Large, climate-sensitive soil carbon stocks mapped with pedology-informed machine learning in the North Pacific coastal temperate rainforest

To cite this article before publication: Gavin McNicol et al 2018 Environ. Res. Lett. in press https://doi.org/10.1088/1748-9326/aaed52

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2018 The Author(s). Published by IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence https://creativecommons.org/licenses/by/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.
Title Page

Article Title: Large, climate-sensitive soil carbon stocks mapped with pedology-informed machine learning in the North Pacific coastal temperate rainforest

List of Authors: Gavin McNicol*, Chuck Bulmerb, David D’Amorec, Paul Sanbornac, Sari Saundersd, Ian Giesbrechtde, Santiago Gonzalez Arriolaa, Allison Bidlackf, David Butmanh, Brian Bumai

a. Alaska Coastal Rainforest Center, University of Alaska Southeast, Juneau, AK-99801, USA
b. B.C. Ministry of Forests, Lands and Natural Resource Operations, Forest Sciences Section, 3401 Reservoir Rd., Vernon, BC V1B 2C7, Canada
c. USDA Forest Service, Pacific Northwest Research Station, Juneau, AK-99801, USA
d. Ecosystem Science and Management Program, University of Northern British Columbia, 3333 University Way, Prince George, BC, V2N 4Z9, Canada
e. Hakai Institute, Tula Foundation, P.O. Box 309, Heriot Bay, BC, V0P 1H0, Canada
f. B.C. Ministry of Forests, Lands and Natural Resource Operations, 103-2100 Labieux Rd., Nanaimo, BC V9T 6E9, Canada
g. School of Resource and Environmental Management, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada
h. School of Environmental and Forest Sciences and the Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA
i. Department of Integrative Biology, University of Colorado, 1151 Arapahoe, SI 4101, Denver, CO 80301, USA

* Now at: Earth System Science Department, Stanford University, Stanford, CA-94305, USA

Corresponding Author Contact Email: gmenicol@stanford.edu
Main Text Words: 4,290  Abstract Words: 279

Abstract

Accurate soil organic carbon (SOC) maps are needed to predict the terrestrial SOC feedback to climate change, one of the largest remaining uncertainties in Earth system modeling. Over the last decade, global scale models have produced varied predictions of the size and distribution of SOC stocks, ranging from 1,000 to > 3,000 Pg of C within the top 1 m. Regional assessments may help validate or improve global maps because they can examine landscape controls on SOC stocks and offer a tractable means to retain regionally-specific information, such as soil taxonomy, during database creation and modeling. We compile a new transboundary SOC stock database for coastal watersheds of the North Pacific coastal temperate rainforest, using soil classification data to guide gap-filling and machine learning approaches used to explore spatial controls on SOC and predict regional stocks. Precipitation and topographic attributes controlling soil wetness were found to be the dominant controls of SOC, underscoring the dependence of C accumulation on high soil moisture. The random forest model predicted stocks of 4.5 Pg C (to 1 m) for the study region, 22% of which was stored in organic soil layers. Calculated stocks of 228 ± 111 Mg C ha⁻¹ fell within ranges of several past regional studies and indicate 11-33 Pg C may be stored across temperate rainforest soils globally. Predictions compared very favorably to regionalized estimates from two spatially-explicit global products (Pearson’s correlation: ρ = 0.73 vs. 0.34). Notably, SoilGrids250m was an outlier for estimates of total SOC, predicting 4-fold higher stocks (18 Pg C) and indicating bias in this global product for the soils of the temperate rainforest. In sum our study demonstrates that CTR ecosystems represent a moisture-dependent hotspot for SOC storage at mid-latitudes.

Social Media Abstract: Large soil carbon stocks track climate gradients in the N. Pacific coastal temperate rainforest.
1. Introduction

Accurate global soil organic carbon (SOC) maps are necessary to validate terrestrial carbon (C) cycle predictions in Earth System Models (ESMs; Todd-Brown et al. 2013) however current SOC models drawing on international pedon (soil profile) databases (SoilGRIDs, HWSD, ISCN, etc.) display considerable differences (Köchy et al. 2015; Sanderman et al. 2017, 2018). Database construction, including filling data gaps, account for some of these discrepancies (Tifafi et al. 2018) while other sources of uncertainty are associated with scaling up spatially from relatively sparse pedon observations (~1 m²) to globally gridded products (Todd-Brown et al. 2013) and the loss of information relevant to SOC storage at intermediate scales (10 m² – 1 km²) such as topography (Mishra and Riley 2015; Siewert et al. 2017). Regional digital soil mapping may help bridge these scale discontinuities and produce finer resolution (< 100 m) predictions that retain information on the spatial drivers of SOC storage (Minasny et al. 2013). For example, Sanderman et al. (2018) compiled mangrove SOC stock measurements and used them in conjunction with a global SOC map (Hengl et al. 2017) and maps of environmental covariates to estimate global mangrove SOC at 30 m resolution. The integration of detailed pedological information with machine learning approaches for large-scale spatial predictions may also enable improvements in SOC mapping (Ramcharan et al. 2017) and regional SOC assessments may help diagnose errors in global products by providing higher resolution information on SOC controls and its distribution.

