008) and with a decrease in permafrost thickness from 15 m in 1980 to 9 m 2009 in one borehole (Åkerman and Johansson, 2008; Johansson et al., 2011).~~

Fig. 1 and Fig. 2~~Figure 2 and 3~~ near here

3 Methods

The workflow of this study is outlined in Fig. 3. First, datasets of environmental predictors and soil pedons are compiled. Second, a LCC of very spatial high-resolution at 1 m (i.e. 1 × 1 m) is generated using a subset of the environmental predictors. Third, digital soil mapping is performed by combining the environmental predictor variables and the soil pedon dataset. For this, four commonly used machine-learning algorithms are tested to generate prediction models of SOC using a regression approach. The best performing model is used to develop high-resolution (1 m) spatial maps of SOC for different depth intervals. The effect of spatial resolution is analyzed by predicting the SOC with a progressively resampled environmental predictor dataset at resolutions of 2 m, 10 m, 30 m, 100 m, 250 m and 1000 m. The LCC is used for stratified extraction of the SOC per land cover class. Lastly, radiocarbon samples are analyzed to understand past SOC aggregation.

Fig. 3 near here

3.1 Field survey and SOC data

Soil sampling was performed in September 2013 and June 2015. In total, 47 sites were sampled following initial field reconnaissance. Sampling was undertaken along 4 main transects with 8–10 points of equal distance between 50 to 300 m (Fig. 1). These transects were laid out as a semi-random sampling scheme~~s~~ to represent major environmental gradients with a restricted amount of sampling points in difficult terrain. The~~se~~ main transects were complemented with smaller transects (n = 2 & 3) and six additional profiles from land covers with small patch sizes otherwise not covered. Some transects ~~are~~are incomplete due to points located in lakes, and to avoid disturbance of experimental installations or wildlife.

The sampling procedure followed (Schoeneberger et al., 2012), with an additional protocol for permafrost-affected soils ~~(Ping et al., 2013; Siewert et al., 2016)~~(Ping et al., 2013; Siewert et al., 2016). Sampling in peat was performed in 5 cm intervals by cutting samples of known volume from the open pit, using fixed volume cylinders or in case of waterlogged conditions using a fixed volume Russian peat corer. ~~The permafrost was sampled by hammering a steel pipe into the frozen ground.~~ Sampling in ~~the surrounding~~ birch forest and tundra was performed according to soil horizons. The organic layer (OL) was sampled completely, while deeper soil horizons were sampled in 5 to 10 cm intervals depending on horizon thickness. Frozen soil (permafrost) and deep soil layers were sampled by hammering a steel pipe into the ground. Soils outside the peat complexes are in general very shallow, ~~and~~ have large volumes of coarse fragments and often become impossible to sample after ~20–50 cm as the fractured lithic contact is reached. The unsampled coarse fraction consisting of unconsolidated bedrock was noted and used to correct the amount of soil material. For the tundra heath, the OL can be discontinuous with patches of vegetation alternating with patches of bare ground. Here the SOC storage was corrected for the proportional coverage of the OL per m². Each point location was recorded in the field using a hand-held GPS device (±5 m location accuracy).

A total of 278 individual soil samples were collected. Dry bulk density (DBD, g cm^{-3}) was calculated from oven dried soil samples ~~at (~~(65°C for 5 days). The loss on ignition (LOI) method was performed on all samples at 550°C for 5 hrs to determine the organic matter (OM) content and at 950°C for 2 hrs to determine the inorganic C content (Heiri et al., 2001). C % was measured for a subset of 73 samples using an EA 1110 Elemental Analyzer (CE Instruments, Italy). A further subset of samples with relatively high inorganic C were acid treated but showed very little reaction.~~, and~~ LOI at 950°C for all samples also indicated very low inorganic C content in the soils ~~of (~~(0.73 ± 0.62%)~~). for all samples and h~~Hence, inorganic C content was not further analyzed. C% values were then used to predict C% for samples where only LOI was available using a third order polynomial regression model (Fuchs et al., 2015; Hugelius et al., 2011; Siewert et al., 2015). The SOC storage was calculated per soil sample using C%, DBD, excluding soil material of the coarse fraction (determined by sieving of the sample ~~(~~≥~~<~~2 mm, CF, %) and sample depth interval. Depth intervals that have not been sampled were gap filled based on soil horizon information. The average pedon depth was 103 ± 29 cm for wetland pedons and 26 ± 27 cm for non-wetland pedons. To estimate the total SOC (SOC–Tot) stored in the landscape, all wetland pedons were processed to a reference depth of 1.5 m and non-wetland pedons to a depth of 1 m. If the pedon did not reach that depth it was extrapolated based on a trend in the pedon or similar pedons or set to zero if the lithic contact was reached. ~~This interpolation is necessary to provide standardized and consistent input data for the predictive modeling approach.~~ Other extracted depth intervals~~ that were extracted~~ are the SOC stored in the organic surface layer (SOC–OL), the SOC stored for the top 30 cm (SOC 0–30) and the SOC for the top 100 cm (SOC 0–100). The depth of the organic surface layer (OL–Depth) was also predicted.

To evaluate~~evalutate~~ long-term SOC dynamics, eight samples were submitted for AMS ^{14}C dating to the Radiocarbon Laboratory in Poznan, Poland. The resulting dates were calibrated to calendar years, cal yr BP (1950) using OxCal 4.2 (Bronk Ramsey, 2016).

3.2

3.3 Environmental datasets

A set of spatially referenced environmental datasets ~~is~~was used as predictors to reflect ecosystem properties in the study area. In the optical domain, an orthophoto of 1 m spatial resolution with RGB-bands from 2008 (© Lantmäteriet, I2014/00691) and a SPOT5 orthorectified multispectral satellite image (Path 045, row 208, acquired 10.08.2013)(© Lantmäteriet, I2014/00691) were available. The spectral bands of the SPOT image include green, red and near-infrared (NIR) at 10 m spatial resolution and a shortwave-infrared (SWIR) band at 20 m spatial resolution. A topographical correction for differential illumination was applied to the orthophoto and the SPOT image to compensate for terrain shadows using a "Minnaert correction with slope" implementation (Law and Nichol, 2004). The ~~illumination corrected~~ SPOT image was used to derive the normalized difference vegetation index (NDVI) (Rouse et al., 1974) and the soil-adjusted vegetation index (SAVI) with an L-value of 0.7 (Huete, 1988). Furthermore, the ratio of NIR/SWIR bands was~~is~~ used. A digital elevation model (DEM) of 2 m spatial resolution (© Lantmäteriet, I2014/00691) was used to generate several derivative topographic datasets. These include slope, aspect, profile and plan curvature, topographic ruggedness index (TRI), topographic position index (TPI) (Wilson et al., 2007), a TPI based landform classification (Guisan et al., 1999), and topographic wetness index (TWI) (Moore et al., 1991). Survey based vector maps with a scale of 1:250 000 were obtained for the geology and quaternary land cover (© SGU, I2014/00691) and for vegetation (© Lantmäteriet, I2014/00691). Geospatial analyses as well as raster and vector processing ~~wa~~were~~s~~ performed using GDAL/OGR (GDAL, 2016), SAGA (Conrad et al., 2015), Orfeo toolbox and R (R Core Team, 2017) softwares (see Code.A.1 in the supplement).

3.4 Land cover classification

An object-based approach was used to generate a detailed LCC (Blaschke, 2010). Object-based classifications avoid miss-classification of individual pixels that can be problematic with pixel-based classifications at very high-resolution. This reduces the need for post-processing, filtering and other generalization methods otherwise necessary (Siewert et al., 2015). The LCC is used as a predictor variable and for stratified extraction of the digital soil mapping results. First, the orthophoto was combined with the DEM ~~at~~resampled to 1 m spatial resolution. A segmentation layer was generated by grouping pixels into homogeneous areas with a minimum region size of 130 m². From this a water mask was classified ~~in a separate step~~ using the red band of the orthophoto and a slope layer. A land cover training set was created by combining field survey information with visual interpretation of the orthophoto and topography. The following layers were used as input for the classification algorithm: the orthophoto, elevation and slope; the SPOT5 4-band satellite image and NDVI (Rouse et al., 1974), SAVI (Huete, 1988) and NIR/SWIR (SPOT5 derivatives). The ratio of NIR/SWIR band can be beneficial to separate bedrock and bare rock areas (Andersson, 2016). The segments were then classified using a support vector machine (SVM; Chang and Lin, 2011) algorithm. Artificial surfaces were hand digitized and masked out. ~~The latter mainly includes a road and railway passing through the study area.~~ The individual thematic classes ~~are~~were adapted from Andersson (2016) and are described in Table A.1 in the supplement.

The quality of the classification was assessed using a set of 108 ground control points. These include the locations of the soil sampling sites and points along pathways collected in equal distance from the starting point. The kappa coefficient and the overall accuracy ~~are~~were calculated for all land covers excluding water and artificial areas (Congalton, 1991).

3.5 Digital soil mapping using machine-learning

~~This article investigates the general applicability of machine-learning in the specific context of regression techniques for the mapping of SOC in high-latitude permafrost and peatland environments.~~ Numerous machine-learning algorithms and approaches exist. A comprehensive general overview on machine-learning techniques is provided by Hastie et al. (2009) and the ~~general~~ use of different machine-learning algorithms for digital soil mapping is thoroughly discussed for instance in McBratney et al. (2003), Li et al. (2011), Were et al. (2015) and Taghizadeh-Mehrjardi et al. (2016). Thus, only a brief description follows. Four commonly used machine-learning techniques were compared: a multiple linear regression (MLR) model, an artificial neural network (ANN) (Ripley, 1996), a support vector machine (SVM) (Chang and Lin, 2011) and random forest (RF) (Breiman, 2001).

Multiple linear regression (MLR) assumes that the regression function defining the soil variable is linear or can be approximated using a linear equation. ~~In a linear regression model t~~The soil variable *f(x)* represents the dependent variable and the environmental predictors the independent variables X_i. Where *a* is the intercept and b_i are regression coefficients.

$$f(X)=a+\sum_{i=0}^{n} b_i x_i \tag{1}$$

The training data is used to define the regression equation and then used to predict the soil variable for unseen occurrences in the environmental space. MLR is a popular technique that is comparatively simple. It is possible for linear regression models to outperform non-linear methods i~~I~~n situations with limited input data and low-signal to-noise ratio~~, linear regression models can sometimes outperform non-linear methods~~ (Forkuor et al., 2017; Hastie et al., 2009).

For the MLR model the *lm* function in R was used (R Core Team, 2017). A model was trained using 10-fold cross-validation with five repetitions~~ was trained~~ to develop a stable model~~s~~ of SOC–Tot using the ‘caret’ R package (Forkuor et al., 2017; Kuhn, 2008b).

Artificial neural network (ANN) is a technique that simulates the biological nervous system. For continuous soil variables, it is a two-stage regression model typically represented by a network diagram with three layers. A layer of input cells represents environmental covariates ~~and transmi~~and is connected ~~ts it~~ to a layer of hidden cells, which is ~~forwards it~~connected ~~again~~ to an output layer representing the soil property to be predicted. ~~The units between the layers are connected by synapses.~~ These connections, also called weights or synapses, form a network that defines the model. The model simulates learning from examples by training the network iteratively with information about the conditions in which a certain value of the soil variable occurs. During each iteration the connection between the input layer, hidden layer and the output unit is adjusted. Finally, the trained model is used to predict soil properties of unvisited pixels (Behrens et al., 2005; Hastie et al., 2009).

The ANN was parameterized using a grid search approach for the variables defining the *size* of the hidden layer and the *decay* of weights in the neural network. The tuning was started with a value of 2 for *size* to avoid a local minimum. The 'caret' R package was used in combination with a 10-fold cross-validation with five repetitions to ~~develop~~fit a stable model of SOC–Tot based on the smallest RMSE value (Forkuor et al., 2017; Kuhn, 2008b).

Support vector machine (SVM) is a technique that generates an optimal separating hyperplane to differentiate classes that overlap and are not separable in a linear way. In this case, a large, transformed feature space is created to map the data with the help of kernel functions to separate it along a linear boundary. While initially developed for classification purposes, this technique can also be used for regression problems (Hastie et al., 2009; Vapnik, 1998).

For SVM a ε-regression with a gaussian radial basis kernel was used. This kernel can be considered a good general purpose kernel (Zeileis et al., 2004). The cost parameter *C* and the *ε* error threshold parameter were determined using the grid-search method. ~~This was combined~~ The final model was developed ~~with~~using ~~a~~ 10-fold cross-validation with five repetitions implemented by the 'caret' package in R (Forkuor et al., 2017; Kuhn, 2008b).

Random Forest (RF) is a tree-based learner that combines decision tree and bagging methods (Breiman, 2001). RF draws a number of bootstrap samples (n_{tree}) from the input dataset representing individual soil samples and grows a large amount of unpruned regression trees (e.g. 500), where at each node random samples (m_{try}) of the environmental predictors are chosen. It then averages the prediction of all trees to predict new data (Liaw and Wiener, 2002).

For RF the 'randomForest' R package was used (Liaw and Wiener, 2002). ~~The default~~ A value~~s~~ of 7 for m_{try} equaling the ~3/predictors, a node size of 5 and n_{tree} = 1~~5~~00 provided stable and visually meaningful ~~sound~~ results. Different variations of the parameters m_{try}, n_{tree}, node size and maximum node number were tested. However, these~~, but~~ generated visually inferior ~~visual~~ results~~,~~ while providing only minor improvements to R^2 ~~but~~ or increased~~ing~~ the dependency of the most important predictive variable indicating overfitting. ~~It is known that RF does not need extensive fine-tuning which can lead to overfitting (Ließ et al., 2016), thus the standard settings were applied.~~ Sampling was performed with ***replacement*** and ***bias correction*** was applied to decrease overestimation for low values and underestimation for high values. To achieve stable model results, a 10-fold cross-validation with five repetitions was applied (Forkuor et al., 2017; Kuhn, 2008).

3.6 Model selection and validation

Originally, all models overestimated SOC contents for bare ground surfaces. These areas ~~lowland such as ,~~include exposed bedrock, blockfields (areas covered by shattered rock fragments with little or no fine substrate; Fig. 2b) and stone beaches along ~~the~~ lake shores~~.~~ ~~A~~(alpine heat tundra with minimal soil development and cryogenic features forms a separate class). To address the overestimation, 10 pseudo-training samples ~~with~~of 0.0 kg C m^{-2} SOC were added at bare ground locations. These were based on field knowledge ~~at bare ground locations~~ and were identified in the orthophoto. They~~se~~ were distributed across the study area and kept to a low number to avoid strong bias of the training dataset. A similar approach has been used by Siewert et al. (2012) to support spatial interpolation of limited line measurements to estimate the sediment thickness of talus cones. ~~The performance of the models is assessed with and without these 10 pseudo-training samples.~~

T~~First, t~~he ~~performance~~ability of all four machine-learning algorithms (MLR, ANN, SVM and RF) to predict SOC–Tot was ~~. For the validation,evaluated~~tested by training each model~~ was trained~~ ~~using~~with ~~a 80% random split of~~ the~~-~~ soil pedon dataset (excluding pseudo-training samples). ~~The models are then assessed using an internal validation of predicted values against the training dataset and an external validation using the remaining 20 % split as an unseen control dataset.~~ The models ~~are~~were ~~compared~~evaluated based on three commonly used error criteria derived by cross-validation (one out method):~~ for the internal and external validation. The error criteria include~~ the coefficient of determination (R^2), the root mean squared error (RMSE) and Lin's concordance correlation coefficient (CCC) (Lin, 1989). R^2 is an indicator of model precision, while RMSE is an indicator of accuracy and CCC combines measures of accuracy and precision to determine the agreement to a 45° line. Maps were developed by applying the predictive models ~~of the entire soil pedon dataset~~ to the environmental datasets. These were then visually examined~~ for further model evaluation~~ and compared to a thematic map of the SOC storage based on the combination of the LCC and average SOC values of soil pedons per LCC class.

~~Originally, all models overestimated SOC contents for bare ground surfaces. These areas lowland such as ,include exposed bedrock, blockfields (areas covered by shattered rock fragments with little or no fine substrate; Fig. 2b) and stone beaches along the lake shores. Alpine heat tundra with minimal soil development and cryogenic features form a separate class. To address the overestimation, 10 pseudo-training samples with 0.0 kg C m^−2 SOC were added based on field knownledge at bare ground locations identified in the orthophoto. These were distributed across the study area and kept to a low number to avoid strong bias of the training dataset. A similar approach has been used by Siewert et al. (2012) to support spatial interpolation of limited line measurements to estimate sediment thickness of talus cones. The performance of the models is assessed with and without these 10 pseudo-training samples.~~

~~-~~For the final analysis, the best performing algorithm (RF) was chosen to model and develop maps for each of the following soil variables~~ depths~~: SOC–OL, SOC 0–30, SOC 0–100, SOC–Tot and the OL–Depth at a spatial resolution of 1 m. To investigate potential circumpolar SOC mapping efforts, the SOC–Tot was also mapped at spatial resolutions of 2 m, 10 m, 30 m, 100 m 250 m and 1000 m by resampling the spatial predictor dataset. These resolutions aim to mimic commonly used and freely available satellite data, e.g. Sentinel-2 (10 m), Landsat (30 m), Moderate Resolution Imaging Spectroradiometer (MODIS 250-1000 m) and Advanced Very High Resolution Radiometer (AVHRR; 1100 m). They also represent common resolutions for Earth System Model applications. The SOC was then extracted using the 1 m resolution LCC. ~~For this the entire soil pedon dataset and all environmental predictors were used.~~ Multicollinearity among the predictor variables was analyzed using a correlation matrix. Variables with a correlation >0.90 were excluded (Kuhn, 2008a) . Finally, ~~P~~predicted SOC values below 0 were set 0.

4 Results

4.1 Land cover classification

The LCC showed good agreement with the classes that have been observed in the field and areas that have been visually identified in the orthophoto (Fig. 4a and b). The accuracy assessment ~~against ground control points collect in the field~~ results in a Kappa value of 0.71 and an overall accuracy of 74% (Table A.2). ~~This does not include any water or artificial surfaces.~~ These values are comparable to other high-latitude LCC accuracy assessments (Schneider et al., 2009; Siewert et al., 2015; Virtanen et al., 2004; Virtanen and Ek, 2014).

The object-based classification has a minimum patch size of 130 m^2~~;~~. ~~hichw~~This size was found to best differentiate areas of homogeneous land cover, while preserving characteristic shapes of landforms ~~-~~relevant for SOC storage in high-latitudes, such as peat plateaus.

