Mapping the maximum peat thickness of cultivated organic soils in the southwest plain of Montreal

Authors: Deragon, Raphaël, Saurette, Daniel D., Heung, Brandon, and Caron, Jean

Source: Canadian Journal of Soil Science, 103(1) : 103-120

Published By: Canadian Science Publishing

URL: https://doi.org/10.1139/cjss-2022-0031

Canadian Journal of Soil Science

OPEN ACCESS | Article

Mapping the maximum peat thickness of cultivated organic soils in the southwest plain of Montreal

Raphaël Deragon[a], **Daniel D. Saurette**[b,c], **Brandon Heung**[d], **and Jean Caron**[a]

[a]Département des sols et de génie agroalimentaire, Université Laval, 2325 Rue de l'Université, Québec, QC G1V 0A6, Canada; [b]Ontario Ministry of Agriculture Food and Rural Affairs, 1 Stone Road West, 3rd Floor SE, Guelph, ON N1G 4Y2, Canada; [c]School of Environmental Sciences, University of Guelph, Guelph, ON, N1G 2W1, Canada; [d]Faculty of Agriculture, Department of Plant, Food, and Environmental Sciences, Dalhousie University, 50 Pictou Rd., Truro, NS B2N 5E3, Canada

Corresponding author: **Raphaël Deragon** (email: raphael.deragon.1@ulaval.ca)

Abstract

Large organic deposits in the southwestern plain of Montreal have been converted to agricultural land for vegetable production. In addition to the variable depth of the organic deposits, these soils commonly have an impermeable coprogenous layer between the peat and the underlying mineral substratum. Estimations of the depth and thickness of these materials are critical for soil management. Therefore, five drained and cultivated peatlands were studied to estimate their maximum peat thickness (MPT)—a potential key soil property that can help identify management zones for their conservation. MPT can be defined as the depth to the mineral layer (DML) minus the coprogenous layer thickness (CLT). The objective of this study was to estimate DML, CLT, and MPT at a regional scale using environmental covariates derived from remote sensing. Three machine-learning models (Cubist, Random Forest, and k-Nearest Neighbor) were compared to produce maps of DML and CLT, which were combined to generate MPT at a spatial resolution of 10 m. The Cubist model performed the best for predicting both features of interest, yielding Lin's concordance correlation coefficients of 0.43 and 0.07 for DML and CLT, respectively, using a spatial cross-validation procedure. Interpretation of the drivers of CLT was limited by the poor predictive power of the final model. More precise data on MPT are needed to support soil conservation practices, and more CLT field observations are required to obtain a higher prediction accuracy. Nonetheless, digital soil mapping using open-access geospatial data shows promise for understanding and managing cultivated peatlands.

Key words: predictive digital soil mapping, machine learning, organic soils, peat thickness, coprogenous soil

1. Introduction

Canada accounts for about a quarter of the world's peatland extent (Vepraskas and Craft 2015). Cultivated organic soils cover only 4% of the Province of Quebec's southern region; yet, they greatly contribute to food production and to the economy of the province with exports to northeastern USA (Groupe AGÉCO 2007; Parent and Gagné 2010). These soils are prized for their vegetable production, but they are affected by intense soil loss processes that are unique to the evolution of peat materials. Soil loss occurs primarily by subsidence, oxidation, and erosion after land conversion for agriculture, and is enhanced by the drainage of peatlands (Kroetsch et al. 2011; Vepraskas and Craft 2015). Over the past decades, with an annual estimated soil loss of 2.5 cm (Ilnicki 2003; Esselami et al. 2014), degraded, shallower fields have transitioned to less productive mineral soils, as if the mineral boundary had been moving toward the surface.

Furthermore, coprogenous material can also be found between the peat and mineral layers (Lamontagne et al. 2014)—also reducing the thickness of the cultivable peaty layer. When the material is found within the first 160 cm and has a minimum thickness of 5 cm, it is referred to as a "limnic" layer (SCWG 1998). This gelatinous, impervious material is unsuitable for agricultural production and specific to lacustrine organic deposits (Kroetsch et al. 2011). When it dries, this material shrinks and does not rewet (SCWG 1998). In Poland, calcareous limnic deposits have been shown to limit root growth (Ilnicki 2003). Limnic materials can be of different types: coprogenous earth (sedimentary peat), diatomaceous earth, or marl (SCWG 1998). In the southwestern plain of Montreal, coprogenous earth is most common and sometimes found with a small layer of marl. The latter is effervescent due to the presence of shells and precipitated $CaCO_3$, while the former can be a mineral (<17% organic carbon), or organic deposit (≥17% organic carbon) enriched by algae or aquatic life plants transformation products (SCWG 1998). Since pedological surveys have typically focused on the spatial extent of peat and descriptions of soil series, little is known about the depth to the mineral layer (DML) and the coprogenous layer thickness (CLT). To mitigate the impacts of long-term soil loss, soil conservation approaches, such as the addition of biomass crop amendments

(Dessureault-Rompre´ et al. 2020), the plantation of tree windbreaks, or water table management, are needed; however, these approaches have a related cost and material requirements to be considered. Therefore, priority management zones need to be defined to guide the application of regional soil conservation plans. The effective peat thickness that can be used for agriculture, hereinafter referred to as the "maximum peat thickness" (MPT), could be used to define these priority management zones. In other words, the MPT can be defined as the thickness of the peaty layer of an organic soil, therefore, excluding coprogenous and mineral materials. This definition would better reflect the real long-term agricultural potential of a field than only mapping the DML as other studies have done. It is necessary to understand the spatial distribution of the MPT to better manage shallower soils.

While pedological surveys and taxonomic map products are available at a global scale, new mapping efforts are targeting soil properties and soil functions (FAO 2020). Modern digital soil mapping (DSM) techniques can provide such information by leveraging technological advances for predicting a soil attribute or class using georeferenced field measurements and a suite of environmental covariates obtained via remote or proximal sensing (McBratney et al. 2003). In peatland-related DSM studies, the thickness of the peat and organic carbon content are the most frequently predicted features (Minasny et al. 2019). To the best of our knowledge, predictive mapping of coprogenous materials has never been explored and the relevant covariates are unknown. Yet, the information that could be acquired through a predictive map on a regional scale is crucial to estimating the MPT, and, therefore, the delineation of priority management zones. Manual probing over large areas is labor intensive and relatively slow (Parry et al. 2014). A regional approach, relying on a field calibration data set, could provide useful maps when combined with a relevant set of covariates. Many combinations of covariates are commonly used in regression and classification mapping of peatlands and could be investigated as potential candidates for predicting the thickness of the coprogenous layer. Peat thickness and its extent are often evaluated by combining a digital elevation model (DEM) and its derivatives, airborne gamma radiometric data, electromagnetic data, and satellite data (Rudiyanto et al. 2018; Gatis et al. 2019; Minasny et al. 2019; Siemon et al. 2020). Most of the covariates are the product of remote sensing techniques, while ground penetrating radar, gamma radiometric data, and soil electrical resistivity or conductivity can be obtained with proximal sensing techniques (Rosa et al. 2009; Parry et al. 2014; Comas et al. 2015; Beucher et al. 2020).

Hence, given the lack of data concerning the DML and CLT for the study area and the abundance of covariates, the main objective of this study was to integrate open-access, remote sensing covariates and field data to predict the spatial distribution of the MPT, including mineral and coprogenous materials' depth. Here, the specific objectives of the study were (*i*) to determine the covariates that most contribute to the prediction of DML and CLT and (*ii*) to derive a map of MPT from the mineral and coprogenous material predictions as a tool to guide soil conservation practices.

2. Methodology

Figure 1 shows the methodological framework for this study.

2.1. Study area

The study area is comprised of five drained and cultivated peatlands in the southwest plain of Montreal, Quebec, covering approximately 90 km^2 (45.1°N to 45.3°N latitude and −73.3°W to −73.7°W longitude). Figure 2 shows the study area and the spatial distribution of sampling sites. Natural forests are still present but are likely affected by agricultural drainage. Most of the soils are used for horticulture, while a small proportion is used for producing bags of gardening soil. The five peatlands are part of three separate watersheds (Lamontagne et al. 2014). The elevation ranges between 45 and 70 m above mean sea level; however, the agricultural fields are leveled across the peatlands.

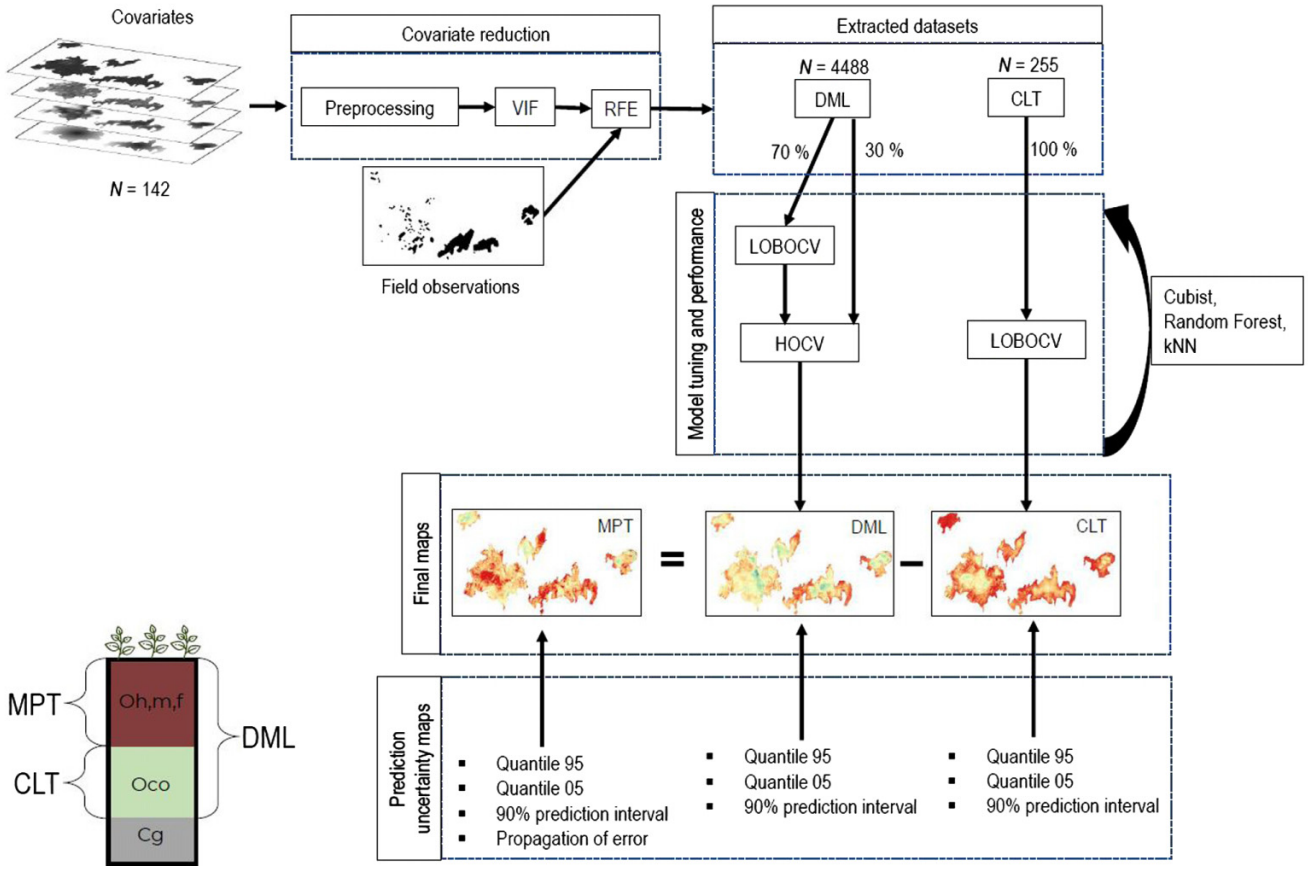
Fieldwork was carried out on 14 partnering research farms, which were converted to agricultural use between 1950 and 2010. Most fields were classified as Humisols and Mesisols, while a small area of recently converted fields was classified as Fibrisols (SCWG 1998). The organic materials found in the southwest plain of Montreal were deposited in channels or in depressions. Basin bogs or shore swamps, the latter having shallower deposits, were the result of hundreds of years of accumulation (Lamontagne et al. 2014). Peatlands in this region are mainly composed of forest peat or herbaceous materials (LaSalle 1963; Grenon 1988). Limnic materials are found throughout the study area but not in every peatland: the northwestern-most peatland appeared to be devoid of limnic material, whereas the other peatlands had regions with limnic layers.