Regional SOC assessments at high-latitudes have, to date, focused on Arctic and boreal permafrost soils (Hugelius et al. 2013, 2014), while coastal temperate rainforests (CTR) have not received similar attention despite their similarly high SOC storage (Carpenter et al. 2014). Globally, temperate rainforests contain the highest density aboveground forest C stocks (up to 1500 Mg ha⁻¹; Keith et al. 2009), and can be found along the coastal margins of North and South America, Japan and Korea, Australasia, and Scandinavia (Alaback 1991). The N. Pacific coastal temperate rainforest (NPCTR) biome is the largest example, and spans 4,000 km of the N. American coast from the Russian River in
California, to Kodiak Island in the Gulf of Alaska (DellaSala 2011). Although several studies have produced regional estimates of SOC stocks in Alaska (Leighty et al 2006; Johnson et al 2011; Mishra et al 2012), no studies to date have produced spatially-explicit SOC stock estimates across the transboundary domain of southeast Alaska (SEAK) and coastal British Columbia (BC).

Soils of the NPCTR can store large quantities of SOC, especially in the wet seasonal and perhumid zones (Carpenter et al 2014), with stocks > 300 Mg ha\(^{-1}\) (to 1 m mineral soil depth) frequently observed in SEAK (Johnson et al 2011) and stocks of > 200 Mg ha\(^{-1}\) common in coastal BC (Shaw et al 2018). These large SOC stocks have accumulated in distinctive soil conditions across the NPCTR’s mosaic of three hydropedologic landscape units (Lin et al., 2006): (1) upland forest soils on well-drained slopes, (2) forested wetlands, and (3) poor (lowland) fens (D’Amore et al 2015). Despite their relatively young age (12-14 cal ka BP; Eamer et al 2017), elevated C concentrations are observed in mineral soils that often exceed 1 m in depth (Chandler, 1943; Michaelson et al., 2013) due to a combination of rapid mineral weathering, high primary production and litter inputs, and the translocation of soluble C into deeper horizons (Alaback 1991). In addition to mineral soils, the perhumid NPCTR also exhibits a variety of vertically-accreting organic soils, including deep (3-5 m) peat-forming bogs and fens (Heusser 1952, 1954; Hansen 1955; Ugolini and Mann 1979), and thick (> 40 cm) forest floor organic horizons that accumulate due to slow decomposition under ubiquitous hydric soil conditions and the rarity of fire (Alaback 1991). In places, organic horizons overlay C-rich mineral soils (known as Folisols, or folistic horizons) and contribute to the highest pedon C stocks in the NPCTR (D’Amore and Lynn 2002; Fox and Tarnocai 2011; Johnson et al 2011; Michaelson et al 2013).

Quantifying total SOC storage across the NPCTR and understanding its environmental controls is necessary to predict the region’s response to global change, including climate feedbacks. Observational data (Buma and Barrett 2015) and ecosystem C models (Genet et al 2018) indicate that the NPCTR is sequestering C. However, changes in the amount and form of precipitation and higher annual
temperatures may increase growing season length and productivity (Buma et al. 2016) while soil warming may lead to more rapid decomposition of soil organic matter (Davidson et al. 2006; Fellman et al. 2017).

In the present study we address the need for a unified SOC model for the NPCTR by compiling a new transboundary pedon database across SEAK and coastal British Columbia that retains relevant pedological details. With this database we train a predictive model to estimate total SOC stocks spatially across the region, to enable meaningful comparisons with other regional and global SOC products, and to explore the environmental controls on SOC in the NPCTR.

2. Methods

2.1 Study Extent and Characteristics

The largest climatic zones within the NPCTR are the seasonal and perhumid forests that form a transboundary extent across SEAK and coastal BC (Alaback 1991). The SOC assessment encompassed all of the perhumid and part of the seasonal zone, spanning 10° of latitude (Figure 1). The study perimeter was defined by the outer boundary of rainforest dominated watersheds mapped using a harmonized transboundary dataset (Gonzalez Arriola et al. 2018) between the Fraser River in Vancouver, British Columbia (BC) and Lituya Bay south of Yakutat, Alaska, excluding the four major river basins (Taku, Stikine, Nass, Skeena) which extend into interior boreal forest and a more continental climate.

Mean annual precipitation across the study domain ranges from 1,800 to >3,000 mm and mean annual temperatures range from 6-9 °C, with monthly means of -5 °C in winter in the north (Farr and Hard 1987) and ~15 °C in summer in the south (Alaback 1991). Forest species diversity is relatively low, reflecting a consistent climate and disturbance regime across the study area, and generally dominated by *Picea sitchensis* (Sitka spruce) and *Tsuga heterophylla* (western hemlock) in SEAK (van Hees 2003). In BC, *Tsuga heterophylla* and *Thuja plicata* (western redcedar) become the dominant conifers. *Callitropsis nootkatensis* (yellow cedar) and *Tsuga mertensiana* (mountain hemlock) are found from sea level in the north to high elevations in the south and *Pinus contorta var. contorta* (shore pine) is a significant
component of bog locations throughout. Geology is varied consisting of granitic, basaltic, and limestone bedrock, the latter of which supports some of the most productive forest, however, much of the surficial geology is dominated by glacial drift including ablation and compact till, alluvial outwash, and glaciomarine sediments (Nowacki et al 2003).