~~Reese et al. (2014, 2015) demonstrated for the Abisko area, that very detailed pixel based vegetation maps are possible by combining laser scanning point clouds and SPOT5 satellite imagery. However, the chosen patch-size seemed more realistic to reflect most-likely homogeneous soil properties at this scale, both as input to predict SOC content and for the stratified extraction of the predicted SOC values as applied here.This avoidsobject-based classifications avoid miss-classification of individual pixels that can be a problematic with pixel-based classification approaches at very high-resolution and it reduces the need for post-processing filtering and other generalization methods otherwise nesecessary (Siewert et al., 2015).~~

Fig. 4~~Figure 4~~ near here

4.2 Performance of four machine-learning algorithms to predict SOC

Four different models were compared to predict SOC–Tot stocks (Fig. 5). ~~Table 1 presents the results of the internal and external validation.~~ There were~~are~~ large discrepancies between the models. RF ~~consistently~~ achieved~~s~~ the highest coefficient of determination (R^2~~)~~; 0.736) and CCC (0.572), while having the lowest root mean squared error (RMSE; 14.131). This indicates high precision (R^2), good agreement with the 45° line (CCC) and high accuracy (RMSE). This was followed by SVM with a slightly ~~lower~~ inferior performance for each error criteria (R^2 = 0.726, CCC = 0.543; RMSE=14.9). MLR and ANN showed a significantly higher deviation of predicted to sampled SOC values (MLR: R^2 = 0.48, CCC = 0.418; RMSE=21.22; ANN: R^2 = 0.533, CCC = 0.483; RMSE=17.614). ~~achieved reasonable results for the internal validation. SVM showed acceptable results, while MLR and ANN fail to match values of the internal validation. The exclusion of the pseudo sampling points influences in particular the external validation for which the results show a significant drop in performance.~~

~~Table1 near hereOnly RF showed tantamount performance for the internal and external validation (R^2 = 0.939 compared to 0.908).~~

All models underestimate large values and overestimate low values of SOC-Tot ~~(Fig. 5)~~ (Fig. 5). This so called regression to mean effect is a known shortcoming. For ~~of~~ the RF algorithm ~~and~~it was addressed using the bias correction option in the ‘randomForest’ package (Liaw and Wiener, 2002; Zhang and Lu, 2012). Yet, a slight overestimation for low values from ~0–25 kg C m^{-2} and an underestimation for SOC–Tot values above ~60 kg C m^{-2} remains.

~~Furthermore, RF cannot forecast values that are beyond the training dataset. Thus, even if the environmental variables suggest higher SOC stocks, there is no trend extrapolation. In this study, the highest SOC value was 90 kg C m^{-2} and cannot be exceeded in the model. This likely underestimates SOC values for some areas with thicker peat deposits than the 138 cm measured in our transect sampling, as the thickness of peat can be up to 3 m in the mire (Malmer and Wallén, 1996). However, this also prevents gross overestimation of SOC.~~

Fig. 5~~Figure 5~~ near here

The developed output maps show significant differences for the four prediction models (Fig. 4 d-g). Major wetland areas can be recognized in all four maps. ~~Wetlands and peat bogs generally represent areas with significantly higher SOC storage compared to surrounding soils (e.g. Hugelius et al., 2011; Siewert et al., 2015).~~ The MLR model (Fig. 4d) shows a strong contrast in SOC storage with sharp transitions between wetland and ~~very strong gradients from~~ non-wetland areas ~~in SOC storage~~. ~~These~~SOC storage in non-wetland areas seems to be ~~under~~overestimated ~~opposed to~~. ~~w~~Wetland areas ~~that~~ are predicted to have a very high SOC storage~~values~~ throughout~~, compared to an expected distribution of SOC following a thematic map~~ corresponding well to the thematic map (Fig. 4c). The ANN model (Fig. 4e) does not generalize well ~~s not reflect well~~to the field situation. While ~~of strong~~ it differentiates well ~~contrasts~~ between wetland and non-wetland areas, it also exhibits a significant amount of noise with unrealistically high values in birch forest. ~~Birch forest areas and blockfields seem in general highly overestimated with SOC–Tot values of ~20-30 kg C m^{-2} compared to 4.2 ± 2.2 kg C m^{-2} for the average of soil pedons in birch forest.~~ Wetland areas show low values compared to the other models. SVM and RF visually correspond best to the field situation of sampled and analyzed soils as exemplified in the thematic map of SOC (Fig. 4~~d~~c, g, f). SVM ~~seems to slightly~~ overestimates low SOC values ~~close~~ for lowland birch forest and bare ground areas (Fig. 4f). RF ~~is the only prediction model that can~~ replicates the extent of elevated peat bogs and plateaus slightly better than SVM and MLR (Fig. 2 and Fig. 4d-g). RF and MLR perform best at ~~manages to~~ representing ~~the contrast~~low SOC values of bare ground areas and blockfields ~~with SOC values typically <1 kg C m^{-2}~~.

4.3 Environmental controls on SOC distribution

As the RF model provided the best performance for the error metrics and the visual evaluation, the remaining analysis proceeds using only RF.~~-~~ The importance of each input variable for the prediction of the SOC–Tot at 1 m resolution is presented in Fig. 6. The LCC is the most important predictive variable. ~~The strong dependence on the land cover classification likely reflects sensitivity to land cover segmentation.~~ This is followed by TWI, a group of SPOT 5 based input variables~~,~~ and elevation~~, TWI and slope~~. The SPOT5 variables include the NIR band, the ratio of NIR/SWIR~~,~~ and NDVI ~~and SAVI~~. ~~The strong dependence on the LCC likely reflects sensitivity to land cover segmentation. The SPOT5 variables variables are related to the sensitivity of the included bands to vegetation, but also to bare ground cover signature (Andersson, 2016; Huete, 1988; Rouse et al., 1974). Elevation likely reflects the lapse rate of the mean annual air temperature gradient in steep mountainous terrain. The contribution of the TWI data is likely to identify waterlogged conditions along streams and in mires. This is likely supported by slope and is particularly evident for alpine willow communities that are located along flow accumulation pathways.~~ Other variables ~~seem to~~ have limited~~equally little~~ predictive power~~.overestimation for the lower range of SOC values. an Exclusion of these predictive variables was tested, but generate~~ and five variables were excluded after testing for multicollinearity.

The variable importance changes for the individual prediction models of SOC 0–30, SOC 0–100, SOC–OL, Depth–OL (Fig.~~-~~ A.1) and SOC–Tot. However, the pattern usually resembles that of SOC–Tot (Fig. 6)**. Land cover is the most important enviro**nmental variable in all models except for SOC 0–30, where NIR/SWIR is the most important variable ~~with the~~followed by ~~LCC~~NDVI, LCC, ~~SAVI,~~ and NIR~~ in one group following, and NDVI~~.

Fig. 6~~Figure 6~~ near here

4.4 SOC stocks and landscape partitioning ~~and age~~

Table 1 shows the landscape partitioning of the sampled SOC pedon values and the predicted SOC values using machine-learning. The ~~L~~landscape mean SOC–Tot storage is predicted to be ~~7.9~~8.3 ± 8.0 kg C m^{-2} and ~~to~~ 7.~~0~~7 ± 6.~~3~~2 kg C m^{-2} for the top meter (~~0–100 cm;~~ SOC 0–100) of soil. This compares to 5.8 ± 0.5 kg C m^{-2} for the SOC–Tot ~~using the land cover class for thematic mapping~~ and 5.3 ± 0.5 kg C m^{-2} for the interval 0–100 cm using the LCC for thematic mapping. The highest SOC stock per class is estimated for the ***Sphagnum*** covered wetland~~s~~ areas (3~~8~~9.0~~5~~ ± ~~9~~8.7~~3~~ kg C m^{-2}) followed by the other wetland classes: peat bog (3~~6~~5.~~3~~5 ± 9.~~4~~0 kg C m^{-2}), lowland shrub wetland (3~~6~~7.~~4~~2 ± 7.~~8~~5 kg C m^{-2}), sedge wetland

(33.~~6~~7 ± 9.~~9~~0 kg C m^{-2}) and forested wetland (~~28~~31.~~6~~3 ± ~~6~~7.~~2~~6 kg C m^{-2}). The alpine willow class stores the highest amount of SOC of the remaining non-wetland classes with ~~9~~10.~~2~~4 ± 3.~~2~~4 kg C m^{-2}, followed by birch forest (7.~~9~~8 ± ~~4~~3.~~3~~8 kg C m^{-2}) and dwarf-shrubs (~~7~~8.~~5~~1 ± ~~4~~3.8~~0~~ kg C m^{-2}). The bare ground class stores the lowest amount of SOC with 1.~~1~~7 ± 2.~~2~~3 kg C m^{-2}. This represents most likely an overestimation and should be close to <0.1 kg C m^{-2}. For SOC–OL, SOC 0–30 and SOC 0–100 similar patterns emerge. Permafrost was encountered in six pedons of which 4 were located in ~~the~~ peat bog with an average depth of 50 ± 20 cm, one in the S~~s~~***phagnum*** wetlands and one alpine tundra heath. The mire permafrost soils were sampled in early September in 2013, while the alpine heath tundra (AL–Depth = 37 cm) pedon was sampled in June and does not represent maximum annual active layer depth. The partition of SOC stored in permafrost is 10.1 ± 13.3 kg C m^{-2} for peat bog soil~~ sample~~s and 8.5 ± 12.1 kg C m^{-2} for ***Sphagnum*** wetland s~~ample~~oils. This equals 0.2 ± 0.0 kg C m^{-2} of the total landscape SOC weighted by area using thematic mapping. Prediction of the SOC in permafrost ~~using digital soil mapping techniques~~ did not yield sound results (data not shown)~~, likely because permafrost is to a large extent a thermal property and needs a dedicated prediction approach for the occurrence and active layer thickness (Riseborough et al., 2008)~~. ~~Such results could however be combined with predicted SOC values.~~ The predicted depth of the OL compares well to sampled depths except for ***Sphagnum*** and forested wetlands, where it underestimates mean depth and for alpine willow class, where it approximately doubles the pedon mean.

~~Table 2~~Table 1 near here

The relative landscape SOC storage partitioning is shown in Fig. 7. Birch forest stores ~~41~~38% of the ~~total SOC~~SOC–Tot covering 41% of the soil area. This is followed by alpine heath tundra that stores ~~12~~14% of the SOC on 17% of the area. Wetlands store 25% of the SOC–Tot while only covering % of the landscape. The individual wetland classes have the highest ratio of stored SOC (5.~~5~~4–~~7~~6.~~8~~1%) compared to the area covered (1.2–1.8%) (except *Sphagnum* wetland, where the ratio is 0.~~7~~6% SOC to 0.1% of the area), while bare ground has the lowest ratio covering 8% of the area and storing only 1.~~2~~6% of the SOC.~~ This is likely an overestimation.~~

Fig. 7~~Figure 7~~ near here

4.5 Effect of reduced spatial resolution

The effect of a reduced spatial resolution on the estimate of landscape mean SOC–Tot storage and partitioning is presented in Fig. 8, Table 1 and Fig.A2. The estimate of SOC–Tot first increases from 8.3 ± 8.0 kg C m^{-2} at a resolution of 1 m to

10±8.0 kg C m^{-2} at a resolution of 30 m. At resolutions of 100 m, 250 m and 1000 m, the estimate drops to values between 7±5.6 and 7.2±7.9 kg C m^{-2}. This decrease is associated with a considerable drop in the estimate of SOC stored in wetland classes, while non-wetland classes do not change ~~drastically~~significantly in SOC estimates. The model at 10 m resolution has the highest R^2 with 0.777 (using 47 + 10 pedons), followed by the models at 1 m (R^2 = 0.759), 2 m (R^2 = 0.752) and 30 m (R^2 = 0.733). The models at resolutions of 100 m, 250 m and 1000 m have an R^2 between 0.520 and 0.538 (Fig. 8). The predicted maps show how SOC values are progressively less detailed and the strong contrast between wetland areas and the surrounding disappears. Each resolution emphasizes different details, e.g. SOC in alpine willow stands are most pronounced at a resolution of 10 m (Fig.A2.).

Fig. 8 near here

4.6 SOC history and age

Eight samples have been radiocarbon-dated to understand long-term C dynamics ~~in the system~~ (Table 2). The oldest sample from the central mire indicates a transition from C enriched mineral sediments to organic peat ~~with an age of~~at 5218 cal yrs BP. A second phase of increased peat accumulation was dated to 2230–1936 cal yrs BP ~~as found in~~based on a profile close to ~~the shores of~~ a lake ~~Mellersta Harrsjön~~ and in a profile in ombrotrophic waterlogged *Sphagnum* patch in the center of Stordalen mire. ~~, For the center of the mire~~Furthermore, one sample indicates a marked change in peat composition at 150.84 ± 0.36 pMC for the mire center, which likely corresponds to a change from poor fen to palsa peat accumulation. This is probably related to permafrost ~~accumulation~~formation (Kokfelt et al., 2010). A shallow soil pedon with ~31 cm depth ~~I~~in ~~the~~alpine birch forest located at mid-slope position was dated with two samples. ~~from a pedon~~The base of the OL w~~ere~~as dated to modern age and ~~-~~ the mineral subsoil at a depth of 22-28 cm was dated to 150 cal yrs BP ~~at the base of the OL on a very shallow soil of ~31 cm depth~~. In ~~-located~~~~lower~~birch forest located at the footslope, the base of the OL ha~~d~~s a modern age and soil at a depth of 20–26 cm was dated to 1345 cal yrs BP.

Table 2~~Table 3~~ near here

5 Discussion

~~Quantitative estimates of SOC, its spatial distribution and longterm dynamics in permafrost environments are a major uncertainty in future predictions of the global C cycle. Therefore, new SOC methods to estimate landscape scale SOC storage need to be investigated and connected to a temporal framework. This article demonstrates the successful prediction of SOC using machine-learning algorithms at very high spatial resolution (2 × 2m) in a sub-Arctic permafrost peatland environment. Four SOC prediction models were tested. The prediction approach is discussed, followed by insights into environmental controls of SOC, present day SOC stocks and their past and future development.~~

5.1 Predicting SOC using machine-learning algorithms

Of the compared prediction algorithms, the RF algorithm clearly performed best. This applies to all three error criteria: R^2, CCC and RMSE, as well as to the visual evaluation ~~(Error: Reference source not found)~~.~~The high R^2 of 0.939 for the internal validation can be confirmed by external validation ($R^2 = 0.908$), a high value for the CCC and lowest RMSE value.~~ On visual comparison, it was the only algorithm that could reflect the expected distribution of SOC following simple thematic mapping, and also replicate SOC storage ~~dominating~~defining landforms, ~~(~~i.e. peat plateaus~~)~~, realistically (Fig. 5). SVM also provided good results, but there was a significant drop in quality for MLR and ANN. The advantage of tree based machine-learning techniques such as RF is that they can cope well with non-linearity and have minimal assumptions about the data (McBratney et al., 2003). This is an important property, as in the northern circumpolar permafrost region only limited environmental datasets of varying quality are available. ANN and SVM are also non-linear methods, but don't seem to match the result of the tree based method ~~did not perform equally well as the RF modelhowever in this case they~~. The RF model has ~~often shows~~shown superior performance for regression applications in many environments (e.g. Forkuor et al., 2017; Li et al., 2011). However, in some environments other models outperformed RF, for example SVM (Were et al., 2015) or ANN (Taghizadeh-Mehrjardi et al., 2016). This indicates that different machine learning algorithms might suit different landscapes and that several algorithms should be compared (Forkuor et al., 2017). ~~Mishra and Riley (2012) showed that cographically weighted regression (GWR) can be successfully used at regional level to predict SOC stocks in Alaska at 60 m spatial resolution. However, GWR is based on the concept of spatial autocorrelation (Fotheringham et al., 2002). In Abisko, very strong environmental gradients of SOC distribution are found and suggest low spatial autocorrelation. In general, machine-learning algorithms are a very promising approach for regression modeling of SOC in peatland and permafrost environments with RF providing the best results.~~

This study highlights two distinctive and related local scale properties of high-latitude permafrost ecosystems~~,~~. ~~that~~These need to be considered when choosing an appropriate spatial prediction method~~machine-learning predictor~~. One factor is the strong land cover fragmentation of tundra environments with very small land cover patch sizes (Virtanen and Ek, 2014). In the study area, this is reflected in the occurrence of blockfields, small mires and peat plateaus (Fig. 2). The second factor is a high spatial variability of soil properties and thus SOC storage, related to the presence or absence of permafrost, peatlands and meter scale periglacial landforms. For example, blockfields are areas with SOC–Tot stocks of <0.1 kg C m^{-2}. These are often surrounded by forested areas with around ~~~4~~2 – 9~~8~~ kg C m^{-2} or in direct neighborhood of mire ecosystems with SOC–Tot stocks of ~~~2~~10–90 kg (Table 1). These sharp transitions in SOC storage between different land covers suggest low spatial autocorrelation at local scale, i.e. little relationship in SOC values between points far apart. Mishra and Riley (2012) showed that geographically weighted regression (GWR), which is based on spatial autocorrelation, can successfully be used to predict SOC stocks at regional scale in Alaska with a spatial resolution of 60 m. However, permafrost environments can be very variable at local scale.~~Analogous, Hugelius et al. (2011) showed for a study area in the European Russian Arctic, that a Quickbird based 2.4 m spatial resolution LCC was necessary to locate distinctive peat plateaus that store 30–58% of the ecosystem C dominated by SOC, while covering less than ~20% of the area.~~ In areas with ice-wedge polygons, common in lowland tundra environments, the local scale variability of SOC can be even higher than in Abisko, with SOC values from almost zero on polygon rims to several tens of kg in polygon centers on distances of a few meters (Ping et al., 2013, 2015, Siewert et al., 2015, 2016). Thus, any applied machine-learning approach must be able to cope with ~~strong~~sharp and ~~potentially~~ non-linear ~~environmental gradients~~transitions in SOC storage, while efforts based on spatial~~-~~ autocorrelation may fail at local scale~~.~~ unless they are supported by highly resolved environmental variables and a substantially higher number of soil pedons than usually available.

In Abisko, the distribution of SOC is defined by the occurrence of peatlands (Fig. 5 and Fig. 8), to the extent that two separate populations of soil pedons can be identified (Table 1). This strong non-linearity may be the reason, why some predictive models perform better. In the future, it should be tested if in such a case separate models for different populations of soil pedons can improve the prediction.