2.2. Soil data

Two data sets were used to build the predictive models. The first data set consisted of 255 sites that were sampled between 2019 and 2021 (shown as white dots in Fig. 2). The CLT and the DML were obtained by manually extracting soil cores using a Macauley corer (Eijkelkamp peat sampler) until the mineral horizon was reached.

The second data set consisted of 4488 observations. Here, 4286 locations were manually probed (shown as grey dots in Fig. 2) between 2010 and 2014, and were combined with the first data set. Manual probing involved inserting a thin metal rod in the soil and when a change of soil resistance to the rod penetration was felt, the depth was recorded as the DML. It should be noted that coprogenous materials did not offer notable resistance when a rod was inserted into the soil in comparison to the apparent change in density when the probe reached the underlying mineral soil. Manual probing was susceptible to measurement errors due to buried wood pieces (Parry et al. 2014). Nonetheless, manual probing could be done rapidly and required less energy to do manually than the Macauley sampling technique. Furthermore, the sites were clustered in three of the five peatlands. It was believed that the addition of this data set to this study could contribute to the predictive power of

Fig. 1. Methodological framework used in this study. Two prediction maps were produced, (1) for the depth to the mineral layer (DML) and (2) for the coprogenous layer thickness (CLT). Then, they were subtracted to obtain the maximum peat thickness map (MPT). Note that the number of covariates and observations in each model differ. VIF = Variance inflation factor, RFE = recursive feature elimination, LOBOCV = leave-one-block-out cross-validation, and HOCV = hold-out cross-validation. Bottom left is a schematic of an organic soil profile. Oh, m, f = Humic, mesic, or fibric peaty layer, Oco = Coprogenous earth layer, and Cg = Mineral soil layer, according to the Canadian System of Soil Classification (SCWG 1998).



the models when generating DML predictions even though manual probing might have been more prone to measurement bias or error. Since some of the sampling locations were close to each other, observations that fell into the same raster cell were averaged. The summary statistics of the two data sets are presented in Table 1 and were also computed for both data sets for each individual peatland (Table A1).

2.3. Environmental covariates

When generating a training data set to be fitted with a predictive model, the georeferenced soil observations are spatially intersected with a suite of geospatial environmental layers (i.e., covariates or predictors). The predictive models are used to establish the relationships between the covariate and soil properties to predict the spatial phenomenon of interest. Here, we present the five categories of covariates used, which were selected based on the SCORPAN model (i.e., soil, climate, organisms, relief, parent material, time, and spatial location) in McBratney et al. (2003) and a review of the digital mapping of peatlands in Minasny et al. (2019). The description and preprocessing details are provided for the final selection of covariates.

2.3.1. Digital elevation model (DEM) data

A DEM is a common data set that can be used to derive a suite of geomorphological and hydrological covariates (McBratney et al. 2003; Minasny et al. 2019). Since organic soils are formed in lowlands, depressions, and water saturated conditions, the topographical covariates are expected to be effective predictors of peat accumulation and the delineation of peatlands. The DEMs were produced and distributed by the Ministère des Forêts, de la Faune et des Parcs, where it was downloaded from the Données Québec repository. Here, the multiple DEM tiles were mosaicked. The DEM data were derived from a LiDAR survey and made available at a 1 m spatial resolution.

The DEM was preprocessed in five steps using the whitebox package in R (Wu 2020). Missing data were first filled-in; following this, a feature preserving smoothing tool was used to remove short-scale noise due to the use of the ultrafine

Fig. 2. Study area and sampling sites categorized by the nature of the information collected. Overlaid is a square grid of 1.5 km size used to spatially partition the data sets to allow spatial cross-validation. Numbers beside the basin limits are peatlands identification numbers used in this study.

Table 1. Summary statistics of the coprogenous layer thickness (CLT) and of the depth to the mineral layer (DML).

	CLT	DML
N	255	4488
Minimum (cm)	0	5
Maximum (cm)	206	392
Mean (cm)	34.16	136.23
Median (cm)	25.00	130.00
Standard deviation (cm)	37.22	70.97
Coefficient of variation (%)	108.83	52.11
Variance (cm^2)	1385.62	5036.64
Skewness	1.26	0.46
Kurtosis	1.74	−0.42

resolution DEM data. Afterward, a mean filter with a 5 m × 5 m window was used to further smooth the DEM and remove artifacts. To reduce the computational demands and memory requirements, the DEM was aggregated to a 10 m resolution. This covariate became the reference to which all other covariates were resampled to match its extent and resolution. Single cell pits were filled to remove local artifacts and to produce a more continuous output. The Breach Depressions Least Cost tool was used to prepare a DEM for deriving the hydrological covariates (Lindsay 2016). This tool breaches and then fills depressions in the DEM to ensure continuous flow paths through depressions using a least cost pathway method based on a breaching algorithm by Lindsay (2016).

After preprocessing, both the RSAGA (Brenning et al. 2018) and whitebox packages were used to generate 70 covariates. Some of these covariates included the difference and deviation from mean elevation, topographic wetness index, multiresolution valley bottom flatness (MRVBF), multiresolution ridge top flatness (MRRTF), catchment areas, hillshade, slope, aspect, and curvature (Beven and Kirkby 1979; Gallant and Dowling 2003; McBratney et al. 2003; Lindsay et al. 2015).

2.3.2. Aerial gamma ray spectroscopy data

Six raster layers were downloaded from Natural Resources Canada's GEOSCAN database (Natural Resources Canada 2019). The layers represented surface concentrations of potassium (K, %), equivalent uranium (URA, ppm), equivalent

thorium (THO, ppm), and ratios URA/THO (RUT), URA/K (RUK), and THO/K (RTK). Radiometric data have proven to be useful for differentiating mineral soils from organic soils given their difference in parent material, porosity, and water content (Beamish 2013; Keaney et al. 2013). Water-filled, peat soil attenuates bedrock geology radioactivity and can provide an indication of the magnitude of the deposit. The six raster layers were resampled from a 250 m to 10 m spatial resolution using a bilinear method to match the extent and resolution of the other covariates.

2.3.3. Sentinel-2 L2-A data

Sentinel-2 is a constellation of two satellites with a combined repeat cycle of 5 days over the same area. Common indices and raw bands were tested as potential covariates. Gholizadeh et al. (2018) obtained significant correlations (r = −0.74 to −0.36) between bands 4, 5, 11, and 12 and soil organic carbon; hence, raw bands could serve as discriminant covariates to delineate a peatland's extent. L2-A level of preprocessing generates a bottom-of-atmosphere corrected reflectance raster for all 13 bands. This level of preprocessing includes radiometric, geometric (including orthorectification), and atmospheric corrections. The raster layers were preprocessed by the European Space Agency and made available via the Copernicus Open Access Hub. All final raster layers were resampled to a 10 m spatial resolution using the bilinear method since the raw bands had a native resolution of 10, 20, or 60 m. Lastly, the raster layers were reprojected to match the other covariates.

A multitemporal approach was adopted to capture moisture gradients over the study area (Fathololoumi et al. 2021). Therefore, median and standard deviation layers were produced for all raw bands and indices that were gathered for the following dates: 31 March 2020, 20 April 2020, 25 April 2020, and 20 May 2020. These dates were selected because they had neither snow nor crops covering the fields and, hence, were more effective in differentiating between the organic and mineral soils, as well as forests. The dates were also selected based on data availability and cloud cover over the study area.

2.3.4. Landsat 8 OLI_TIRS C2 Level-2 data

Landsat 8 is a satellite with a 16-day repeat cycle over the same area. C2 Level-2 data are referred to as "analysis ready" data—being preprocessed by the USGS Earth Resources Observation and Science (EROS) Center and then made freely available. After Level-2 processing, surface reflectance (bottom of atmosphere) values were obtained. More information on the Landsat missions and preprocessing can be found in Young et al. (2017).

This satellite provides information on the Earth's temperature and land surface. The data acquired from two thermal infrared sensors (100 m spatial resolution) and nine spectral bands (30 m spatial resolution with one at 15 m resolution) are sensitive to different wavelength ranges. Similar to the Sentinel-2 data, the Landsat 8 raw bands and indices were resampled to a 10 m spatial resolution. A multitemporal approach was also used to capture seasonal and moisture gradients. Three series of satellite images were chosen based on cloud percentage, the absence of snow and crops. Images from 6 May 2015, 14 May 2018, and 30 May 2018 were used. The median and standard deviation over the three dates were used as covariates. The remote sensing indices were generated from both satellites, using their respective bands and were summarized in Table 2. All indices were computed in QGIS (QGIS.org, version 3.10).

2.3.5. Distance and directional covariates

The center of a peatland is often the point where peat accumulation is at its maximum, whereas near the border of the peatland, one can expect thinner deposits. Therefore, project-specific distance layers made in QGIS with the Multi-Distance Buffer plugin were included (Tveite 2018). The first layer was produced by locating a peatland's centroid, then generating 10 m buffers around it until the peatland's border was reached. The product was then rasterized to a 10 m spatial resolution. The second layer was produced by generating 10 m inner buffers from the peatland's border, which was then rasterized. One might think the results would be the same; however, the resulting covariates exhibit significantly different patterns due to the shape of each peatland. Aspect reflects the topographic shape and directionality but is a circular measure and cannot be used directly as a covariate; therefore, we generated raster layers that reflected the eastness and northness of the study area. In addition to these covariates, Euclidean distance fields were generated for the study area to provide spatial context (Behrens et al. 2018).

2.4. Variance inflation factor analysis

The 142 covariates produced were reduced using a stepwise variance inflation factor (VIF) procedure implemented with the vifstep function (Naimi et al. 2014) to reduce multicollinearity (O'Brien 2007). Since it required the use of standardized covariates, the covariates were scaled using the raster package. The VIF approach takes each of the standardized covariate and uses the remaining ones as independent predictors in a multiple linear regression via ordinary least squares. Then, the coefficient of determination of the covariate acting as the dependent variable (R_i^2) is calculated. This process is repeated for every covariate. In eq. 1, the VIF is computed for every i^{th} predictor, where a high proportion of variance explained by a given combination of predictors will result in a higher VIF score. The model then evaluates if one or more covariates has a VIF score that exceeds a predetermined threshold value. In our case, a score of 5 was selected although 10 is also a common threshold (O'Brien 2007; James et al. 2014; Bian et al. 2020). The covariate with the highest VIF is removed, and the process is repeated until all remaining covariates are below the threshold.

