

Modeling of total and active organic carbon dynamics in agricultural soil using digital soil mapping: a case study from Central Nova Scotia

Authors: Paul, Siddhartho S., Heung, Brandon, and Lynch, Derek H.

Source: Canadian Journal of Soil Science, 103(1) : 64-80

Published By: Canadian Science Publishing

URL: <https://doi.org/10.1139/cjss-2022-0012>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Modeling of total and active organic carbon dynamics in agricultural soil using digital soil mapping: a case study from Central Nova Scotia

Siddhartho S. Paul ^{a,b}, Brandon Heung^b, and Derek H. Lynch ^b

^aSt. John's Research and Development Centre, Agriculture and Agri-Food Canada, St. John's, NL A1E 6J5, Canada; ^bDepartment of Plant, Food and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Truro, NS, B2N 5E3, Canada

Corresponding author: Siddhartho S. Paul (email: siddhartho.paul@canada.ca)

Abstract

Monitoring the changes in soil organic carbon (SOC) pools is critical for sustainable soil and agricultural management. This case study models total and active organic carbon dynamics (2015/2016 to 2019/2020) using digital soil mapping (DSM) techniques. Model predictors include topographic variables generated from light detection and ranging data; soil and vegetation indices derived from Landsat satellite images; and soil and crop inventory information from Agriculture and Agri-Food Canada to predict total organic carbon (TOC) and permanganate oxidizable carbon (POXc) at the 0–15 cm depth increment for a 37 km² study area in Truro, Nova Scotia. Quantile Regression Forest and stochastic Gradient Boosting Model were utilized for prediction. Although both models performed equally well for predicting TOC and POXc, the accuracy of TOC predictions (e.g., concordance correlation coefficient (CCC) = 0.67) was better than POXc predictions (e.g., CCC = 0.53). The Landsat variables and crop inventory were dominant predictors, while topographic variables across the relatively homogeneous terrain had relatively little influence. During the study period, changes in POXc were predicted across 98% of the study area, with a mean absolute loss of 5.77 (± 11.48) mg/kg/year, and in TOC on 27% of the area, with a mean absolute loss of 0.15 (± 0.09) g/kg/year. While the annual crop fields observed the highest loss of TOC and POXc, the decline in pasture–grassland–forage fields was relatively low. The study reinforced the effectiveness of DSM for modeling multiple SOC pools at the farm to landscape scales.

Key words: crop inventory, LiDAR, machine learning, soil organic carbon, remote sensing

1. Introduction

Soil organic carbon (SOC) is a key soil quality parameter that influences a range of physical, chemical, and biological properties of soil (Wiesmeier et al. 2019); therefore, monitoring and maintaining SOC contents are critical for agricultural productivity. Furthermore, SOC is recognized as the largest terrestrial carbon pool and plays a vital role in the global C cycle, providing important benefits for climate change mitigation and adaptation (Minasny et al. 2017). Yet, in Atlantic Canada, the intensity of cropping systems is leading to continued declines in SOC. In a national-scale assessment, Clearwater et al. (2016) estimated that most of Atlantic Canada experienced a moderate (25–90 kg/ha/year) to large (>90 kg/ha/year) decrease in SOC from 1981 to 2011. Hence, developing an effective SOC sequestration strategy has emerged as a fundamental need for this region; however, for the successful implementation of such strategies, spatiotemporal assessments of SOC under different environmental stresses and land-management conditions are critical. Recently, the direct financial returns of SOC sequestration, or “carbon credit”, have been in discussion to offset the associated cost of SOC sequestration in farm lands, but

this also requires accurate and cost-effective quantification of SOC (Paustian et al. 2019). In this context, the site-specific mapping of SOC has gained wide attention among producers, land managers, and policy-makers.

Within the SOC pool, active carbon, which can be measured as permanganate oxidizable carbon (POXc), is a carbon fraction that can be easily oxidized and lost due to changes in agricultural management and land use practices (Culman et al. 2012). The monitoring of POXc may, therefore, provide more immediate information on the soil's response to shifts in management practices compared to the total organic carbon (TOC). Assessing the dynamics of TOC and POXc can provide important cues for effective agricultural soil management and enhance agroecosystem resilience against climatic perturbations. Furthermore, combining the evaluation of TOC and POXc is critical in tracking the spatial distribution and changes in overall soil health (Zebarth et al. 2019; Norris et al. 2020).

Despite the importance of monitoring SOC dynamics, spatiotemporal modeling of SOC remains a consistent challenge due to the complex, biogeochemical processes within the soil ecosystem. The widely adopted approaches for

investigating SOC dynamics include (i) the statistical comparison of plot-level data from long-term research trials that incorporate different management treatments (VandenBygaert et al. 2010) and (ii) the use of dynamic, process-based models (e.g., Century, DeNitrification-DeComposition, Rothamsted Carbon Model), which can simulate the impacts of agricultural management and land use changes on various SOC pools over a certain period (Li et al. 2016; Dimassi et al. 2018). Plot-level measurement may provide accurate assessments of SOC dynamics; however, long-term research trials may not always replicate the management conditions of operational agricultural farms. Furthermore, setting up such trials may become prohibitively time and cost intensive. Conversely, the dynamic process-based models usually require an extremely large set of complex input data for model parameterization to accurately replicate the soil biogeochemical processes; thus, the predicted outcomes may include sizeable uncertainties when there is poor or inadequate information for landscape-scale analysis (Adhikari et al. 2019; Huang et al. 2019).

Due to the continuous advancements in computation power, remote sensing technologies, and machine-learning techniques, the field of digital soil mapping (DSM) has provided the means for modeling and predicting the spatial patterns of soil properties at various scales (McBratney et al. 2003). In DSM, environmental variables, derived from remote sensing and other geospatial data sources, are correlated with soil observations using empirical models (e.g., machine learning and geostatistical) to predict soil properties. Here, DSM practitioners utilize the generic SCORPAN framework, proposed by McBratney et al. (2003), to generate spatial predictions of soil properties based on geospatial layers that represent the intrinsic properties of soils (S), climate (C), organisms (O), relief (R), parent materials (P), soil age (A), and spatial location (N). The SCORPAN factors are correlated with the target soil variables to generate a predictive model, which may then be used to simulate the impacts related to temporal changes in climate, management, and land use practices. When simulating the spatiotemporal patterns of soil, such as the prediction of SOC dynamics, time-series, remote sensing data may be used for capturing the temporal changes (Taghizadeh-Mehrjardi et al. 2021; Fathizad et al. 2022). Thus, this pragmatic modeling technique has widened the scope for predicting spatiotemporal dynamics of soil and has recently been used in various regional-scale studies. For example, Yigini and Panagos (2016) and Paul et al. (2020b) both evaluated changes in SOC for Europe and the Lower Fraser Valley of British Columbia, respectively. Similarly, Fathizad et al. (2020) simulated the changes in soil quality due to land use and land cover changes for the central desert of Iran, while Taghizadeh-Mehrjardi et al. (2021) modeled the changes in heavy metals in soils for the same region. Although DSM studies, such as Paul et al. (2020b) and Fathizad et al. (2022), predicted the spatiotemporal patterns of SOC at the landscape scale, there has yet to be any research that explicitly predicted active C within the DSM literature. Given the highly volatile nature of the active C fraction of soil, applying DSM techniques for predicting POXc may provide a fast and cost-effective assessment of active C dynamics in response to changes in management practices.

The spatial resolution and types of environmental predictors are key considerations in DSM. Lamichhane et al. (2019) identified that environmental variables representing agricultural management and land use activities were the most frequently used covariates for SOC mapping, followed by climate and topographic variables—although climate variables were usually influential for large, regional-scale studies. For landscape-scale studies, agricultural management and land use information were typically derived from satellite images at various spatial resolutions (Maynard and Levi 2017; Lamichhane et al. 2019; Fatholouloumi et al. 2020), which may have spatial resolutions as fine as 10 m (e.g., Sentinel-2 imagery); however, variables obtained from Landsat satellite images (30 m spatial resolution) have been used globally (Boettinger et al. 2008; Minasny and McBratney 2016). Landsat data have been readily available since 1972 and hence, the integration of Landsat-derived variables into predictive models has the unique ability to track spatiotemporal changes. Specifically, Landsat imagery can be used to derive map products, such as crop inventories, which may be an important determinant for modeling TOC and POXc (Wang et al. 2019). Within Canada, for example, Agriculture and Agri-Food Canada (AAFC) has used optical-based (e.g., Landsat, AWiFS, and DMC) and radar-based (e.g., Radarsat-2) satellite images to produce the annual crop inventory since 2009, which consists of 72 classes and is distributed at a 30 m spatial resolution (Fisette et al. 2013, 2014). Despite its comprehensive coverage, the use of this product remains relatively unexplored in the Canadian DSM literature. Another source of data, which has greatly transformed DSM in recent years, is the increasing availability of fine-resolution light detection and ranging (LiDAR) data as it significantly enhances our ability to capture field-scale topographic variability in DSM. Blackford et al. (2020) demonstrated a workflow using LiDAR for DSM application and successfully predicted multiple soil attributes, including soil moisture regimes and texture in Hearst Forest, Ontario. Similarly, Kasraei et al. (2021) predicted soil pH, soil thickness, and depth to carbonates for the Kamloops region of British Columbia, and predicted soil pH and SOC for the Ottawa region of Ontario using LiDAR. Therefore, the integration of environmental predictors of variable spatial resolutions derived from satellite images, crop inventory, and LiDAR data can be an effective approach for predicting the dynamics of multiple SOC pools.

For this study, DSM techniques were applied to Truro, Nova Scotia, for modeling the dynamics of TOC and POXc. Specifically, the objectives were (i) to estimate the changes in TOC and POXc from 2015/2016 to 2019/2020 across the study area using DSM and machine learning models; (ii) to evaluate the performance of various environmental variables derived from multiple sources for predicting TOC and POXc; and (iii) to understand the impacts of agricultural management on the spatiotemporal variation of TOC and POXc.

2. Methods

This study required the acquisition and preprocessing of environmental variables from multiple sources; the calibration and validation of the machine learning models; the

application of variable importance analysis; and the spatial prediction of TOC and POXc. **Figure 1** shows the methodological framework of this study. The R statistical software (version 3.3.2, **R Core Team 2018**) was used for all modeling activities.

2.1. Study area

The study was conducted on a 37 km² area within the Truro region of Nova Scotia (45°22'53.4"N, 63°16'55.1"W; **Fig. 2**). Cropping in the study area comprises predominately pasture, forage, corn, and soybean production, with some additional small grain cereals to support the dairy farms and other non-ruminant livestock. The area is characterized by a cool, humid, temperate climate with mean annual precipitation ranging from 1058 to 1443 mm, while the mean temperature usually reaches as high as 19.2 °C in July and as low as -7.3 °C in January. The elevation ranges from 10 to 238 m (mean 25 m) above mean sea level. The surficial materials in the study area are predominantly moderately coarse-textured to moderately fine-textured glacial tills, derived from shale and sandstone (**Webb et al. 1991**). The soils are mainly Gleyed Podzols and Orthic Gleysols of the Pupash and Hansford associations, and Gleyed Gray Luvisols and Luvic Gleysols of the Queens soil association (**Webb et al. 1991**). Additionally, there are small inclusions of Regosolic soils (i.e., reclaimed marshland) in the Acadia soil association.

2.2. Soil sampling and laboratory analysis

In this study, soil samples (**Fig. 2**) were acquired at the 0–15 cm depth increment while ensuring that the collected soil samples represented the different types of agricultural production or crop types. Over the course of two separate sampling campaigns, the first set of samples ($n = 120$) was collected as part of an Atlantic regional soil health survey in 2019, and a second set of samples ($n = 150$) was collected in 2020. The samples acquired in 2020 were intended to improve the spatial distribution of the sample locations and were determined using the conditioned Latin hypercube sampling design (cLHS; **Minasny and McBratney 2006**). Here, cLHS is a maximally stratified random sampling technique that captures the variation in the feature space from a suite of input environmental covariates. This automated sample location selection procedure utilized the topographic indices as well as soil survey and annual crop inventory variables described in Section 2.3.1. A principal component analysis was also performed to minimize the number of covariates before using them in the *clhs* package (**Roudier 2014**) within the R software (version 3.3.2, **R Core Team 2018**).

