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

Gavin McNicol1, Chuck Bulmer2, David D’Amore3, Paul Sanborn3,5, Sari Saunders4, Ian Giesbrecht5,6,7,8,9, Santiago Gonzalez Arriola1, Allison Bidlack8, David Butman1,2 and Brian Buma1

1 Alaska Coastal Rainforest Center, University of Alaska Southeast, Juneau, AK 99801, United States of America
2 B. C. Ministry of Forests, Lands and Natural Resource Operations, Forest Sciences section, 103-2100 Labieux Rd., Nanaimo, BC V9T 6E9, Canada
3 USDA Forest Service, Pacific Northwest Research Station, Juneau, AK 99801, United States of America
4 Ecosystem Science and Management Program, University of Northern British Columbia, 3333 University Way, Prince George, BC, V2N 4Z9, Canada
5 Hakai Institute, Tula Foundation, PO Box 309, Heriot Bay, BC, V0P 1H0, Canada
6 B. C. Ministry of Forests, Lands and Natural Resource Operations, 103-2100 Labieux Rd., Nanaimo, BC V9T 6E9, Canada
7 School of Resource and Environmental Management, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada
8 School of Environmental and Forest Sciences and Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, United States of America
9 Department of Integrative Biology, University of Colorado, 1151 Arapahoe, SI 4101, Denver, CO 80301, United States of America

E-mail: gmcnicol@stanford.edu

Keywords: machine learning, digital soil mapping, coastal temperate rainforest, soil organic carbon, biogeochemistry, soil science

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 1000 to >3000 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 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: $\rho = 0.73$ versus 0.34). Notably, SoilGrids 250 m 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.

1. Introduction

Accurate global soil organic carbon (SOC) maps are necessary to validate terrestrial carbon (C) cycle predictions in Earth System Models (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, 2018a). Database construction, including filling data gaps,
account for some of these discrepancies (Tilfai 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 landscape topography (Mishra and Riley 2015, Siewert 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. (2018b) 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 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 4000 km of the N. American coast from the Russian River in California, to Kodiak Island in the Gulf of Alaska (Della-Sala 2011). Although several studies have produced regional estimates of SOC stocks in Alaska (Lighty et al. 2006, Johnson et al. 2011, Mishra and Riley 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 distinct soil conditions across the NPCTR’s mosaic of three hydromorphic landscape units (Lin et al. 2006): (1) upland forest soils on well-drained slopes, (2) forested wetlands, and (3) poor (lowland) fens (Neiland 1971, 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 peHumid 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, deep (>40 cm) organic horizons overlay C-rich mineral soils (known as Folisols, or folistic horizons) and contribute to the highest 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. Changes in the amount and form of precipitation and higher temperatures may increase growing season length and productivity (Buma et al. 2016) however soil warming may lead to more rapid decomposition of soil organic matter (Davidson and Janssens 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 BC 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, 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 1800 to >3000 mm and mean annual temperatures range from 6 °C to 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 > 1300 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 (supplementary tables 1–5). The database is available online at: https://datadryad.org/resource/doi:10.5061/dryad.5j6j1r. In brief, US soil classification was converted to Canadian where necessary and gaps filled with published values 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 (supplementary 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 supplementary material) 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 supplementary 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 (www.R-project.org)) using the R-package randomForest (4.6; Liaw and Wiener 2002). Random forests grow a large number of regression trees (Breiman et al 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 5000 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 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). 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—the final SOC map is available at: https://datadryad.org/resource/doi:10.5061/dryad.5jf6j1r (Dryad)).

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 Lal (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 (available at: http://sis.agr.gc.ca/cansis/interpretations/carbon/index.html). 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. SoilGrids 250 m 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 accuracy and bias of the random forest model, is compared to two global SOC products (SoilGrids250m and the FAO GSOC map).