2.2 Transboundary SOC Database

We compiled a transboundary database of soil profile descriptions (pedons) across SEAK and BC from published and archive data sources. For each pedon we calculated SOC stocks for the top 1 m of mineral soil plus surface organic horizons using data harmonization and gap-filling procedures that are detailed in the Supplementary Information (Sup. Table 1-5). In brief, US soil classification was converted to Canadian where necessary and gaps filled by published or modeled estimates grouped by soil class, horizon, and lithology. In contrast to some other regional and global C assessments, this approach avoided use of generalized empirical relationships between soil properties and missing variables, such as between soil C and soil bulk density, or soil C and depth.

2.3 Environmental Covariates

Environmental covariates were selected (Sup. Table 6) to predict SOC stock due to their relationship with soil forming factors (climate, organisms, relief, parent material, and time; Jenny 1994). Covariate data were extracted from the rasters at the pedon coordinates and appended to the final SOC stocks (in Supplemental data) to use in all further analyses. Further details of the 12 selected environmental covariates along with justification for inclusion and pre-processing steps are listed in Sup. Table 6. Briefly, only high quality and spatially continuous data products were used. Curating covariates based upon knowledge of regional soil development facilitates clearer interpretation and reduces the risk of autocorrelation between variables.

2.4 Random Forest Model
A random forest model was trained to predict stocks of SOC across the NPCTR in R (v. 3.4; R Core Team 2018) using the R-package randomForest (4.6; Liaw and Wiener 2002). Random forests grow a large number of regression trees (Breiman 1984) from different random subsets of training data and predictor variables, thereby reducing variance relative to single trees, and greatly reducing the risk of over-fitting model predictions and non-optimal solutions – though at the cost of interpretability (Breiman 2001). The transboundary database SOC stocks and associated covariates were first split into training (80%) and testing (20%) data and the model was parameterized to grow 5,000 trees. For each tree, a subsample equivalent to ¼ of the total sample size was utilized (with replacement). Node size was set at 4 to minimize the out-of-bag (OOB) error based on preliminary testing. Model performance was measured from goodness-of-fit, distributions of residuals, and predictions of test SOC stocks. Confidence intervals were computed using an infinitesimal jack-knife procedure (Wager et al. 2013). Predictive performance was measured from test data goodness-of-fit as well as the distribution of the root mean squared error (RMSE). Predictions were made across the NPCTR study extent using R-package raster (v2.6; Hijmans 2017) which produced a SOC map at 90.5 m resolution. All lakes > 10 ha were clipped from the final map (HydroLakes; Messager et al. 2016) and glacier area was clipped using the Randolph Glacier Inventory 5.0 (GLIMS; Raup et al. 2007) database. Final SOC stocks were adjusted for topography by scaling the SOC map with actual land surface area calculated from cell slope values. The random forest model was re-run for the three gap-filling sensitivity analyses (see SI).

2.5 Comparison to Regional and Global Maps

Stocks of SOC were compared with two previous Alaskan studies, two regional/national models, and two global models (Table 1). Published summary statistics for NPCTR regions were either referred to directly (Johnson et al. 2011) or estimated from published data (Michaelson et al. 2013). The Canadian SOC map produced by Tarnocai and Lacelle (1996) was regionalized, rasterized, and resampled to extract pixels that overlapped with the study boundary and methods to calculate mean and total SOC stocks were replicated. Two global SOC maps, SoilGrids250m (Hengl et al. 2017) and the Global Soil Organic Carbon...
map (GSOC; FAO and ITPS 2018), were downloaded as rasters and resampled from 250 m and 1 km resolutions, respectively. SoilGrids250m was built from a database of ca. 150,000 soil profiles and a stack of 158 covariates to produce a continuous global surface of SOC stock to 1 m, whereas the FAO GSOC map is a composite of national SOC stock assessments and covers a depth of 0-30 cm. Genet et al. (2018) estimated SOC across the N. Pacific Landscape Conservation Cooperative using pedons from relevant forest cover types in SEAK. Differences in SOC stocks are explored quantitatively in the context of different extents, gap-filling procedures, and data sources. Finally, the predictive capacity of the random forest model, measured by the Pearson’s correlation coefficient and RMSE of observations versus predictions, is compared to two global SOC products: SoilGrids250m and the FAO GSCO map.

2.6 Spatial Controls of SOC

To explore interactions between controls on SOC stocks across the NPCTR, classification and regression trees (CART) was applied to the transboundary SOC database using R-package `rpart` (v4.1; Therneau and Atkinson 2011). Unlike weak learner regression trees grown in random forests, CART analyses fit to entire datasets provide readily interpretable outputs. CART is also well suited to interpretation of complex data with many interacting variables, non-normally distributed data, and can identify key covariate interactions and thresholds.