All samples were analyzed at the Atlantic Soil Health Lab at Dalhousie University where TOC and POXc were measured. Samples were air dried, sieved to a <2 mm particle size, and for TOC, also finely ground using a roller grinder before analysis. To determine TOC, a combustion elemental analysis with a Vario Max Elemental Analyzer (Elementar, Langensfeld, Germany) was performed on the samples. POXc was measured with a Jenway 6505 UV-Vis spectrophotometer (Jenway, Staffordshire, UK) using a 0.2 M potassium permanganate (KMnO₄) solution (**Culman et al. 2012**).

2.3. Environmental variables

A total of 112 environmental variables were acquired from a combination of Landsat 8 satellite imagery, LiDAR-derived digital elevation model (DEM), soil survey data, and annual crop inventory (**Table 1**).

2.3.1. Landsat variables

A suite of soil and vegetation indices and image textural variables from Landsat images was captured at multiple time steps in 2015 and 2019. To represent the entire growing season, Landsat images were obtained from two pre-growing periods (i.e., April and May), one peak growing period (i.e., July or August), and one postgrowing period (i.e., late September). Landsat 8 Level-2 surface reflectance images (Path 8/Row 28 or Path 7/Row 29) were downloaded from the United States Geological Survey Earth Explorer satellite image inventory. Although obtaining cloud-free imagery captured at the desired time frame was a consistent challenge, it was ensured that images of the same growing season from 2015 to 2019 were captured within a period of 15 days to reduce the impacts of seasonal variability. However, the variability in annual crop types cultivated in 2015 and 2019 might have caused some discrepancies in our image analysis. The description of the Landsat variables used can be found in **Paul et al. (2020a)**. The Landsat images were also resampled from a 30 m spatial resolution to 10 m to match the spatial resolution of the topographic variables described below.

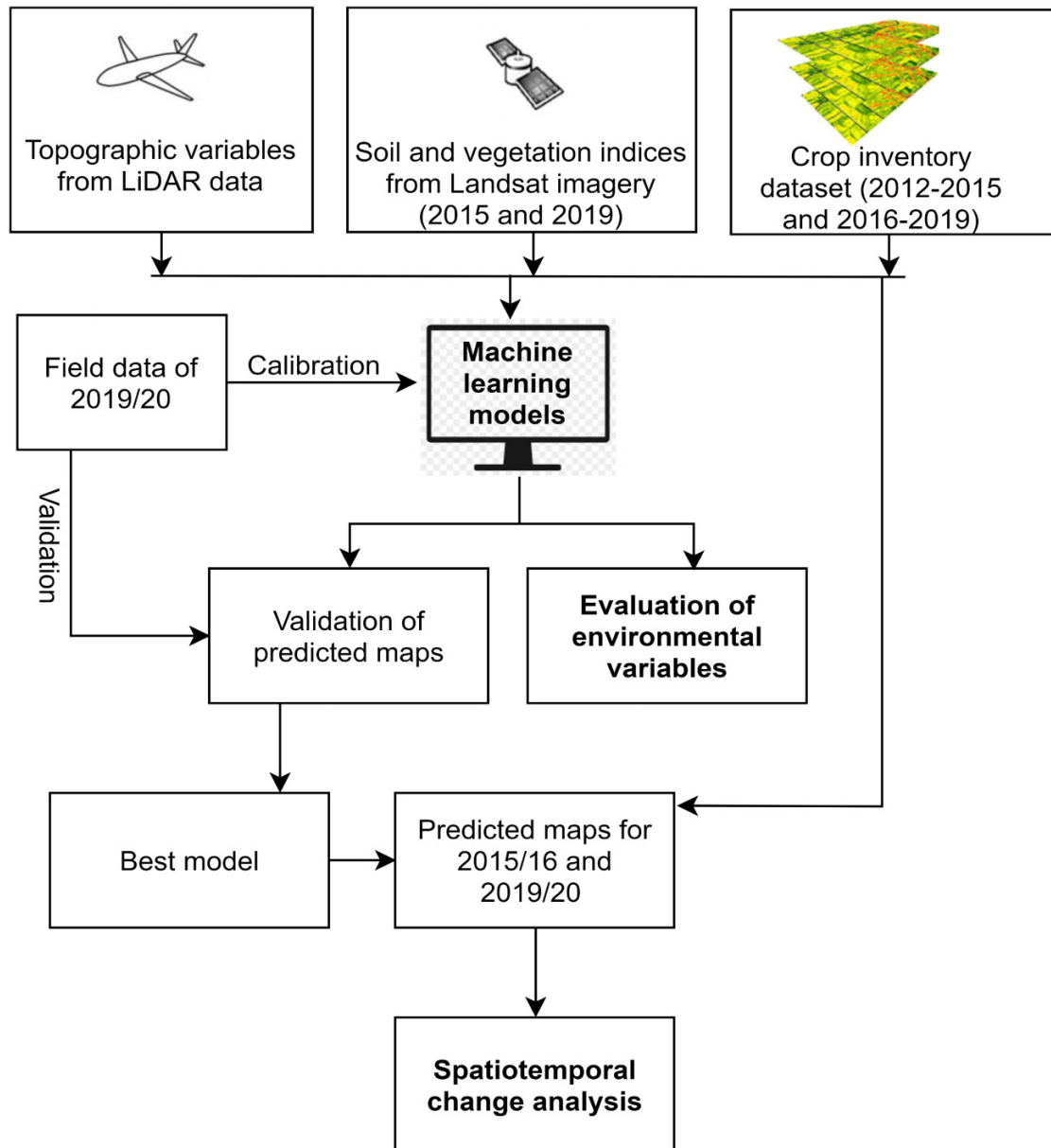
2.3.2. Topographic variables

A set of topographic variables from a LiDAR-derived DEM (10 m spatial resolution) to represent local (e.g., slope and aspect) and landscape-scale (e.g., multiscale topographic position index) morphometry, as well as hydrological characteristics (e.g., catchment area, valley depth, and wetness index) was generated. **Table 1** provides a complete list of the topographic variables produced for this analysis. The details of these indices may be found in http://www.saga-gis.org/saga_tool_doc/2.1.3/a2z.html and in **Blackford et al. (2020)**.

2.3.3. Soil survey data

Information on soil texture (i.e., percent of sand, silt, and clay) and SOC was extracted from the existing digitized version of the Soils of Colchester County, Nova Scotia, soil survey. The soil survey was originally produced at a 1:50 000 map scale and digitally distributed by the Canadian Soil Information Service. Here, the polygons were assigned the soil attribute values for the 0–15 cm depth increment of the dominant soil series within each map unit (i.e., the soil series that occupied >50% of the map unit area). Because the soil attributes were reported on a horizon basis for each soil series, the horizon information was harmonized to the common 0–15 cm depth increment using an equal-area quadratic spline function (**Bishop et al. 1999**). After assigning soil attribute

Fig. 1. Schematic diagram showing the general framework of the prediction of total organic carbon (TOC) and permanganate oxidizable carbon (POXc) in 2015/2016 and 2019/2020.



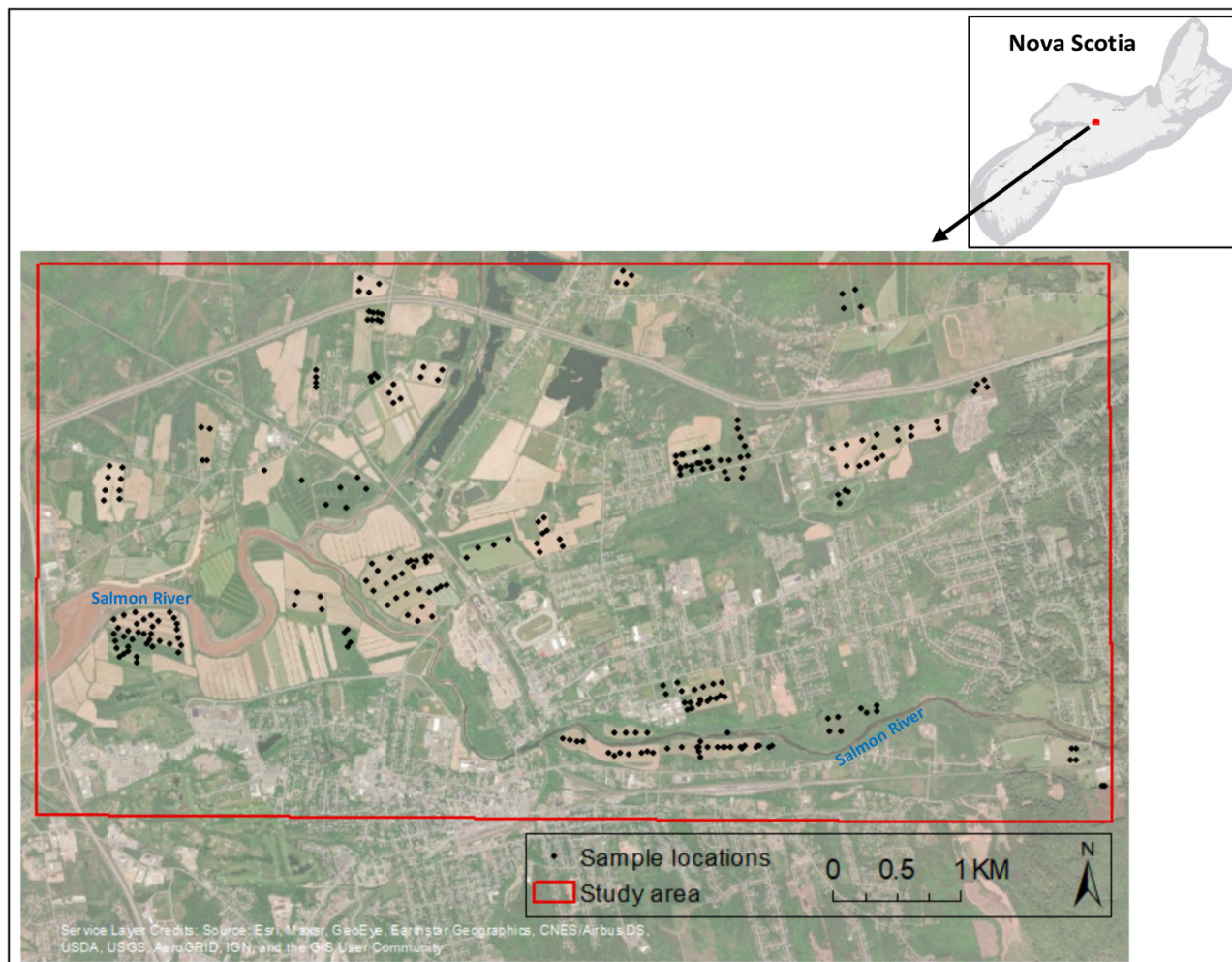
values to each map unit, the polygons were rasterized to a 10 m spatial resolution.

2.3.4. Annual crop inventory data

The AAFC annual crop inventory (30 m spatial resolution) was used to produce a crop rotation map for the study area. There were 72 classes in the AAFC crop inventory, which included different crop types and other land use categories (e.g., water, urban, and wetland). For the subsequent analysis, these classes were grouped into four categories: annual crops, pasture–grassland–forage, perennial crop (i.e., mainly berry), and non-agricultural land uses. Forages are typically annual crop in the study but were grouped with perennial pasture and grassland because of the spectral similarities of

forage lands, pastures, and grasslands in the Landsat imagery. A pixel-by-pixel correlation was performed on the annual crop inventories from 2012 to 2019 to create the crop rotation layer. For example, if a pixel had pasture–grassland–forage in 2012–2014 and was then converted to other annual crops in 2015, it was marked with a crop rotation class of “pasture–grassland–forage to other annual crop” for the SOC prediction in 2015/2016. A similar approach was performed on the rest of the crop inventory pixels across the study area. For the 2015/2016 predictions, the crop rotation layer was generated from the inventory maps of 2012–2015, while the 2019/2020 predictions used the 2016–2019 crop inventory maps. The resulting crop rotation layers were examined against the observed crop history data set and pixels were manually corrected where necessary. A majority filter with a 3×3 moving

Fig. 2. Study area in Truro, Nova Scotia, Canada (45°22'53.4"N, 63°16'55.1"W) with soil sample locations. The samples were collected in 2019 and 2020. The sampling scheme did not consider non-agriculture areas. The inset map shows the province of Nova Scotia.



window was applied to remove noise from both crop rotation layers. Finally, the crop rotation layers were resampled to 10 m spatial resolution.