2.6. Spatial controls of SOC

To explore 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 2018). Unlike weak learner regression trees grown in random forests, CART analyses fits 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 (supplementary figures 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 supplementary 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; supplementary figures 2(e) and (f)). The distributions of the environmental covariate data extracted at pedons
| Study                  | Scale          | Region                                      | Method                               | Mean SOC $^a$ (Mg ha$^{-1}$) | Total SOC $^b$ (Pg C) |
|-----------------------|----------------|---------------------------------------------|--------------------------------------|------------------------------|-----------------------|
| Present Study         | Regional       | NPCTR Watersheds                           | Random Forest                        | 228 ± 111                    | 4.5                   |
| Present Study—Sens. Test | Regional | + High Bulk Density                       | Database Summary Stats.             | 329 ± 160                    | 6.5                   |
| Present Study—Sens. Test | Regional | + High Mineral %C                         | Database Summary Stats.             | 249 ± 125                    | 4.9                   |
| Present Study—Sens. Test | Regional | + No Coarse Fra.                           | Database Summary Stats.             | 224 ± 118                    | 4.5                   |
| Johnson et al (2011)  | Regional       | Coastal Rainforests (Upland)               | Database Summary Stats.             | 240 ± 132                    | —                     |
| Johnson et al (2011)  | Regional       | Coastal Rainforests (Lowland)              | Database Summary Stats.             | 258 ± 174                    | —                     |
| Michaelson et al (2013)| Regional       | SE Alaska                                  | Database Summary Stats.             | 587 ± 379                    | —                     |
| Tarnocai and Lacelle (1996) | National | Regionalized to NPCTR (BC only)           | Soil Map Unit Assignment             | 294 ± 180                    | 3.4$^c$                |
| Leighty et al (2006)  | State          | Tongass Nat. Forest(AK)                    | Econometric Flow Chart               | —                            | 1.9$^c$                |
| Genet et al (2018)    | State          | North Pacific Coast LLC (AK)               | Land Cover Class                     | —                            | 4.8$^c$                |
| SoilGRIDS250m (Hengl et al 2017) | Global | Regionalized to NPCTR                     | Random Forest                        | 486 ± 149                    | 18.0                  |
| GSOC (FAO and ITPS 2018) | Global | Regionalized to NPCTR                     | US: Soil Map Unit BC: Ensemble       | 136 ± 60                     | 2.5$^d$                |

$^a$ Map of database mean ±SD; a dash denotes no means reported.

$^b$ SOC stock for overlap extent with present study (unless otherwise noted); a dash denotes no regional prediction.

$^c$ 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.

$^d$ The FAO Global SOC map extends only to 30 cm depth.
were generally very similar to the distributions of covariates across the region (supplementary figures 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 predictive performance on test data by the random forest model was low ($R^2 = 0.32$), the model covariates were representative of the region (supplementary figure 2), the mean of the residuals was zero, and largest errors were under-estimations in areas otherwise correctly predicted to have higher than typical SOC (supplementary figure 3). We therefore have high confidence in model predictions for the regional scale patterns in SOC, with less confidence for variation at finer spatial scales. The predictions of the random forest in this study were more accurate (figure 2) 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 30% after increasing organic horizon bulk density ca. 3-fold to 0.33 g cm$^{-3}$ (supplementary 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 (<2147 mm MAP), southeastern 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 (2147–2833 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 (>2833 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 ($228 \pm 111$ Mg ha$^{-1}$) to global CTR extent (ca. $9.7 \times 10^5$ 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 (supplementary figures 5 and 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 matched exactly our estimate of the proportion of SOC in organic soil (peatlands and surface organic accumulations). The 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 aboveground 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 ($228 \pm 111$ 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 2018b). 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 (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...
Figure 2. Predicted versus 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 (\(\rho\)) 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).
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 to 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 variability 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 overestimate organic layer (high %C) bulk density by up to 4-fold relative to published values (supplementary figure 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

![Figure 3](image-url)
global model comparison found much larger SOC stock estimates using the SoilGrids database than from other global databases (Tifafi et al 2018). Our study similarly suggests SoilGrids250m overestimates SOC stocks within the NPCTR, and possibly in other organic and/or high-latitude soils (figure 2(c)). We cannot however explain the differences with organic soil bulk density alone, based upon our sensitivity analysis where bulk density was tripled. We therefore propose that the juxtaposition of highly contrasting soils in the NPCTR (supplementary figures 5 and 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 organic from subsurface mineral 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 2013). 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 central SEAK (figures 3 and 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 does not distinguish between upland soils and forested wetland coverage. Though lithology has been shown to be important across Alaska generally as a predictor of SOC stocks (Mishra and Riley 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 CTRs of Alaska, driven by projected average temperature increases of 3.5 °C ± 1.5 °C by 2100. Based upon the primacy of MAP and topographic wetness in our analyses, both higher predicted MAP and a reduction in the proportion 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 distinctly large component of NPCTR ecosystem C budgets (Oliver et al. 2017). Similarly, effects of temperature may be bi-directional. Elevated temperatures lead to rapid decomposition of NPCTR soil organic matter under laboratory conditions (Fallman 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. 2001), 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 (supplementary figure 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 to a 100 m resolution. Similarly, Siewert (2017) compared a wide range of resolutions (2–1000 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.