3. Results

3.1 Summary of PCTR Observations

Pedon SOC stocks and depths were log-normally distributed (Sup. Fig 2 a-d). Median soil depth across all the samples was 66 cm and median calculated SOC density was 168.4 Mg ha\(^{-1}\). Other database summary statistics are provided in Sup. Table 7. Soil classes in the pedon database were mostly Spodosols (426), Inceptisols (214), and Entisols (84), with fewer Histosols (70) and Folists (9).
Sample locations were generally well distributed across the study extent with some clustering around S Vancouver Is., N BC, and central SEAK (Figure 1; Sup. Fig. 2 e-f). The distributions of the environmental covariate data extracted at pedons were generally very similar to the distributions of covariates across the region (Sup. Fig 2 e-n). Samples were slightly biased to lower and less steep areas, and the presence of large icefields and high alpine (not sampled) explained discrepancies in percent forest cover and land cover classes.

3.2 Random Forest SOC Model Performance

Model performance was strongest for larger scale patterns in SOC. Though predictions on test data by the random forest model was low ($R^2 = 0.32$), the predictive model performance on all observations was high and considerably better than two global SOC products ($\rho = 0.73$; Figure 2). In addition, model covariates were representative of the region (Sup. Fig. 2), the mean of the residuals was zero, and largest errors were under-estimations concentrated in areas otherwise correctly predicted to have higher than typical SOC (Sup. Fig. 3). We therefore have high confidence in model predictions for median values and the regional scale patterns in SOC, with less confidence for extreme values and variation at finer spatial scales. The predictions of the random forest in this study were much more accurate ($R^2 = 0.53$ vs. 0.11) compared to those extracted from two global products SoilGrids250 and the FAO GSOC map at the same locations (Figure 2).

3.3 Estimates of SOC Stocks

Total SOC within the NPCTR of SEAK and BC was estimated at 4.5 Pg C (Table 1) with highest stocks (> 500 Mg ha$^{-1}$) found in the central islands of SE Alaska and westerly locations, and lower stocks (< 200 Mg ha$^{-1}$) predicted for more southerly and easterly locations. Sensitivity analyses indicated that SOC stock estimates were most sensitive to bulk density gap-filling assumptions as estimates increased by approximately 50% after increasing organic horizon bulk density ca. 3-fold to 0.33 g cm$^{-3}$ (Sup. Table
5). From the fractional increase in SOC caused by the tripling of organic horizon bulk density, we computed that 22% of the predicted NPCTR SOC stocks must be stored in organic soil horizons.

3.4 Environmental Covariates of SOC Stock

CART analysis (Figure 3) showed that the lowest SOC stocks ranging from 128.5 to 194.8 Mg ha\(^{-1}\) were associated with the driest (< 2,147 mm MAP), southeasterly locations. Intermediate stocks (252.7-442.9 Mg ha\(^{-1}\)) were assigned to wetter climates at higher topographic positions (upslope). Very high SOC stocks (336.0-523.3 Mg ha\(^{-1}\)) were also associated with wet climate areas (2,147 – 2,833 mm) on foot-slope (downslope) landscape positions. Finally, exceptionally high SOC stocks of 446.2 - 708.6 Mg ha\(^{-1}\) were assigned to the wettest climates (> 2,833 mm) at relatively low elevations (< 189 m).

4. Discussion

4.1 SOC Stocks in the Global Context

The estimated 4.5 Pg C stored within perhumid and the northern seasonal NPCTR watersheds indicates the region contains approximately 2% of North American SOC within less than 1% of its surface area (Köchy et al 2015). Using a simple upscaling from study region mean stocks (228 ± 111 Mg ha\(^{-1}\)) to global CTR extent (ca. 9.7 x 10\(^5\) km\(^2\); Alaback 1991) we can estimate that 22 ± 11 Pg C may be stored globally in CTR ecosystems. These estimates are likely conservative due to our 1 m depth range, our assumption that the coarse fraction is entirely mineral (Zabowski et al 2011), the abundance of deep (3-5 m) peat-forming fens that can form in wet landscape depressions (Sup. Fig. 5, 6) that are smaller than our spatial resolution (Heusser 1952; D’Amore and Lynn 2002), as well as the likely occurrence of cryptic wetlands hidden within forests (Creed et al 2003).

In a review of global SOC, Jackson et al (2017) calculated the first biome-specific SOC stocks, estimating that 64 Pg C is stored in ca. 6 M km\(^2\) of non-permafrost soils in temperate conifer forests. Our results disaggregate this result further, suggesting CTRs within the temperate conifer forest biome contain
one-third of the total SOC, while representing less than one-sixth of the biome’s area. Jackson et al. (2017) also estimated that 22% (14 Pg) of the SOC was stored in organic peatlands, which, remarkably, matched exactly our estimate of the proportion of SOC in organic soil (peatlands and surface organic accumulations). The exact agreement, while remarkable, is not truly scalable, but it does likely reflect a common suite of C input and stabilization mechanisms in cool, wet temperate conifer forest ecosystems that may lead to consistent partitioning of stocks between mineral and organic soils.