2.4. Variance inflation factor analysis

Given that multiple covariates were calculated from the same input data (e.g., DEM), issues related to multicollinearity were handled using variance inflation factor (VIF) analysis. When carrying out VIF analysis, each predictor is regressed against all other predictors using a linear model and fitted via ordinary least squares regression, where the coefficient of determination (R^2) is calculated and used to calculate VIF in the following equation:

$$(1) \quad VIF_i = 1 / (1 - R_i^2)$$

where R_i^2 is the coefficient of determination from fitting a linear regression between the i^{th} independent variable and all other independent variables in the model. A VIF value is calculated for each predictor; the predictor with the highest VIF is removed; VIF is recalculated for all remaining

predictors, and the process is repeated until all predictors have a VIF that is below a threshold. Here, a $VIF = 5$ was used as the stopping criteria for the analysis, and the remaining predictors were retained for modeling processes (O'Brien 2007; Saurette 2022). The VIF analysis was performed using the `onsoilsurvey` package in R (Saurette 2021).

2.5. Predictions of TOC and POXc

The VIF analysis resulted in the retention of 37 continuous predictors, which were spatially intersected with the sample locations. It should be noted that the annual crop inventory data, which was a categorical variable, was not included in the VIF analysis as VIF is only suitable for continuous predictors. This study applied two machine learning models: Quantile Regression Forest (QRF; Meinshausen and Ridgeway 2006) and stochastic Gradient Boosting Model (GBM; Friedman 2002). QRF and GBM are tree-based ensemble models that combine the outcomes from a large number of trees to improve the prediction performance and thus, these models can handle a complex mixture of environmental variables required to represent the soil-landscape

Table 1. List of environmental variables used for digital soil mapping.

Variables	Characteristics
Landsat variables	
Normalized difference vegetation index	Vegetation
Soil adjusted vegetation index	Vegetation
Normalized difference moisture index	Vegetation
Soil brightness index	Soil
Normalized difference tillage index	Soil/vegetation
Clay minerals ratio	Soil
Bare soil index	Soil
Tasseled cap (greenness, brightness, and wetness)	Soil/vegetation
Image texture (homogeneity, contrast, and dissimilarity)	Soil/vegetation
Topographic variables (derived from LiDAR imagery)	
Aspect	Local relief
Downslope curvature	Local relief
General curvature	Local relief
Local curvature	Local relief
Local downslope curvature	Local relief
Local upslope curvature	Local relief
Maximum curvature	Local relief
Mid-slope position	Local relief
Minimum curvature	Local relief
Normalized height	Local relief
Plan curvature	Local relief
Profile curvature	Local relief
Real surface area	Local relief
Slope	Local relief
Standardized height	Local relief
Tangential curvature	Local relief
Terrain ruggedness index	Local relief
Terrain surface concavity	Local relief
Terrain surface convexity	Local relief
Terrain surface texture	Local relief
Topographic negative openness	Local relief
Topographic positive openness	Local relief
Total curvature	Local relief
Upslope curvature	Local relief
Upslope height	Local relief
Multiresolution index of ridge top flatness	Landscape relief
Multiresolution index of valley bottom flatness	Landscape relief
Multiscale topographic position index	Landscape relief
Valley depth	Landscape relief
Catchment area	Hydrology
Catchment slope	Hydrology
Modified catchment area	Hydrology
Topographic wetness index	Hydrology
AAFC inventory variables (derived from soil survey and crop inventory)	
Sand	Soil
Silt	Soil
Clay	Soil
Soil organic carbon	Soil
Annual crop inventory	Crop/management

variability (Wadoux et al. 2020). It should be noted that the QRF model is an extension of the Random Forest model (Breiman 2001); however, it can generate estimates of uncertainty using the marginal distributions from the terminal nodes of the model ensemble. Hence, the QRF model may be used to calculate the 90% prediction interval widths by subtracting the 5% lower limit from the 95% upper limit predictions. These models were implemented via the caret package in R (Kuhn 2009).

2.5.1. Accuracy metrics

The accuracy of the final prediction was tested using multiple metrics, including coefficient of determination (R^2), concordance correlation coefficient (CCC), root mean square error (RMSE), and normalized RMSE (nRMSE). nRMSE is the RMSE that has been normalized by the range of the observed data. These accuracy measures were derived using repeated cross-validation with 20 repeats of 10-fold cross-validation. The average accuracy metrics and their corresponding standard deviation were reported.

2.5.2. Recursive feature elimination

To minimize complexity, improve model interpretation, and remove irrelevant predictors, recursive feature elimination was carried out using the caret package in R (Kuhn 2009). Recursive feature elimination operates by fitting an initial model with all the predictors; calculating variable importance and accuracy metrics; removing the least important predictor; and repeating the process until one predictor remains. Here, the model that resulted in the lowest RMSE was selected as the final model, where the RMSE was calculated based on 20 repeats of 10-fold cross-validation. To evaluate the effects of variable reduction via the use of VIF to remove redundant predictors and the use of recursive feature elimination to remove irrelevant predictors, the machine learners were trained and validated using all predictors and the reduced predictor set.

2.5.3. Evaluating changes in TOC and POXc

The final models, generated from the 2019/2020 data, were then used to predict the TOC and POXc for 2015/2016 by substituting the dynamic variables (e.g., Landsat variables and the crop rotation layers). It should be noted, however, that the 2015/2016 predictions could not be validated due to the unavailability of field data. Next, the predicted outcomes of both time steps were compared pixel by pixel to extract the changes in TOC and POXc. The absolute and relative changes in TOC and POXc were calculated as shown in eqs. 2 and 3.

$$(2) \quad \text{Absolute change} = \left(\text{TOC} \frac{\text{g}}{\text{kg}} \text{ or POXc} \frac{\text{mg}}{\text{kg}} \text{ of year 2} \right) - \left(\text{TOC} \frac{\text{g}}{\text{kg}} \text{ or POXc} \frac{\text{mg}}{\text{kg}} \text{ of year 1} \right)$$

Table 2. Summary statistics of total organic carbon (TOC) and permanganate oxidizable carbon (POXc) derived from soil sample data collected in 2019/20.

Soil property	Minimum	Maximum	Mean	Standard deviation	Range
TOC (g/kg)	8.30	80.90	21.92	9.10	72.60
POXc (mg/kg)	190.04	994.17	575.09	166.40	804.13

$$(3) \quad \text{Relative change} = \frac{\left(\text{TOC } \frac{\text{g}}{\text{kg}} \text{ or POXc } \frac{\text{mg}}{\text{kg}} \text{ of year 2} \right) - \left(\text{TOC } \frac{\text{g}}{\text{kg}} \text{ or POXc } \frac{\text{mg}}{\text{kg}} \text{ of year 1} \right)}{\left(\text{TOC } \frac{\text{g}}{\text{kg}} \text{ or POXc } \frac{\text{mg}}{\text{kg}} \text{ of year 1} \right)} \times 100$$

Finally, the TOC and POXc dynamics (2015/2016 to 2019/2020) against the changes in agricultural management practices were evaluated using a pixel-by-pixel comparison. For this purpose, the crop rotation layers were employed to identify the TOC and POXc changes in response to changes in agricultural management.

3. Results and discussion

3.1. Summary statistics

The TOC and POXc concentrations in the study area varied considerably while the mean POXc was about 3% of the mean TOC (Table 2). Nevertheless, the within-field variability for TOC and POXc was low. The variability in TOC and POXc may be due to the heterogeneous cropping and soil management practices. The highest values were mainly obtained from fields consistently managed as pasture, while low values were distributed across different cropping management. Consistent pasture management was reported to associate with higher SOC concentration in many previous studies, whereas soils on annual cropping exhibited lower SOC. For example, a long-term case study conducted by Martens et al. (2004) found continuous pasture retained 25% more SOC compared to annually cropped fields.

3.2. Prediction accuracy of QRF and GBM models

The prediction accuracy varied across the soil properties and machine learning models (Fig. 3). However, the prediction accuracy was similar between the models with all variables and with the top variables extracted from recursive feature elimination. For instance, the R^2 and CCC of QRF-TOC model with all variables were only 2% and 1%, respectively, better than the model with the top variables, while the nRMSE value was improved by 4% for the latter model. Similar results were observed for QRF-POXc model, where R^2 increased by 3%, CCC decreased by 1%, and nRMSE increased by 2% using the top variables. The predictions using GBM also resulted in similar findings for both TOC and POXc.

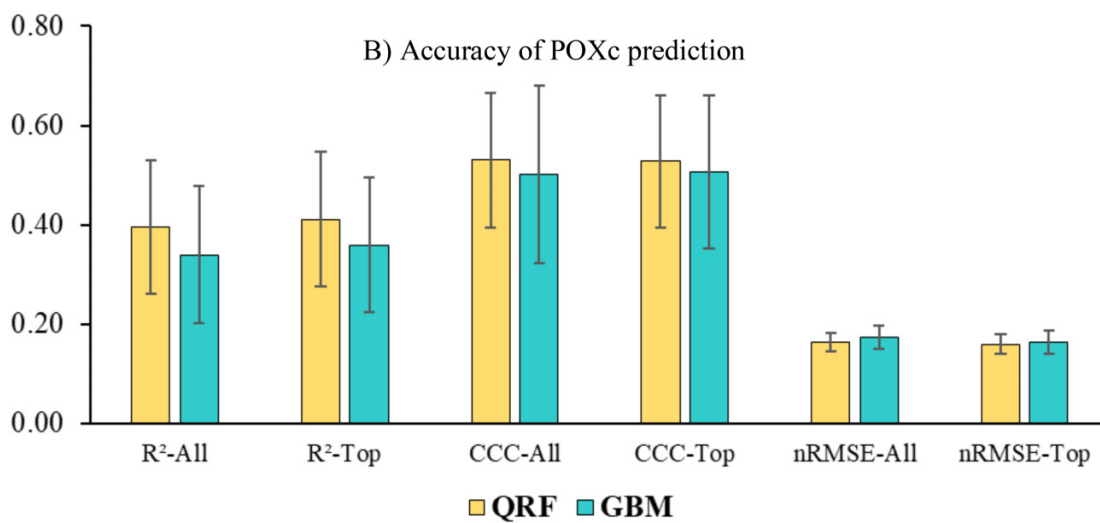
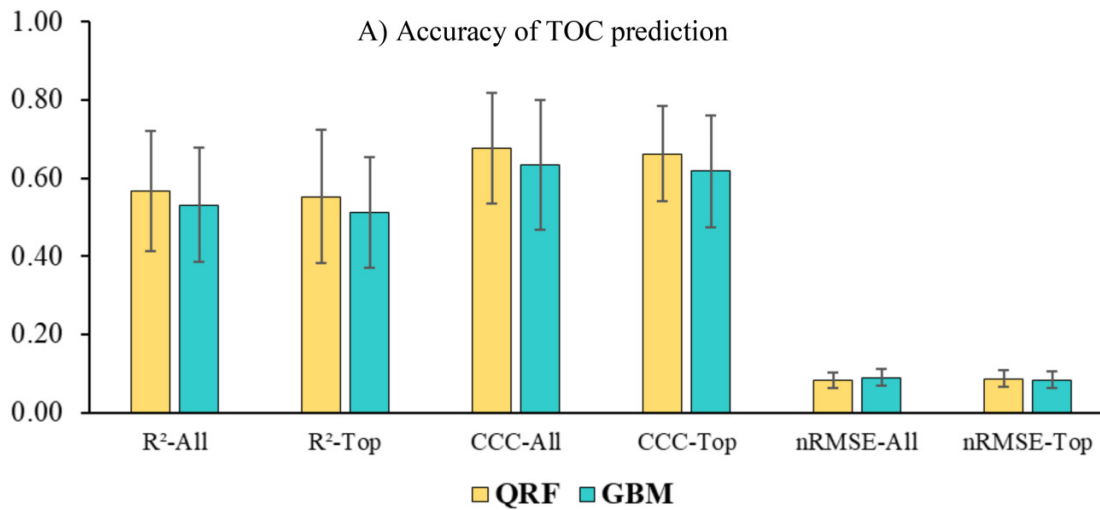
QRF and GBM models performed equally well for TOC and POXc predictions, but QRF model was marginally better for some accuracy measures. For example, the R^2 , CCC, and

nRMSE of QRF-TOC prediction (with top variables) were 8%, 7% and 3% better than those of the GBM-TOC prediction. Likewise, QRF model achieved better R^2 (increase by 16%) and CCC (increase by 4%) but slightly worse nRMSE (decrease by 2%) compared to GBM model for predicting POXc (with top variables). Overall, the TOC prediction was more accurate than the POXc prediction in terms of all accuracy measures and using both QRF and GBM models, such as, CCC of 67% for the best TOC model compared to CCC of 53% for the best POXc model.