5.2 Limitations, sources of error and uncertainty

Several sources of error for the predicted SOC values can be identified. One source of uncertainty is related to the amount of sampled soil pedons and limitations of transect sampling as opposed to ideal random sampling. Given the difficult nature of Arctic and mountainous environments, transect sampling is the most time efficient data collection method providing sufficient amounts of soil pedons for regression analysis. Most sub-Arctic and Arctic research facilities are located in remote areas and logistics are difficult, thus the main source of uncertainty of many local and regional scale studies on SOC distribution is the low amount of available soil data with only ~10–50 pedons. The results show that machine-learning predictions using few soil pedons (n = 47~~ + ~~10) can provide sound results~~ using machine-learning models~~. However, the amount of soil pedons should ideally not be lower. Also, ~~T~~the importance to sample~~cover~~ the entire environmental gradient of a study area, including end-members with very low and very high-values in soil surveys ~~is~~needs to be underlined. The original dataset included only ~~very~~ few data points for the lower range of SOC–Tot values close to <0.1 kg C m^{-2} on blockfields resulting in overestimation. This was compensated by introducing 0.0 kg C m^{-2} SOC pseudo-training examples. At the same time, ~~T~~the SOC–Tot may be underestimated in small parts of the mires as the applied RF model cannot predict values beyond the range of the input data, e.g. in case the peat accumulation exceeds the maximum thickness of 138 cm that was sampled. Also, potential shifts in environmental gradients that have not been sampled, i.e. changes in geology, may have strong influence on SOC storage ~~possibly not covered by this study~~. These demands on ~~the data~~pedon databases are analogous to the ones outlined by Hugelius (2012) for making credible thematic map~~ping and supports the idea of stratified sampling schemes that explicitly include all landscape types~~.

~~, but risks the omission of important environmental gradients not covered by transects~~ ~~Sources of error for the predicted SOC values in this article include limitations of transect sampling as opposed to ideal random sampling. However, given the nature of Arctic environments, this is the most time effective data collection method providing sufficient amounts of soil pedons for regression analysis.~~Further, sources of uncertainty include ~~In this article~~the strong dependence of the RF model to the LCC that could result from overfitting. Yet, ~~good external~~the error metrics derived from cross-validation ~~of the projection~~ indicate~~s~~ reliable results. ~~and a~~A single dominating predictor variable has also been reported by other authors (e.g. Hengl et al., 2015). The fast degradation of permafrost in the environment has likely le~~d~~t to temporal differences in the high-resolution predictive datasets. For instance, areas visible as peat bog in the orthophoto from 2008 have since been submerged by water due to permafrost degradation and resulting in a mismatch with soil pedons collected in 2013-15~~may have affected some sampling points~~. Error propagation from the generation of the LCC and the use of the LCC as input variable and for stratified extraction cannot be ~~out-~~ruled out. Different spatial resolutions of the input data can reduce the spatial accuracy of higher resolution input layers. Future applications of machine-learning methods in this context should investigate spatial resolution optimization of input variables ~~, an automated forward selection of the input variables (Ließ et al., 2016) or alternatively and dimension reduction using principal component analysis (PCA) to reduce processing time and multicollinearity among the environmental variables (Howley et al., 2006).~~(Behrens et al., 2010; Drăguţ et al., 2009) and ~~Q~~quantitative uncertainty estimates with confidence intervals for the predicted SOC distribution ~~are an important next step~~ (Hugelius, 2012; Hugelius et al., 2014; Zhu, 2000).

5.3 Environmental controls of SOC distribution

A set of ~~23~~environmental variables with varying quality and spatial resolutions was used for the prediction of SOC–Tot at 1 m resolution. Land cover is the most important variable ~~to predict the total SOC stock~~, followed by TWI, a group of input variables based on the SPOT 5 image, including NIR, NIR/SWIR~~, NIR~~ and ~~SAVI~~NDVI, and ~~slopeDEM derivatives including TWI and elevation and~~ elevation. The relevance of land cover for the prediction model likely reflects its ability to segment distinct soil bodies and sharp transitions between land covers at high resolution. This seems to be particularly important to map high SOC stocks related to deep OL depths and peat deposits, as the importance of the LCC is diminished for the estimate of SOC–30 and SOC–100 (Fig.A.1). Furthermore, the LCC is an integrated parameter combining information from several data sources. The vegetation sensitive SPOT 5 bands and derived indices complement the LCC with information on vegetation productivity and the fraction of bare ground cover (Andersson, 2016; Huete, 1988; Rouse et al., 1974). The contribution of the TWI data is to identify soil moisture gradients. This is supported by curvature and is particularly evident for alpine willow communities that are located along flow accumulation pathways. Elevation likely reflects vegetation productivity as a function of the lapse rate of the mean annual air temperature gradient in steep mountainous terrain. Five other variables were excluded due to multicollinearity. The importance of NDVI in combination with TWI to predict SOC has also been reported by Taghizadeh-Mehrjardi et al. (2016). ~~important. notwere of the orthophoto sresolution, the bandbetter the or because ofDespite All other variables showed comparatively little relevance.~~

It is known that the scale of an environmental variable can have significant influence on the prediction accuracy and the highest resolution is not always providing the highest accuracy (Behrens et al., 2005; Drăguț et al., 2009). Despite the highest spatial resolution of 1 m, the red band of the orthophoto was only the 6th most important prediction variable.~~potentially catenary position. slope and soil moisture variabilityreflectsTWI and complement this with information on vegetation productivitycomposites The vegetation sensitive SPOT5 bands and . to the next land coverfrom one to reproduce distinct soil bodies and sharp transitions possibilityits contained in the LCC and alreadyThe relevance of land cover for the prediction model likely reflects the amount of information Micro-site effects such as cryoturbation patterns, wind erosion of the OL on small ridges, or increased accumulation in small pits in the alpine soils (Becher et al., 2013; Klaminder et al., 2009) likely occur at a finer resolution than can be resolved in this study. The input environmental variables were not selected or organized according to specific soil forming factors (McBratney et al., 2003). For example no climatic dataset is included. Yet, Klaminder et al. (2009) find a clear connection of SOC accumulation in dry tundra soils and mean annual precipitation along a transect from Abisko towards the more humid western coast. While this study can be considered to be representative for a mountainous area with discontinuous permafrost, it will be necessary to includeTherefore, climatic datasets would be necessary for a as environmental variables to correctly model SOC storage along a larger regional climatic gradientlarger study area.~~

While the LCC was constructed at 1 m resolution, each patch has a minimum size of 130 m², which is similar to the pixel size of the SPOT 5 variables at 10–20 m resolution. The DEM has an original resolution of 2 m and its derivative TWI is the second most important variable. While there is no clear trend for the variable importance,

~~From the~~the visual inspection ~~it~~shows that ~~seems~~ lower resolution environmental variables ~~reduce the spatial accuracy of the mapping result~~ creat~~ing~~e ~~a~~ pixel artifacts. ~~A~~An exclusion of all SPOT 5 input variables was tested, but significantly reduced model performance (result not shown). This indicates that even lower resolution environmental variables can improve the final prediction if they support higher resolution datasets. ~~Yet,~~ However, Samuel-Rosa et al. (2015) found that more detailed environmental variables only improved model performance incrementally and the cost may outweigh the benefits. Instead, efforts should be placed in more soil pedon data, which should also be a priority in permafrost environments. ~~Most sub-Arctic and Arctic research facilities are located in remote areas and logistics are difficult, thus the main source of uncertainty of many local and regional scale studies on SOC distribution is the low amount of available soil data with only ~10–50 pedons. The latest circumpolar SOC estimate for the top meter is based on only 1778 pedons and reports substantial regional gaps in pedon data, particularly for areas in the High Arctic with thin sediment overburden, and for cryoturbated soils and for peatland soils (Hugelius et al., 2014). Similarly, it was shown for the SOC storage in Alaska, that despite 556 existing soil pedons, >300 additional soil pedons arew necessary to reflect the entire environmental space (Vitharana et al., 2017).Future research should also investigate the use of new predictive variables, such as synthetic aperture radar remote sensing data sensitive to soil moisture, which has recently been used to continuously map SOC at circumpolar scale north of the treeline (Bartsch et al., 2016).~~

5.4 Comparing the present day SOC storage

This study provides the first landscape scale estimate and partitioning of SOC for the Abisko ~~area~~region. A mean landscape storage of 8~~7~~.3~~9~~ ± 8.0 kg C m^{-2} for the SOC–Tot and 7.~~0~~7 ± 6.~~3~~2 kg C m^{-2} for the SOC 0–100 is predicted. Overall, the predicted estimates are in line with previous studies for individual land cover types. In the mires of the study area, Klaminder et al. (2008) found organic matter stocks of 30–80 kg m^{-2} on hummocks (peat bogs) and 35–110 kg m^{-2} in hollows (wetlands) which translates into similar amounts of SOC as estimated in this study. Some publications have emphasized higher amounts of SOC in tundra heath ~7–9 kg C m^{-2} compared to birch forest ~4–5 kg C m^{-2} (Hartley et al., 2012; Parker et al., 2015). The sampled and predicted values for these classes are in the same range in this study, but higher amounts of SOC in tundra heath cannot be confirmed. ~~Potentially because the sampling included a wider range of birch forest sites, including areas with a longer time for SOC accumulation. C stocks in lakes were not included, but sediment cores published by Kokfelt et al. (2010) indicate that lakes in the study area could contain similar amounts of C as the sedge wetland class. Overall, the predicted estimates are in line with previous studies for individual land cover types.~~

Few studies have estimated ~~mean~~ landscape SOC stocks ~~at local scale~~ in mountainous environments in ~~the~~ permafrost regions. Low storage of SOC was documented by Fuchs et al. (2015) in ~~Tarfala,~~ a sub-Arctic alpine valley 50 km south of Stordalen. The SOC stocks vary from 0.05 kg C m^{-2} to 8.4 kg C m^{-2} for different land cover classes, with a landscape mean SOC storage of 0.9 ± 0.2 kg C m^{-2}. Dörfer et al. (2013) estimate ~~the mean landscape~~ SOC stocks for two study areas on the Tibetan plateau to 3.4 and 10.4 kg C m^{-2} for the top 0–30 cm of soil. For the Abisko area, a value of 3.9 ± 1.~~7~~8 kg C m^{-2} is calculated for the same depth interval showing a good agreement. A study with a similar environmental setting is presented by Palmtag et al. (2015) for~~who investigated SOC stocks in~~ Zackenberg, NE Greenland. This landscape features a combination of higher located barren alpine areas and lower located wetland areas including palsas. They found a mean landscape SOC storage of 8.3 ± 1.8 kg C m^{-2}; for the top meter. This~~which~~ compares well to this study, despite the location in the High Arctic. This could~~ potentially~~ be explained by decreased decomposition rates and slower C turnover in Greenland (Hobbie et al., 2000).

5.5 From local to circumpolar scale

The ~~SOC~~ estimate~~ for Abisko~~d 7.7 ± 6.2 kg C m^{-2} for the SOC–100 in Abisko is considerably lower than the 26.1 kg C m^{-2} ~~for SOC 0–100~~ estimated for the same area at circumpolar scale in the ~~Northern Circumpolar Soil Carbon Databasein the~~ NCSCD ~~indicated for this area~~ (Hugelius et al., 2013, 2014). A similar discrepancy has been noted for example by Fuchs et al. (2015) and Palmtag et al. (2016) for study areas in Sweden and Siberia. The NCSCD is based on thematic mapping using soil polygons for northern Europe with an average area of 205 ± 890 km² and averages many soil pedons for individual soil types across the entire Arctic. The Abisko study area falls in a single polygon in the NCSCD. Clearly, the large generalization and the thematic mapping approach cannot reflect local scale soil properties in~~–~~ this and other highly diverse permafrost environments.

Using digital soil mapping in combination with machine-learning methods can in the future provide improved, pixel-based regional to circumpolar estimates of SOC. The spatial resolution of such products is largely restricted by the available spatial input data and by computing resources. The ~~to map SOCeffectiveness~~ spatial prediction approach~~ at to resolutions~~ provides stable results at resolutions of 2 m, 10 m and 30 m for the SOC–Tot and for the SOC estimated under different land cover classes (Fig. 8). A significant drop in quality of the estimate occurs between a resolution of 30 m and 100 m. This results in an underestimation of the SOC–Tot at resolutions of 100 m, 250 m and 1000 m, as wetland classes can no longer be resolved by the spatial predictor variables~~ necessary to assess local or regional C dynamics.~~. This is associated with a drop in R^2 from ~0.7 at resolutions <30m to ~0.5 at resolutions >30m.

The importance of wetlands for SOC global stocks has long been pointed out. ~~Peatlands~~While they store large amounts of SOC~~ while~~, they typically occupy~~ing~~ only a small fraction of the ~~total~~ landscape~~. Northern peatlands are highly relevant to the global C cycle and store large amounts of SOC in boreal forests and tundra regions~~ (Gorham, 1991)~~. Peatlands are restricted to waterlogged conditions. In mountainous terrain they therefore mostly occupy valley bottoms. Subsetting the SOC–Tot map by areas that have a OL–Depth ≥40 cm, corresponding to a common definition of peatlands (Tarnocai and Stolbovoy, 2006), it is found that peatlands represent 3.2% of the total soil area, but 13.9% of the SOC–Tot at a resolution of 1 m (Fig. 8). According to the LCC, the area covered by all wetland classes is 6.8 %, but this area stores 25.0% of the SOC. For comparison, the Swedish CORINE land cover dataset (© Lantmäteriet, I2014/00691) based on Landsat TM imagery with a resolution of 25 m, indicates a wetland area of only 3.3%. Extracting the SOC for these areas results in only 11.0% of the landscape SOC–Tot. The difference can be attributed to many small scale wetlands in forest areas not captured in the lower resolution datasets. This means an underestimation of wetland areas using a country scale LCC and SOC stored in~~

~~wetlands by 3.4% of the total soil area and 14.1% of the SOC respectively. ~~. In Abikso, wetlands store 25% of the carbon, but cover only 6% of the area (Fig. 7). ~~While the m~~Major wetland complexes ~~are~~can be mapped at coarse resolution, but small ones are~~ clearly~~ often omitted and high-resolution approaches are necessary to extract this information~~The effect of spatial resolution on the mapping of wetlands and fens has 75% of the landscape SOC is not stored in wetlands, but to a large extent in the birch forest with 40.8% and in alpine tundra heath with 12.1% (Fig. 7). It has to be pointed out that . also been pointed out by .~~(Hugelius, 2012; Virtanen and Ek, 2014). For example, Hugelius et al. (2011) showed for a study area in the European Russian Arctic, that a Quickbird based 2.4 m spatial resolution LCC was necessary to map peat plateaus that store 30–58% of the ecosystem C dominated by SOC, while covering less than ~20% of the area. Subsetting the SOC–Tot map for Abisko to show only wetland classes highlights the fragmentation and dispersion of minor wetlands detached from major wetland complexes (Fig.A2h). ~~tYet,~~The significance of these smaller wetland~~ areas~~s for C cycling at landscape scale has so far found little attention in the literature~~.~~.

Future updates at circumpolar scale would clearly benefit from including high resolution data derived from Sentinel-2 (10 m), Landsat (30 m) satellites or the ArcticDEM (2-5 m; 2017). At such resolutions, SOC estimates may be considered reliable in environments similar to Abisko. However, areas with a significant amount of ice-wedge polygons may require~~even~~ higher-resolution mapping approaches (Siewert et al., 2015, 2016). At the same time, this will need considerable amounts of computing power. Estimates using data from satellites like MODIS (250-1000 m) and AVHRR (1100 m) are likely to underestimate SOC stocks and may not reflect extreme values. At regional to circumpolar scale, it will also be necessary to include climatic datasets as environmental variables to correctly model the SOC storage. For example, Klaminder et al. (2009) find a clear connection between mean annual precipitation and SOC stored in tundra soils along a transect from Abisko towards the more humid western coast (not covered here). Future research should also investigate the use of new predictive variables, such as synthetic aperture radar remote sensing data. This product is sensitive to soil moisture and has recently been used to continuously map SOC at circumpolar scale north of the treeline (Bartsch et al., 2016). A priority should however be the collection of soil pedons. The latest circumpolar SOC estimate for the top meter is based on only 1778 pedons and reports substantial regional gaps in pedon data. This is particularly the case for areas in the High Arctic with thin sediment overburden, for cryoturbated soils and for peatland soils (Hugelius et al., 2014). Similarly, it was shown for the SOC storage in Alaska, that despite 556 existing soil pedons, >300 additional soil pedons are necessary to reflect the entire environmental space (Vitharana et al., 2017).

5.6 SOC age, past and future development

A major research question is whether Arctic environments have in the past ~~and will be in the future a~~been a sink or a source of C and how this will develop in the future. The Abisko ~~area~~region was deglaciated around 9500 cal. yrs BP (Berglund et al., 1996), leaving a ***glacier forefield*** like landscape with no SOC. This study finds that initial peat inception in Stordalen took place around 5200 yrs ago in the center part of the mire, while other studies indicate ~~a range~~peat inception between 6000 and 4700 yrs for different parts of the mire (Sonesson, 1972; Kokfelt et al., 2010). Other ~~amples~~profiles from ~~other~~different parts of the mire indicate an initial peat deposition and a transition to pure peat between 1900 and 2200 cal BP. This peat ~~that may have~~has likely accumulated due to peat erosion from already developed peat deposits in the surrounding, followed by continuous peat production. Similarly, Kokfelt et al. (2010) found a change towards ombrotrophic conditions and potential permafrost aggradation ~~in the mire~~ taking place around 2800 cal BP and prevailing permafrost conditions between 2650–2100 cal BP. This was followed by a phase of thermokarst and peat erosion and SOC accumulation in surrounding lakes. After 700 cal BP permafrost conditions ~~prevailed~~reappeared and palsa formation took place in the northern part of the mire around 120 cal BP (Kokfelt et al., 2010). This interpretation with relatively stable accumulation of SOC in the past 2000 is coherent with the results in this article.

The majority of the sampled SOC in the birch forest has an age of less than 1350 years. No signs of significant SOC burial due to solifluction or cryoturbation processes w~~as~~ereas found ~~in the transect sampling~~. ~~Yet, t~~However, these processes can store significant amounts of SOC in alpine and tundra terrain (Palmtag et al., 2015; Siewert et al., 2016) and are a common phenomenon on some slopes of the area. Becher et al. (2013) found three major periods of burial of SOC in non-sorted circles near Abisko, that coincide with transitions from colder to warmer conditions~~. These were~~ dated to 0–100 A.D., 900–1250 A.D. and 1650–1950 A.D.

~~Overall~~In summary, the bulk of the present day SOC in the study area ~~has~~must have accumulated during the past 2000 yrs~~, both for the peatlands and the birch forest~~. Despite some episodes of palsa and peat plateau degradation and peat erosion, it can be assumed that the study area has over the time period of the Holocene ~~likely~~ been a C sink, with a significant portion of the SOC stored in labile and temperature sensitive peat plateaus.