As has been demonstrated for above-ground biomass (Keith et al. 2009), SOC densities in the CTR appear to rank among the highest globally. The mean SOC stock estimate from this study (228 ± 111 Mg ha⁻¹) positions the NPCTR below estimates for permafrost soils (178-691 Mg ha⁻¹; 5th – 95th percentile), but substantially higher SOC densities than grasslands (56-289 Mg ha⁻¹), evergreen broadleaf forests (83-223 Mg ha⁻¹), and croplands (60-200 Mg ha⁻¹), and within a similar range as permanent wetlands (114-474 Mg ha⁻¹; Sanderman et al. 2018). Our results also suggest SOC densities to 1 m in temperate rainforests are higher than in tropical rainforests (85-271 Mg ha⁻¹) perhaps in part due to litter accumulations which are typically absent in tropical rainforest floors due to very favorable conditions for decomposition (Parton et al. 2007).

4.2 Model Comparisons

Estimated SOC stocks agreed with some past estimates of SOC storage in Alaskan coastal rainforests (Johnson et al. 2011) but were much lower than the SEAK SOC estimates from Michaelson et al. 2013 (Table 1). The two regional/national studies that approximate SEAK (Tongass National Forest; Leighty et al. 2006) and BC (Tarnocai and Lacelle 1996), when summed, produced an estimate of 5.3 Pg C compared to our estimate of 4.5 Pg C. However, Tarnocai and Lacelle (1996) integrated SOC to the full observed depth of organic soils which was found to be approximately ~1.51 m, or ~50% greater than the reference depth in this study (1 m), which may explain the larger estimate. The GSOC map estimated lower SOC stocks (2.5 Pg C) because it only considers the top 30 cm of soil, but if it is assumed that ca.
50% of the SOC stock is stored from 30-100 cm (James et al. 2014), then the estimates (~ 5 Pg C) align well with the present study.

For one study and two global SOC products we identified large discrepancies with our SOC predictions. The global model SoilGrids250m and the regional Alaska database produced by Michaelson et al. (2013) were outliers in our comparison, predicting 4-fold higher total SOC for the region, and 2 or 3-fold higher mean SOC, respectively (Table 1). Our model also more accurately predicted the spatial variation in SOC across the NPCTR relative to the FAO GSOC map and SoilGrids250m (Figure 2). Both global products showed strong bias for the region, with over-estimates where we predict lower stocks and weak correlation with the observed stocks overall. Finally, unrealistic spatial discontinuities are present in the FAO GSOC map at the US-Canada border that did not exist in our transboundary assessment (Figure 5). Global SOC maps created from a mosaic of national inventories clearly benefit where nations conduct quality SOC assessments, evidenced by the reasonable summed stock estimates of GSOC for the NPCTR (Table 1). However, we propose that biome-specific assessments are better than national inventories because spatial discontinuities that form within global mosaics will fall along ecologically significant, rather than arbitrary political, boundaries (Ramcharan et al. 2017).

The bulk density gap-filling procedure applied by Michaelson et al. (2013) was not replicated in this study due to the observation that the pedotransfer functions (Van Looy et al. 2017) relating %C to bulk density over-estimate organic layer (high %C) bulk density by up to 4-fold relative to published values (Sup. Fig. 4; Shaw et al. 2005). Using these pedotransfer functions would bias the SOC estimates too high in the NPCTR where organic soils are common (D’Amore et al. 2010) and forest floor organic horizons are often deep (Kranabetter and Banner 2000). Direct comparison of our database with that of Michaelson et al. (2013) illustrates how gap-filling procedures can lead to very different SOC estimates. Similar issues may underlie discrepancies observed with SoilGrids250m (18 Pg C; Table 1). Models built using the Harmonized World Soil Database and SoilGrids250m have been gap-filled using pedotransfer
functions that may overestimate organic soil bulk density and lead to overestimated SOC stocks (Köchy 
*et al* 2015). A recent global model comparison found much larger SOC stock estimates using the 
SoilGrids database than from other global databases (Tifafi *et al* 2017). Our study similarly suggests 
SoilGrids250m overestimates SOC stocks within the NPCTR, and possibly in other organic and/or high-
latitude soils (Figure 2c). We cannot explain the differences with organic soil bulk density alone, based 
upon our sensitivity analysis where bulk density was tripled. We propose that the juxtaposition of highly 
contrasting soils in the NPCTR (Sup. Fig. 5, 6) may make the region particularly susceptible to SOC 
errors when aggregating and modeling pedon observations. In the NPCTR, C-rich litter layers and organic 
soils lie adjacent (vertically and laterally) to mineral soils (Michaelson *et al* 2013; Shaw *et al* 2018) and, 
without separate representation during gap-filling or modeling steps values may be artificially inflated. 
Populating the database and calculating SOC on the basis of pedological information, including 
distinguishing surface from subsurface soil horizons, may have improved SOC variable estimation in this 
study.

4.3 Covariates of NPCTR SOC Stocks

Digital soil mapping assumes properties, such as SOC, can be predicted spatially across 
landscapes from the distribution of geospatial covariates related to the classical factors of soil formation 
(Jenny 1994; Minasny 2011). We found that high precipitation is the primary control on SOC storage in 
the NPCTR; SOC stocks tracked regional gradients in MAP, and longitude, with the highest stocks in the 
north coast of BC and especially central SEAK (Figure 3; Figure 4). Topographic attributes including 
elevation, wetness, and slope position, which modulate temperature and soil moisture conditions, also 
emerged as important controls. Land cover was not a strong predictor, however both the region and pedon 
database are dominated by conifer forest which did not distinguish between upland and cryptic forested 
wateland coverage. Though lithology has been shown to be important across Alaska generally as a 
predictor of SOC stocks (Mishra *et al* 2012), we did not find support for lithology as a broadly important
covariate. However, lithology is an imperfect indicator of parent material for soil formation across the region due to the extensive presence of glacial deposits.