The accuracy of the TOC predictions was similar or better compared to other DSM analyses conducted in various landscapes. The studies conducted by Kempen et al. (2019) and Song et al. (2020), for example, acquired R^2 of 50%–60% for SOC prediction in landscape-scale assessments in Tanzania and China, respectively. These accuracies were in agreement with the results achieved from QRF and GBM models for TOC prediction. In addition, other previous studies had SOC predictions that had $R^2 < 40\%$ (Adhikari et al. 2019; Mulder et al. 2016; Dharumarajan et al. 2021) and were notably less accurate than this study. However, most of these SOC predictions were performed at national or larger regional scales with far more spatial complexity in the landscapes. Although our study landscape was comprised of widely heterogeneous soil textures, it was uniform in terms of climate, topographic variability, and agricultural management practices. Moreover, this study utilized a dense sampling network for model parameterization (i.e., $n = 270$ for 37 km² area) and an extensive suite of environmental variables from various sources. All these factors may have resulted in the accurate prediction of TOC. In comparison, the POXc predictions, however, were relatively less accurate compared to TOC. To the best of our knowledge, previous studies that predicted POXc were not available in the literature and thus, the results could not be compared.

Despite the relatively poorer performance of the POXc models, the prediction accuracies were still in the accepted range for DSM analysis as reported in the literature for other related soil properties (Forkuor et al. 2017; Khaledian and Miller 2020). The highly volatile nature of POXc could be a reason for the relatively weaker performance of the POXc model. In addition, our field observation included data from two different years to provide a large enough data set for DSM prediction, which might have contributed to some mi-

Fig. 3. Accuracy of (A) total organic carbon (TOC) prediction and (B) permanganate oxidizable carbon (POXc) prediction in terms of coefficient of determination (R^2), concordance correlation coefficient (CCC), and normalized root mean square error (nRMSE) using Quantile Regression Forest (QRF; Meinshausen and Ridgeway 2006) and stochastic Gradient Boosting Model (GBM; Friedman 2002). The error bars in (A) and (B) show the standard deviation derived from 10-fold cross-validation with 20 repeats. (C) Observed values were plotted against predicted values for both soil properties.



C) Observed vs predicted (based on QRF model) TOC and POXc

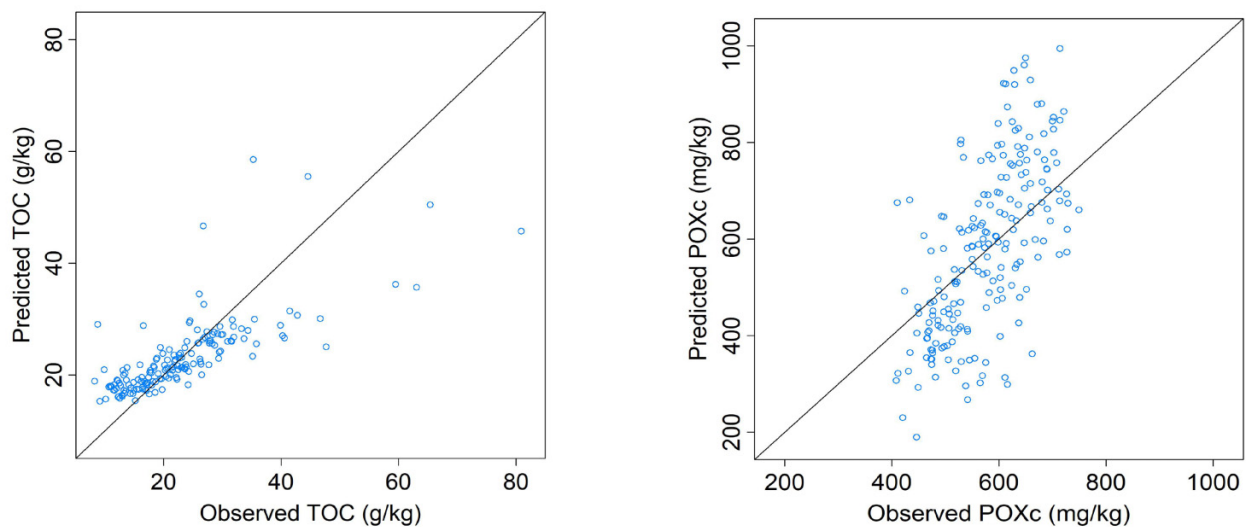
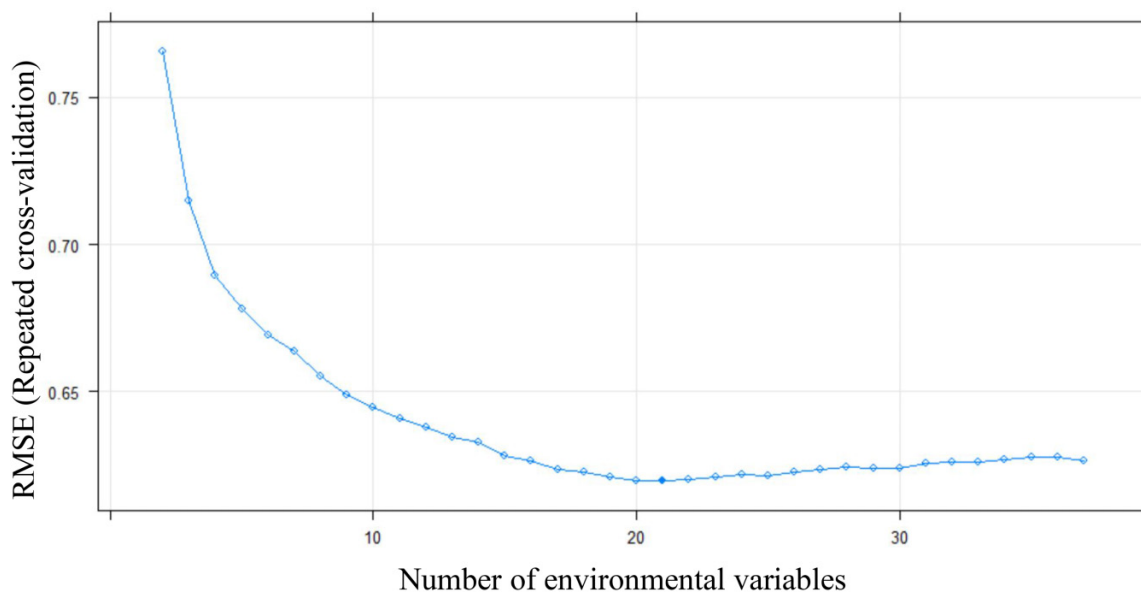


Fig. 4. Recursive feature elimination for predicting total organic carbon using Quantile Regression Forest showing prediction accuracy plateaued after 21 variables. Prediction accuracy is represented as root mean square error (RMSE) from 20-repeat, 10-fold cross-validation.



nor uncertainties in POXc prediction. Since the samples were collected within the timeframe of a year and there were no major changes in management practices, we expected that the temporal changes in POXc during the 2019/2020 period were negligible. The integration of more on-farm management information, such as tillage, fertilizer, and manure applications, could potentially improve the prediction accuracy of POXc.

Lastly, it was observed that both QRF and GBM models performed equally well in our analysis, which was different from the findings of Wang et al. (2018) and Paul et al. (2020a) who reported better performance of QRF compared to GBM. Yet, model accuracy is highly case specific and largely relies on the variability of feature space, environmental conditions, and soil property to be modeled (Ließ et al. 2016).

3.3. Variable reduction and importance

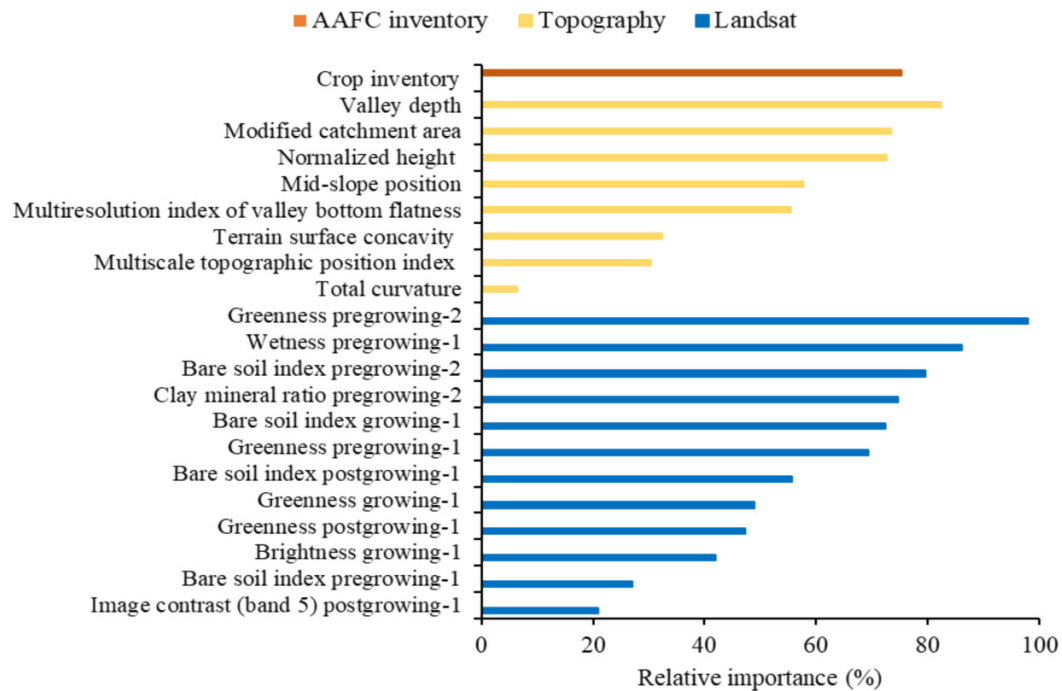
The VIF and recursive feature elimination analyses substantially reduced the number of environmental variables required for the optimum model performance. The VIF analysis only retained 37 of the 111 continuous environmental variables. These 37 variables along with the crop inventory categorical variable were then used for recursive feature elimination using both the QRF and GBM models. This process further reduced the variable set by 40%–45% for predicting TOC and POXc. For example, the QRF-TOC model finally retained 21 variables as top predictors since adding more variables after this point did not equally improve the model accuracy (Fig. 4). The top variables were mainly dominated by topographic and Landsat indices. The Landsat variables derived from the pre- and postgrowing season images were found to be among the strongest predictors for all models (Fig. 5). During these seasons, soils in the annual crop fields are typically exposed and can also be differentiated from perennial crop

fields. Particularly, the greenness and wetness of the surface and the bare soil index powerfully contributed to the prediction using both machine learning models. For example, the greenness and wetness were the most dominant predictors for QRF-TOC and QRF-POXc models (Fig. 5).