At present, permafrost of the peat plateaus of northern Fennoscandia seems to be at a critical thermal limit and close to collapse. In the Abisko region, permafrost is warming rapidly (Johansson et al., 2011). In Stordalen, a decrease of dry peat plateau areas by 10% and an increase of wet graminoid dominated water areas by 17% has been documented for the period from 1970 to 2000 and was likely caused by permafrost degradation (Malmer et al., 2005). ~~rs (Johansson et al, (2015). Snow manipulation experiments have shown a rapid increase of ground temperatures leading to permafrost degradation and changes in vegetation within only seven yea~~In Tavvavuoma, a peat plateau/thermokarst lake complex located in the sporadic permafrost zone in northern Sweden, the same trend is observed. Here significant thermokarst lake formation, drainage and infilling with fen vegetation has occurred from 1963 to 2003 (Sannel and Kuhry, 2011). These are significant landscape changes that affect the C balance in peat plateau areas and the permafrost in this area is clearly not in equilibrium with the present day warmer climate causing dynamic ~~Unless pronounced cooling sets in permafrost degradation in the peat plateau will occur (Sannel et al., 2015)..~~adjustments.

~~Past degradation of permafrost in Abisko resulted in redistribution of C rich organic sediments from eroding peat plateaus . (Kokfelt et al., 2010) and partial loss as dissolved organic carbon (DOC) suggest that much of the eroded peat from Stordalen has been lost as dissolved organic carbon (DOC) (Malmer et al., 2005). Analogous to paraglacial sediment systems (Ballantyne, 2002; Church and Ryder, 1972), the decay of permafrost will trigger and condition a set of processes that likely result in a significant geomorphic impact. While in the alpine heath ecosystem cryoturbation followed periods of climate interruptions causing burial of SOC (Becher et al., 2013, 2015).P Analogous to paraglacial sediment systems (Ballantyne, 2002; Church and Ryder, 1972), the decay of permafrost will trigger and condition a set of processes that likely result in a significant geomorphic impact.~~

How these adjustments will influence the carbon balance in the future seems unclear. Lundin et al. (2016) find that the Stordalen catchment is unlikely to be a present day C sink, but rather acts as a source of C. However, Fuchs et al. (2015) argue for the alpine Tarfala valley with very low SOC stocks, that these landscapes will under future climatic changes turn into a sink of C, despite degradation of the permafrost, as biomass will increase and soils develop. In Abisko most of the SOC is stored in birch forest soils. A further shrubification of alpine areas and a rise of the tree line combined with a positive priming effect could bind more SOC in the above ground vegetation (Hartley et al., 2012). Yet, for most permafrost environments this is unlikely to offset SOC releases from the permafrost and deeper soil layers by mass (Siewert et al., 2015) and may even lead to a release of SOC (Hartley et al., 2012). In wetlands, an increase of sedge and *Eriophorum* dominated water submerged areas will be important to regulate DOC and SOC dynamics (Tang et al., 2018) .~~Also~~Furthermore, the role of many minor wetland areas revealed by high resolution mapping has not received sufficient attention. Indeed, a holistic perspective will be necessary to predict how SOC storages will evolve in post-permafrost landscapes in the future.

6 Conclusions

F~~Very f~~ew studies have applied ~~regression methods or~~ digital soil mapping techniques to map soil organic carbon (SOC) in sub-Arctic and Arctic permafrost environments ~~to map soil organic carbon (SOC)~~. In general thematic mapping approaches have been favored. P~~The results show p~~romising results have been obtained using machine-learning techniques to predict SOC in a mountainous sub-Arctic peatland environment with typical periglacial landforms such as permafrost raised peat plateaus. A random forests prediction model showed the best results in terms of ~~)908coefficient of determination (R^2 = 0.~~error metrics and upon visual inspection. ~~typical for tundra environments.very strong environmental gradientsThis is~~

~~followed by a support vector machine, while multiple linear regression and artificial neural networks could not sufficiently reflect the fragmented SOC distribution with ~~Digital soil mapping of SOC is a significant improvement over upscaling methods using thematic maps such as land cover classifications~~ or soil maps~~. Yet, carefully generated thematic maps remain essential to our understanding of a landscape, its partitioning and patchiness. Thematic maps can be used as a predictor variable and for stratified extraction of soil properties such as SOC. This will ~~and ~~help to understand these variables at landscape level.~~Future field surveys must pay attention to sample the entire environmental gradient, including low C storage end-members. This applies to local as well as circumpolar scale, where large regional soil pedon gaps remain especially for low SOC areas.This study shows good initial results for the use of machine-learning algorithms to project SOC stocks even with limited ground data typical for remote and less accessible study areas in the Arctic. ~~

For the Abisko study area, the mean landscape total SOC storage is estimated to ~~7~~8.~~9~~3 ± 8.0 kg C m^{-2} and the 0–100 cm storage to 7.~~0~~7 ± 6.~~3~~2 kg C m^{-2}~~;~~ at a spatial resolution of 1 m. This estimate is significantly lower than the estimates from the Northern Circumpolar ~~li~~Soil Carbon Database, but in line with other high-resolution mapping results of SOC storage in similar environments, indicating the value of high-resolution upscaling and partitioning studies. Reducing the spatial resolution of the environmental input data reveals a significant drop in mapping accuracy for resolutions coarser than 30 m and a tendency to underestimate SOC stocks. The amount of soil pedon data is clearly the limiting factor to map SOC in permafrost environments. Future field surveys must pay attention to sample the entire environmental gradient, including low C storage end-members. This applies to local as well as circumpolar scale, where large regional soil pedon gaps remain especially for areas with low SOC stocks. Many small wetlands~~ that~~ are not ~~resolved in~~mapped at ~~lower~~coarse resolutions ~~studies are mapped ~~and the results highlight the importance of peatlands and peat plateaus for the total SOC stocks in sub-Arctic environments. Most SOC in the study area was accumulated during the past 2000 years. The landscape history emphasis that present day SOC stocks represent a snapshot in time in an ecosystem that is subject to continuous environmental ~~change~~adaptation associated with ~~complex biotic and abiotic ecological adoptions and interactions~~permafrost degradation caused by climate change. ~~. The development of alpine heath and birch forest SOC stocks remains unclear. and will likely release these into the carbon cycleinto lakes – of organic sediments and transportRapid future permafrost degradation in peatlands may lead to erosion~~A holistic approach will be necessary to understand ~~‘post-~~how permafrost~~’ processes~~ degradation ~~and~~will affect the landscape distribution of SOC in the future.

7 Acknowledgements

The soil data used in this study was collected as part of the European Union FP7—ENVIRONMENT project PAGE21 (grant number 282700), the NordForsk DEFROST (grant number 23001) project and with the support of the Bolin Centre for climate research (grant to Juri Palmtag). I would like to thank Matthias Fuchs, Jannike Andersson, Juri Palmtag and Robin

Wojcik for their assistance during fieldwork. Matthias Fuchs and Robin Wojcik for the analysis of the soil samples. I thank Peter Kuhry for his valuable advices to my work. I greatly acknowledge Gustaf Hugelius for his insightful comments on the manuscript and his continuous support.

8 References

Åkerman, H. J. and Johansson, M.: Thawing permafrost and thicker active layers in sub-arctic Sweden, Permafr. Periglac. Process., 19(3), 279–292, doi:10.1002/ppp.626, 2008.

Andersson, J.: Assessing the ecosystem carbon storage in two subarctic areas - Phytomass and soil organic carbon stock at Abisko in Sweden and Utsjoki in Finland, Master Thesis, Stockholm University, Stockholm, 22 January., 2016.

ArcticDEM: ArcticDEM, [online] Available from: https://www.pgc.umn.edu/data/arcticdem/, 2017.

Bartsch, A., Widhalm, B., Kuhry, P., Hugelius, G., Palmtag, J. and Siewert, M. B.: Can C-band synthetic aperture radar be used to estimate soil organic carbon storage in tundra?, Biogeosciences, 13(19), 5453–5470, doi:10.5194/bg-13-5453-2016, 2016.

Baughman, C. A., Mann, D. H., Verbyla, D. L. and Kunz, M. L.: Soil surface organic layers in Arctic Alaska: Spatial distribution, rates of formation, and microclimatic effects, J. Geophys. Res. Biogeosciences, 120(6), 2015JG002983, doi:10.1002/2015JG002983, 2015.

Becher, M., Olid, C. and Klaminder, J.: Buried soil organic inclusions in non-sorted circles fields in northern Sweden: Age and Paleoclimatic context, J. Geophys. Res. Biogeosciences, 118(1), 104–111, doi:10.1002/jgrg.20016, 2013.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D. and Goldschmitt, M.: Digital soil mapping using artificial neural networks, J. Plant Nutr. Soil Sci., 168(1), 21–33, doi:10.1002/jpln.200421414, 2005.

Behrens, T., Zhu, A.-X., Schmidt, K. and Scholten, T.: Multi-scale digital terrain analysis and feature selection for digital soil mapping, Geoderma, 155(3–4), 175–185, doi:10.1016/j.geoderma.2009.07.010, 2010.

Berglund, B. E., Barnekow, L., Hammarlund, D., Sandgren, P. and Snowball, I. F.: Holocene forest dynamics and climate changes in the Abisko area, northern Sweden: the Sonesson model of vegetation history reconsidered and confirmed, Ecol. Bull., 15–30, 1996.

Blaschke, T.: Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens., 65(1), 2–16, doi:10.1016/j.isprsjprs.2009.06.004, 2010.

Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E. and Kienast-Brown, S., Eds.: Digital Soil Mapping, Springer Netherlands, Dordrecht. [online] Available from: http://link.springer.com/10.1007/978-90-481-8863-5 (Accessed 23 May 2015), 2010.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, 2001.

Bronk Ramsey, C.: OxCal 4.2, [online] Available from: https://c14.arch.ox.ac.uk/embed.php?File=oxcal.html (Accessed 28 April 2016), 2016.

Brown, J., Ferrians, O. J., Heginbottom, J. A. and Melnikov, E. S.: Circum-Arctic map of permafrost and ground-ice conditions, National Snow and Ice Data Center, Boulder, Colorado. [online] Available from: http://pubs.usgs.gov/cp/45/report.pdf (Accessed 6 June 2013), 1997.

Callaghan, T. V., Jonasson, C., Thierfelder, T., Yang, Z., Hedenas, H., Johansson, M., Molau, U., Van Bogaert, R., Michelsen, A., Olofsson, J., Gwynn-Jones, D., Bokhorst, S., Phoenix, G., Bjerke, J. W., Tommervik, H., Christensen, T. R., Hanna, E., Koller, E. K. and Sloan, V. L.: Ecosystem change and stability over multiple decades in the Swedish subarctic: complex processes and multiple drivers, Philos. Trans. R. Soc. B Biol. Sci., 368(1624), 20120488–20120488, doi:10.1098/rstb.2012.0488, 2013.

Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. TIST, 2(3), 27, 2011.

Christensen, T. R., Johansson, T., Åkerman, H. J., Mastepanov, M., Malmer, N., Friborg, T., Crill, P. and Svensson, B. H.: Thawing sub-arctic permafrost: Effects on vegetation and methane emissions, Geophys. Res. Lett., 31(4), L04501, doi:10.1029/2003GL018680, 2004.

Congalton, R. G.: A review of assessing the accuracy of classifications of remotely sensed data, Remote Sens. Environ., 37(1), 35–46, doi:doi:10.1016/0034-4257(91)90048-B, 1991.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci Model Dev, 8(7), 1991–2007, doi:10.5194/gmd-8-1991-2015, 2015.

Davidson, E. A. and Janssens, I. A.: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change, Nature, 440(7081), 165–173, 2006.

Ding, J., Li, F., Yang, G., Chen, L., Zhang, B., Liu, L., Fang, K., Qin, S., Chen, Y., Peng, Y., Ji, C., He, H., Smith, P. and Yang, Y.: The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores, Glob. Change Biol., 22(8), 2688–2701, doi:10.1111/gcb.13257, 2016.

Dörfer, C., Kühn, P., Baumann, F., He, J.-S. and Scholten, T.: Soil Organic Carbon Pools and Stocks in Permafrost-Affected Soils on the Tibetan Plateau, edited by C. P. Slomp, PLoS ONE, 8(2), e57024, doi:10.1371/journal.pone.0057024, 2013.

Drăguț, L., Schauppenlehner, T., Muhar, A., Strobl, J. and Blaschke, T.: Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling, Comput. Geosci., 35(9), 1875–1883, doi:10.1016/j.cageo.2008.10.008, 2009.

FAO: World reference base for soil resources 2014 international soil classification system for naming soils and creating legends for soil maps Update 2015, FAO, Rome., 2015.

Forkuor, G., Hounkpatin, O. K. L., Welp, G. and Thiel, M.: High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models, PLoS ONE, 12(1), doi:10.1371/journal.pone.0170478, 2017.

Fuchs, M., Kuhry, P. and Hugelius, G.: Low below-ground organic carbon storage in a subarctic Alpine permafrost environment, The Cryosphere, 9(2), 427–438, doi:10.5194/tc-9-427-2015, 2015.

GDAL: GDAL - Geospatial Data Abstraction Library: Version 1.11.2, Open Source Geospatial Foundation., 2016.

Gorham, E.: Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming, Ecol. Appl., 1(2), 182–195, doi:10.2307/1941811, 1991.

Guisan, A., Weiss, S. B. and Weiss, A. D.: GLM versus CCA spatial modeling of plant species distribution, Plant Ecol., 143(1), 107–122, 1999.

Hartley, I. P., Garnett, M. H., Sommerkorn, M., Hopkins, D. W., Fletcher, B. J., Sloan, V. L., Phoenix, G. K. and Wookey, P. A.: A potential loss of carbon associated with greater plant growth in the European Arctic, Nat. Clim. Change, 2(12), 875–879, doi:10.1038/nclimate1575, 2012.

Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning, Springer New York, New York, NY. [online] Available from: http://link.springer.com/10.1007/978-0-387-84858-7 (Accessed 7 August 2016), 2009.

Heiri, O., Lotter, A. F. and Lemcke, G.: Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results, J. Paleolimnol., 25(1), 101–110, doi:10.1023/A:1008119611481, 2001.

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., Jesus, J. M. de, Tamene, L. and Tondoh, J. E.: Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions, PLOS ONE, 10(6), e0125814, doi:10.1371/journal.pone.0125814, 2015.

Hobbie, S. E., Schimel, J. P., Trumbore, S. E. and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude soils, Glob. Change Biol., 6(S1), 196–210, 2000.

Huete, A. R.: A soil-adjusted vegetation index (SAVI), Remote Sens. Environ., 25(3), 295–309, doi:10.1016/0034-4257(88)90106-X, 1988.

Hugelius, G.: Spatial upscaling using thematic maps: An analysis of uncertainties in permafrost soil carbon estimates, Glob. Biogeochem. Cycles, 26(2), doi:10.1029/2011GB004154, 2012.

Hugelius, G. and Kuhry, P.: Landscape partitioning and environmental gradient analyses of soil organic carbon in a permafrost environment, Glob. Biogeochem. Cycles, 23(3), GB3006, doi:10.1029/2008GB003419, 2009.

Hugelius, G., Kuhry, P., Tarnocai, C. and Virtanen, T.: Soil organic carbon pools in a periglacial landscape: a case study from the central Canadian Arctic, Permafr. Periglac. Process., 21(1), 16–29, doi:10.1002/ppp.677, 2010.

Hugelius, G., Virtanen, T., Kaverin, D., Pastukhov, A., Rivkin, F., Marchenko, S., Romanovsky, V. and Kuhry, P.: High-resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic, J. Geophys. Res., 116(G3), doi:10.1029/2010JG001606, 2011.

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P. and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5(1), 3–13, doi:10.5194/essd-5-3-2013, 2013.

Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C.-L., Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J. and Kuhry, P.: Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps, Biogeosciences, 11(23), 6573–6593, doi:10.5194/bg-11-6573-2014, 2014.

IPCC: Climate change 2013: The physical science basis, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Intergov. Panel Clim. Change Work. Group Contrib. IPCC Fifth Assess. Rep. AR5Cambridge Univ Press N. Y. [online] Available from: http://www.climatechange2013.org/images/report/WG1AR5_Frontmatter_FINAL.pdf (Accessed 26 November 2014), 2013.

Johansson, M., Akerman, J., Keuper, F., Christensen, T. R., Lantuit, H. and Callaghan, T. V.: Past and Present Permafrost Temperatures in the Abisko Area: Redrilling of Boreholes, Ambio, 40(6), 558–565, doi:10.1007/s13280-011-0163-3, 2011.

Jonasson, C., Sonesson, M., Christensen, T. R. and Callaghan, T. V.: Environmental Monitoring and Research in the Abisko Area—An Overview, AMBIO, 41(3), 178–186, doi:10.1007/s13280-012-0301-6, 2012.

Klaminder, J., Yoo, K., Rydberg, J. and Giesler, R.: An explorative study of mercury export from a thawing palsa mire, J. Geophys. Res. Biogeosciences, 113(G4), G04034, doi:10.1029/2008JG000776, 2008.

Klaminder, J., Yoo, K. and Giesler, R.: Soil carbon accumulation in the dry tundra: Important role played by precipitation, J. Geophys. Res. Biogeosciences, 114(G4), G04005, doi:10.1029/2009JG000947, 2009.

Köchy, M., Hiederer, R. and Freibauer, A.: Global distribution of soil organic carbon – Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world, SOIL, 1(1), 351–365, doi:10.5194/soil-1-351-2015, 2015.

Kokfelt, U., Reuss, N., Struyf, E., Sonesson, M., Rundgren, M., Skog, G., Rosén, P. and Hammarlund, D.: Wetland development, permafrost history and nutrient cycling inferred from late Holocene peat and lake sediment records in subarctic Sweden, J. Paleolimnol., 44(1), 327–342, doi:10.1007/s10933-010-9406-8, 2010.

Kuhn, M.: Building Predictive Models in R Using the caret Package, J. Stat. Softw., 28(5), doi:10.18637/jss.v028.i05, 2008a.

Kuhn, M.: Caret package, J. Stat. Softw., 28(5), 1–26, 2008b.

Law, K. H. and Nichol, J.: Topographic correction for differential illumination effects on IKONOS satellite imagery, Int Arch Photogramm Remote Sens Spat Inf. Sci, 641–646, 2004.

Li, J. and Heap, A.: A review of spatial interpolation methods for environmental scientists, Geoscience Australia, Canberra., 2008.

Li, J., Heap, A. D., Potter, A. and Daniell, J. J.: Application of machine learning methods to spatial interpolation of environmental variables, Environ. Model. Softw., 26(12), 1647–1659, doi:10.1016/j.envsoft.2011.07.004, 2011.

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R News, 2(3), 18–22, 2002.

Lin, L. I.-K.: A Concordance Correlation Coefficient to Evaluate Reproducibility, Biometrics, 45(1), 255–268, doi:10.2307/2532051, 1989.