4.4 Vulnerabilities of NPCTR SOC Stocks

Stocks of SOC in the NPCTR may be sensitive to several climate-related changes in the coming century, but the overall direction of effects is uncertain. Wolken et al (2011) highlighted loss of winter snow and ice as the most important biophysical change in the coastal temperate rainforests of Alaska, driven by projected average temperature increases of $3.5 \pm 1.5 \degree C$ by 2100. Based upon the primacy of MAP and topographic wetness in our analyses, both higher predicted MAP and a reduction in the percentage as snow (Shanley et al. 2015) may expand the spatial and temporal domain for high SOC accumulations in the NPCTR. However, this may be balanced by increases in lateral exports of terrestrial DOC, which is already a distinctively large component of NPCTR ecosystem C budgets (Oliver et al 2017). Similarly, effects of temperature may also be bi-directional. Elevated temperatures lead to rapid decomposition of NPCTR soil organic matter under laboratory conditions (Fellman et al 2017) however it is unclear to what degree this effect will be limited by the saturated soil conditions in situ, which constrain decomposition rates (Freeman et al 2008), or offset by concurrent increases in SOC inputs via enhanced primary productivity and litterfall (Buma et al 2016; Genet et al 2018).

4.5 SOC Modeling Considerations

Our study shows that digital soil mapping can be valuable across the NPCTR where soil survey and conventional soil mapping is challenging (Carpenter et al 2014), however a baseline of high quality pedon data is still essential for accurate predictions. Vitharana et al (2017) found that existing data for SE Alaska well represented environmental variability and our covariate data distributions (Sup. Fig. 2) agree with this conclusion, however, much of the central and northern BC coastline is less well sampled (Figure 1) and those data we did obtain required extensive gap-filling. Our model also under-predicted the largest
SOC stocks which, as random forests subset data to grow each tree, may be due to relatively few observations of very high SOC.

Model improvements may also be possible if input covariate datasets and the final map are obtained at finer spatial resolution. For example, the NPCTR displays complex topography that may not be fully resolved at 90 m. Mishra and Riley (2015) found that soil wetness (derived from landform) and aspect were lost as significant predictors of Alaskan C stocks when moving from a 50 m to a 100 m resolution. Similarly, Siewert (2017) compared a wide range of resolutions (2 – 1,000 m) for random forest predictions of a sub-Arctic peatland SOC stocks in Sweden and found resolutions > 30 m led to underestimates. Building models using more accurately georeferenced pedon data and more finely resolved (< 50 m) covariate surfaces may improve spatial predictions of SOC.

5. Conclusions

Regional SOC stock assessments can validate and improve global maps by considering drivers, and compiling datasets, in greater detail. We compiled a SOC database for the NPCTR, using pedology data to guide gap-filling and predictive modeling. Regression tree models predicted high SOC stocks in wet coastal watersheds, indicating that the coastal temperate rainforest represents a moisture-dependent hotspot for SOC at mid-latitudes.

References


Breiman L 2001 Random forests Mach. Learn. 45 5–32
Breiman L, Friedman J H, Olshen R A and Stone C G 1984 *Classification and Regression Trees* (Belmont, California, USA: Wadsworth International Group)

Buma B and Barrett T M 2015 Spatial and topographic trends in forest expansion and biomass change, from regional to local scales *Glob. Chang. Biol.* 21 3445–54


Chandler R F 1943 The time required for Podzol profile formation as evidenced by the Mendenhall glacial deposits near Juneau, Alaska *Soil Sci. Soc. Am. J.* 7 454–9 Online: [https://www.soils.org/publications/sssaj/abstracts/7/C/SS00700C0454](https://www.soils.org/publications/sssaj/abstracts/7/C/SS00700C0454)


Hartmann J and Moosdorf N 2012 The new global lithological map database GLiM: A representation of rock properties at the Earth surface Geochemistry, Geophys. Geosystems 13 1–37


Heusser C J 1952 Pollen Profiles from Southeastern Alaska Ecol. Monogr. 22 331–52


Liaw A and Wiener M 2002 Classification and Regression by randomForest *R news* 2 18–22


Advances in Agronomy vol 118 (Elsevier) pp 1–47 Online: http://dx.doi.org/10.1016/B978-0-

12-405942-9.00001-3

Mishra U and Riley W J 2015 Scaling impacts on environmental controls and spatial heterogeneity of

soil organic carbon stocks Biogeosciences 12 3993–4004

Mishra U and Riley W J 2012 Alaskan soil carbon stocks: spatial variability and dependence on

environmental factors Biogeosciences 9 3637–45

NALCMS 2017 Land cover of North America at 30 meters Online: http://www.cec.org/tools-and-

resources/north-american-environmental-atlas/land-cover-and-land-cover-change

Neiland B J 1971 The forest-bog complex of southeast Alaska Veg. Acta Geobot. 22 1–64