The top topographic predictors for all models included variables representing local relief, landscape relief, and hydrology. The valley depth and modified catchment area, for example, were the most dominant topographic predictors for QRF-TOC and QRF-POXc models (Fig. 5). The crop inventory variable strongly contributed to all predictions, however, the soil clay content was found as a top predictor only for POXc. Although the topographic covariates were derived from a high-resolution LiDAR DEM, most of the topographic covariates, except a few, contributed weakly or moderately to predicting TOC and POXc. Such contribution by the topographic covariates may be related to the relatively homogeneous terrain of the study area. This outcome was in contrast to Grimm et al. (2008), who identified topographic covariates as the most important predictor of SOC in a study conducted on Barro Colorado Island, Panama. Interestingly, soil and vegetation indices derived from time-series Landsat images, as well as the crop inventory data, which was also produced from Landsat and other coarser-resolution remote sensing images, performed remarkably well for predicting TOC and POXc in our analysis. These remote sensing indices, particularly those derived from pre- and postgrowing season images, could effectively differentiate between various agricultural management and land use practices and hence, contributed strongly to the prediction of TOC and POXc. Rial et al. (2017) and Lamichhane et al. (2019) also concluded that remote sensing-derived agricultural management and land use information could be important predictors for SOC dynamics at local scales. Furthermore, the better performance

Fig. 5. Environmental variables used for the final prediction of (A) total organic carbon (TOC, $n = 21$) and (B) permanganate oxidizable carbon (POXc, $n = 22$) predictions using Quantile Regression Forest. The variables are of three different categories—(i) topographic variables—derived from the digital elevation model, (ii) AAFC inventory variables—derived from Agriculture and Agri-Food Canada crop rotation (2012–2019) and soil survey data, and (iii) Landsat variables—derived from Landsat imagery of 2015 and 2019 for pregrowing, growing, and postgrowing seasons.

A)



B)

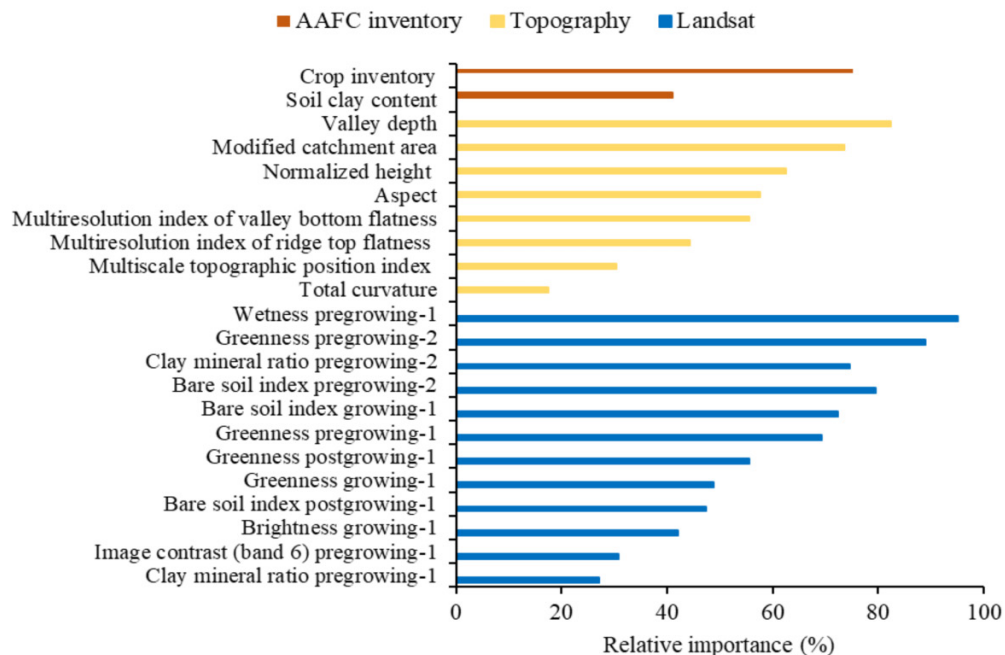
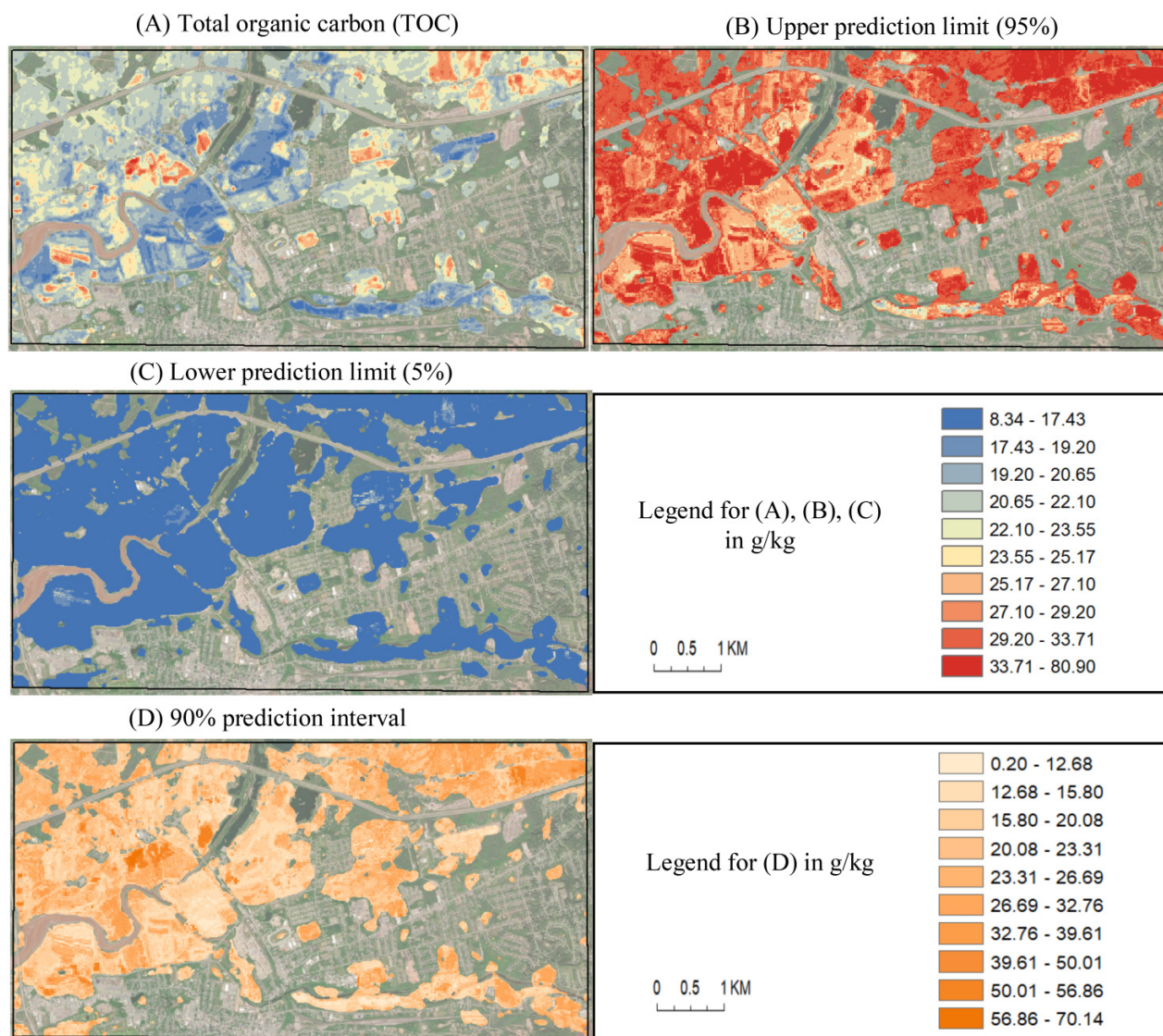


Fig. 6. (A) Distribution of total organic carbon (TOC) across the study area in 2019/2020. (B) Upper and (C) lower prediction limits derived using Quantile Regression Forest (Meinshausen and Ridgeway 2006). (D) 90% prediction interval or uncertainty calculated from upper and lower prediction limits. Nonagriculture areas were not predicted for TOC. (Base map sources: Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.)



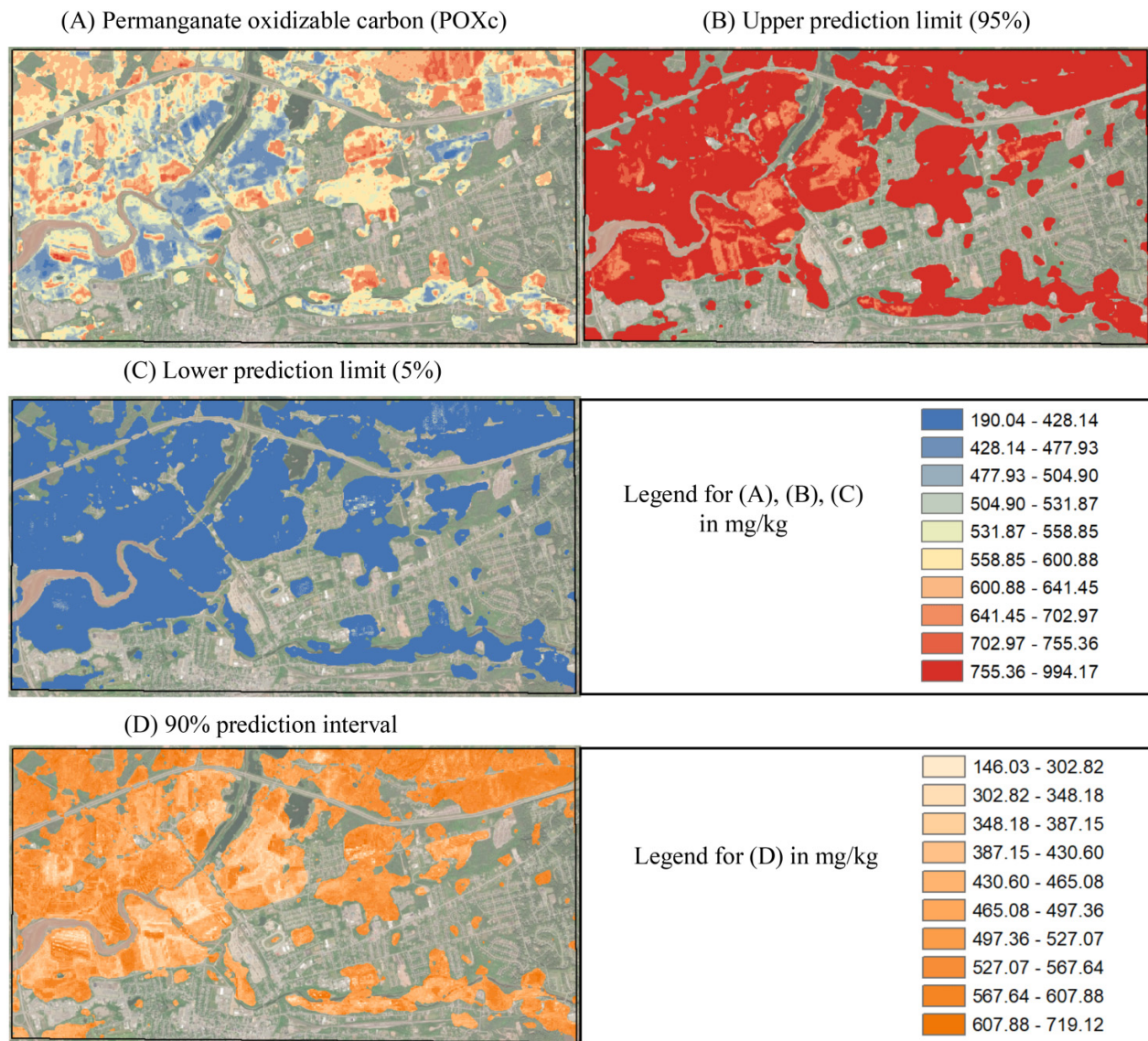
of the pre- and postgrowing season Landsat indices was supported by the findings of Bartholomeus et al. (2011) and Paul et al. (2020b) who identified that surface reflectance of bare soil during these seasons was highly correlated to SOC concentration. Landsat imagery captured during these seasons can also efficiently discriminate between annual and perennial vegetation, which likely contributes to better prediction performance in heterogeneous agricultural landscapes (Paul et al. 2020b).