Lundin, E. J., Klaminder, J., Giesler, R., Persson, A., Olefeldt, D., Heliasz, M., Christensen, T. R. and Karlsson, J.: Is the subarctic landscape still a carbon sink? Evidence from a detailed catchment balance, Geophys. Res. Lett., 43(5), 2015GL066970, doi:10.1002/2015GL066970, 2016.

Malmer, N., Johansson, T., Olsrud, M. and Christensen, T. R.: Vegetation, climatic changes and net carbon sequestration in a North-Scandinavian subarctic mire over 30 years, Glob. Change Biol., 11(11), 1895–1909, doi:10.1111/j.1365-2486.2005.01042.x, 2005.

McBratney, A. B., Mendonça Santos, M. L. and Minasny, B.: On digital soil mapping, Geoderma, 117(1–2), 3–52, doi:10.1016/S0016-7061(03)00223-4, 2003.

McGuire, A. D., Anderson, L. G., Christensen, T. R., Dallimore, S., Guo, L., Hayes, D. J., Heimann, M., Lorenson, T. D., Macdonald, R. W. and Roulet, N.: Sensitivity of the carbon cycle in the Arctic to climate change, Ecol. Monogr., 79(4), 523–555, 2009.

Mishra, U. and Riley, W. J.: Alaskan soil carbon stocks: spatial variability and dependence on environmental factors, Biogeosciences, 9(9), 3637–3645, doi:10.5194/bg-9-3637-2012, 2012.

Mishra, U. and Riley, W. J.: Active-Layer Thickness across Alaska: Comparing Observation-Based Estimates with CMIP5 Earth System Model Predictions, Soil Sci. Soc. Am. J., 78(3), 894, doi:10.2136/sssaj2013.11.0484, 2014.

Mishra, U., Jastrow, J. D., Matamala, R., Hugelius, G., Koven, C. D., Harden, J. W., Ping, C. L., Michaelson, G. J., Fan, Z., Miller, R. M., McGuire, A. D., Tarnocai, C., Kuhry, P., Riley, W. J., Schaefer, K., Schuur, E. A. G., Jorgenson, M. T. and Hinzman, L. D.: Empirical estimates to reduce modeling uncertainties of soil organic carbon in permafrost regions: a review of recent progress and remaining challenges, Environ. Res. Lett., 8(3), 035020, doi:10.1088/1748-9326/8/3/035020, 2013.

Moore, I. D., Grayson, R. B. and Ladson, A. R.: Digital terrain modelling: a review of hydrological, geomorphological, and biological applications, Hydrol. Process., 5(1), 3–30, 1991.

Palmtag, J., Hugelius, G., Lashchinskiy, N., Tamtorf, M. P., Richter, A., Elberling, B. and Kuhry, P.: Storage, landscape distribution, and burial history of soil organic matter in contrasting areas of continuous permafrost, Arct. Antarct. Alp. Res., 47(1), 71–88, doi:http://dx.doi.org/10.1657/AAAR0014-027, 2015.

Palmtag, J., Ramage, J., Hugelius, G., Gentsch, N., Lashchinskiy, N., Richter, A. and Kuhry, P.: Controls on the storage of organic carbon in permafrost soil in northern Siberia: Soil organic carbon storage in permafrost terrain, northern Siberia, Eur. J. Soil Sci., 67(4), 478–491, doi:10.1111/ejss.12357, 2016.

Parker, T. C., Subke, J.-A. and Wookey, P. A.: Rapid carbon turnover beneath shrub and tree vegetation is associated with low soil carbon stocks at a subarctic treeline, Glob. Change Biol., 21(5), 2070–2081, doi:10.1111/gcb.12793, 2015.

Pastick, N. J., Rigge, M., Wylie, B. K., Jorgenson, M. T., Rose, J. R., Johnson, K. D. and Ji, L.: Distribution and landscape controls of organic layer thickness and carbon within the Alaskan Yukon River Basin, Geoderma, 230, 79–94, 2014.

Ping, C. L., Jastrow, J. D., Jorgenson, M. T., Michaelson, G. J. and Shur, Y. L.: Permafrost soils and carbon cycling, SOIL, 1(1), 147–171, doi:10.5194/soil-1-147-2015, 2015.

Ping, C.-L., Clark, M. H., Kimble, J. M., Michaelson, G. J., Shur, Y. and Stiles, C. A.: Sampling Protocols for Permafrost-Affected Soils, Soil Horiz., 54(1), 13, doi:10.2136/sh12-09-0027, 2013.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [online] Available from: http://www.R-project.org (Accessed 25 July 2014), 2017.

Ridefelt, H., Etzelmüller, B., Boelhouwers, J. and Jonasson, C.: Statistic-empirical modelling of mountain permafrost distribution in the Abisko region, sub-Arctic northern Sweden, Nor. Geogr. Tidsskr. - Nor. J. Geogr., 62(4), 278–289, doi:10.1080/00291950802517890, 2008.

Ripley, B. D.: Pattern recognition and neural networks, Cambridge university press. [online] Available from: https://www.google.com/books?hl=sv&lr=&id=m12UR8QmLqoC&oi=fnd&pg=PR9&dq=Ripley+Pattern+Recognition+and+Neural+Networks+Cambridge.&ots=aNNmeD-JXh&sig=Xo5yVY294wsdwB_2Xkr7GtJehVU (Accessed 29 April 2016), 1996.

Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W.: Monitoring vegetation systems in the Great Plains with ERTS, in Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, pp. 301–317, Nasa, Greenbelt, MD. [online] Available from: http://adsabs.harvard.edu/abs/1974NASSP.351..309R (Accessed 26 March 2016), 1974.

Samuel-Rosa, A., Heuvelink, G. B. M., Vasques, G. M. and Anjos, L. H. C.: Do more detailed environmental covariates deliver more accurate soil maps?, Geoderma, 243–244, 214–227, doi:10.1016/j.geoderma.2014.12.017, 2015.

Sannel, A. B. K. and Kuhry, P.: Warming-induced destabilization of peat plateau/thermokarst lake complexes, J. Geophys. Res. Biogeosciences, 116(G3), 16, doi:10.1029/2010JG001635, 2011.

Schneider, J., Grosse, G. and Wagner, D.: Land cover classification of tundra environments in the Arctic Lena Delta based on Landsat 7 ETM+ data and its application for upscaling of methane emissions, Remote Sens. Environ., 113(2), 380–391, doi:doi:10.1016/j.rse.2008.10.013, 2009.

Schoeneberger, P. J., Wysocki, D. A., Benham, E. C. and Soil Survey Staff: Field book for describing and sampling soils, Version 3.0, Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE., 2012.

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C. and Vonk, J. E.: Climate change and the permafrost carbon feedback, Nature, 520(7546), 171–179, doi:10.1038/nature14338, 2015.

Siewert, M. B., Krautblatter, M., Christiansen, H. H. and Eckerstorfer, M.: Arctic rockwall retreat rates estimated using laboratory-calibrated ERT measurements of talus cones in Longyeardalen, Svalbard, Earth Surf. Process. Landf., 37(14), 1542–1555, doi:10.1002/esp.3297, 2012.

Siewert, M. B., Hanisch, J., Weiss, N., Kuhry, P., Maximov, T. C. and Hugelius, G.: Comparing carbon storage of Siberian tundra and taiga permafrost ecosystems at very high spatial resolution: Ecosystem carbon in taiga and tundra, J. Geophys. Res. Biogeosciences, (120), 1973–1994, doi:10.1002/2015JG002999, 2015.

Siewert, M. B., Hugelius, G., Heim, B. and Faucherre, S.: Landscape controls and vertical variability of soil organic carbon storage in permafrost-affected soils of the Lena River Delta, CATENA, 147, 725–741, doi:10.1016/j.catena.2016.07.048, 2016.

Sonesson, M.: Cryptogams, Int. Biol. Program. Tundra Biome Proj. Tech. Rep., (9), 1972.

Taghizadeh-Mehrjardi, R., Nabiollahi, K. and Kerry, R.: Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran, Geoderma, 266, 98–110, doi:10.1016/j.geoderma.2015.12.003, 2016.

Tang, J., Yurova, A. Y., Schurgers, G., Miller, P. A., Olin, S., Smith, B., Siewert, M. B., Olefeldt, D., Pilesjö, P. and Poska, A.: Drivers of dissolved organic carbon export in a subarctic catchment: Importance of microbial decomposition, sorption-desorption, peatland and lateral flow, Sci. Total Environ., 622–623, 260–274, doi:10.1016/j.scitotenv.2017.11.252, 2018.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G. and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Glob. Biogeochem. Cycles, 23(2), 11, doi:10.1029/2008GB003327, 2009.

Vapnik, V. N.: Statistical learning theory, Springer., 1998.

Virtanen, T. and Ek, M.: The fragmented nature of tundra landscape, Int. J. Appl. Earth Obs. Geoinformation, 27, Part A, 4–12, doi:10.1016/j.jag.2013.05.010, 2014.

Virtanen, T., Mikkola, K., Nikula, A., Christensen, J. H., Mazhitova, G. G., Oberman, N. G. and Kuhry, P.: Modeling the location of the forest line in northeast European Russia with remotely sensed vegetation and GIS-based climate and terrain data, Arct. Antarct. Alp. Res., 36(3), 314–322, 2004.

Vitharana, U. W. A., Mishra, U., Jastrow, J. D., Matamala, R. and Fan, Z.: Observational needs for estimating Alaskan soil carbon stocks under current and future climate: Capturing Heterogeneity of Soil Carbon, J. Geophys. Res. Biogeosciences, 122(2), 415–429, doi:10.1002/2016JG003421, 2017.

Were, K., Bui, D. T., Dick, O. B. and Singh, B. R.: A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, Ecol. Indic., 52, 394–403, doi:10.1016/j.ecolind.2014.12.028, 2015.

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C. and Grehan, A. J.: Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope, Mar. Geod., 30(1–2), 3–35, doi:10.1080/01490410701295962, 2007.

Zeileis, A., Hornik, K., Smola, A. and Karatzoglou, A.: kernlab-an S4 package for kernel methods in R, J. Stat. Softw., 11(9), 1–20, 2004.

Zhang, G. and Lu, Y.: Bias-corrected random forests in regression, J. Appl. Stat., 39(1), 151–160, doi:10.1080/02664763.2011.578621, 2012.

Zhu, A.-X.: Mapping soil landscape as spatial continua: The Neural Network Approach, Water Resour. Res., 36(3), 663–677, doi:10.1029/1999WR900315, 2000.

Zubrzycki, S., Kutzbach, L., Grosse, G., Desyatkin, A. and Pfeiffer, E.-M.: Organic carbon and total nitrogen stocks in soils of the Lena River Delta, Biogeosciences, 10(6), 3507–3524, doi:10.5194/bg-10-3507-2013, 2013.~~Åkerman, H. J. and Johansson, M.: Thawing permafrost and thicker active layers in sub-arctic Sweden, Permafr. Periglac. Process., 19(3), 279–292, doi:10.1002/ppp.626, 2008.~~

~~Andersson, J.: Assessing the ecosystem carbon storage in two subarctic areas - Phytomass and soil organic carbon stock at Abisko in Sweden and Utsjoki in Finland, Master Thesis, Stockholm University, Stockholm, 22 January., 2016.~~

~~Bartsch, A., Widhalm, B., Kuhry, P., Hugelius, G., Palmtag, J. and Siewert, M. B.: Can C-band synthetic aperture radar be used to estimate soil organic carbon storage in tundra?, Biogeosciences, 13(19), 5453–5470, doi:10.5194/bg-13-5453-2016, 2016.~~

~~Baughman, C. A., Mann, D. H., Verbyla, D. L. and Kunz, M. L.: Soil surface organic layers in Arctic Alaska: Spatial distribution, rates of formation, and microclimatic effects, J. Geophys. Res. Biogeosciences, 120(6), 2015JG002983, doi:10.1002/2015JG002983, 2015.~~

~~Becher, M., Olid, C. and Klaminder, J.: Buried soil organic inclusions in non-sorted circles fields in northern Sweden: Age and Paleoclimatic context, J. Geophys. Res. Biogeosciences, 118(1), 104–111, doi:10.1002/jgrg.20016, 2013.~~

~~Becher, M., Olofsson, J. and Klaminder, J.: Cryogenic disturbance and its impact on carbon fluxes in a subarctic heathland, Environ. Res. Lett., 10(11), 114006, doi:10.1088/1748-9326/10/11/114006, 2015.~~

~~Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D. and Goldschmitt, M.: Digital soil mapping using artificial neural networks, J. Plant Nutr. Soil Sci., 168(1), 21–33, doi:10.1002/jpln.200421414, 2005.~~

~~Behrens, T., Zhu, A.-X., Schmidt, K. and Scholten, T.: Multi-scale digital terrain analysis and feature selection for digital soil mapping, Geoderma, 155(3–4), 175–185, doi:10.1016/j.geoderma.2009.07.010, 2010.~~

~~Berglund, B. E., Barnekow, L., Hammarlund, D., Sandgren, P. and Snowball, I. F.: Holocene forest dynamics and climate changes in the Abisko area, northern Sweden: the Sonesson model of vegetation history reconsidered and confirmed, Ecol. Bull., 15–30, 1996.~~

~~Blaschke, T.: Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens., 65(1), 2–16, doi:10.1016/j.isprsjprs.2009.06.004, 2010.~~

~~Boettinger, J. L., Howell, D. W., Moore, A. C., Hartemink, A. E. and Kienast-Brown, S., Eds.: Digital Soil Mapping, Springer Netherlands, Dordrecht. [online] Available from: http://link.springer.com/10.1007/978-90-481-8863-5 (Accessed 23 May 2015), 2010.~~

~~Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, 2001.~~

~~Bronk Ramsey, C.: OxCal 4.2, [online] Available from: https://c14.arch.ox.ac.uk/embed.php?File=oxcal.html (Accessed 28 April 2016), 2016.~~

~~Brown, J., Ferrians, O. J., Heginbottom, J. A. and Melnikov, E. S.: Circum-Arctic map of permafrost and ground-ice conditions, National Snow and Ice Data Center, Boulder, Colorado. [online] Available from: http://pubs.usgs.gov/cp/45/report.pdf (Accessed 6 June 2013), 1997.~~

~~Callaghan, T. V., Jonasson, C., Thierfelder, T., Yang, Z., Hedenas, H., Johansson, M., Molau, U., Van Bogaert, R., Michelsen, A., Olofsson, J., Gwynn-Jones, D., Bokhorst, S., Phoenix, G., Bjerke, J. W., Tommervik, H., Christensen, T. R., Hanna, E., Koller, E. K. and Sloan, V. L.: Ecosystem change and stability over multiple decades in the Swedish subarctic: complex processes and multiple drivers, Philos. Trans. R. Soc. B Biol. Sci., 368(1624), 20120488–20120488, doi:10.1098/rstb.2012.0488, 2013.~~

~~Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. TIST, 2(3), 27, 2011.~~

~~Christensen, T. R., Johansson, T., Åkerman, H. J., Mastepanov, M., Malmer, N., Friborg, T., Crill, P. and Svensson, B. H.: Thawing sub-arctic permafrost: Effects on vegetation and methane emissions, Geophys. Res. Lett., 31(4), L04501, doi:10.1029/2003GL018680, 2004.~~

~~Congalton, R. G.: A review of assessing the accuracy of classifications of remotely sensed data, Remote Sens. Environ., 37(1), 35–46, doi:doi:10.1016/0034-4257(91)90048-B, 1991.~~

~~Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci Model Dev, 8(7), 1991–2007, doi:10.5194/gmd-8-1991-2015, 2015.~~

~~Davidson, E. A. and Janssens, I. A.: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change, Nature, 440(7081), 165–173, 2006.~~

~~Ding, J., Li, F., Yang, G., Chen, L., Zhang, B., Liu, L., Fang, K., Qin, S., Chen, Y., Peng, Y., Ji, C., He, H., Smith, P. and Yang, Y.: The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores, Glob. Change Biol., 22(8), 2688–2701, doi:10.1111/gcb.13257, 2016.~~

~~Dörfer, C., Kühn, P., Baumann, F., He, J.-S. and Scholten, T.: Soil Organic Carbon Pools and Stocks in Permafrost-Affected Soils on the Tibetan Plateau, edited by C. P. Slomp, PLoS ONE, 8(2), e57024, doi:10.1371/journal.pone.0057024, 2013.~~

~~Drăguț, L., Schauppenlehner, T., Muhar, A., Strobl, J. and Blaschke, T.: Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling, Comput. Geosci., 35(9), 1875–1883, doi:10.1016/j.cageo.2008.10.008, 2009.~~

~~FAO: World reference base for soil resources 2014 international soil classification system for naming soils and creating legends for soil maps Update 2015, FAO, Rome., 2015.~~

~~Forkuor, G., Hounkpatin, O. K. L., Welp, G. and Thiel, M.: High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models, PLoS ONE, 12(1), doi:10.1371/journal.pone.0170478, 2017.~~

~~Fuchs, M., Kuhry, P. and Hugelius, G.: Low below-ground organic carbon storage in a subarctic Alpine permafrost environment, The Cryosphere, 9(2), 427–438, doi:10.5194/tc-9-427-2015, 2015.~~

~~GDAL: GDAL - Geospatial Data Abstraction Library: Version 1.11.2, Open Source Geospatial Foundation., 2016.~~

~~Gorham, E.: Northern Peatlands: Role in the Carbon Cycle and Probable Responses to Climatic Warming, Ecol. Appl., 1(2), 182–195, doi:10.2307/1941811, 1991.~~

~~Guisan, A., Weiss, S. B. and Weiss, A. D.: GLM versus CCA spatial modeling of plant species distribution, Plant Ecol., 143(1), 107–122, 1999.~~

~~Hartley, I. P., Garnett, M. H., Sommerkorn, M., Hopkins, D. W., Fletcher, B. J., Sloan, V. L., Phoenix, G. K. and Wookey, P. A.: A potential loss of carbon associated with greater plant growth in the European Arctic, Nat. Clim. Change, 2(12), 875–879, doi:10.1038/nclimate1575, 2012.~~

~~Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning, Springer New York, New York, NY. [online] Available from: http://link.springer.com/10.1007/978-0-387-84858-7 (Accessed 7 August 2016), 2009.~~

~~Heiri, O., Lotter, A. F. and Lemcke, G.: Loss on ignition as a method for estimating organic and carbonate content in sediments: reproducibility and comparability of results, J. Paleolimnol., 25(1), 101–110, doi:10.1023/A:1008119611481, 2001.~~

~~Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., Jesus, J. M. de, Tamene, L. and Tondoh, J. E.: Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions, PLOS ONE, 10(6), e0125814, doi:10.1371/journal.pone.0125814, 2015.~~

~~Hobbie, S. E., Schimel, J. P., Trumbore, S. E. and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude soils, Glob. Change Biol., 6(S1), 196–210, 2000.~~