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \quad (1)$$

Table 2. Bands and indices from Landsat 8 and Sentinel-2 satellites explored as potential covariates.

Bands and indices	Landsat	Sentinel	Equation	Source
Brightness Index	x	x	$((\text{Red})^2 + (\text{NIR})^2)^{0.5}$	Escadafal 1994
Normalized Difference Vegetation Index (NDVI)	x	x	(NIR − Red)/(NIR + Red)	Rouse et al. 1974
Soil Color Index	x	x	3 × NIR + Red − Green − 3 × Blue	Poggio and Gimona 2017
Soil Moisture Index	x	x	NIR/Blue	Poggio and Gimona 2017
Bare soil index	x		((Red + SWIR) − (NIR + Blue))/((Red + SWIR) + (NIR + Blue))	Rikimaru et al. 2002
Combined Spectral Response Index	x		(Blue + Green)/(Red + NIR) × NDVI	Taghizadeh-Mehrjardi et al. 2021
Normalized Difference Moisture Index	x		(Red − NIR)/(Red + NIR)	Gao 1996
Salinity Index	x		Green/Red	Khan et al. 2005
Raw bands 2–7	x			
Normalized Difference Salinity Index		x	(Red − NIR)/(Red + NIR)	Khan et al. 2005
Soil Adjusted Vegetation Index		x	(NIR − Red)/(NIR + Red + L) × (1 + L)	Huete 1988
Transformed Vegetation Index		x	$((\text{NIR} - \text{Red})/(\text{NIR} + \text{Red}) + 0.5)^{0.5}$	Nellis and Briggs 1992
Raw bands 2–8, 8A, 9, 11, and 12		x		

Note: $L = 0.5$ (medium density of vegetation) was used in the Soil Adjusted Vegetation Index equation.

2.5. Modeling approaches

Three commonly used machine learning algorithms in DSM were compared for predicting CLT and DML (Minasny et al. 2019; Rudiyanto et al. 2018): Cubist, Random Forest (RF), and k-Nearest Neighbor (kNN). Although this section will briefly summarize each model, details concerning the learners are provided in Heung et al. (2016), where the models were reviewed and compared. All the modeling was carried out using the R statistical software (R Core Team 2020) and the caret package, which included all the tested models (Kuhn 2020). Final maps were generated in QGIS at a 10 m spatial resolution.

2.5.1. Cubist

Cubist is a rule-based model that produces a regression tree based on the M5 model (Quinlan 1992), and was subsequently adapted by Kuhn and Johnson (2013) in R. At each node of the tree, the data set is split based on a set of rules using the value of one or many covariates, and form groups that minimize the within-node variability. The terminal nodes (i.e., leaves) of the resulting tree consist of multivariate linear models, which are applied to the covariates to make predictions. The hybridization of piecewise linear models and the hierarchical tree models allow Cubist to capture both linear and nonlinear relationships between predictors and the response variables (Malone et al. 2017). This model has two hyperparameters to be tuned when training the model (Kuhn and Johnson 2013). Overpredictions and underpredictions can be accounted for with a boosting technique using the committees hyperparameter. It specifies the number of similar trees to be sequentially produced and aggregated to optimize the set of rules. This process is explained in greater details in Kuhn and Johnson (2013). The second hyperparameter is the number of neighbors to be considered in a nearest-neighbor search through the training data set to find the closest observation to the predicted value (Quinlan 1993). This final step is useful for adjusting the value at the predicted location based on the average of its nearest neighbors.

2.5.2. Random Forest

Based on ensemble theory (Zhang and Ma 2012), RF is a nonparametric, decision-tree model, where a set of uncorrelated trees are combined (Breiman 2001). Multiple trees are generated with the use of bagging, where each tree is built on a random bootstrap sample of the original data set (with replacement). Using the bootstrap sample, a series of node-splitting rules are generated with respect to the covariates with the objective of maximizing the within-node homogeneity and the between node heterogeneity. In the caret package, m_{try} is the main tuning parameter and is used to control the number of randomly selected covariate at each node, where the values of m_{try} range from 1 to m number of predictors. Higher values of m_{try} results in a higher likelihood of correlation between the trees; however, it lowers the prediction variance at the same time (Hastie et al. 2009). The trees from the forest are then aggregated to obtain an average prediction for each new observation (Genuer and Poggi 2020). Predictions can be regarded as more effective than those from single tree-learners due to RF's ability to mitigate bias (Kuhn and Johnson 2013).

2.5.3. K-Nearest Neighbors

kNN is a distance-based learner with one hyperparameter, k, which represents the number of neighbors of the unobserved location to be used for the prediction (Hastie et al. 2009). The rationale behind this learner is that observations that have similar properties (i.e., covariates values) will tend to have a similar value for the response variable of interest. Therefore, kNN predictions are made using k observations from training data that have similar values to those of the predicted site in the covariates multivariate feature space. To find neighbors of the predicted value, a distance metric must be used (i.e., Euclidean distance) to assess the closeness

between observations. In a regression context, the k closest training data will be averaged, but if $k = 1$, the predicted value will be assigned the value of the nearest training observation. Since the covariates had different value ranges, they were standardized to ensure nonbiased calculation of the distance metrics (Hastie et al. 2009).

2.6. Model tuning and validation

2.6.1. Spatial cross-validation

As stated in a review of peat mapping studies (Minasny et al. 2019), 64 % of the 90 reviewed studies did not perform validation, which may result in overfitting and overoptimistic models' performance metrics. Therefore, each of the machine learners were trained to find the optimal hyperparameters for DML and CLT models. Considering recent papers that had overoptimistic performances of machine learners in DSM applications, spatial cross-validation was used to evaluate these models (Pohjankukka et al. 2017; Meyer et al. 2018; Schratz et al. 2019; Ploton et al. 2020). This approach has proven to be more effective at measuring internal error and controlling for spatial autocorrelation in the data set. When the data are clustered, the model is more likely to correctly predict nearby observations during the training and validation process and, therefore, estimates of model performance can be biased.

Both data sets had clustered observations over the study area. To allow a fair assessment of model performance, leave-one-block-out cross-validation (LOBOCV) was used (Roberts et al. 2017). This form of spatial cross-validation is similar to leave-one-out cross-validation, where one observation from a data set is partitioned out during model training and is used as a test observation. During LOBOCV, one spatial block is put aside as a validation set. The model is trained using all observations except for those from the testing block. The model is then applied to this unseen data and error measures are derived. The process is repeated by changing each validation block for a new one at each fold. The number of folds is equal to the number of blocks, respectively, 45 (CLT) and 44 (DML) for this study. To compare LOBOCV with the conventional k-fold cross-validation approach, the model performance of both final models at the testing stage was also evaluated with a 10-fold cross-validation with 10 repeats.

The CLT data set was used without modification in LOBOCV because of its low number of observations ($N = 255$), whereas the higher number of observations in the DML data set ($N = 4488$) made partitioning possible. The traditional method to split a data set for training and testing purposes involves a 70%–30% split, respectively, with the use of random sampling. Since the DML data were highly clustered, random sampling could not divide observations evenly in training and testing data sets for peatlands with less data. Thus, conditioned Latin hypercube sampling on the X and Y coordinates (Minasny and McBratney 2006) was preferred over random sampling. Before merging the two data sets ($255 + 4286$), they were split based on their spatial coordinates to ensure a similar spatial coverage of the study area between the training and test partitions. The LOBOCV was performed to tune the model with 70% of the data, and to compute internal performance assessment, the test data set (i.e., the remaining 30% yet unseen by the model) was used in a hold-out cross-validation procedure.

To identify the suitable size for each spatial block unit, the range from the theoretical variograms for the DML and CLT were calculated using the approach described in Oliver and Webster (2014). The experimental and theoretical variograms were computed using the gstat package (Pebesma 2004; Gräler et al. 2016). Weighted least residual sum of squares and visual assessment of the variogram fit were used to select the best model (Oliver and Webster 2014). If all models had an adequate fit via visual assessment, the one with the lowest weighted sum of squared errors was selected. Four outliers were identified in the DML data set; however, they corresponded to deep peat deposits and were kept in the model. Four outliers were identified in the CLT and the data were skewed (Table 1). Since outliers were not mistakes nor did they belong to another population, they were kept in the model and a square root transformation was applied (Oliver and Webster 2014). This transformation was more effective at reducing skewness than a logarithmic transformation. No outliers remained after transformation. The grid was then applied to the area and can be seen in Fig. 2. The same grid was applied to both features prediction under the assumption that CLT and DML would have shared similar autocorrelation if CLT data set had more observations.

2.6.2. Model performance

Two accuracy metrics were used to select the best hyperparameters based on accuracy and precision: root mean square of error (RMSE, eq. 2) and Lin's concordance correlation coefficient (CCC, eq. 3):

$$\text{RMSE} = \sqrt{\sum_{i=1}^{N} \frac{(O_i - P_i)^2}{N}} \tag{2}$$

where O is the observed value and P the corresponding predicted value and N is the total number of observations. CCC was calculated as follows:

$$\text{CCC} = \frac{2s_{op}}{(\bar{o} - \bar{p})^2 + s_o^2 + s_p^2} \tag{3}$$

where s_{op} is the covariance between predicted and observed values, s^2 are their corresponding variance, $\bar{o}$ is the mean of observed values, and $\bar{p}$ is the mean of predicted values.

CCC can vary between -1 and 1, while a value near 0 indicates a lack of concordance between two variables. This index is similar to R^2, because it measures agreement between two variables but with the notable difference that it corrects for systematic bias if the relationship departs from the 1:1 line (Lawrence and Lin 1989). CCC can never be higher than the absolute value of the Pearson correlation coefficient in the presence of a bias. The CCC and RMSE statistics were obtained using the onsoilsurvey package in R (Saurette 2021).

2.7. Model simplification and variable importance analysis

Recursive feature elimination (RFE) was performed with the rfe function in the *caret* package (Kuhn 2020) to further reduce the number of covariates for training the final models. This model simplification procedure was needed to aid in interpreting soil–environmental relationships and improving generalizability. Briefly, RFE incorporates resampling to measure model performance with a reduced selection of covariates. It is a backward feature selection method (i.e., the model initially includes all predictors), whereby the predictors are iteratively removed to simplify the model, and in some cases, improve its performance. To fit the RFE models, the caretFuncs function was used with 5-fold cross-validation repeated five times. This allowed the computation of RMSE (see eq. 2) that was used as the metric to determine the best model. The most notable difference between VIF and RFE is that VIF is carried out independently of the response variable to address multicollinearity, whereas the RFE reduces covariates as a function of the model performance in predicting the response variable to remove irrelevant predictors. In other words, the VIF was performed once, but RFE was performed six times (i.e., two response variables with three machine learners). To assess the contribution of the remaining selection of covariates, relative variable importance plots were made for both final models using the VarImp function in the caret package.

2.8. Prediction uncertainty

Many peat mapping studies do not include local estimates of the prediction uncertainty (Minasny et al. 2019). To produce a prediction uncertainty map, a bootstrapping approach was adapted from Malone et al. (2017). Here, 100 realizations were produced for both CLT and DML final models. Each raster was generated by randomly sampling with replacement the original data set with an equal number of observations. Afterward, the 0.05 and 0.95 quantiles for each cell were obtained to generate the lower and upper bound of the 90% confidence interval of the predicted property.