3.4. Spatial distribution of TOC and POXc 2019/2020

The study area exhibited relatively higher variability for POXc compared to TOC (Figs. 6 and 7). Overall, the POXc

concentrations were about 2%–3% of the TOC concentrations in the study area. In 2019/2020, TOC concentration ranged from 8 to 80 g/kg, with a mean of 21 g/kg, while POXc concentration varied from 190 to 994 mg/kg, with a mean of 563 mg/kg. The farm fields in the northern portion of the area showed comparatively higher TOC concentration; however, high POXc values were distributed across the study area. The farms with the lowest TOC and POXc were mostly concentrated on the eastern bank of the Salmon River. An uncertainty analysis was conducted for TOC (Fig. 6B–6D) and POXc (Fig. 7B–7D) predictions using the best-performing QRF model. Figure 6D shows a map of the TOC 90% prediction interval width (i.e., uncertainty), which ranged from 0.20 to 70 g/kg for the whole study area. Similarly, Fig. 7D exhibits POXc prediction interval width that varies from 146

Fig. 7. (A) Distribution of permanganate oxidizable carbon (POXc) across the study area in 2019/20. (B) Upper and (C) lower prediction limits derived using Quantile Regression Forest (Meinshausen and Ridgeway 2006). (D) 90% prediction interval or uncertainty calculated from upper and lower prediction limits. Nonagriculture areas were not predicted for POXc. (Base map sources: Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.)

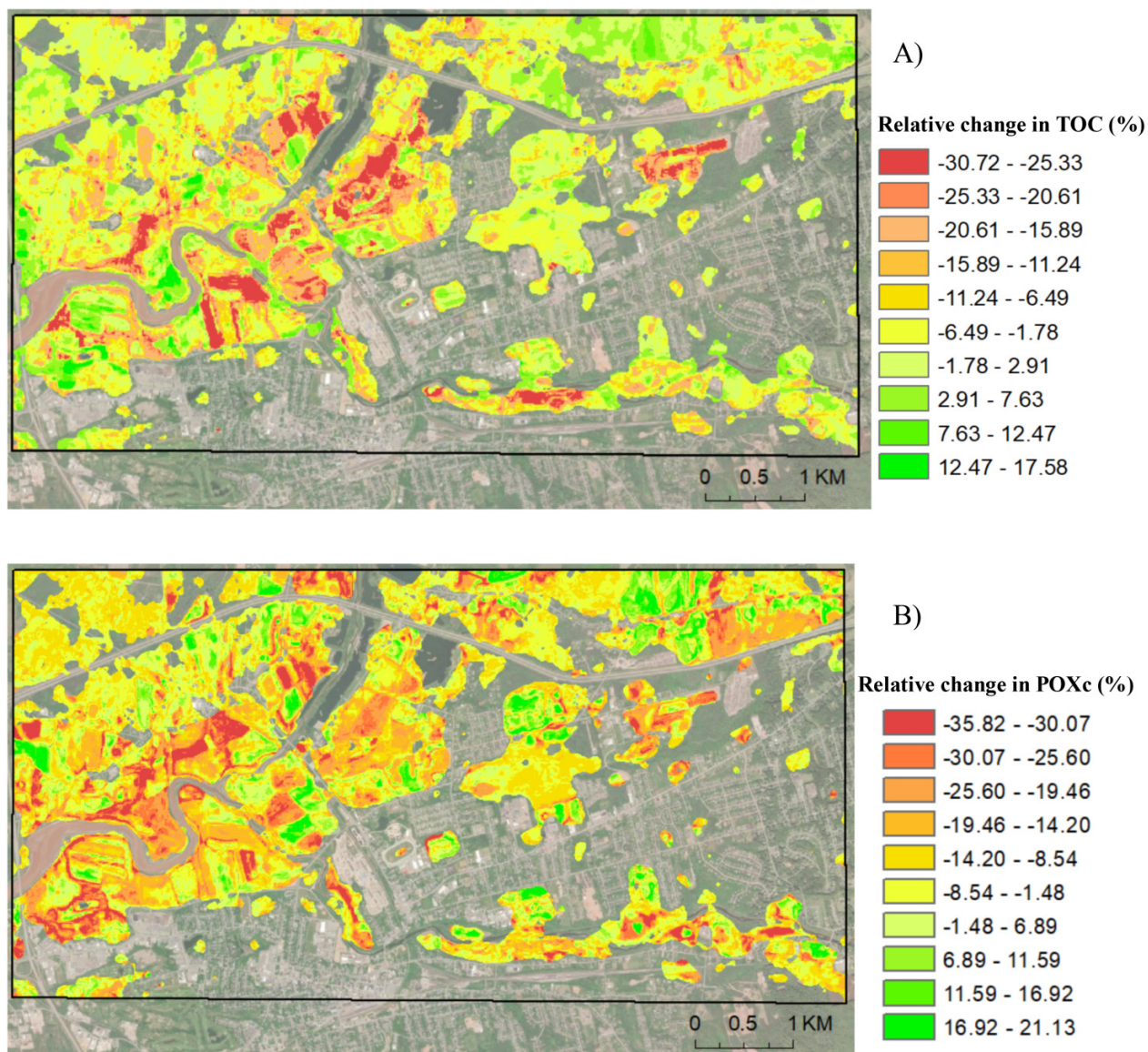


to 719 mg/kg at the 90% confidence level across the study area. The prediction uncertainty for TOC and POXc showed a similar spatial pattern; however, POXc prediction had relatively higher uncertainty. In addition, the crop fields along the riverbanks were predicted with greater uncertainties, especially for TOC. Interestingly, areas with higher TOC and POXc values were generally associated with high prediction uncertainties.

The fields that were consistently either on pasture or grassland or forage production had the highest mean TOC of 22 g/kg in 2019/2020. For the fields with rotation of grassland–forage and other annual crops, the mean TOC was found to be 19 g/kg in 2019/2020, while the fields consistently on other annual crops had a mean TOC of 18 g/kg. A similar

trend was detected for POXc concentration in 2019/2020 with the highest mean value (i.e., 580 mg/kg) associated with the fields constantly used for pasture, grassland, and forage production. The fields on the rotation of grassland/forage and other annual crops had a mean POXc of 510 mg/kg while the fields that were always used for annual crops had a slightly higher mean of 530 mg/kg. A small fraction of crop fields were on perennial berry production and had a mean TOC and POXc of 21 g/kg and 540 mg/kg, respectively. Some substantially high and low TOC and POXc values were observed for all these crop types; however, the lowest TOC and POXc values were associated with the annual crop fields and the highest values were concentrated in the fields consistently on pasture.

Fig. 8. Relative changes in (A) total organic carbon (TOC) and (B) permanganate oxidizable carbon (POXc) from 2015/2016 to 2019/2020. Non-agriculture areas were not predicted for TOC and POXc. (Base map sources: Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community.)



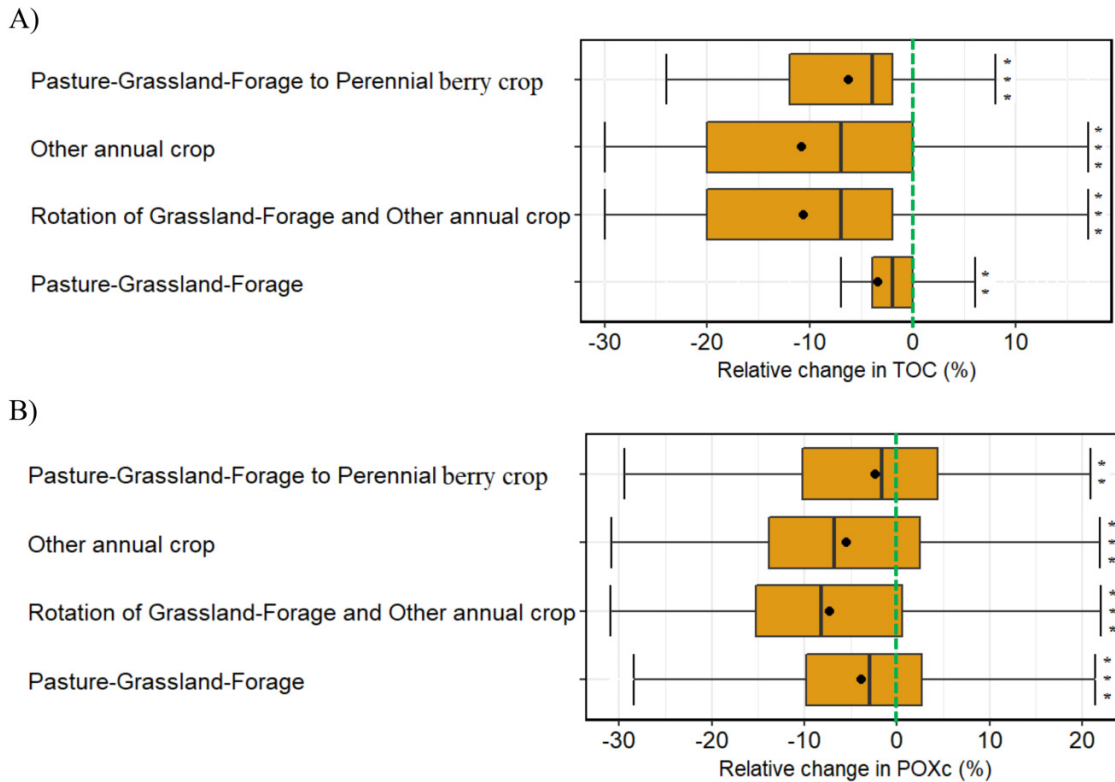
3.5. Changes in TOC and poxc (2015/2016 to 2019/2020)

This study area exhibited highly variable SOC dynamics from 2015/2016 to 2019/2020 with gains, losses, and no change in TOC and POXc concentrations. As expected, POXc was more dynamic compared to TOC during the study period where 65% of the area observed absolute losses in POXc, while absolute gains were detected in 33% of the area. Therefore, 98% of the study area exhibited some degree of absolute change in POXc; however, only 27% of the area (loss in 22% and gain in 5%) experienced an absolute change in TOC. Furthermore, the relative change in TOC (2015/2016 to 2019/2020) ranged from -30% to 17%, while for POXc, it varied from -35% to 21% (Fig. 8). For both soil properties, the highest negative changes were mainly observed along the banks of the watercourses. The crop fields with high gains in

TOC and POXc were scattered over the study area; however, some fields in the northeast corner of the area exhibited considerable gains in TOC and POXc. It is, however, important to note that our study could not validate the predictions for 2015/2016 year due to the unavailability of field data, which might have incorporated some degree of uncertainty in the change analysis.