Nowacki G J, Spencer P, Fleming M, Brock T and Jorgenson T 2003 Unified Ecoregions of Alaska:

2001 Online: http://pubs.er.usgs.gov/publication/ofr2002297

Oliver A A, Tank S E, Giesbrecht I, Korver M C, Floyd W C, Sanborn P, Bulmer C and Lertzman K P

2017 A global hotspot for dissolved organic carbon in hypermaritime watersheds of coastal

British Columbia Biogeosciences 14 3743–62


L A, Hart S C and Fasth B 2007 Global-scale similarities in nitrogen release patterns during

long-term decomposition Science. 5064 361–4


101–10

Sanderman J, Hengl T, Fiske G, Solvik K, Adame M F, Benson L, Bukoski J J, Carnell P, Cifuentes-

Jara M, Donata D, Duncan C, Eid E M, zu Ermgassen P, Lewis C J E, Glass L, Gress S,


and Landis E 2018 Correction for Sanderman et al., Soil carbon debt of 12,000 years of human
land use Proc. Natl. Acad. Sci. 115 E1700 LP-E1700 Online:

http://www.pnas.org/content/115/7/E1700.abstract


Shaw C H, Banfield E and Kurz W A 2008 Stratifying soils into pedogenically similar categories for modeling forest soil carbon Can. J. Soil Sci. 88 501–16


Shaw C, Bhatti J and Sabourin K 2005 An ecosystem carbon database for Canadian forests Online:


Tifafi M, Guenet B and Hatté C 2017 Large differences in global and regional total soil carbon stock
estimates based on SoilGrids, HWSD and NCSCD: Intercomparison and evaluation based on

S D 2013 Causes of variation in soil carbon simulations from CMIP5 Earth system models
and comparison with observations *Biogeosciences* 10 1717–36 Online:

http://www.biogeosciences.net/10/1717/2013/

Ugolini F C and Mann D H 1979 Biopedological origin of peatlands in South East Alaska *Nature* 281
366–8 Online: http://www.nature.com/articles/281366a0

USGS (2006) https://www.sciencebase.gov/catalog/item/4f4e477be4b07f02db47fa14; accessed
03/13/2018)

van Hees W W S 2003 *Forest resources of southeast Alaska, 2000* (Juneau, AK)

L, Zacharias S, Zhang Y and Vereecken H 2017 Pedotransfer functions in Earth system
science: challenges and perspectives *Rev. Geophys.* 1–58 Online:

http://doi.wiley.com/10.1002/2017RG000581

Vitharana U W A, Mishra U, Jastrow J D, Matamala R and Fan Z 2017 Observational needs for

Wager S, Hastie T and Efron B 2013 Confidence intervals for random forests: The jackknife and the
infinitesimal jackknife *J. Mach. Learn. Res.* 15 1625–51 Online:

http://arxiv.org/abs/1311.4555

Mcguire A D, Euskirchen E S, Hennon P E, Beever E A, Conn J S, Crone L K, D’Amore D


**Contributions**

BB, DB, and AB conceived of the project. PS, CB, SS, and DD provided data for, and supervised GM during the compilation, and gap-filling of, the transboundary NPCTR SOC database. BB supervised GM on random forest and CART analyses. IG, SA, and PS created NPCTR digital surface products and provided unpublished data from digital soil mapping efforts by the Hakai Institute in 2014 and 2016. CB and SS provided BEC and BCSIS datasets and CB supervised GM in gap-filling of BEC pedon data. GM and BB authored the first draft of the manuscript and CB, DD, PS, IG, AB, and DB contributed to later drafts.

**Acknowledgements**

We thank Frances Biles at the USFS Pacific Northwest Forestry Station for advice on geospatial data products. We thank Cindy Shaw, Marty Kranabetter, and Gary Michaelson for sharing published and unpublished data for the PCTR database. We also thank Umakant Mishra for discussion and advice relating to state-wide carbon stock modeling and Marwa Tifafi for discussion of global models. This work was supported by the National Science Foundation under Grant No. 1557186, USDA McIntire Stennis funding through the Alaska Agricultural and Forestry Experiment Station, and Tula Foundation funding of the Hakai Institute and associated unpublished field data.
Figure Captions

Figure 1: The extent of the NPCTR displaying the distribution of soil profile descriptions (light green circles) and the study extent (dark green pixels) used in the SOC stock assessment. Inset maps show global (small) and continental (large) extent of the full NPCTR along with the boundaries of climatic sub-regions.

Figure 2: Predicted vs. observed SOC stocks (Mg ha\(^{-1}\)) for individual database profile locations compared for (a) this study’s spatial predictions, (b) the FAO GSOC map, and (c) SoilGrids250m, with smooth spline fits (solid line) and 1:1 line (dashed). Pearson’s correlation coefficients (ρ) and root mean squared error (RMSE) between predicted and observed values are reported for each model. Note FAO predictions (for 0-30 cm) are doubled assuming approximately 50% of SOC stocks are stored from 30-100 cm depth (James et al 2014).

Figure 3: Regression tree results for environmental covariates controlling NPCTR SOC stocks.

Figure 4: Soil organic carbon stock predictions to 1 m (Mg C ha\(^{-1}\)) at 90.5 m resolution for small NPCTR watersheds across BC and SE Alaska.