Since the MPT map was the result of the difference between the DML and CLT maps, propagation of error had to be considered to adequately capture uncertainty. As such, the mean, variance, and covariance of bootstrap predictions were calculated, assuming a normal distribution of the statistics for each cell. Equation 4 describes how the standard deviation value was calculated for each cell of the final map, using the bootstrap-produced maps:

$$\sigma_{\mathrm{MPT}} = \sqrt{\sigma_M^2 + \sigma_C^2 - 2\sigma_{MC}} \quad (4)$$

where σ_{MPT} is the standard deviation of the MPT for a single cell, σ_M^2 is the variance of the mineral predictions, and σ_C^2 is the variance of the coprogenous predictions for 100 bootstraps. It is assumed that both variables were dependent (i.e., correlated); therefore, a term corresponding to the covariance between the features was added to the generic equation (Ku 1966). Based on Malone et al. (2017), the mean square error estimate from the validation data was added to the bootstrapping variance σ_M^2 and σ_C^2 to account for systematic and random errors in the models. Here, σ_{MC} is the covariance, as defined by eq. 5:

$$\sigma_{MC} = \sum_{i=1}^{N} \frac{(M_i - \bar{M})(C_i - \bar{C})}{N - 1} \quad (5)$$

where M_i and C_i are the values of the i^{th} cell of one bootstrap raster, and $\bar{M}$ and $\bar{C}$ are the mean values of the 100 rasters for that cell, respectively, for mineral and coprogenous maps.

Finally, to obtain the upper bound (95$^{\mathrm{th}}$ percentile) and lower bound (5$^{\mathrm{th}}$ percentile) of the 90% prediction interval, the corresponding *z*-value of 1.645 was multiplied to the standard deviation of the MPT. It was then subtracted and added to the predicted MPT value of a given cell (eq. 6).

$$\text{Prediction limits} = \overline{X_{\mathrm{MPT}}} \pm 1.645 \times \sigma_{\mathrm{MPT}} \quad (6)$$

where $\overline{X_{\mathrm{MPT}}}$ is the mean MPT of a given cell obtained from the bootstrapped maps.

3. Results

3.1. Variance inflation factor and recursive feature elimination analyses

The VIF procedure reduced the number of covariates from 142 to 59, while the RFE procedure further reduced the number of covariates for the DML and CLT models to numbers between 5 and 25 covariates. Table 3 lists the final selection of covariates for all models. Only two Landsat 8 covariates and four Sentinel-2 covariates were retained in the final selection, while a larger number of DEM derivatives were retained. The distance and gamma radiometric layers were also retained in the final selection. Both selected kNN models show the lowest number of covariates after the RFE procedure.potassium (K, %), equivalent uranium (URA, ppm), equivalent thorium (THO, ppm), and ratios URA/THO (RUT), URA/K (RUK), and THO/K (RTK).

3.2. Leave-one-block-out cross-validation

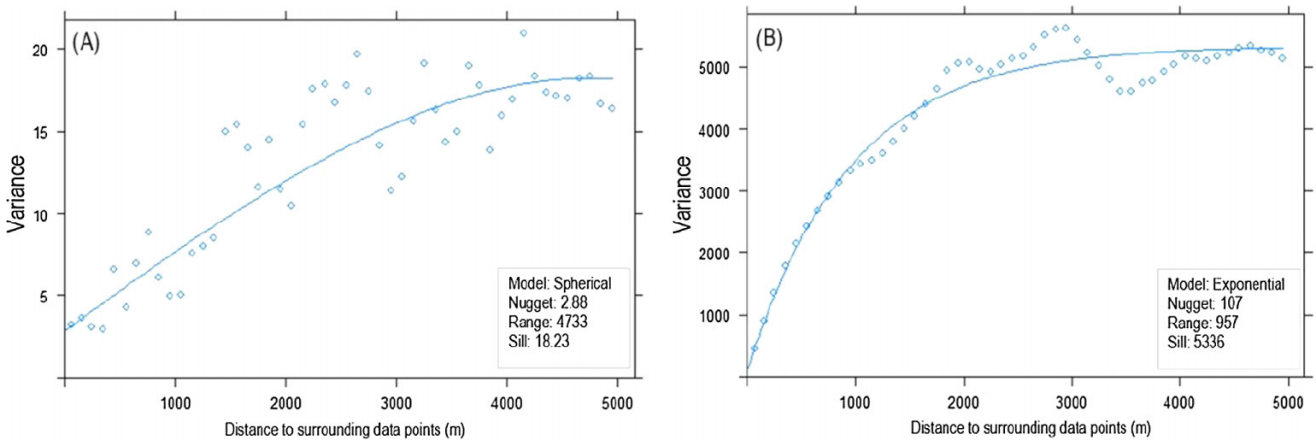
A spherical model was used to fit the CLT data, while an exponential model was used to fit the DML data to evaluate spatial autocorrelation (Fig. 3). The range for the CLT was 4733 m and for the DML was 957 m. The range provides the maximum distance at which two measurements of the same property are related. The results indicated the presence of autocorrelation among both features that needed to be captured during model tuning. One should note that for the exponential model, it was the effective range (i.e., lag corresponding to 95 % of the sill) that was defined.

To account for spatial autocorrelation during the cross-validation procedure, the ranges obtained above were used to generate spatial clusters of sites. The study area was divided in squares using a grid of 1000, 1500, and 2000 m for comparison. For each model, LOBOCV was performed. This approach was compared to a non-spatial 10-fold cross-validation. The latter returned overoptimistic values of

Table 3. Final selection of covariates for the six models after stepwise variance inflation factor and recursive feature selection. DEM = digital elevation model; kNN = k-Nearest Neighbor; RF = Random Forest.

				Depth to the mineral layer			Coprogenous layer thickness		
Covariate name	Description	Provider	SCORPAN factor	Cubist	RF	KNN	Cubist	RF	KNN
RTK	Ratio of equivalent thorium to potassium from airborne gamma ray spectrometry	Natural Resources Canada	P	x	x		x	x	
RUT	Ratio of equivalent uranium to equivalent thorium from airborne gamma ray spectrometry	Natural Resources Canada	P	x	x		x	x	
THO	Surface concentration of equivalent thorium (ppm)	Natural Resources Canada	P	x	x	x	x	x	x
URA	Surface concentration of equivalent uranium (ppm)	Natural Resources Canada	P	x	x		x		
Buffer	Distance layer of 10 m buffers generated from each peatland's boundaries	This study	N	x	x		x	x	x
Center	Distance layer of 10 m buffers generated from each peatland's center	This study	N	x	x	x	x	x	
DIST_MID	Euclidean distance to the middle of the study area	Derived from DEM	N	x	x	x	x	x	x
DIST_Y	Euclidean distance to the northern-most coordinates of the study area	Derived from DEM	N	x	x		x	x	x
ElevPercent219	Elevation percentile from a 2190 m filter kernel	Derived from DEM	R	x	x				
LB2_med	Multitemporal median of Landsat's band 2 over 3 dates	USGS Earth Explorer	O	x	x		x	x	x
LB7_sd	Multitemporal standard deviation of Landsat's band 7 over 3 dates	USGS Earth Explorer	S				x	x	
MaxDiffMean2187	Maximum difference from mean elevation for a maximum search neighbourhood radius of 21 870 m	Derived from DEM	R	x	x	x		x	x
MaxDiffMeanScale2187	Scaled maximum difference from mean elevation for a maximum search neighbourhood radius of 21 870 m	Derived from DEM	R				x	x	x
MaxDiffMeanScale656	Scaled maximum difference from mean elevation for a maximum search neighbourhood radius of 6560 m	Derived from DEM	R		x		x		
MaxElevDevScale2187	Scaled maximum elevation deviation for a maximum search neighbourhood radius of 21 870 m	Derived from DEM	R	x	x		x	x	x
MaxElevDevScale656	Scaled maximum elevation deviation for a maximum search neighbourhood radius of 6560 m	Derived from DEM	R	x	x		x		
MRRTF	Multiresolution index of the ridge top flatness	Derived from DEM	R	x	x		x	x	x
MRVBF	Multiresolution index of valley bottom flatness	Derived from DEM	R	x	x	x	x	x	x
MSP	Mid-slope position	Derived from DEM	R	x	x		x	x	x
RSP	Relative slope position	Derived from DEM	R	x	x		x		
SB12_med	Multitemporal median of Sentinel-2 band 12	Copernicus Open Access Hub	S				x	x	
SSAVI_med	Multitemporal median of the soil adjusted vegetation index using Sentinel-2 images	Copernicus Open Access Hub	S		x		x		
SSoil_color_sd	Multitemporal median of the soil colour index using Sentinel-2 images	Copernicus Open Access Hub	S				x	x	x
SSoil_moist_med	Multitemporal median of the soil moisture over 4 dates using Sentinel-2 images	Copernicus Open Access Hub	S	x	x		x	x	
SLength	Slope length	Derived from DEM	R				x		
SlopeH	Slope height	Derived from DEM	R	x	x		x	x	
VDepth	Valley depth	Derived from DEM	R	x	x		x		
			Total:	20	22	5	25	19	12

Fig. 3. Experimental variograms (dots) of the two predicted features and their best fit model (line). (A) Variogram of the coprogenous layer thickness using 255 square-root-transformed observations. (B) Variogram of the depth to the mineral layer using 4488 observations. The lag was 100 m for both variograms.



concordance and RMSE compared to LOBOCV (data not shown), reinforcing the need of a spatial cross-validation approach. It was also determined that a distance of 1500 m produced better results according to the accuracy metrics from the cross-validation procedure and was used to form spatial blocks. The spatial blocks are shown in Fig. 2.

3.3. Accuracy assessment and model selection

Internal validation results of the three machine learners are summarized in Table 4. Models were compared based on their CCC and RMSE obtained after LOBOCV for the CLT and after hold-out cross-validation for DML. It was clear that CLT models underperformed compared to DML models; CLT models had lower RMSE, but far lower CCC compared to DML models. The Cubist model had the best performance for CLT predictions, with a CCC = 0.07 and RMSE = 30 cm. For the DML predictions, model performance of the Cubist model (CCC = 0.43 and RMSE 48 cm) was similar to the RF model (CCC = 0.40 and RMSE = 44 cm). The difference between the observed and predicted range also led to the selection of the Cubist model as the final DML model. The range of predicted values by Cubist were closer than RFs to the actual range of observed values, although Cubist overpredicted slightly.

Furthermore, a plot of observed and predicted values was generated for both final models (Fig. 4). DML predictions seemed to follow the 1:1 line; however, a shift in the trend was observed for values above 200 cm (Fig. 4A). Concerning the CLT, the datapoints did not follow the 1:1 trendline. Moreover, many sites without a coprogenous layer (observed = 0 cm) had a predicted CLT up to 120 cm (Fig. 4B).

3.4. CLT and the DML predictions and uncertainty estimates

The final CLT map and its 90% confidence interval bounds were produced (Fig. 5). According to Fig. 5B, thicker layers of coprogenous soil were predicted in the two lower peatlands and tended to be thicker towards the center of those peatlands. This trend was consistent with field observations. Nevertheless, no uniform gradient was found, indicating localized accumulation sites across the peatlands. Furthermore, while the peatland in the northwest was reportedly exempt of coprogenous material according to our field sampling survey, the Cubist model predicted a CLT between 0 and 55 cm. The prediction uncertainty maps derived from the bootstrapping technique showed a 90% confidence interval range of 11–304 cm.