This study identified four dominant change categories for agricultural management (2015/2016 to 2019/2020): (i) consistently managed for either pasture or grassland or forage; (ii) consistently managed for annual crops other than forage; (iii) rotation of grassland-forage and other annual crops; and (iv) conversion of pasture-grassland-forage to perennial berry crop (see Section 2.3.4 for the description of agricultural management and crop rotation classes). 75% of the study area was constantly either on pasture or grassland or forage,

Fig. 9. Relative changes in (A) total organic carbon (TOC) and (B) permanganate oxidizable carbon (POXc) from 2015/2016 to 2019/2020 for different agricultural management categories. Boxplots showing the first quartile, median (bar), third quartile, and mean (circles). Error bars represent maximum and minimum values. Significant test results are indicated by ** ($p < 0.01$) and *** ($p < 0.001$).



while 8% of the area was managed consistently for annual crops (other than forage), 16% for rotation of grassland–forage and other annual crops, and only 1% was converted to perennial berry crop. An overall decline was predicted for mean TOC and POXc for all agricultural management categories whether they were changed or not during the study period (Fig. 9). However, the crop fields that were consistently managed for pasture–grassland–forage observed the lowest decline in predicted mean TOC and POXc (i.e., 3% and 4%, respectively). The fraction of the croplands converted from pasture–grassland–forage to perennial berry crop faced losses of 6% and 2% for mean TOC and POXc, respectively. On the contrary, the predicted mean TOC declined by slightly over 10% for the fields consistently cultivated for annual crops (other than forage) and the fields in rotation with grassland–forages and other annual crops. These two agricultural management categories, however, experienced relatively smaller losses in mean predicted POXc (6% and 8%, respectively). Interestingly, substantial increases in TOC and POXc were also observed in some areas for all management categories. For example, some annual crop fields exhibited TOC increases as high as 17% during the study period. Similar results were also predicted for POXc under different management categories. We applied a statistical test—Kruskal–Wallis test (Vargha and Delaney 1998) to evaluate whether the changes of TOC and POXc differ between different agricultural management

categories. It was observed that TOC and POXc changes were significantly different between the management categories ($p < 0.01$ or $p < 0.001$, Fig. 9).

This study successfully applied the DSM approach for tracking changes in SOC pools within a 5-year period (2015/2016 to 2019/2020) across the study area. This spatially explicit change analysis allowed detection of key areas, where SOC was being lost, gained, or remained unchanged. Our analysis also identified specific agricultural management that was likely responsible for these changes. Overall, our analysis predicted a mean absolute TOC change of $-0.15 (\pm 0.09)$ g/kg/year across the study area while POXc experienced a mean absolute change of $-5.77 (\pm 11.48)$ mg/kg/year. The larger decline was found in annual crop fields, which might be attributed to intensive agricultural management, such as heavy tillage or aggressive low residue crop (soybeans and corn silage) production (Grandy et al. 2002; Haddaway et al. 2017). Such management practices might also contribute to overall SOC loss in the crop fields rotated between grassland–forage and other annual crops. However, a marginal TOC and POXc loss were also observed in the crop fields constantly on pasture, grassland, or forage production. A variety of reasons may have contributed to these losses in this management category, such as improper nutrient management or organic amendment, animal stocking rate, and pattern. Studies reported that excess nitrogen input to the soil as manure

or synthetic fertilizer may enhance microbial respiration or even suppress the microbial population and hence, result in loss of SOC (McCarty and Meisinger 1997; Christopher and Lal 2007). Over-application of nitrogen is a common challenge in pasture and forage management and this might have contributed to SOC losses in our study area as well. In addition, Franzluebbers and Stuedemann (2010) found that haying or high-intensity cattle stocking when grazing pastures could lead to SOC decline in the topsoil layers over 12 years.

Although POXc exhibited strong responses to changes in management practices in our short-term analysis, Pérez-Guzmán et al. (2021) reported highly variable POXc responses between management treatments in Ontario, Canada, and highlighted the use of more sensitive indicators, such as the combined enzyme assay technique. In addition, our study could not differentiate between pasture, grassland, and forage fields due to the inadequate spatial resolution of the satellite imagery and lack of reliable agricultural land use information. Since management practices can substantially vary between these production systems, having more precise information could help determine the specifics of SOC dynamics in these production systems. Given that, three-fourths of the study area were included in the consistent pasture-grassland-forage category, careful identification of SOC gains and losses within this category would be highly important for devising an informed SOC management strategy for this region. Although we identified that the changes in TOC and POXc between different management categories were statistically significant, investigating the magnitude of these statistical differences for each pixel across the study area was out of scope for our analysis, mainly due to data insufficiency. Incorporating such pixel-level statistical tests in future change analysis would reinforce the outcomes of DSM-based spatiotemporal SOC modeling.

4. Conclusions

This study presented a cost-effective approach for modeling spatiotemporal dynamics of multiple SOC pools in agricultural landscapes. The application of freely available time-series satellite images, high-resolution LiDAR data, and historical crop inventory information produced an accurate prediction of TOC and POXc at multiple time steps. This study also demonstrated the first application of the AAFC crop inventory data set for DSM of multiple SOC pools. The highly accurate soil maps generated in this study could serve as a baseline for modeling greenhouse gas emissions and farm-level carbon budgeting across the study area, but with a caveat that the unavailability of field observation from 2015/2016 year might have contributed to some uncertainties in the change analysis. In addition, future analysis should consider incorporating additional on-farm management information for a more accurate prediction of SOC dynamics, particularly for identifying changes in POXc. The study approach should also be tested in other agricultural landscapes and compared against robust process-based modeling techniques to verify and enhance the utility.

Acknowledgements

Funding for this research was provided by the Natural Sciences and Engineering Research Council, and the Agricultural Greenhouse Gas Program of Agriculture and Agri-Food Canada.

Article information

History dates

Received: 24 January 2022

Accepted: 25 July 2022

Accepted manuscript online: 27 July 2022

Version of record online: 16 January 2023

Notes

This paper is part of a Collection entitled “Advances in Soil Survey & Classification in Canada”.

Copyright

© 2022 Authors B. Heung and D. Lynch and Her Majesty the Queen in Right of Canada, represented by the Minister of Agriculture and Agri-Food Canada. This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Data availability

The primary research data are held by the authors and may be requested from the corresponding author.

Author information

Author ORCIDs

Siddhartho S. Paul <https://orcid.org/0000-0001-5243-8416>

Derek H. Lynch <https://orcid.org/0000-0001-8279-2934>

Author contributions

Conceptualization: SSP, BH, DHL

Formal analysis: SSP, BH

Funding acquisition: DHL

Supervision: DHL

Writing – original draft: SSP, BH, DHL

Writing – review & editing: SSP, BH, DHL

Competing interests

The authors declare there are no competing interests.

References

- Adhikari, K., Owens, P.R., Libohova, Z., Miller, D.M., Wills, S.A., and Nemecek, J. 2019. Assessing soil organic carbon stock of wisconsin, USA and its fate under future land use and climate change. *Sci. Total Environ.*, **667**: 833–845. Elsevier. doi:[10.1016/j.scitotenv.2019.02.420](https://doi.org/10.1016/j.scitotenv.2019.02.420). PMID: [30852437](https://pubmed.ncbi.nlm.nih.gov/30852437/).
- Bartholomeus, H., Kooistra, L., Stevens, A., van Leeuwen, M., van Weesmael, B., Ben-Dor, E., and Tychon, B. 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.*, **13**: 81–88. Elsevier.