Figure 5: (a) Sharp spatial discontinuities at the border of Alaska (USA) and BC (Canada) in the FAO Global Soil Organic Carbon (GSOC) map are, in the current study (b), smooth landscape-to-regional gradients, due to harmonized data compilation, gap-filling, and modeling approaches across the transboundary extent. Note the FAO GSOC map stock estimates are lower due to a shallower depth range (0-30 cm).
Figures

Figure 1
Figure 2

(a) $\rho = 0.73$, RMSE = 119 Mg ha$^{-1}$

(b) $\rho = 0.34$, RMSE = 175 Mg ha$^{-1}$

(c) $\rho = 0.34$, RMSE = 274 Mg ha$^{-1}$
Figure 3

SOC Stock
(Mg C ha⁻¹)

- MAP > 2147 mm
  - MAP > 1379 mm
    - TWI > 1.9
      - 128.5
        - n = 210
      - 194.8
        - n = 246
    - Longitude < -130.3°
      - 147.5
        - n = 36
      - 252.5
        - n = 176
  - MAP > 2833
    - TPI < -1.5
      - 271.7
        - n = 8
    - Elevation < 189 m
      - Contributing Area > 16
        - 252.7
          - n = 49
      - Contributing Area < 3.5
        - 336.0
          - n = 14
      - 523.3
        - n = 10
    - Elevation > 122 m
      - 446.2
        - n = 34
      - 708.6
        - n = 12
Figure 4
Figure 5

(a) SOC (Mg/ha) distribution map of the study area.

(b) SOC (Mg/ha) distribution map of the study area.

BC
### Table 1: Comparison of methods and results of NPCTR SOC stocks estimated in this study and other regional and global SOC studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Scale</th>
<th>Region</th>
<th>Method</th>
<th>Mean SOC (Mg ha(^{-1}))</th>
<th>Total SOC (Pg C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present Study</td>
<td>Regional</td>
<td>PCTR Watersheds</td>
<td>Random Forest</td>
<td>228 ± 111</td>
<td>4.5</td>
</tr>
<tr>
<td>Present Study – Sens. Test</td>
<td>Regional</td>
<td>Coastal Rainforests (Upland)</td>
<td>+ High Bulk Density</td>
<td>329 ± 160</td>
<td>6.5</td>
</tr>
<tr>
<td>Present Study – Sens. Test</td>
<td>Regional</td>
<td>Coastal Rainforests (Lowland)</td>
<td>+ High Mineral %C</td>
<td>249 ± 125</td>
<td>4.9</td>
</tr>
<tr>
<td>Present Study – Sens. Test</td>
<td>Regional</td>
<td></td>
<td>+ No Coarse Fra.</td>
<td>224 ± 118</td>
<td>4.5</td>
</tr>
<tr>
<td>Johnson et al (2011)</td>
<td>Regional</td>
<td>Coastal Rainforests</td>
<td>Database Summary Stats.</td>
<td>240 ± 132</td>
<td>-</td>
</tr>
<tr>
<td>Johnson et al (2011)</td>
<td>Regional</td>
<td>Coastal Rainforests (Lowland)</td>
<td>Database Summary Stats.</td>
<td>258 ± 174</td>
<td>-</td>
</tr>
<tr>
<td>Michaelson et al (2013)</td>
<td>Regional</td>
<td>SE Alaska</td>
<td>Database Summary Stats.</td>
<td>587 ± 379</td>
<td>-</td>
</tr>
<tr>
<td>Tarnocai and Lacelle (1996)</td>
<td>National</td>
<td>Regionalized to NPCTR (BC only)</td>
<td>Soil Map Unit Assignment</td>
<td>294 ± 180</td>
<td>3.4 (^3)</td>
</tr>
<tr>
<td>Leighty et al (2006)</td>
<td>State</td>
<td>Tongass Nat. Forest (AK)</td>
<td>Econometric Flow Chart</td>
<td>-</td>
<td>1.9 (^3)</td>
</tr>
<tr>
<td>Genet et al (2018)</td>
<td>State</td>
<td>North Pacific Coast LLC (AK)</td>
<td>Land Cover Class</td>
<td>-</td>
<td>4.8 (^3)</td>
</tr>
<tr>
<td>SoilGRID250m (Hengl et al 2017)</td>
<td>Global</td>
<td>Regionalized to NPCTR</td>
<td>Random Forest</td>
<td>486 ± 149</td>
<td>18.0</td>
</tr>
<tr>
<td>GSOC (FAO and ITPS 2018)</td>
<td>Global</td>
<td>Regionalized to NPCTR</td>
<td>US: Soil Map Unit BC: Ensemble</td>
<td>136 ± 60</td>
<td>2.5 (^4)</td>
</tr>
</tbody>
</table>

\(^1\) Map of database mean ± SD; a dash denotes no means reported. \(^2\) SOC stock for overlap extent with present study; a dash denotes no regional prediction. \(^3\) Partially overlap extent with this study. Genet et al (2018) extent excludes BC but includes maritime upland and alder forest cover across southeast and south central Alaska. \(^4\) The FAO Global SOC map extends only to 30 cm.