Concerning the predicted DML final map (Fig. 6B), the spatial pattern of the thickness of organic material deposits shared similarities with the final CLT map. Yet, high DML values did not always correspond to high CLT values. Furthermore, higher DML values were not exclusively found near the center of each peatland (i.e., the southwest and the center peatlands). Spatial artifacts can be seen, mainly around forested areas, and are related to the covariates used in the models. The southeast peatland was the shallowest on average, with a higher concentration of low DML predictions. Compared to CLT predictions, the DML 90% confidence interval width was narrower (i.e., 2–190 cm).

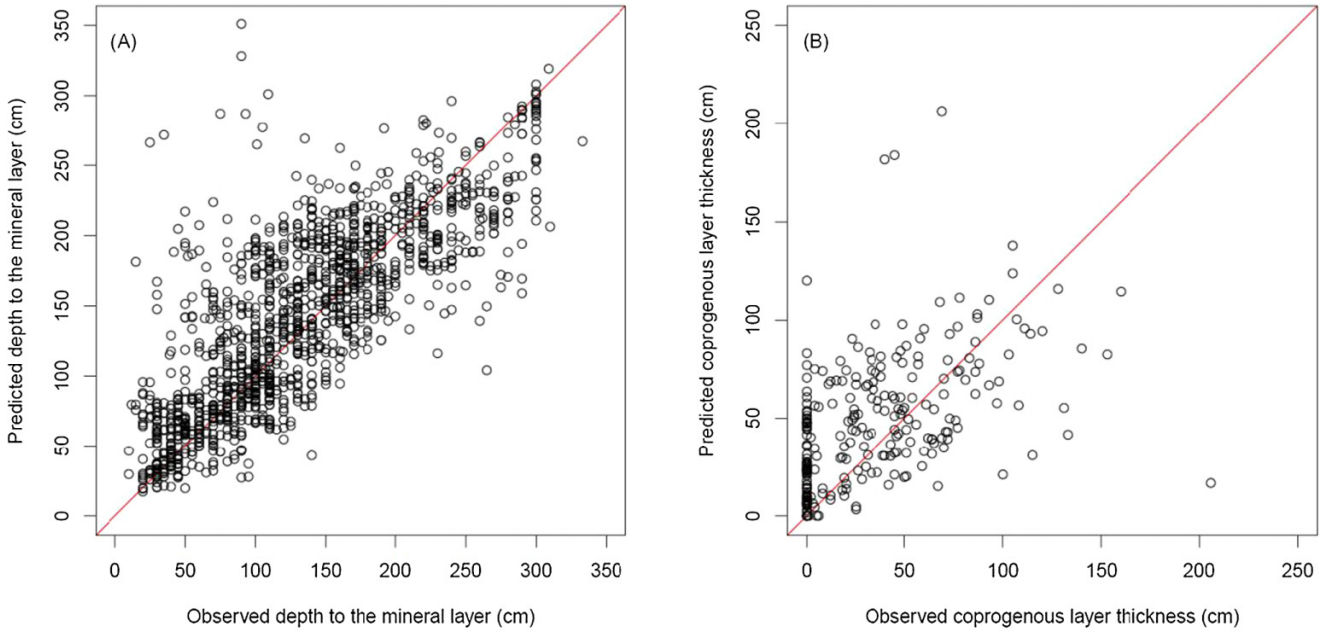
3.5. Variable importance analysis

The relative variable importance plots for both final Cubist models are shown in Fig. 7. These plots show covariate importance in all condition and (or) linear model included in the Cubist tree. They provide information concerning the number of times that a covariate was used in the final model's tree (Kuhn 2020). The contribution of each variable was evaluated with model specific metrics since the final models were both Cubist. For the CLT, MRRTF (100 %) contributed the most to the final model, followed by DIST_MID (86 %), MRVBF (81 %), MSP (69%), and MaxElevDevScale2187 (64 %) for the top 5 covariates. Following that, a gradual decline in variable importance can be seen. Dist_MID (100 %) was the most important covariate in the final DML model, followed by THO (86 %), MRRTF (65 %), DIST_Y (62 %), and Center (49%) for the top 5

Table 4. Tuning and leave-one-block-out cross validation results for the coprogenous layer thickness (CLT) and hold-out cross validation for the depth to the mineral layer (DML) models. Standard deviation of the concordance and of the root mean square of error (RMSE) are in parentheses. RMSE, observed, and predicted ranges are in centimeter. kNN = k-Nearest Neighbor; RF = Random Forest.

N	Feature	Model	Hyperparameter	Concordance	RMSE	Observed range	Predicted range
255	CLT	Cubist	Committees = 10 Neighbors = 1	0.07 (0.24)	30 (32)	[0, 206]	[0, 216]
		kNN	$k = 3$	0.01 (0.11)	34 (26)		[0, 151]
		RF	$m_{try} = 8$	0.08 (0.18)	30 (26)		[0, 138]
4488	DML	Cubist	Committees = 50 Neighbors = 5	0.43 (0.32)	48 (33)	[5, 392]	[5, 417]
		kNN	$k = 4$	0.18 (0.26)	62 (31)		[11, 346]
		RF	$m_{try} = 7$	0.40 (0.23)	44 (22)		[13, 336]

Fig. 4. Plots of the correlation between observed and predicted values for the Cubist model (A) after hold-out cross-validation for the depth to the mineral layer ($N = 1346$), and (B) after leave-one-block-out cross-validation for the coprogenous layer thickness ($N = 255$). A perfect fit is represented by the red 1:1 line.



covariates. A sharp decline in relative importance was observed for the other covariates.

3.6. Predictions of maximum peat thickness and uncertainty estimates

The MPT map at a regional scale was derived by subtracting the CLT map from the DML map to reflect the real thickness of the potentially arable peaty layer (Fig. 8B). MPT ranged from −79 to 367 cm and approximately 0.27% (or 32.05 ha) of the MPT map are cells with a negative MPT. The lower MPT prediction limit ranged between −167 and 271 cm (Fig. 8A), while the upper prediction limit ranged between 9 and 464 cm (Fig. 8C). With the bootstrapping technique and the propagation of error, uncertainty tended to be relatively high and variable across the study area (Fig. 8D). Indeed, the 90% prediction interval varied between 130 and 231 cm.

4. Discussion

4.1. Interpreting model performance

The use of regional, open access covariates has proven useful for predicting DML; however, an underperformance of the CLT models was observed in this study. This was likely due to size and spatial distribution differences between the data sets for the two predicted soil properties. Table A1 further highlights differences found across each peatland with

Fig. 5. Maps of the predicted coprogenous layer thickness (cm) across the study area using the Cubist model and 90% prediction limits derived using bootstrapping ($N = 100$). (A) Lower prediction limit (i.e., 5th percentile), (B) prediction, (C) upper prediction limit (i.e., 95th percentile), and (D) prediction interval width.

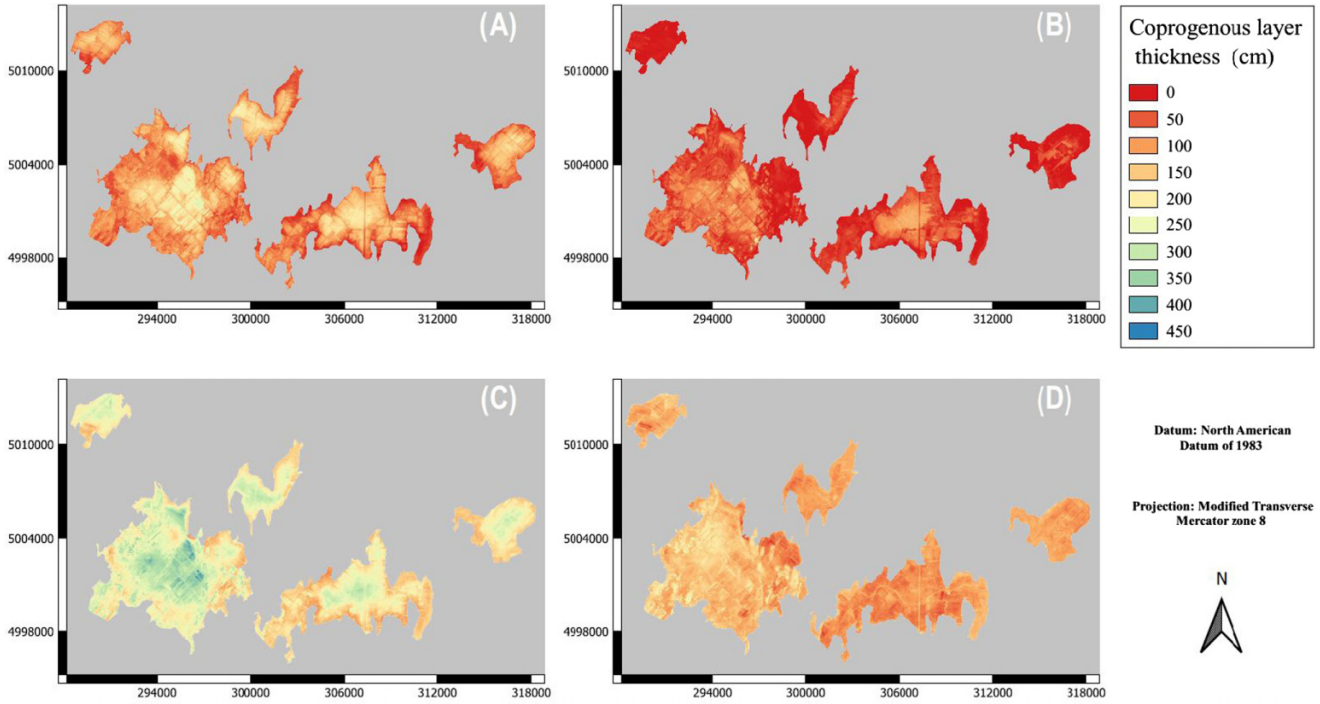


Fig. 6. Maps of the predicted depth to the mineral layer (cm) across the study area using the Cubist model and 90% prediction limits derived using bootstrapping ($N = 100$). (A) Lower prediction limit (i.e., 5th percentile), (B) prediction, (C) upper prediction limit (i.e., 95th percentile), and (D) prediction interval width.