- Bishop, T.F.A., McBratney, A.B., and Laslett, G.M. 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**: 27–45. Elsevier. doi:[10.1016/S0016-7061\(99\)00003-8](https://doi.org/10.1016/S0016-7061(99)00003-8).
- Blackford, C., Heung, B., Baldwin, K., Fleming, R., Hazlett, P. Morris, D.M., et al. 2020. Digital soil mapping workflow for forest resource applications: a case study in the hearst forest, ontario. *Can. J. For. Res.*, **51**: 1–66. NRC Research Press.
- Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S. Nield, S.J., et al. 2008. Landsat spectral data for digital soil mapping. In *Digital soil mapping with limited data*. Springer, pp. 193–202.
- Breiman, L. 2001. Random Forests. *Machine learning*, **45**(1): 5–32. doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- Christopher, S.F., and Lal, R. 2007. Nitrogen management affects carbon sequestration in North American cropland soils. *CRC. Crit. Rev. Plant Sci.*, **26**: 45–64. Taylor & Francis. doi:[10.1080/07352680601174830](https://doi.org/10.1080/07352680601174830).
- Clearwater, R.L., Martin, T., and Hoppe, T.A. 2016. Environmental sustainability of Canadian agriculture: agri-environmental indicators report series (report #4). Agriculture and Agri-Food Canada, Ottawa, Canada.
- Culman, S.W., Snapp, S.S., Freeman, M.A., Schipanski, M.E., Beniston, J. Lal, R., et al. 2012. Permanganate oxidizable carbon reflects a processed soil fraction that is sensitive to management. *Soil Sci. Soc. Am. J.*, **76**: 494–504. Wiley Online Library. doi:[10.2136/sssaj2011.0286](https://doi.org/10.2136/sssaj2011.0286).
- Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Vasundhara, R. Kumar, K.S.A., et al. 2021. Digital soil mapping of soil organic carbon stocks in western ghats, South India. *Geoderma Reg.*, **25**: e00387. Elsevier. doi:[10.1016/j.geodrs.2021.e00387](https://doi.org/10.1016/j.geodrs.2021.e00387).
- Dimassi, B., Guenet, B., Saby, N.P.A., Munoz, F., Bardy, M., Millet, F., and Martin, M.P. 2018. The impacts of CENTURY model initialization scenarios on soil organic carbon dynamics simulation in French long-term experiments. *Geoderma*, **311**: 25–36. Elsevier. doi:[10.1016/j.geoderma.2017.09.038](https://doi.org/10.1016/j.geoderma.2017.09.038).
- Fathizad, H., Ardakani, M.A.H., Heung, B., Sodaiezhadeh, H., Rahmani, A. Fathabadi, A., et al. 2020. Spatio-temporal dynamic of soil quality in the central iranian desert modeled with machine learning and digital soil assessment techniques. *Ecol. Indic.*, **118**: 106736. Elsevier. doi:[10.1016/j.ecolind.2020.106736](https://doi.org/10.1016/j.ecolind.2020.106736).
- Fathizad, H., Taghizadeh-Mehrjardi, R., Hakimzadeh Ardakani, M.A., Zeraatpisheh, M., Heung, B., and Scholten, T. 2022. Spatiotemporal assessment of soil organic carbon change using machine-learning in arid regions. *Agronomy*, **12**: 628. MDPI. doi:[10.3390/agronomy12030628](https://doi.org/10.3390/agronomy12030628).
- Fatholouloumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., and Biswas, A. 2020. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: a case study in Iran. *Sci. Total Environ.*, **721**: 137703. Elsevier. doi:[10.1016/j.scitotenv.2020.137703](https://doi.org/10.1016/j.scitotenv.2020.137703). PMID: [32172111](https://pubmed.ncbi.nlm.nih.gov/32172111/).
- Fisette, T., Davidson, A., Daneshfar, B., Rollin, P., Aly, Z., and Campbell, L. 2014. Annual space-based crop inventory for Canada: 2009–2014. In *2014 IEEE Geoscience and Remote Sensing Symposium*. IEEE, pp. 5095–5098.
- Fisette, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B. Filyer, P., et al. 2013. AAFC annual crop inventory. In *2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*. IEEE, pp. 270–274.
- Forkuor, G., Hounkpatin, O.K.L., Welp, G., and Thiel, M. 2017. 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**: e0170478. Public Library of Science San Francisco, CA USA. doi:[10.1371/journal.pone.0170478](https://doi.org/10.1371/journal.pone.0170478). PMID: [28114334](https://pubmed.ncbi.nlm.nih.gov/28114334/).
- Franzluuebbers, A.J., and Stuedemann, J.A. 2010. Surface soil changes during twelve years of pasture management in the Southern piedmont USA. *Soil Sci. Soc. Am. J.*, **74**: 2131–2141. Wiley Online Library. doi:[10.2136/sssaj2010.0034](https://doi.org/10.2136/sssaj2010.0034).
- Friedman, J.H. 2002. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **38**: 367–378. Elsevier. doi:[10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- Grandy, A.S., Porter, G.A., and Erich, M.S. 2002. Organic amendment and rotation crop effects on the recovery of soil organic matter and aggregation in potato cropping systems. *Soil Sci. Soc. Am. J.*, **66**: 1311–1319. Wiley Online Library. doi:[10.2136/sssaj2002.1311](https://doi.org/10.2136/sssaj2002.1311).
- Grimm, R., Behrens, T., Märker, M., and Elsenbeer, H. 2008. Soil organic carbon concentrations and stocks on barro colorado island—digital soil mapping using random forests analysis. *Geoderma*, **146**: 102–113. Elsevier. doi:[10.1016/j.geoderma.2008.05.008](https://doi.org/10.1016/j.geoderma.2008.05.008).
- Haddaway, N.R., Hedlund, K., Jackson, L.E., Kätterer, T., Lugato, E. Thomssen, I.K., et al. 2017. How does tillage intensity affect soil organic carbon? A systematic review. *Environ. Evid.*, **6**: 30. Springer. doi:[10.1186/s13750-017-0108-9](https://doi.org/10.1186/s13750-017-0108-9).
- Huang, J., Hartemink, A.E., and Zhang, Y. 2019. Climate and land-use change effects on soil carbon stocks over 150 years in wisconsin, USA. *Remote Sens.*, **11**: 1504. Multidisciplinary Digital Publishing Institute. doi:[10.3390/rs11121504](https://doi.org/10.3390/rs11121504).
- Kasraei, B., Heung, B., Saurette, D.D., Schmidt, M.G., Bulmer, C.E., and Bethel, W. 2021. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. *Environ. Model. Softw.*, **144**: 105139. Elsevier. doi:[10.1016/j.envsoft.2021.105139](https://doi.org/10.1016/j.envsoft.2021.105139).
- Kempen, B., Dalsgaard, S., Kaaya, A.K., Chamuya, N., Ruipérez-González, M., Pekkarinen, A., and Walsh, M.G. 2019. Mapping topsoil organic carbon concentrations and stocks for tanzania. *Geoderma*, **337**: 164–180. Elsevier. doi:[10.1016/j.geoderma.2018.09.011](https://doi.org/10.1016/j.geoderma.2018.09.011).
- Khaledian, Y., and Miller, B.A. 2020. Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model.*, **81**: 401–418. Elsevier. doi:[10.1016/j.apm.2019.12.016](https://doi.org/10.1016/j.apm.2019.12.016).
- Kuhn, M. 2009. The caret package. *J. Stat. Softw.*, **28**. Citeseer.
- Lamichhane, S., Kumar, L., and Wilson, B. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review. *Geoderma*, **352**: 395–413. Elsevier. doi:[10.1016/j.geoderma.2019.05.031](https://doi.org/10.1016/j.geoderma.2019.05.031).
- Li, S., Li, J., Li, C., Huang, S., Li, X., Li, S., and Ma, Y. 2016. Testing the RothC and DNDC models against long-term dynamics of soil organic carbon stock observed at cropping field soils in North China. *Soil Tillage Res.*, **163**: 290–297. Elsevier. doi:[10.1016/j.still.2016.07.001](https://doi.org/10.1016/j.still.2016.07.001).
- Ließ, M., Schmidt, J., and Glaser, B. 2016. Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. *PLoS One*, **11**: e0153673. Public Library of Science. doi:[10.1371/journal.pone.0153673](https://doi.org/10.1371/journal.pone.0153673). PMID: [27128736](https://pubmed.ncbi.nlm.nih.gov/27128736/).
- Martens, D.A., Reedy, T.E., and Lewis, D.T. 2004. Soil organic carbon content and composition of 130-year crop, pasture and forest land-use managements. *Glob. Chang. Biol.*, **10**: 65–78. Wiley Online Library. doi:[10.1046/j.1529-8817.2003.00722.x](https://doi.org/10.1046/j.1529-8817.2003.00722.x).
- Maynard, J.J., and Levi, M.R. 2017. Hyper-temporal remote sensing for digital soil mapping: characterizing soil-vegetation response to climatic variability. *Geoderma*, **285**: 94–109. Elsevier. doi:[10.1016/j.geoderma.2016.09.024](https://doi.org/10.1016/j.geoderma.2016.09.024).
- McBratney, A.B., Mendonça Santos, M.L., and Minasny, B. 2003. On digital soil mapping. *Geoderma*, **117**: 3–52. doi:[10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
- McCarty, G.W., and Meisinger, J.J. 1997. Effects of n fertilizer treatments on biologically active n pools in soils under plow and no tillage. *Biol. Fertil. Soils*, **24**: 406–412. Springer. doi:[10.1007/s003740050265](https://doi.org/10.1007/s003740050265).
- Meinshausen, N., and Ridgeway, G. 2006. Quantile regression forests. *J. Mach. Learn. Res.*, **7**: 983–999.
- Minasny, B., Malone, B.P., Mcbratney, A.B., Angers, D.A., Arrouays, D. Chambers, A., et al. 2017. Soil carbon 4 per mille. *Geoderma*, **292**: 59–86. The Authors. doi:[10.1016/j.geoderma.2017.01.002](https://doi.org/10.1016/j.geoderma.2017.01.002).
- Minasny, B., and McBratney, A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.*, **32**: 1378–1388. doi:[10.1016/j.cageo.2005.12.009](https://doi.org/10.1016/j.cageo.2005.12.009).
- Minasny, B., and McBratney, A.B. 2016. Digital soil mapping: a brief history and some lessons. *Geoderma*, **264**: 301–311. Elsevier. doi:[10.1016/j.geoderma.2015.07.017](https://doi.org/10.1016/j.geoderma.2015.07.017).
- Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., and Arrouays, D. 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma*, **263**: 16–34. Elsevier. doi:[10.1016/j.geoderma.2015.08.035](https://doi.org/10.1016/j.geoderma.2015.08.035).
- Norris, C.E., Bean, G. Mac, Cappellazzi, S.B., Cope, M., Greub, K.L.H. Liptzin, D., et al. 2020. Introducing the North American project to evaluate soil health measurements. *Agron. J.*, **112**: 3195–3215. Wiley Online Library. doi:[10.1002/agj2.20234](https://doi.org/10.1002/agj2.20234).
- O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.*, **41**: 673–690. Springer. doi:[10.1007/s1135-006-9018-6](https://doi.org/10.1007/s1135-006-9018-6).
- Paul, S.S., Coops, N.C., Johnson, M.S., Krzic, M., Chandna, A., and Smukler, S.M. 2020a. Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change

- adaptation. *Geoderma*, **363**: 114177. Elsevier. doi:[10.1016/j.geoderma.2020.114177](https://doi.org/10.1016/j.geoderma.2020.114177).
- Paul, S.S., Dowell, L., Coops, N., Johnson, M., Krzic, M., Geesing, D., and Smukler, S. 2020b. Tracking changes in soil organic carbon across the heterogeneous agricultural landscape of the lower Fraser valley of British Columbia. *Sci. Total Environ.*, **732**: 138994. doi:[10.1016/j.scitotenv.2020.138994](https://doi.org/10.1016/j.scitotenv.2020.138994).
- Paustian, K., Collier, S., Baldock, J., Burgess, R., Creque, J. DeLonge, M., et al. 2019. Quantifying carbon for agricultural soil management: from the current status toward a global soil information system. *Carbon Manag.*, **10**: 567–587. Taylor & Francis. doi:[10.1080/17583004.2019.1633231](https://doi.org/10.1080/17583004.2019.1633231).
- Pérez-Guzmán, L., Phillips, L.A., Seuradge, B.J., Agomoh, I., Drury, C.F., and Acosta-Martínez, V. 2021. An evaluation of biological soil health indicators in four long-term continuous agroecosystems in Canada. *Agrosyst. Geosci. Environ.*, **4**: e20164. Wiley Online Library. doi:[10.1002/agg2.20164](https://doi.org/10.1002/agg2.20164).
- R Core Team 2018. R: A language and environment for statistical computing. : R Foundation for Statistical Computing, Vienna, Austria. [Online] Available: <https://www.r-project.org/>.
- Rial, M., Cortizas, A.M., and Rodríguez-Lado, L. 2017. Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils. *Sci. Total Environ.*, **609**: 1411–1422. Elsevier. doi:[10.1016/j.scitotenv.2017.08.012](https://doi.org/10.1016/j.scitotenv.2017.08.012). PMID: 28797147.
- Roudier, P. 2014. *clhs-package: conditioned Latin hypercube sampling*. Citeseer.
- Saurette, D. 2021. *onsoilsurvey: making PDSM in Ontario better*. [Online] Available from: <https://rdrr.io/github/newdale/onsoilsurvey/#vignettes>.
- Saurette, D.D. 2022. Comparing direct and indirect approaches to predicting soil texture class. *Can. J. Soil Sci.*, **102**(4): 835–851. doi:[10.1139/CJSS-2022-0040](https://doi.org/10.1139/CJSS-2022-0040).
- Song, X.-D., Wu, H.-Y., Ju, B., Liu, F., Yang, F. Li, D.-C., et al. 2020. Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma*, **363**: 114145. Elsevier. doi:[10.1016/j.geoderma.2019.114145](https://doi.org/10.1016/j.geoderma.2019.114145).
- Taghizadeh-Mehrjardi, R., Fathizad, H., Ali Hakimzadeh Ardakani, M., Sodaiezadeh, H., Kerry, R., Heung, B., and Scholten, T. 2021. Spatio-temporal analysis of heavy metals in arid soils at the catchment scale using digital soil assessment and a random forest model. *Remote Sens.*, **13**: 1698. Multidisciplinary Digital Publishing Institute. doi:[10.3390/rs13091698](https://doi.org/10.3390/rs13091698).
- VandenBygaart, A.J., Bremer, E., McConkey, B.G., Janzen, H.H., Angers, D.A. Carter, M.R., et al. 2010. Soil organic carbon stocks on long-term agroecosystem experiments in Canada. *Can. J. Soil Sci.*, **90**: 543–550. NRC Research Press. doi:[10.4141/cjss10028](https://doi.org/10.4141/cjss10028).
- Vargha, A., and Delaney, H.D. 1998. The Kruskal-Wallis test and stochastic homogeneity. *J. Educ. Behav. Stat.*, **23**: 170–192. Sage Publications Sage CA: Los Angeles, CA. doi:[10.3102/10769986023002170](https://doi.org/10.3102/10769986023002170).
- Wadoux, A.M.-C., Minasny, B., and McBratney, A.B. 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. *Earth-Sci. Rev.*, **210**: 103359. Elsevier. doi:[10.1016/j.earscirev.2020.103359](https://doi.org/10.1016/j.earscirev.2020.103359).
- Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A. Li Liu, D., et al. 2018. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecol. Indic.*, **88**: 425–438. Elsevier. doi:[10.1016/j.ecolind.2018.01.049](https://doi.org/10.1016/j.ecolind.2018.01.049).
- Wang, Y., Wang, S., Adhikari, K., Wang, Q., Sui, Y., and Xin, G. 2019. Effect of cultivation history on soil organic carbon status of arable land in northeastern China. *Geoderma*, **342**: 55–64. Elsevier. doi:[10.1016/j.geoderma.2019.02.007](https://doi.org/10.1016/j.geoderma.2019.02.007).
- Webb, K.T., Thompspon, R.L., Beke, G.J., and Nowland, J.L. 1991. *Soils of Colchester County, Nova Scotia*. Ottawa, Canada. [Online] Available from: http://sis.agr.gc.ca/cansis/publications/surveys/ns/ns19b/ns19b_report.pdf.
- Wiesmeier, M., Urbanski, L., Hobbey, E., Lang, B., von Lützw, M. Marin-Spiotta, E., et al. 2019. Soil organic carbon storage as a key function of soils - a review of drivers and indicators at various scales. *Geoderma*, **333**: 149–162. Elsevier. doi:[10.1016/j.geoderma.2018.07.026](https://doi.org/10.1016/j.geoderma.2018.07.026).
- Yigini, Y., and Panagos, P. 2016. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Sci. Total Environ.*, **557–558**: 838–850. The Authors. doi:[10.1016/j.scitotenv.2016.03.085](https://doi.org/10.1016/j.scitotenv.2016.03.085). PMID: 27082446.
- Zebarth, B.J., Islam, M.M., Cambouris, A.N., Perron, I., Burton, D.L., Comeau, L., and Moreau, G. 2019. Spatial variation of soil health indices in a commercial potato field in Eastern Canada. *Soil Sci. Soc. Am. J.*, **83**: 1786–1798. Wiley Online Library. doi:[10.2136/sssaj2019.03.0087](https://doi.org/10.2136/sssaj2019.03.0087).