~~Huete, A. R.: A soil-adjusted vegetation index (SAVI), Remote Sens. Environ., 25(3), 295–309, doi:10.1016/0034-4257(88)90106-X, 1988.~~

~~Hugelius, G.: Spatial upscaling using thematic maps: An analysis of uncertainties in permafrost soil carbon estimates, Glob. Biogeochem. Cycles, 26(2), doi:10.1029/2011GB004154, 2012.~~

~~Hugelius, G. and Kuhry, P.: Landscape partitioning and environmental gradient analyses of soil organic carbon in a permafrost environment, Glob. Biogeochem. Cycles, 23(3), GB3006, doi:10.1029/2008GB003419, 2009.~~

~~Hugelius, G., Kuhry, P., Tarnocai, C. and Virtanen, T.: Soil organic carbon pools in a periglacial landscape: a case study from the central Canadian Arctic, Permafr. Periglac. Process., 21(1), 16–29, doi:10.1002/ppp.677, 2010.~~

~~Hugelius, G., Virtanen, T., Kaverin, D., Pastukhov, A., Rivkin, F., Marchenko, S., Romanovsky, V. and Kuhry, P.: High-resolution mapping of ecosystem carbon storage and potential effects of permafrost thaw in periglacial terrain, European Russian Arctic, J. Geophys. Res., 116(G3), doi:10.1029/2010JG001606, 2011.~~

~~Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P. and Swanson, D. K.: The Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5(1), 3–13, doi:10.5194/essd-5-3-2013, 2013.~~

~~Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C.-L., Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J. and Kuhry, P.: Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps, Biogeosciences, 11(23), 6573–6593, doi:10.5194/bg-11-6573-2014, 2014.~~

~~IPCC: Climate change 2013: The physical science basis, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Intergov. Panel Clim. Change Work. Group Contrib. IPCC Fifth Assess. Rep. AR5Cambridge Univ Press N. Y. [online] Available from: http://www.climatechange2013.org/images/report/WG1AR5_Frontmatter_FINAL.pdf (Accessed 26 November 2014), 2013.~~

~~Johansson, M., Akerman, J., Keuper, F., Christensen, T. R., Lantuit, H. and Callaghan, T. V.: Past and Present Permafrost Temperatures in the Abisko Area: Redrilling of Boreholes, Ambio, 40(6), 558–565, doi:10.1007/s13280-011-0163-3, 2011.~~

~~Jonasson, C., Sonesson, M., Christensen, T. R. and Callaghan, T. V.: Environmental Monitoring and Research in the Abisko Area—An Overview, AMBIO, 41(3), 178–186, doi:10.1007/s13280-012-0301-6, 2012.~~

~~Klaminder, J., Yoo, K., Rydberg, J. and Giesler, R.: An explorative study of mercury export from a thawing palsa mire, J. Geophys. Res. Biogeosciences, 113(G4), G04034, doi:10.1029/2008JG000776, 2008.~~

~~Klaminder, J., Yoo, K. and Giesler, R.: Soil carbon accumulation in the dry tundra: Important role played by precipitation, J. Geophys. Res. Biogeosciences, 114(G4), G04005, doi:10.1029/2009JG000947, 2009.~~

~~Köchy, M., Hiederer, R. and Freibauer, A.: Global distribution of soil organic carbon – Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world, SOIL, 1(1), 351–365, doi:10.5194/soil-1-351-2015, 2015.~~

~~Kokfelt, U., Reuss, N., Struyf, E., Sonesson, M., Rundgren, M., Skog, G., Rosén, P. and Hammarlund, D.: Wetland development, permafrost history and nutrient cycling inferred from late Holocene peat and lake sediment records in subarctic Sweden, J. Paleolimnol., 44(1), 327–342, doi:10.1007/s10933-010-9406-8, 2010.~~

~~Kuhn, M.: Building Predictive Models in R Using the caret Package, J. Stat. Softw., 28(5), doi:10.18637/jss.v028.i05, 2008a.~~

~~Kuhn, M.: Caret package, J. Stat. Softw., 28(5), 1–26, 2008b.~~

~~Law, K. H. and Nichol, J.: Topographic correction for differential illumination effects on IKONOS satellite imagery, Int Arch Photogramm Remote Sens Spat Inf. Sci, 641–646, 2004.~~

~~Li, J. and Heap, A.: A review of spatial interpolation methods for environmental scientists, Geoscience Australia, Canberra., 2008.~~

~~Li, J., Heap, A. D., Potter, A. and Daniell, J. J.: Application of machine learning methods to spatial interpolation of environmental variables, Environ. Model. Softw., 26(12), 1647–1659, doi:10.1016/j.envsoft.2011.07.004, 2011.~~

~~Liaw, A. and Wiener, M.: Classification and regression by randomForest, R News, 2(3), 18–22, 2002.~~

~~Lin, L. I.-K.: A Concordance Correlation Coefficient to Evaluate Reproducibility, Biometrics, 45(1), 255–268, doi:10.2307/2532051, 1989.~~

~~Lundin, E. J., Klaminder, J., Giesler, R., Persson, A., Olefeldt, D., Heliasz, M., Christensen, T. R. and Karlsson, J.: Is the subarctic landscape still a carbon sink? Evidence from a detailed catchment balance, Geophys. Res. Lett., 43(5), 2015GL066970, doi:10.1002/2015GL066970, 2016.~~

~~Malmer, N., Johansson, T., Olsrud, M. and Christensen, T. R.: Vegetation, climatic changes and net carbon sequestration in a North-Scandinavian subarctic mire over 30 years, Glob. Change Biol., 11(11), 1895–1909, doi:10.1111/j.1365-2486.2005.01042.x, 2005.~~

~~McBratney, A. B., Mendonça Santos, M. L. and Minasny, B.: On digital soil mapping, Geoderma, 117(1–2), 3–52, doi:10.1016/S0016-7061(03)00223-4, 2003.~~

~~McGuire, A. D., Anderson, L. G., Christensen, T. R., Dallimore, S., Guo, L., Hayes, D. J., Heimann, M., Lorenson, T. D., Macdonald, R. W. and Roulet, N.: Sensitivity of the carbon cycle in the Arctic to climate change, Ecol. Monogr., 79(4), 523–555, 2009.~~

~~Mishra, U. and Riley, W. J.: Alaskan soil carbon stocks: spatial variability and dependence on environmental factors, Biogeosciences, 9(9), 3637–3645, doi:10.5194/bg-9-3637-2012, 2012.~~

~~Mishra, U. and Riley, W. J.: Active-Layer Thickness across Alaska: Comparing Observation-Based Estimates with CMIP5 Earth System Model Predictions, Soil Sci. Soc. Am. J., 78(3), 894, doi:10.2136/sssaj2013.11.0484, 2014.~~

~~Mishra, U., Jastrow, J. D., Matamala, R., Hugelius, G., Koven, C. D., Harden, J. W., Ping, C. L., Michaelson, G. J., Fan, Z., Miller, R. M., McGuire, A. D., Tarnocai, C., Kuhry, P., Riley, W. J., Schaefer, K., Schuur, E. A. G., Jorgenson, M. T. and Hinzman, L. D.: Empirical estimates to reduce modeling uncertainties of soil organic carbon in permafrost regions: a review of recent progress and remaining challenges, Environ. Res. Lett., 8(3), 035020, doi:10.1088/1748-9326/8/3/035020, 2013.~~

~~Moore, I. D., Grayson, R. B. and Ladson, A. R.: Digital terrain modelling: a review of hydrological, geomorphological, and biological applications, Hydrol. Process., 5(1), 3–30, 1991.~~

~~Palmtag, J., Hugelius, G., Lashchinskiy, N., Tamtorf, M. P., Richter, A., Elberling, B. and Kuhry, P.: Storage, landscape distribution, and burial history of soil organic matter in contrasting areas of continuous permafrost, Arct. Antarct. Alp. Res., 47(1), 71–88, doi:http://dx.doi.org/10.1657/AAAR0014-027, 2015.~~

~~Palmtag, J., Ramage, J., Hugelius, G., Gentsch, N., Lashchinskiy, N., Richter, A. and Kuhry, P.: Controls on the storage of organic carbon in permafrost soil in northern Siberia: Soil organic carbon storage in permafrost terrain, northern Siberia, Eur. J. Soil Sci., 67(4), 478–491, doi:10.1111/ejss.12357, 2016.~~

~~Parker, T. C., Subke, J.-A. and Wookey, P. A.: Rapid carbon turnover beneath shrub and tree vegetation is associated with low soil carbon stocks at a subarctic treeline, Glob. Change Biol., 21(5), 2070–2081, doi:10.1111/gcb.12793, 2015.~~

~~Pastick, N. J., Rigge, M., Wylie, B. K., Jorgenson, M. T., Rose, J. R., Johnson, K. D. and Ji, L.: Distribution and landscape controls of organic layer thickness and carbon within the Alaskan Yukon River Basin, Geoderma, 230, 79–94, 2014.~~

~~Ping, C. L., Jastrow, J. D., Jorgenson, M. T., Michaelson, G. J. and Shur, Y. L.: Permafrost soils and carbon cycling, SOIL, 1(1), 147–171, doi:10.5194/soil-1-147-2015, 2015.~~

~~Ping, C.-L., Clark, M. H., Kimble, J. M., Michaelson, G. J., Shur, Y. and Stiles, C. A.: Sampling Protocols for Permafrost-Affected Soils, Soil Horiz., 54(1), 13, doi:10.2136/sh12-09-0027, 2013.~~

~~R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [online] Available from: http://www.R-project.org (Accessed 25 July 2014), 2017.~~

~~Ridefelt, H., Etzelmüller, B., Boelhouwers, J. and Jonasson, C.: Statistic-empirical modelling of mountain permafrost distribution in the Abisko region, sub-Arctic northern Sweden, Nor. Geogr. Tidsskr. - Nor. J. Geogr., 62(4), 278–289, doi:10.1080/00291950802517890, 2008.~~

~~Ripley, B. D.: Pattern recognition and neural networks, Cambridge university press. [online] Available from: https://www.google.com/books?hl=sv&lr=&id=m12UR8QmLqoC&oi=fnd&pg=PR9&dq=Ripley+Pattern+Recognition+and+Neural+Networks+Cambridge.&ots=aNNmeD-JXh&sig=Xo5yVY294wsdwB_2Xkr7GtJchVU (Accessed 29 April 2016), 1996.~~

~~Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W.: Monitoring vegetation systems in the Great Plains with ERTS, in Proceedings of the Third Earth Resources Technology Satellite-1 Symposium, pp. 301–317, Nasa, Greenbelt, MD. [online] Available from: http://adsabs.harvard.edu/abs/1974NASSP.351..309R (Accessed 26 March 2016), 1974.~~

~~Samuel-Rosa, A., Heuvelink, G. B. M., Vasques, G. M. and Anjos, L. H. C.: Do more detailed environmental covariates deliver more accurate soil maps?, Geoderma, 243–244, 214–227, doi:10.1016/j.geoderma.2014.12.017, 2015.~~

~~Sannel, A. B. K. and Kuhry, P.: Warming-induced destabilization of peat plateau/thermokarst lake complexes, J. Geophys. Res. Biogeosciences, 116(G3), 16, doi:10.1029/2010JG001635, 2011.~~

~~Schneider, J., Grosse, G. and Wagner, D.: Land cover classification of tundra environments in the Arctic Lena Delta based on Landsat 7 ETM+ data and its application for upscaling of methane emissions, Remote Sens. Environ., 113(2), 380–391, doi:doi:10.1016/j.rse.2008.10.013, 2009.~~

~~Schoeneberger, P. J., Wysocki, D. A., Benham, E. C. and Soil Survey Staff: Field book for describing and sampling soils, Version 3.0, Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE., 2012.~~

~~Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C. and Vonk, J. E.: Climate change and the permafrost carbon feedback, Nature, 520(7546), 171–179, doi:10.1038/nature14338, 2015.~~

~~Siewert, M. B., Krautblatter, M., Christiansen, H. H. and Eckerstorfer, M.: Arctic rockwall retreat rates estimated using laboratory-calibrated ERT measurements of talus cones in Longyeardalen, Svalbard, Earth Surf. Process. Landf., 37(14), 1542–1555, doi:10.1002/esp.3297, 2012.~~

~~Siewert, M. B., Hanisch, J., Weiss, N., Kuhry, P., Maximov, T. C. and Hugelius, G.: Comparing carbon storage of Siberian tundra and taiga permafrost ecosystems at very high spatial resolution: Ecosystem carbon in taiga and tundra, J. Geophys. Res. Biogeosciences, (120), 1973–1994, doi:10.1002/2015JG002999, 2015.~~

~~Siewert, M. B., Hugelius, G., Heim, B. and Faucherre, S.: Landscape controls and vertical variability of soil organic carbon storage in permafrost-affected soils of the Lena River Delta, CATENA, 147, 725–741, doi:10.1016/j.catena.2016.07.048, 2016.~~

~~Sonesson, M.: Cryptogams, Int. Biol. Program. Tundra Biome Proj. Tech. Rep., (9), 1972.~~

~~Taghizadeh-Mehrjardi, R., Nabiollahi, K. and Kerry, R.: Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran, Geoderma, 266, 98–110, doi:10.1016/j.geoderma.2015.12.003, 2016.~~

~~Tang, J., Yurova, A. Y., Schurgers, G., Miller, P. A., Olin, S., Smith, B., Siewert, M. B., Olefeldt, D., Pilesjö, P. and Poska, A.: Drivers of dissolved organic carbon export in a subarctic catchment: Importance of microbial decomposition, sorption-desorption, peatland and lateral flow, Sci. Total Environ., 622–623, 260–274, doi:10.1016/j.scitotenv.2017.11.252, 2018.~~

~~Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G. and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Glob. Biogeochem. Cycles, 23(2), 11, doi:10.1029/2008GB003327, 2009.~~

~~Vapnik, V. N.: Statistical learning theory, Springer., 1998.~~

~~Virtanen, T. and Ek, M.: The fragmented nature of tundra landscape, Int. J. Appl. Earth Obs. Geoinformation, 27, Part A, 4–12, doi:10.1016/j.jag.2013.05.010, 2014.~~

~~Virtanen, T., Mikkola, K., Nikula, A., Christensen, J. H., Mazhitova, G. G., Oberman, N. G. and Kuhry, P.: Modeling the location of the forest line in northeast European Russia with remotely sensed vegetation and GIS-based climate and terrain data, Arct. Antarct. Alp. Res., 36(3), 314–322, 2004.~~

~~Vitharana, U. W. A., Mishra, U., Jastrow, J. D., Matamala, R. and Fan, Z.: Observational needs for estimating Alaskan soil carbon stocks under current and future climate: Capturing Heterogeneity of Soil Carbon, J. Geophys. Res. Biogeosciences, 122(2), 415–429, doi:10.1002/2016JG003421, 2017.~~

~~Were, K., Bui, D. T., Dick, O. B. and Singh, B. R.: A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape, Ecol. Indic., 52, 394–403, doi:10.1016/j.ecolind.2014.12.028, 2015.~~

~~Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C. and Grehan, A. J.: Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope, Mar. Geod., 30(1–2), 3–35, doi:10.1080/01490410701295962, 2007.~~

~~Zeileis, A., Hornik, K., Smola, A. and Karatzoglou, A.: kernlab-an S4 package for kernel methods in R, J. Stat. Softw., 11(9), 1–20, 2004.~~

~~Zhang, G. and Lu, Y.: Bias-corrected random forests in regression, J. Appl. Stat., 39(1), 151–160, doi:10.1080/02664763.2011.578621, 2012.~~

~~Zhu, A.-X.: Mapping soil landscape as spatial continua: The Neural Network Approach, Water Resour. Res., 36(3), 663–677, doi:10.1029/1999WR900315, 2000.~~

~~Zubrzycki, S., Kutzbach, L., Grosse, G., Desyatkin, A. and Pfeiffer, E.-M.: Organic carbon and total nitrogen stocks in soils of the Lena River Delta, Biogeosciences, 10(6), 3507–3524, doi:10.5194/bg-10-3507-2013, 2013.~~

~~*Fig. 1. Workflow diagram for this study. Environmental predictor variables are used to generate a land cover classification. Digital soil mapping is performed using soil pedons (supplemented with pseudo sampling points) combined with environmental predictor variables to train prediction models for SOC-Total. The best performing model (RF) is used to develop maps of different SOC depth increments and the OL depth. The results are discussed, with the help of eight radio carbon samples, in the context of present day SOC storage, past and future SOC dynamics and the relevance of wetlands for the SOC storage in permafrost environments.*~~

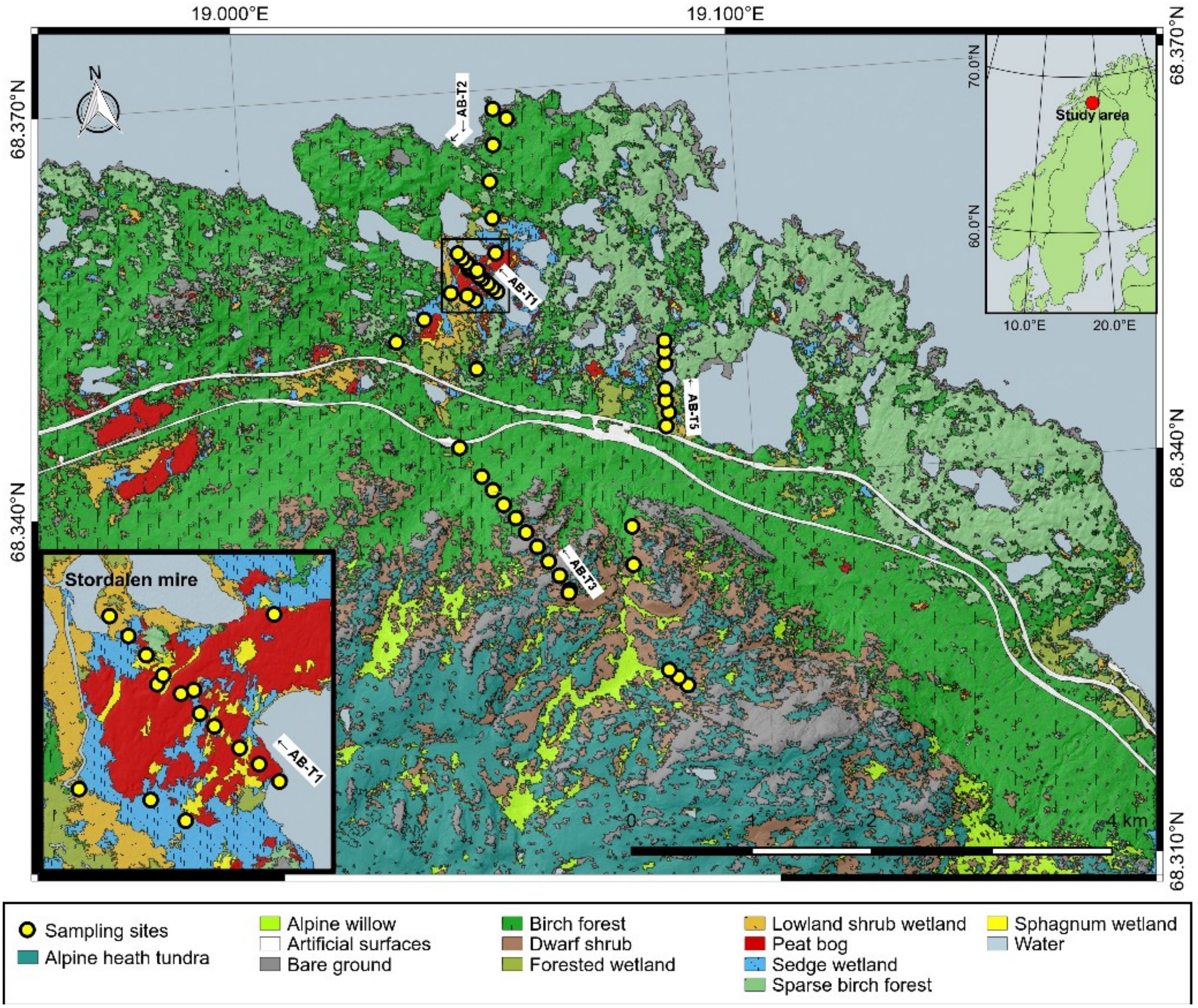

Fig. 1. Top right inset showing the location of the study area. Main view showing the land cover classification for the entire mapping extent. The bottom left inset shows a closeup of Stordalen mire. The beginning and counting direction of the four main sampling transects are marked.