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).
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 CTR represents a moisture-dependent hotspot for SOC at mid-latitudes.

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.

Acknowledgments

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. The BEC plot data used in this analysis was provided by the Research Program of the B.C. Ministry of Forests, Lands, and Natural Resource Operations. We also thank Uma-kant Mishra for discussion and advice relating to state-wide carbon stock modeling and Marwa Tifafi and Hélène Genet for discussion of global and Alaskan models, respectively. 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.

ORCID iDs

Gavin McNicol @ https://orcid.org/0000-0002-6655-8045
Ian Giesbrecht @ https://orcid.org/0000-0001-7013-4766
Brian Buma @ https://orcid.org/0000-0003-2402-7737

References

Alaback P B 1991 Comparative ecology of temperate rainforests of the Americas along analogous climatic gradients Rev. Chil. Hist. Nat. 64 399–412
Breiman L 2001 Random forests Mach. Learn. 45 35–53
Breiman L, Friedman J H, Olshen R A and Stone C G 1984 Classification and Regression Trees (Belmont, CA: 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. Change Biol. 21 3445–54
Buma B, Krapek J and Edwards R T 2016 Watershed-scale forest biomass distribution in a perhumid temperate rainforest as driven by topographic, soil, and disturbance variables Can. J. For. Res. 46 844–54
Carpenter D N, Bockheim J G and Reich P F 2014 Soils of temperate rainforests of the North American pacific coast Geoderma 230–231 250–64
Chandler R F 1943 The time required for Podzol formation as evidenced by the Mendenhall glacial deposits near Juneau, Alaska Soil Sci. Soc. Am. J. 7 454–9
Credt J F, Sanford S E, Beall F D, Molot L A and Dillon P J 2003 Cryptic wetlands: integrating hidden wetlands in regression models of the export of dissolved organic carbon from forested landscapes Hydrol. Process. 17 3629–48
D’Amore D V, Fallman J B, Edwards R T and Hood E 2010 Controls on dissolved organic matter concentrations in soils and streams from a forested wetland and sloping bog in southeast Alaska Ecolhydrology 3 249–61
D’Amore D V and Lynn W C 2002 Classification of forested Histosols in southeast Alaska Soil Sci. Soc. Am. J. 66 554
D’Amore D V, Ping C-L and Herendeen P A 2015 Hydromorphic soil development in the coastal temperate rainforest of Alaska Soil Sci. Soc. Am. J. 79 698
Davidson E A and Janssens I A 2006 Temperature sensitivity of soil carbon decomposition and feedbacks to climate change Nature 440 165–73
DellaSala D A 2011 Temperate and boreal rainforests of the Pacific coast of North America Temperate and Boreal Rainforests of the World: Ecology and Conservation ed D A DellaSala (Washington, DC: Island Press) pp 42–82
Eamer J B R, Shugar D H, Walker I J, Lian O B, Neudorf C M and Telka A M 2017 A glacial readvance during retreat of the Cordilleran Ice Sheet, British Columbia central coast Quat. Res. 87 468–81
FAO and ITTPS 2018 Global Soil Organic Carbon Map (GSOCmap) Technical Report 162 FAO and ITTPS (Rome)
Farr W A and Hard J S 1987 Multivariate analysis of climate along the southern coast of Alaska—some forecasting implications 38 US Department of Agriculture, Forest Service, Pacific Northwest Research Station p 372
Fallman J B, D’Amore D V, Hood E and Cunningham P 2017 Vulnerability of wetland soil carbon stocks to climate warming in the perhumid coastal temperate rainforest Biogeochemistry 133 165–79
Fox C and Tarnocai C 2011 Organic soils of Canada: II. Upland organic soils Can. J. Soil Sci. 91 823–42
Freeman C, Ostle N and Kang H 2001 An enzymic ‘latch’ on a global carbon store Nature 409 149
Genet H et al 2018 The role of driving factors in historical and projected carbon dynamics of upland ecosystems in Alaska Ecol. Appl. 28 5–27
Gonzalez Arriola S, Giesbrecht I, Biles J J W and D’Amore D V 2018 Watersheds of the northern Pacific coastal temperate rainforest margin: internationally harmonized watershed polygons at the finest scale available (Hakai Inst. Data Package) (https://doi.org/10.21966/1.715755)
Hansen H 1935 Postglacial forests in south-central and central British Columbia Am. J. Sci. 253 640–58
Hengst T et al 2017 SoilGrids250m: global gridded soil information based on machine learning PLoS ONE 12 e0169748