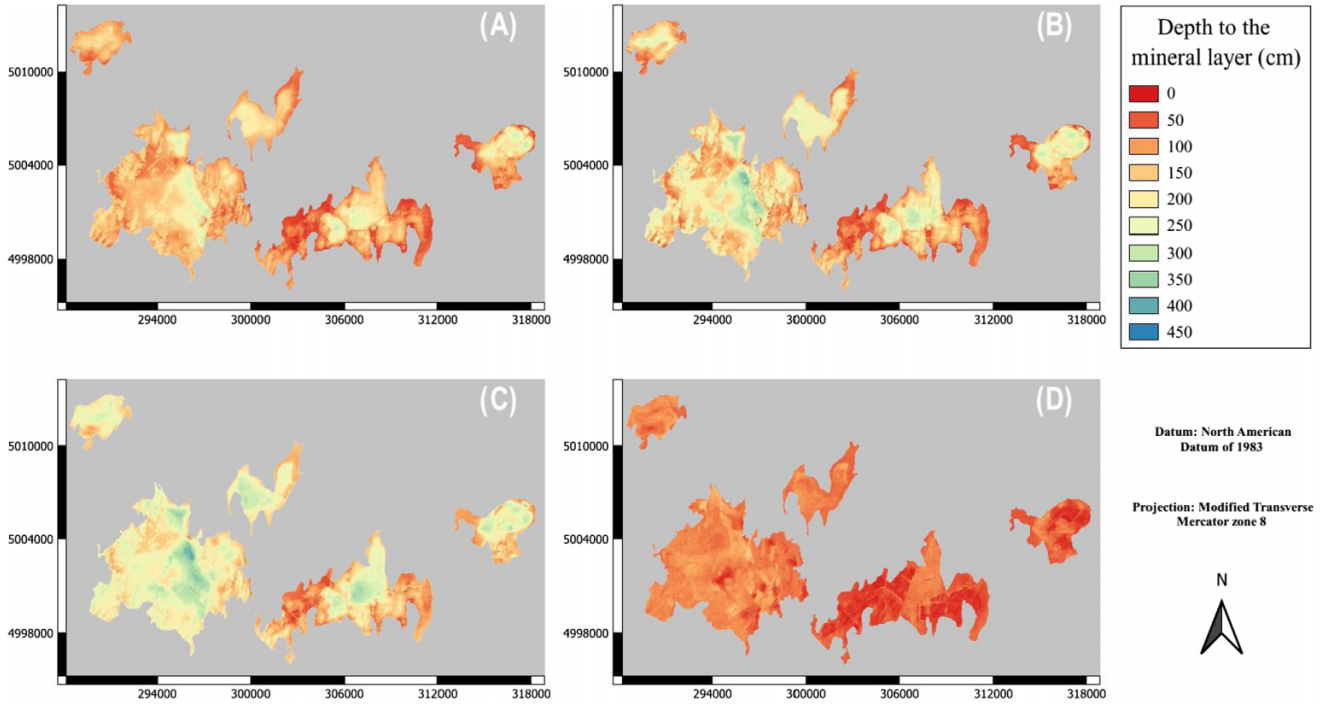


Fig. 7. Relative importance of each covariate (%) on the final Cubist model (A) of the depth to the mineral layer and (B) of the coprogenous layer thickness. Refer to Table 3 for a description of covariates' abbreviations.

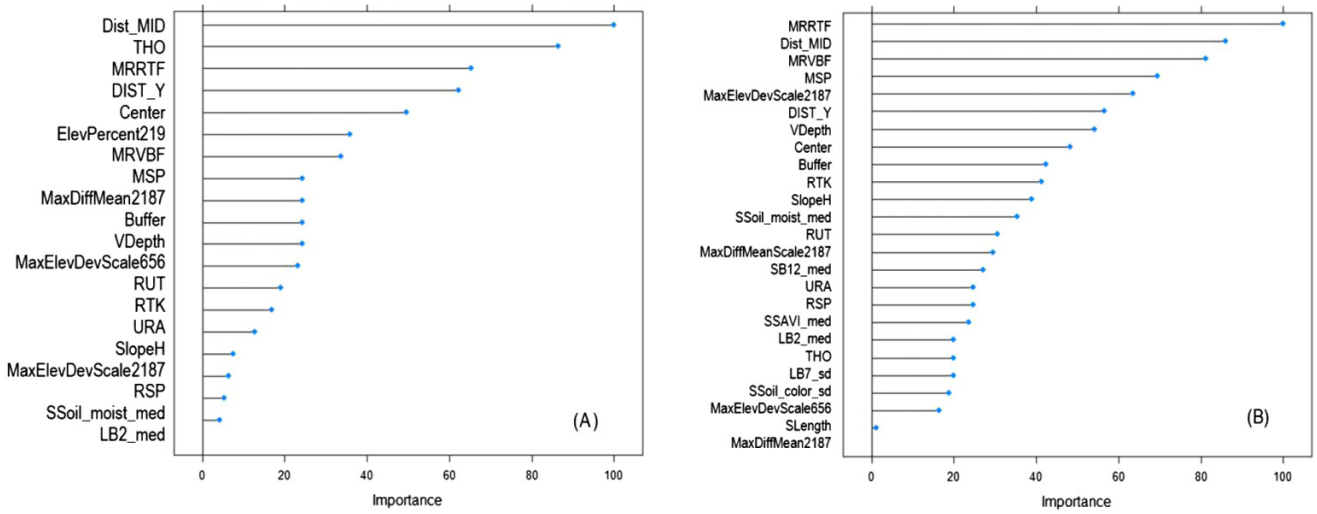
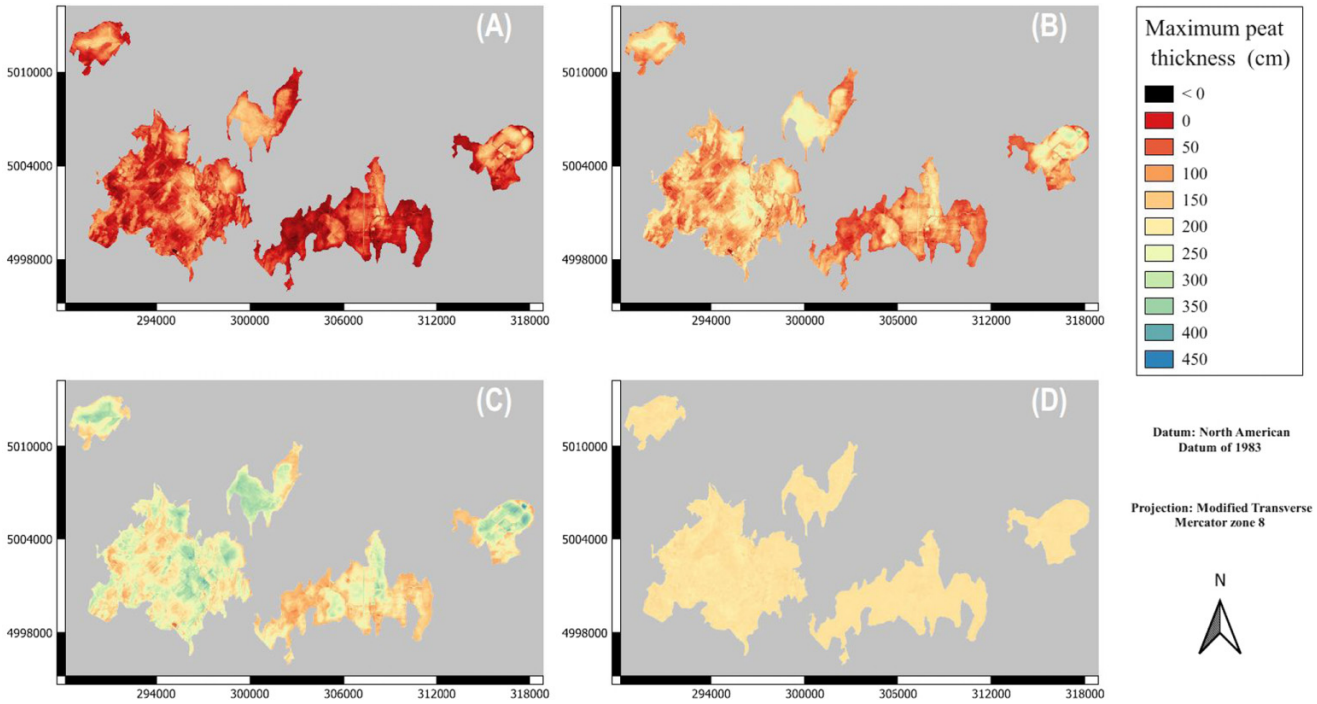


Fig. 8. Maximum peat thickness map obtained by subtracting the coprogenous layer thickness to the depth to the mineral layer and 90% prediction limits derived using bootstrapping ($N = 100$). (A) Lower prediction limit (i.e., 5th percentile), (B) prediction, (C) upper prediction limit (i.e., 95th percentile), and (D) prediction interval width.



regards to the descriptive statistics of both predicted features. Here, the number of observations, the range, and the coefficient of variation of each peatland may have affected the model performance. For the sake of simplicity, a single model was trained for the entire study area, although the CLT and DML spatial gradients differed between peatlands. The DML model was able to moderately generalize across all peatlands (CCC = 0.43 and RMSE = 48 cm), which was not the case for the CLT model (CCC = 0.07 and RMSE = 30 cm). The high coefficient of variation of the CLT (109%, Table 1) indicated high heterogeneity in the data set, compared to that of the DML (52%). This can be explained by the fact that CLT could be considered as a zero-inflated variable with a wide range of values. A large proportion of the data set was composed of sites without a coprogenous layer, as seen in Fig. 4*b*. Furthermore, data clustering and poor spatial coverage of the study

area's extent could have limited the model performance in areas with a lower sampling density, especially when considering variable importance which showed that spatial distance-based predictors were relatively important.

Nonetheless, our prediction errors were comparable to other studies. Rudiyanto et al. (2018) tested 14 machine-learning models to map the peat thickness in Indonesian peatlands, where DML was predicted with a RMSE ranging from 1.8 to 2.8 m at a regional scale (50 000 ha), while Gatis et al. (2019) obtained a RMSE of 0.31 m for a 40 600 ha study area. These RMSE values were of the same magnitude as the RMSE values observed in this study (Table 4). Moreover, Rudiyanto et al. (2018) had better results with Cubist and RF models compared to the other tested models. Considering the number of observations (1779 and 159) and the peat depth range (0–7 and 0–12 m) in Gatis et al. (2019) and Rudiyanto et al. (2018), respectively, our CLT and DML models showed similar model performances when treated separately. Since both studies reported R^2 and we used CCC, we cannot make direct comparisons, but both studies clearly outperformed ours. This may be due to the use of spatial cross-validation instead of standard cross-validation. As recently described in Wadoux et al. (2021), spatial cross-validation of autocorrelated data are not always the best solution to obtaining a representative assessment of the bias of a model. Standard cross-validation might yield better results without being overoptimistic. This could partially explain why the CCC obtained with LOBOCV for the CLT model was as low as 0.07, but standard cross-validation yielded a CCC = 0.65, and RMSE = 51.6 cm. The DML model had a higher CCC = 0.95 and a lower RMSE = 21.6 cm using standard cross-validation than when using LOBOCV. The latter cross-validation technique was preferred due to its more conservative estimates of model performance and to account for spatial autocorrelation of the data.

4.2. Variable importance analysis

The most important DML predictors aligned with other studies, while the interpretation of CLT predictors was limited by the model performance. Results from Rudiyanto et al. (2018) showed that MRVBF was ranked as the fourth most important covariate among those tested. This supports the importance of this DEM-derived covariate on the prediction of deposited materials. Some soil-forming factors affect limnic and organic deposits differently depending on the scale at which phenomenon are studied (Behrens et al. 2018). As stated by Gallant and Dowling (2003), MRRTF and MRVBF can be important predictors for hydrologic and geomorphic processes that are related to valley, depressions, and slopes, like sedimentary soil deposits (i.e., coprogenous material and subsequently peat deposits). Multiresolution covariates can provide this information to the model to allow a better understanding of the magnitude of the deposits. This was likely why so many topographic and scale-related covariates were retained in the final models of this study.

The Euclidean distance to the center of the study area (DIST_MID), the Euclidean distance to the north of the study area (DIST_Y) and the distance to the center of a given peatland (Center) were also important predictors, meaning higher deposits are generally found near the center and thinner deposits near the peatland border, but also that there is a general gradient from a peatland to another. Furthermore, we expected that gamma radiometric covariates would contribute to delineating the peat extent and predicting peat thickness (Minasny et al. 2019). Indeed, the most important gamma-radiometric covariate was ranked as the second most important variable (Fig. 7A) in the DML model. The same authors also suggested that multi-temporal satellite covariate might provide useful information on peat thickness. In this study, most of the Landsat 8 and Sentinel-2 predictors were highly correlated, possibly due to a short period from when the imagery was acquired. Moreover, many Landsat and Sentinel covariates were not retained from RFE, possibly due to their low effectiveness as predictors. While the multi-temporal median of the soil moisture index from Sentinel-2 had a variable importance of 35%, little can be interpreted from this due to the low accuracy of the CLT model. A weak relationship between CLT and the suite of covariates could suggest that the model failed to capture the variability of CLT at a regional scale. Other sampling surveys should be made at a smaller spatial extent (i.e., individual farm) and the potential benefits of proximal soil sensing tools for producing more relevant covariates should be evaluated.