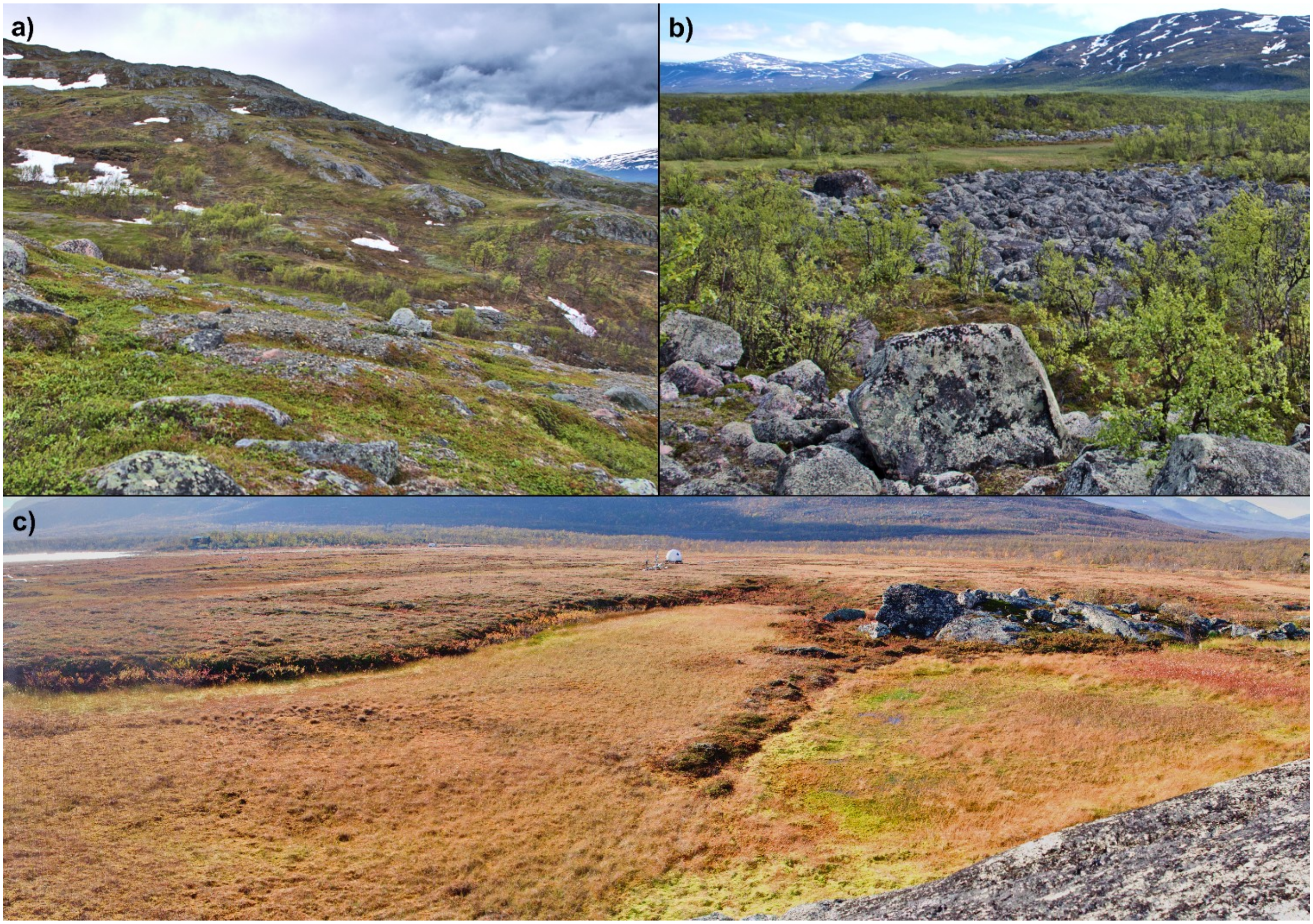

Fig. 2: Photographs exemplifying the study area. a) Alpine landscape mosaic showing several land cover classes including bare ground and alpine tundra heath transitioning into birch forest. b) Lowland landscape mosaic showing sparse birch forest, blockfields and a small wetland in direct neighborhood. c) Raised permafrost peat plateau (left) with sharp transition to Sphagnum dominated wetland areas and exposed bedrock areas (right).

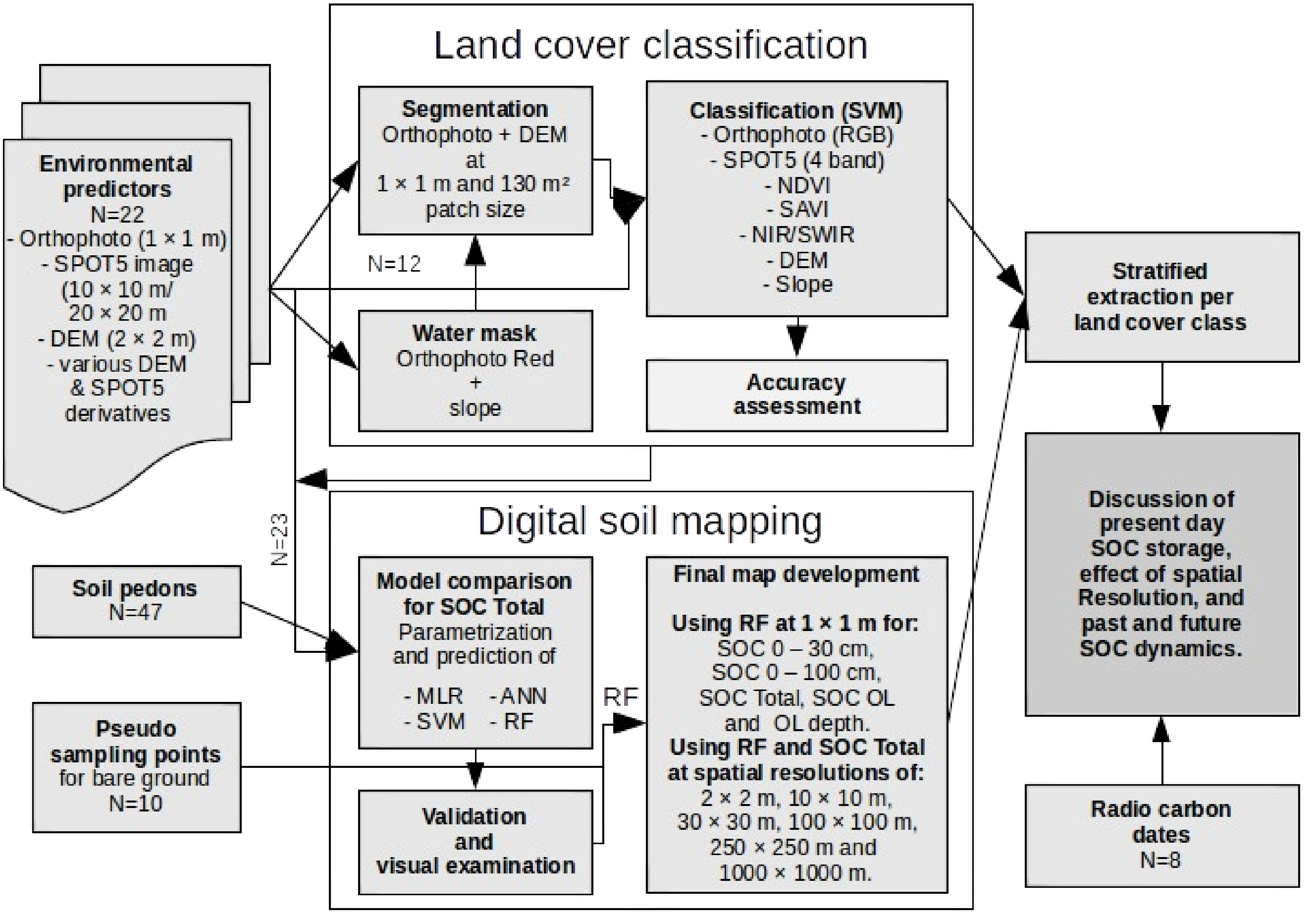

Fig. 3. Workflow diagram for this study. Environmental predictor variables are used to generate a land cover classification. Digital soil mapping is performed using soil pedons combined with environmental predictor variables to train prediction models for SOC-Total. The best performing model (RF) is used to develop maps of different SOC depth increments and the OL depth (supplemented with pseudo sampling points). The results are discussed, with the help of eight radio carbon samples, in the context of present day SOC storage, the effect of spatial resolution, past and future SOC dynamics and the relevance of wetlands for the SOC storage in permafrost environments.

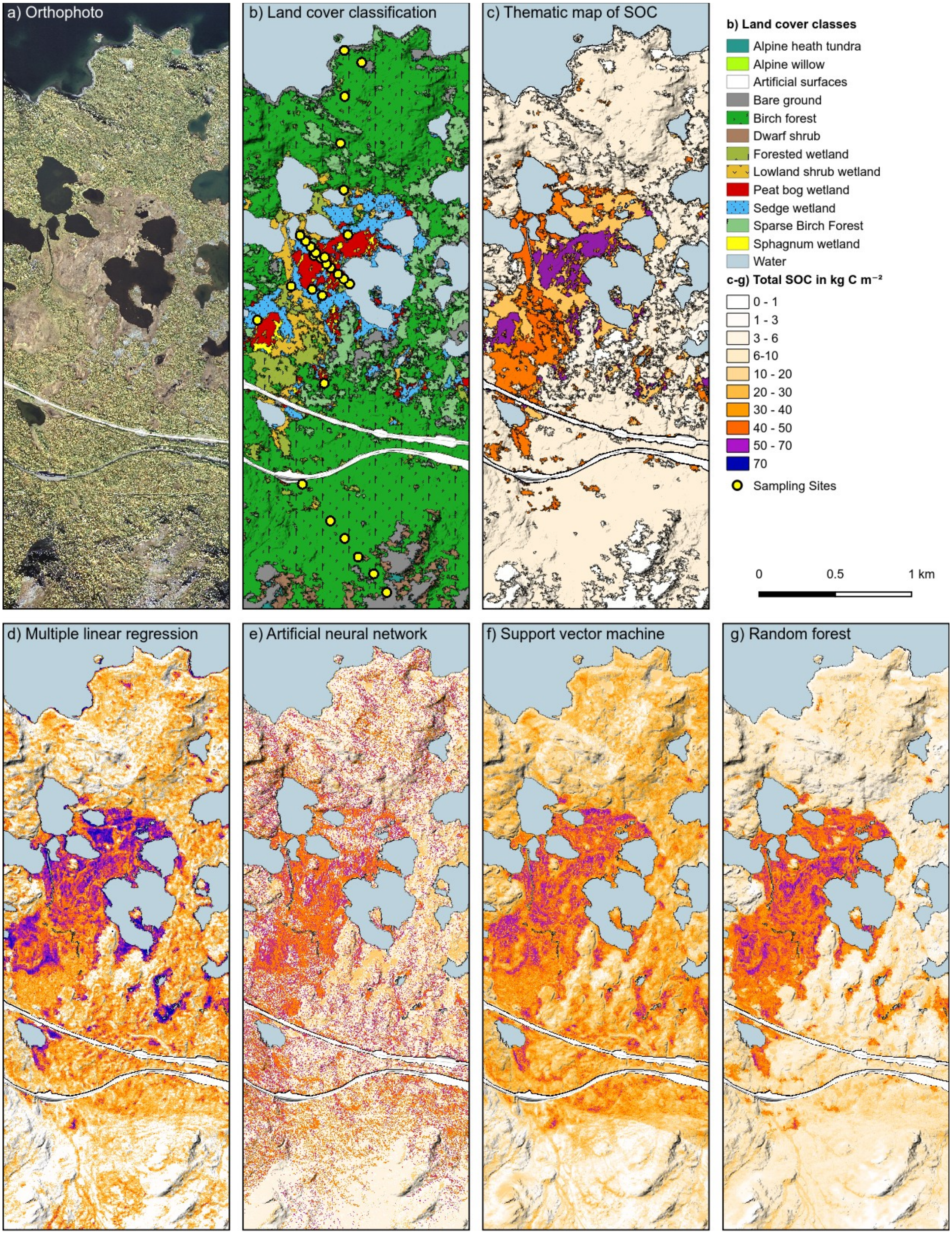
a) Orthophoto
b) Land cover classification
c) Thematic map of SOC
b) Land cover classes
Alpine heath tundra
Alpine willow
Artificial surfaces
Bare ground
Birch forest
Dwarf shrub
Forested wetland
Lowland shrub wetland
Peat bog wetland
Sedge wetland
Sparse Birch Forest
Sphagnum wetland
Water
c-g) Total SOC in kg C m⁻²
0 - 1
1 - 3
3 - 6
6-10
10 - 20
20 - 30
30 - 40
40 - 50
50 - 70
70
Sampling Sites
0
0.5
1 km
d) Multiple linear regression
e) Artificial neural network
f) Support vector machine
g) Random forest

Fig. 4: A close up of the area near Stordalen mire comparing different maps. a) Illumination corrected orthophoto (© Lantmäteriet, I2014/00691). b) Land cover classification, c) Soil organic carbon storage using a thematic mapping approach. d-g) Maps developed using different machine-learning models. Each map uses all input soil pedons and pseudo sampling points d) multiple linear regression model (MLR), e) artificial neural network (ANN), f) support vector machine (SVM) and g) random forest (RF).

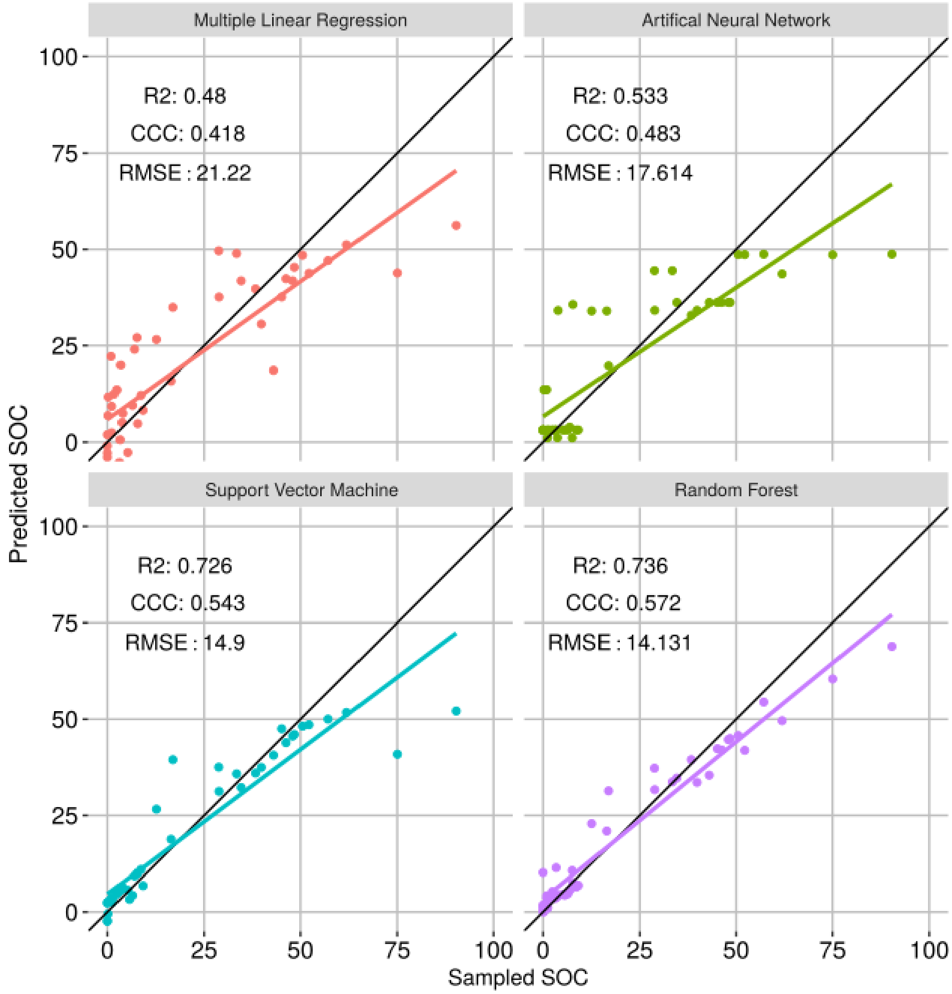

Fig. 5. *Performance of different prediction models developed using the soil pedon dataset excluding pseudo-training samples. The evaluation is based on observed against predicted total soil organic carbon values. The performance metrics are based on cross-validation. (R2: coefficient of determination; CCC: Lin's concordance correlation coefficient (CCC); RMSE: root mean squared error.* ~~*Performance of different prediction models comparing observed against predicted total soil organic carbon values (all values in kg C m−2) using the full soil pedon dataset including 10 pseudo-training samples for bare ground set to 0.0 kg C m−2.*~~

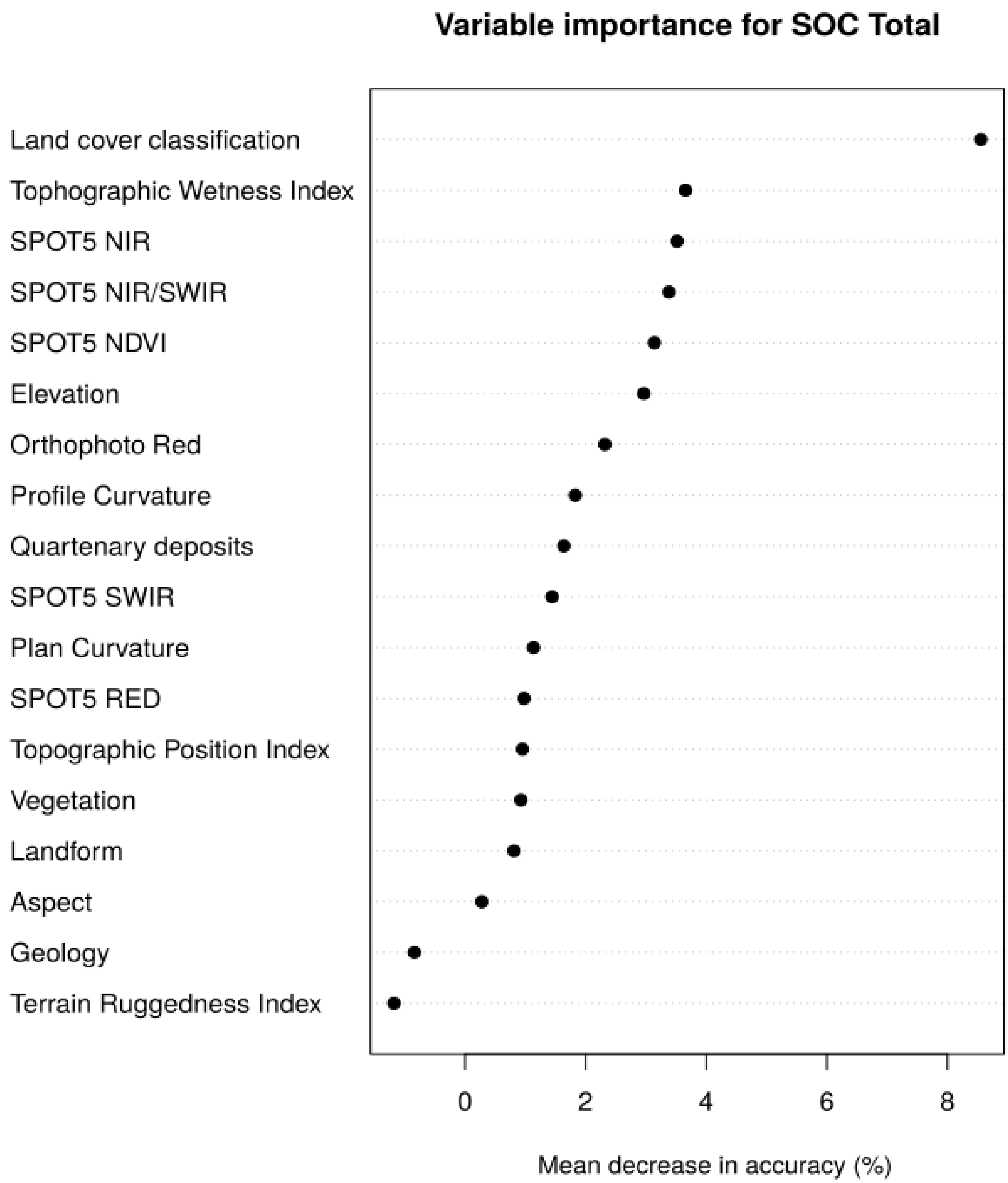

Fig. 6. Variable importance for the prediction of total SOC measured as mean decrease in accuracy of the random forest model if the variable is excluded~~as a result of permutation of the input variable~~. The higher the value the more important is the variable.

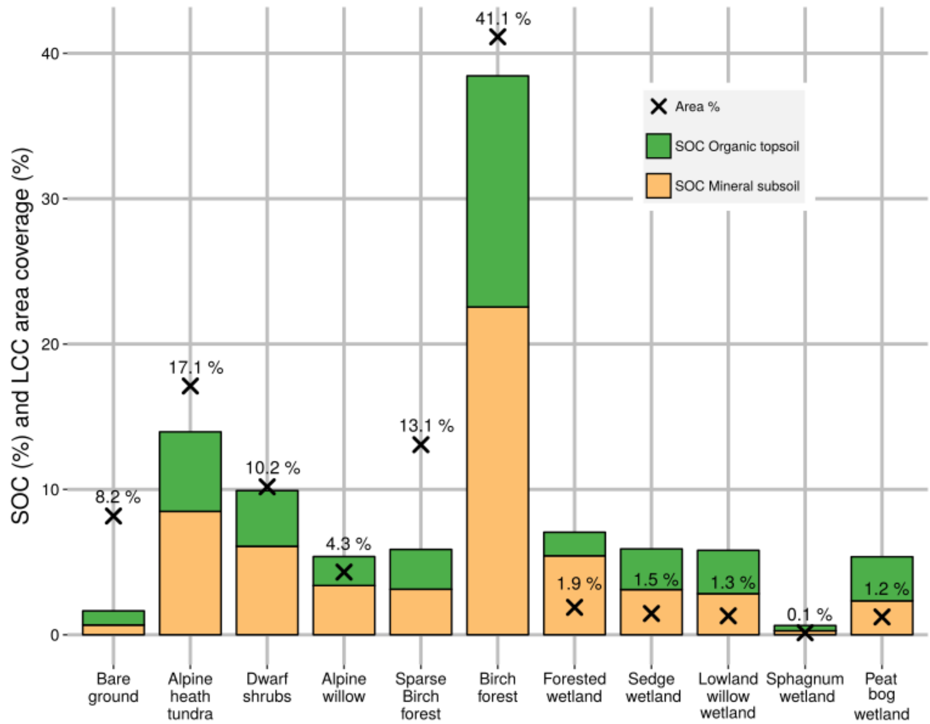

Fig. 7. Partitioning of modeled total soil organic carbon (SOC) storage and respective land cover class coverage in % using a random forest predictor. Height of the column represent the fraction of the SOC–Tot. The mineral SOC is the amount of SOC–OL subtracted from the SOC–Tot. Crosses indicate the percentage areal coverage of the respective land cover class of the total landscape soil area.

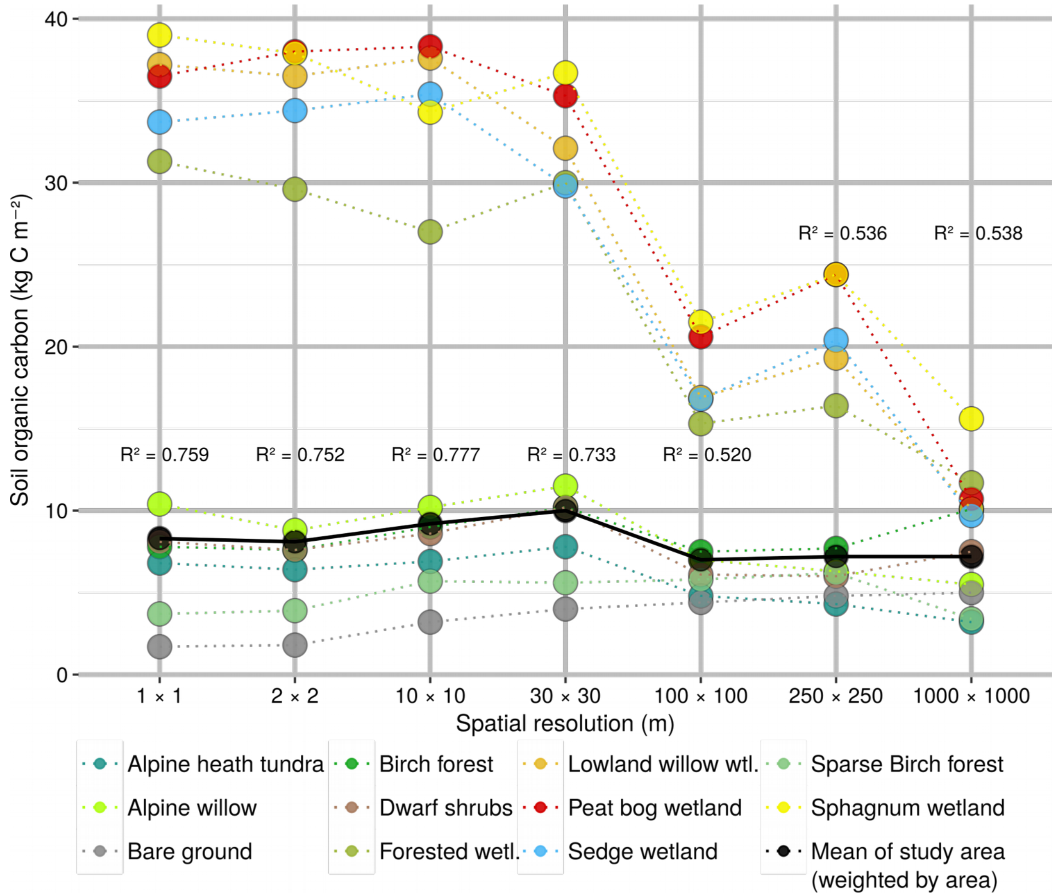

Fig. 8. ~~Map of the predicted soil organic carbon storage using the random forests model. a) Total soil organic carbon,~~Summary of the predicted SOC at different spatial resolutions. The dots show the development of the SOC–Tot while the colored dots show the SOC stored in individual classes extracted using the LCC at 1 m spatial resolution. ~~b) SOC stocks in wetlands classes of the high-resolution land cover classification, c) SOC stocks in peatlands as revealed from modeling the organic layer depth, d) SOC stocks in wetland areas according to the CORINE dataset~~The R^2 is estimated using the full pedon dataset including pseudo-sampling points.

~~*Table1: Validation of SOC-Tot prediction models. The validation is based on the internal (80%) training data subset and on an external (20%) control subset of the SOC input data. The validation is performed with and without pseudo sampling points for bare ground areas.*~~

	~~Including Pseudo sampling points~~			~~Excluding Pseudo sampling points~~		
	~~R^2~~	~~CCC~~	~~RMSE~~	~~R^2~~	~~CCC~~	~~RMSE~~
~~**MLR Int.**~~	~~0.812~~	~~0.879~~	~~9.347~~	~~0.825~~	~~0.882~~	~~9.074~~
~~**MLR Ext.**~~	~~0.143~~	~~0.424~~	~~25.051~~	~~0.022~~	~~0.299~~	~~27.343~~
~~**ANN Int.**~~	~~0.843~~	~~0.872~~	~~9.1~~	~~0.823~~	~~0.858~~	~~9.560~~
~~**ANN Ext.**~~	~~0.352~~	~~0.547~~	~~19.739~~	~~0.229~~	~~0.440~~	~~23.362~~
~~**SVM Int.**~~	~~0.898~~	~~0.908~~	~~7.512~~	~~0.793~~	~~0.837~~	~~10.331~~
~~**SVM Ext.**~~	~~0.698~~	~~0.552~~	~~17.755~~	~~0.377~~	~~0.438~~	~~21.783~~
~~**RF Int.**~~	~~0.939~~	~~0.925~~	~~6.482~~	~~0.949~~	~~0.945~~	~~4.986~~
~~**RF Ext.**~~	~~0.908~~	~~0.650~~	~~15.392~~	~~0.470~~	~~0.536~~	~~19.992~~

~~*R^2: coefficient of determination*~~

~~*RMSE: root mean squared error*~~

~~*CCC: Lin's concordance correlation coefficient*~~

Table 1: Soil pedon properties by land cover class and p*redicted SOC carbon stocks using a random forest model.*

				Soil pedon data									Random forest predicted values										
				Mean ± StD cm		Mean ± StD kg C m^{-2}							Mean ± StD cm	Mean ± StD kg C m^{-2}									
Land cover class	Sites (n)	Area in km²	% of Soil Area	Depth Organic layer	Depth Active layer	SOC Organic layer	SOC Mineral	SOC Permafrost	SOC 0 - 30 cm	SOC 0 - 100 cm	SOC Total	SOC Total Min.–Max.	Depth Organic layer 1 x 1 m	SOC Organic layer 1 x 1 m	SOC 0 - 30 cm 1 x 1 m	SOC 0 - 100 cm 1 x 1 m	SOC Total 1 x 1 m	SOC Total 2 x 2 m	SOC Total 10 x 10 m	SOC Total 30 x 30 m	SOC Total 100 x 100 m	SOC Total 250 x 250 m	SOC Total 1000 x 1000 m
Alpine heath tundra	7	8	17.1	5±3.4	37	1.4±1.3	2.6±2.9	0.1±0.3	3.5±2.8	4±3.1	4±3.1	1.1–9.2	8±4.9	2.7±2.1	3.6±1.6	6.5±3.6	6.8±4.1	6.4±3.2	6.9±7	7.8±3.6	4.8±3.1	4.3±2.9	3.2±3.9
Alpine willow	1	2	4.3	5	–	1.1	6.5	0	7.5	7.6	7.6	7.6–7.6	13.6±5.1	3.9±2	5.4±0.7	9.7±2.6	10.4±3.4	8.8±2.8	10.2±6.7	11.5±3.2	7±2.8	6.3±2.8	5.5±4.7
Bare ground	4	3.8	8.2	0.1±0	–	0±0	0±0	0±0	0±0	0±0	0±0	0–0	2.5±3.8	1±1.9	0.8±1.2	1.8±2.1	1.7±2.3	1.8±2.1	3.2±4.5	4±3.4	4.4±5.2	4.8±5	5±6.3
Birch forest	6	19.3	41.1	9.5±3.9	–	2.4±0.4	1.8±2.4	0±0	4±2	4.2±2.2	4.2±2.2	2.4–8.6	10.4±4.7	3.2±1.8	4.5±0.9	7.4±2.8	7.8±3.8	7.6±3.6	9±6.9	10.2±4.9	7.5±4.5	7.7±6.1	10.1±8.7
Dwarf shrubs	2	4.8	10.2	6±2.8	–	1.9±1.3	1.4±1.9	0±0	3.4±3.2	3.4±3.2	3.4±3.2	1.1–5.6	10.5±4.6	3.1±1.9	4.8±1.1	8.1±3	8.1±3.8	7.6±3.5	8.6±6.8	10.2±4.3	6.1±3.7	6±4.2	7.5±6.6
Forested wetland	4	0.9	1.9	39.2±34.7	–	15.6±22.8	24.7±14.1	0±0	8.9±5.7	32.1±15.9	40.3±18.7	16.4–61.9	26.5±9.8	7.2±3.4	6±0.9	24±4.6	31.3±6.6	29.6±6.8	27±11.2	30±11.8	15.3±8.2	16.4±12.5	11.7±12.3
Lowland willow wetl.	2	0.6	1.3	46±38.2	–	17.5±17.3	28.2±18.1	0±0	9.6±2.5	35±6.7	45.7±0.8	45.1–46.3	47.6±13.7	19.1±6.1	6.7±1	29.5±5.2	37.2±7.5	36.5±8.2	37.6±11.1	32.1±13.3	16.9±9.4	19.3±14	10.1±11
Peat bog wetland	9	0.6	1.2	56.9±27.2	50±20	28.3±20.6	23.9±10.4	10.1±13.3	9.9±4.5	40.6±17.9	52.2±20.3	28.8–90.3	44.7±12.9	20.6±6.8	7±1.4	31.3±6.8	36.5±9	38±8.9	38.3±13.8	35.3±13.1	20.6±10.5	24.4±14.5	10.7±10.3
Sedge wetland	4	0.7	1.5	48.8±51.1	–	13.3±16.2	8±2.4	0±0	4.6±2.7	14.5±8.5	21.3±18.2	7.7–48	44.1±14.4	15.9±6.1	5.3±1.6	26.9±5.9	33.7±9	34.4±9.6	35.4±13.1	29.8±15	16.8±12.2	20.4±16.2	9.7±12.3
Sparse Birch forest	6	6.1	13.1	5±3.7	–	1.7±1.9	1.4±1.5	0±0	3±3.4	3.1±3.4	3.1±3.4	0.2–7.8	5.7±3.8	1.7±1.9	2.2±1.4	3.7±2.4	3.7±2.9	3.9±2.8	5.7±6.7	5.6±4.5	5.8±5.2	6.2±7.4	3.4±5.2
Sphagnum wetland	2	0.1	0.1	93.5±62.9	73	22.1±15.4	12±14.6	8.5±12.1	6.6±2.5	22.3±6	34±0.8	33.4–34.6	54.2±17.5	22.1±7.2	7.4±1.1	29.1±6.1	39±8.7	37.9±8.3	34.3±11.5	36.7±12.4	21.5±11.6	24.4±15	15.6±15
Anthropogenic	-	0.8	1.6	Study area mean ± StD:[a]									Study area mean ± StD:[a]										
Water	-	17.4	37.1																				
Mean of study area (weighted by area)	47	650.3	100	9.1±1.1	-	2.8±0.3	3±0.2	0.2±0	3.8±0.5	5.3±0.5	5.8±0.5	0–90.3	10.6±9.7	3.5±4	3.9±1.8	7.7±6.2	8.3±8	8.1±7.9	9.2±9.7	10±8	7±5.6	7.2±7.2	7.2±7.9

[a] Weighted by area, excluding Artificial surfaces and water areas.

Table 2: Summary of radiocarbon dating.

Soil pedon	Depth	Sample description	Lab. no.[a]	Age ^{14}C	Age cal BP (yrs)[b]
AB-T1-06	14–15 cm	Palsa, base of marked change in peat composition	Poz-59879	150.84 ± 0.36 pMC[c]	modern
AB-T1-06	94–95 cm	Palsa, base of pure peat organics	Poz-59880	4565 ± 30 BP	5218
AB-T1-10	72–73 cm	Lowland shrub wetland, base of OL	Poz-59882	2215 ± 30 BP	2230
AB-T2-06	48–49 cm	***Sphagnum*** patch, base of OL	Poz-59883	1985 ± 30 BP	1936
AB-T3-07	11–12 cm	Alpine birch forest, base of OL	Poz-59884	119.16 ± 0.33 pMC	modern
AB-T3-07	22–28 cm	Alpine birch forest, mineral subsoil	Poz-59885	150 ± 30 BP	150
AB-T3-09	18–19 cm	Birch, base of OL	Poz-59886	100.82 ± 0.29 pMC	modern
AB-T3-09	20–26 cm	Birch forest, mineral subsoil	Poz-59887	1455 ± 30 BP	1345

[a] Laboratory number of the Radiocarbon Laboratory in Poznan, Poland.

[b] Mean age at 95.4% probability expressed in calendar years before 1950.

[c] Percent Modern Carbon.