The study area encompassed four pedological surveys for which digital maps were freely available online. Legacy soil data were considered at first due to their popular use in DSM but were not included as covariates due to temporal incoherence between the surveys (1950, 2000, 2001, and 2014). The evolution of organic soils over time can lead to substantial biases and render maps less useful, even though they can be updated with modern techniques (Kempen et al. 2009). Otherwise, this covariate could have been useful to predict areas with coprogenous soil and to delineate peat extent.

4.3. Predictions of maximum peat thickness and uncertainty estimates

The results suggested a potential limited use of the MPT map produced at a regional scale for soil management purposes at the field-scale given the high uncertainty of the combined CLT and DML predictions. Artifacts in the DML and CLT maps created sharp transitions between neighbouring cells that were not observed in nature and may require further field verifications. These issues may affect interpretation of the final MPT map for fields near the artifacts and the propagation of error may occur when combining maps with such artifacts. These errors may limit the use of such mapping products for precision agriculture and conservation projects. Moreover, the accuracy was not sufficient to predict at which depth tile drainage could be installed or if soil conservation MPT thresholds for management are reached (for details, see Deragon et al. 2022).

Negative values were observed in a small portion of the final map. This was due to a predicted coprogenous layer being thicker than the predicted depth to the mineral layer. As previously stated, higher imprecision in the CML model may

have been linked to a poor coverage of the study area; furthermore, a weak relationship between covariates and the predicted feature could be responsible for model uncertainty. Peatlands differed quite significantly on their average MPT. For instance, the southeast peatland showed the lowest average MPT compared to the other ones. This has major implications from a soil conservation perspective, because degraded and shallow soils are not uniformly distributed nor confined to a peatlands' border. Therefore, farmers from different peatlands are not affected equally by the shallowness of their soil and will have to use different soil conservation measures accordingly to their situation.

Many sources of uncertainty were present in this study (Heuvelink 2017). Two methods of sampling with a different precision were combined. In addition, the data set from 2010 was mainly clustered in two peatlands. Since model tuning inevitably fits the model on a majority of points coming from those peatlands, the final model might not be as effective in predicting the features found in the other three peatlands if their formation and evolution differs significantly (i.e., botanical origin, soil management, groundwater level variations based on differences related to the watershed and elevation, average number of years since conversion to agriculture, etc.). Preprocessing of the original data to produce the covariates may have introduced error in the model as well (i.e., resampling, projecting, and smoothing raster layers). As stated in Samuel-Rosa et al. (2015), the number and the location of calibration points matter; furthermore, the covariates are only approximations of the real-world soil-forming conditions and thereby inherently prone to errors. These errors are complicated if not impossible to quantify and are propagated at each new step of the workflow. A better sampling design would have been needed from the start to cover the full feature space of DML and CLT of the study area, and a better spatial distribution of observation points. Unfortunately, little was known about the DML and CLT in three of the peatlands prior to the start of the study. Therefore, the MPT map should be used with caution, as one decision tool among others to manage organic soil conservation. Yet, not considering coprogenous materials would greatly overestimate the remaining soil resource that can be used to produce crops. Moreover, despite important errors in depth estimates, digital mapping results were consistent with the observed distribution of the DML and CLT.

5. Conclusion

The use of freely available covariates and DSM techniques provided the first maps of the depth to the mineral layer, coprogenous layer thickness, and maximum peat thickness covering five peatlands at a regional scale. Although their prediction error was comparable to other DSM studies, the CLT model did not achieve a sufficient accuracy to produce CLT and MPT maps of similar precision to site sampling for now and further work is needed. Thus, we were not able to elaborate on the contribution of the individual covariates on the CLT model performance. Although the DML map provides more accurate information as a tool to determine priority intervention zones, MPT still remains a key metric to guide soil conservation practices. The produced MPT map should serve as a baseline to be built upon and improved in future research. For instance, proximal sensing tools could be investigated as a more relevant source of covariate data to produce coprogenous layer thickness maps at a local scale instead of relying on remote sensing tools at a regional scale. Such could provide new insight on the CLT spatial variability and would also improve the accuracy of the MPT map.

Acknowledgements

The authors are thankful for the financial support from a Canadian Graduate Scholarship program by the Natural Sciences and Engineering Research Council of Canada (NSERC) and a Master's Scholarship program (B1X) by the Fonds de recherche du Québec—Nature et technologies granted to R. Deragon.

We also acknowledge the financial support of the NSERC through an Industrial Research Chair Grant in Conservation and Restoration of Cultivated Organic Soils (IRCPJ 411630–17) in partnership with Delfland Inc., Productions maraîchères Breizh inc., La Production Barry inc., Les Fermes R.R. et fils inc., Le Potager Montréalais ltée, R. Pinsonneault et fils ltée, Patate Isabelle inc., Les Fermes du Soleil Inc., Les Jardins A. Guérin et Fils inc., Le Potager Riendeau inc., Vert Nature Inc., Fermes Hotte et Van Winden Inc., Production Horticole Van Winden, and Maraichers J.P.L. Guerin & fils.

The authors would also like to thank Michaël Leblanc and Lucie Grenon for their early contribution to this project and John Lindsay for his advice on DEM preprocessing.

Article information

History dates

Received: 24 February 2022
Accepted: 10 July 2022
Accepted manuscript online: 5 January 2023
Version of record online: 5 January 2023

Notes

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

Copyright

Data availability

Data generated or analyzed during this study are not available due to the nature of this research. This data are of strategic importance for the research partners and is considered sensitive information.

Author information

Author ORCIDs

Raphaël Deragon https://orcid.org/0000-0002-2912-386X
Daniel D. Saurette https://orcid.org/0000-0002-1971-1238

Author notes

An earlier version of this article is available online in a dissertation that was posted to CorpusUL, an institutional repository (permalink: http://hdl.handle.net/20.500.11794/71034).
Daniel D. Saurette and Brandon Heung served as a Guest Editor at the time of manuscript review and acceptance; peer review and editorial decisions regarding this manuscript were handled by Angela Bedard-Haughn.

Author contributions

Conceptualization: RD, DDS, BH, JC
Data curation: RD, DDS, BH, JC
Formal analysis: RD, DDS, BH
Funding acquisition: RD, JC
Investigation: RD, DDS, BH
Methodology: RD, DDS, BH
Project administration: RD, DDS, BH, JC
Resources: RD, DDS, JC
Software: RD, DDS, BH
Supervision: DDS, BH, JC
Validation: RD, DDS, BH, JC
Visualization: RD, DDS, BH, JC
Writing – original draft: RD, DDS, BH, JC
Writing – review & editing: RD, DDS, BH, JC

Competing interests

The authors declare there are no competing interests.

References

Beamish, D. 2013. Gamma ray attenuation in the soils of northern ireland, with special reference to peat. J. Environ. Radioact. **115**: 13–27. doi:10.1016/j.jenvrad.2012.05.031. PMID: 22858640.

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., and MacMillan, R.A. 2018. Spatial modelling with euclidean distance fields and machine learning. Eur. J. Soil Sci. **69**(5): 757–770. doi:10.1111/ejss.12687.

Beucher, A., Koganti, T., Iversen, B.V., and Greve, M.H. 2020. Mapping of peat thickness using a multi-receiver electromagnetic induction instrument. Remote Sens., **12**(15): 2458–2458. doi:10.3390/rs12152458.

Beven, K.J., and Kirkby, M.J. 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. **24**: 43–69. doi:10.1080/02626667909491834.

Bian, Z., Guo, X., Wang, S., Zhuang, Q., Jin, X., Wang, Q., and Jia, S. 2020. Applying statistical methods to map soil organic carbon of agricultural lands in northeastern coastal areas of China. Arch. Agron. Soil Sci. **66**(4): 532–544. doi:10.1080/03650340.2019.1626983.

Breiman, L. 2001. Random forests. Mach. Learn., **45**: 5–32. doi:10.1023/A:1010933404324.

Brenning, A., Bangs, D., and Becker, M. 2018. RSAGA: SAGA Geoprocessing and Terrain Analysis. R package version 1.3.0. Available from: https://CRAN.R-project.org/package=RSAGA

Comas, X., Terry, N., Slater, L., Warren, M., Kolka, R. Kristijono, A., et al. 2015. Imaging tropical peatlands in indonesia using ground penetrating radar (GPR) and electrical resistivity imaging (ERI): implications for carbon stock estimates and peat soil characterization. Biogeosciences, **11**(10): 2995–3007. doi:10.5194/bg-12-2995-2015.

Deragon, R., Julien, A.-S., Dessureault-Rompre´, J., and Caron, J. 2022. Using cultivated organic soils' depth to form soil conservation management zones. Can. J. Soil Sci. **102**(3): 633–650. doi:10.1139/cjss-2021-0148.

Dessureault-Rompre´, J., Libbrecht, C., and Caron, J. 2020. Biomass crops as a soil amendment in cultivated histosols: can we reach carbon equilibrium? Soil Sci. Soc. Am. J. **84**(2): 597–608. doi:10.1002/saj2.20051.

Escadafal, R. 1994. Soil spectral properties and their relationships with environmental parameters-examples from arid regions. In *Imaging Spectrometry—A Tool for Environmental Observations*. Springer: Dordrecht, The Netherlands, pp. 71–87.

Esselami, D., Boudache, M., and Grenon, L. 2014. L'évolution des terres noires et le problème de la compaction. Prisme Consortium. Présentation donnée aux journées horticoles. [In French] Available from: https://www.mapaq.gouv.qc.ca/SiteCollectionDocuments/Regions/Monteregie-Ouest/Journees_horticoles_2014/4_decembre/Terres_noires/9h05_b_JH2014_profil_compaction_DEssalami.pdf.

FAO 2020. Soil maps and databases: other global soil maps and databases. [Web page] Available from: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/other-global-soil-maps-and-databases/en/.

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., and Biswas, A. 2021. Effect of multi-temporal satellite images on soil moisture prediction using a digital soil mapping approach. Geoderma, **385**: 114901. doi:10.1016/j.geoderma.2020.114901.

Gallant, J.C., and Dowling, T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. **39**/**12**: 1347–1359.

Gao, B.-c. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. **58**(3): 257–266. doi:10.1016/S0034-4257(96)00067-3.

Gatis, N., Luscombe, D.J., Carless, D., Parry, L.E., Fyfe, R.M. Harrod, T.R., et al. 2019. Mapping upland peat depth using airborne radiometric and lidar survey data. Geoderma, **335**: 78–87. doi:10.1016/j.geoderma.2018.07.041.

Genuer, R., and Poggi, J.-M. 2020. Random Forests with R (Ser. Use r!). Springer. doi:10.1007/978-3-030-56485-8.

Gholizadeh, A., Daniel, Ž., Saberioon, M., and Luboš, B. 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and sentinel-2 spectral imaging. Remote Sens. Environ. **218**: 89–103. doi:10.1016/j.rse.2018.09.015.

Gräler, B., Pebesma, E., and Heuvelink, G. 2016. Spatio-temporal interpolation using gstat. R Journal, **8**(1): 204–218. doi:10.32614/RJ-2016-014.

Grenon, L. 1988. Répartition des terres humides dans la plaine du saint-laurent. Équipe pédologique du québec, centre de recherches sur les terres. Direction générale de la recherche, agriculture canada, sainte-foy. 3 cartes à l'échelle, **1**: 250000.

Groupe AGÉCO. 2007. Portrait et priorités du secteur maraîcher québecois; Rapport final. [En ligne] Available from: https://www.mapaq.gouv.qc.ca/fr/Publications/Portrait_secteurmaraicher.pdf.

Hastie, T., Tibshirani, R., and Friedman, J.H. 2009. The elements of statistical learning: data mining, inference, and prediction (Second edition, corrected 7th printing, Ser. Springer series in statistics). Springer.

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., and Schmidt, M.G. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma, **265**: 62–77. doi:10.1016/j.geoderma.2015.11.014.

Heuvelink, G. 2017. *Uncertainty. Soil Organic Carbon Mapping Cookbook*. 1st ed. Rome, Food and Agriculture Organisation of the United Nations, pp. 109–121.

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

Ilnicki, P. 2003. *Agricultural Production Systems for Organic Soil Conservation. Organic Soils and Peat Materials for Sustainable Agriculture*. CRC Press, Boca Raton, Florida, pp. 209–221.

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2014. *An Introduction to Statistical Learning : with Applications in R (Corrected at 4th printing 2014, Ser. Springer texts in statistics)*. Springer.

Keaney, A., McKinley, J., Graham, C., Robinson, M., and Ruffell, A. 2013. Spatial statistics to estimate peat thickness using airborne radiometric data. Spat. Stat., **5**: 3–24. doi:10.1016/j.spasta.2013.05.003.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., and Stoorvogel, J.J. 2009. Updating the 1:50,000 dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma, **151**(3-4): 311–326. doi:10.1016/j.geoderma.2009.04.023.

Khan, N.M., Rastoskuev, V.V., Sato, Y., and Shiozawa, S. 2005. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. Agric. Water Manag. **77**: 96–109. doi:10.1016/j.agwat.2004.09.038.

Kroetsch, D.J., Geng, X., Chang, S.X., and Saurette, D.D. 2011. Organic soils of canada: part 1. Wetland organic soils. Can. J. Soil Sci. **91**(5): 807–822. doi:10.4141/cjss10043.

Ku, H.H. 1966. Notes on the use of propagation of error formulas. J. Res. Natl. Bur. Stand. **70**(4): 263–273.

Kuhn, M. 2020. caret: Classification and Regression Training. R package version 6.0-86. Available from: https://CRAN.R-project.org/package=caret

Kuhn, M., and Johnson, K. 2013. *Applied Predictive Modeling*. Springer. doi:10.1007/978-1-4614-6849-3.

Lamontagne, L., Martin, A., and Nolin, M.C. 2014. Étude pédologique du comté de Napierville (Québec). Laboratoires de pédologie et d'agriculture de précision, Centre de recherche et de développement sur les sols et les grandes cultures, Direction générale des sciences et de la technologie, Agriculture et Agroalimentaire Canada, Québec (Québec).

LaSalle, P. 1963. Géologie de la région de verchères. Dépôts meubles. Rapport préliminaire. Ministère des Richesses Naturelles du Québec. 9 pages + 2 cartes au **1**: 50 000.

Lawrence, I., and Lin, K. 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics, **45**: 255–268. PMID: 2720055.

Lindsay, J.B. 2016. Efficient hybrid breaching-filling sink removal methods for flow path enforcement in digital elevation models. Hydrol. Processes, **30**(6): 846–857. doi:10.1002/hyp.10648.

Lindsay, J.B., Cockburn, J.M.H., and Russell, H.A.J. 2015. An integral image approach to performing multi- scale topographic position analysis. Geomorphology, **245**: 51–61. doi:10.1016/j.geomorph.2015.05.025.

Malone, B.P., Minasny, B., and McBratney, A.B. 2017. Using r for digital soil mapping (Ser. Progress in soil science). Springer Nature. doi:10.1007/978-3-319-44327-0.

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

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., and Nauss, T. 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environmental Modelling & Software, **101**: 1–9.

Minasny, B., and McBratney, A.B. 2006. A conditioned latin hypercube method for sampling in the presence of ancillary information. Computers and Geosciences, **32**(9): 1378–1388. doi:10.1016/j.cageo.2005.12.009.

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., Vries, de F. Gimona, A., et al. 2019. Digital mapping of peatlands – a critical review. Earth Sci. Rev. **196**: 102870. doi:10.1016/j.earscirev.2019.05.014.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., and Toxopeus, A.G. 2014. Where is positional uncertainty a problem for species distribution modelling?, Ecography, **37**(2): 191–203. doi:10.1111/j.1600-0587.2013.00205.x.

Natural Resources Canada. 2019. Geoscience data repository for geophysical data. Magnetic-Radiometric- EM datasets [WWW Document]. URL. doi: http://gdr.agg.nrcan. gc.ca/gdrdap/dap/search-eng.php.

Nellis, M.D., and Briggs, J.M. 1992. Transformed vegetation index for measuring spatial variation in drought impacted biomass on konza prairie, kansas. Transactions of the Kansas Academy of Science (1903-), **95**(1-2): 93–99. doi:10.2307/3628024.

O'Brien, R. 2007. A caution regarding rules of thumb for variance inflation factors. Quality and Quantity, **41**(5): 673–690. doi:10.1007/s11135-006-9018-6.

Oliver, M.A., and Webster, R. 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. Catena, **113**: 56–69. doi:10.1016/j.catena.2013.09.006.

Parent, L.-É., and Gagné., G. 2010. Guide de référence en fertilisation (2e éd.). Québec, Canada: Centre de référence en agriculture et agroalimentaire du Québec (CRAAQ).

Parry, L.E., West, L.J., Holden, J., and Chapman, P.J. 2014. Evaluating approaches for estimating peat depth. Journal of Geophysical Research: Biogeosciences, **119**(4): 567–576. doi:10.1002/2013JG002411.

Pebesma, E.J. 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. **30**: 683–691.

Ploton, P., Mortier, F., Re´jou-Me´chain, M., Barbier, N., Picard, N. Rossi, V., et al. 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. **11**(1): 4540. doi:10.1038/s41467-020-18321-y. PMID: 32917875.

Poggio, L., and Gimona, A. 2017. Assimilation of optical and radar remote sensing data in 3d mapping of soil properties over large areas. Sci. Total Environ. **579**: 1094–1110. doi:10.1016/j.scitotenv.2016.11.078. PMID: 27923574.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., and Heikkonen, J. 2017. Estimating the prediction performance of spatial models via spatial k-fold cross validation. International Journal of Geographical Information Systems, **31**(10): 2001–2019. doi:10.1080/13658816.2017.1346255.

QGIS.org, 2021. QGIS Geographic Information System. QGIS Association. http://www.qgis.org.

Quinlan, J.R. 1992. Learning with Continuous Classes. Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart 16-18 November 1992, 343-348.

Quinlan, J.R. 1993. Combining instance-based and model-based learning. Proceedings of the Tenth International Conference on Machine Learningpp. 236–243.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rikimaru, A., Roy, P.S., and Miyatake, S. 2002. Tropical forest cover density mapping. Trop. Ecol. **43**: 39–47. [Google Scholar]

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J. Guillera-Arroita, G., et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, **40**(8): 913–929. doi:10.1111/ecog.02881.

Rosa, E., Larocque, M., Pellerin, S., Gagné, S, and Fournier, B. 2009. Determining the number of manual measurements required to improve peat thickness estimations by ground penetrating radar. Earth Surf. Processes Landforms, **34**(3): 377–383. doi:10.1002/esp.1741.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W. 1974. Monitoring vegetation systems in the great plains with ERTS. NASA special publication, **351**(1974): 309.

Rudiyanto, Minasny, B., Setiawan, B.I., Saptomo, S.K., and McBratney, A.B. 2018. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. Geoderma, **313**: 25–40. doi:10.1016/j.geoderma.2017.10.018.

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

Saurette, D. 2021. onsoilsurvey: Making PDSM in Ontario Better. R package version 0.0.0.9000.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A. 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol. Modell. **406**: 109–120. doi:10.1016/j.ecolmodel.2019.06.002.

Siemon, B., Ibs-von Seht, M., and Frank, S. 2020. Airborne electromagnetic and radiometric peat thickness mapping of a bog in northwest germany (ahlen-falkenberger moor). Remote Sensing, **12**(2): 203–203. doi:10.3390/rs12020203.

Soil Classification Working Group (SCWG). 1998. The Canadian System of Soil Classification, 3rd edition. Agriculture and Agri-Food Canada, Publication **1646**: 187pp.

Taghizadeh-Mehrjardi, R., Emadi, M., Cherati, A., Heung, B., Mosavi, A., and Scholten, T. 2021. Bio-Inspired hybridization of artificial neural networks: an application for mapping the spatial distribution of soil texture fractions. Remote Sens. **13**: 1025. doi:10.3390/rs13051025.

Tveite, H. 2018. The QGIS Multi-distance buffer Plugin, Version 3.2.4. http://plugins.qgis.org/plugins/MultiDistanceBuffer/.
Vepraskas, M.J., and Craft, C.B. 2015. Wetland Soils: Genesis, Hydrology, Landscapes, and Classification(2nd ed.): CRC Press.
Wadoux, A.M.J.C., Heuvelink, G.B.M., Bruin, D.S., and Brus, D.J. 2021. Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Modell. **457**: 109692. doi:10.1016/j.ecolmodel.2021.109692.
Wu, Q. 2020. whitebox: 'WhiteboxTools' R Frontend. R package version 1.4.0/r27. https://R-Forge.R-project.org/projects/whitebox/.
Young, N.E., Anderson, R.S., Chignell, S.M., Vorster, A.G., Lawrence, R., and Evangelista, P.H. 2017. A survival guide to landsat preprocessing. Ecology, **98**(4): 920–932. doi:10.1002/ecy.1730. PMID: 28072449.
Zhang, C., and Ma, Y. 2012. Ensemble machine learning: methods and applications. Springer. doi:10.1007/978-1-4419-9326-7.

Appendix A

Table A1. Summary statistics of the coprogenous layer thickness (CLT) and of the depth to the mineral layer (DML) for each of the five peatlands. Refer to Fig. 2 for peatland identification numbers.

Peatland:	1		2		3		4		5	
Feature:	CLT	DML	CLT	DML	CLT	DML	CLT	DML	CLT	DML
N	22	22	7	7	9	814	157	451	60	3194
Minimum (cm)	0	34	0	192	0	30	0	17	0	5
Maximum (cm)	0	260	281	284	223	310	287	392	172	300
Mean (cm)	0	131	74	236	89	178	81	181	88	119
Median (cm)	0	112	0	246	95	185	71	170	92	110
Standard deviation (cm)	NA	63.68	126.38	33.25	94.12	64.14	76.00	78.74	40.73	63.98
Coefficient of variation (%)	NA	48.78	171.78	14.10	105.36	36.09	93.39	43.52	46.24	53.69
Variance (cm^2)	NA	4055	15 971	1105	8859	2888	5776	6630	1659	2584
Skewness	NA	0.64	0.78	−0.03	0.25	−0.02	0.55	0.17	−0.22	0.58
Kurtosis	NA	−0.64	−1.51	−1.68	−1.81	−0.93	−0.81	−0.73	−0.40	−0.11

Note: No coprogenous layer was found in peatland 1. Therefore, no summary statistics can be computed.