

## Article

# Digital Mapping of Soil Organic Matter and Cation Exchange Capacity in a Low Relief Landscape Using LiDAR Data

Shams R. Rahmani <sup>1</sup>, Jason P. Ackerson <sup>2</sup>, Darrell Schulze <sup>1</sup>, Kabindra Adhikari <sup>3</sup> and Zamir Libohova <sup>4,\*</sup>

<sup>1</sup> Department of Agronomy, Purdue University, 915 West State Street, West Lafayette, IN 47907, USA; rahmani@purdue.edu (S.R.R.); dschulze@purdue.edu (D.S.)

<sup>2</sup> Soil Health Institute, 2803 Slater Road Suite 115, Morrisville, NC 27560, USA; jackerson@soilhealthinstitute.org

<sup>3</sup> Grassland, Soil & Water Research Laboratory, United States Department of Agriculture, Agricultural Research Service, 808 East Blackland Road, Temple, TX 76502, USA; kabindra.adhikari@usda.gov

<sup>4</sup> Dale Bumpers Small Farms Research Center, United States Department of Agriculture, Agricultural Research Service, 6883 South State Highway 23, Booneville, AR 72927, USA

\* Correspondence: zamir.libohova@usda.gov; Tel.: +1-(402)-318-1146

**Abstract:** Soil organic matter content (SOM) and cation exchange capacity (CEC) are important agronomic soil properties. Accurate, high-resolution spatial information of SOM and CEC are needed for precision farm management. The objectives of this study were to: (1) map SOM and CEC in a low relief area using only lidar elevation-based terrain attributes, and (2) compare the prediction accuracy of SOM and CEC maps created by universal kriging, Cubist, and random forest with Soil Survey Geographic (SSURGO) database. For this study, 174 soil samples were collected from a depth from 0 to 10 cm. The topographic wetness index, topographic position index, multi resolution valley bottom flatness, and multi resolution ridge top flatness indices generated from the lidar data were used as covariates in model predictions. No major differences were found in the prediction performance of all selected models. For SOM, the predictive models provided results with coefficient of determination ( $R^2$ ) (0.44–0.45), root mean square error (RMSE) (0.8–0.83%), bias (0–0.22%), and concordance correlation coefficient ( $\rho_c$ ) (0.56–0.58). For CEC, the  $R^2$  ranged from 0.39 to 0.44, RMSE ranged from 3.62 to 3.74  $\text{cmol}_c \text{kg}^{-1}$ , bias ranged from 0–0.17  $\text{cmol}_c \text{kg}^{-1}$ , and  $\rho_c$  ranged from 0.55 to 0.57. We also compared the results to the USDA Soil Survey Geographic (SSURGO) data. For both SOM and CEC, SSURGO was comparable with our predictive models, except for few map units where both SOM and CEC were either under or over predicted.

**Keywords:** soil property modeling; precision agriculture; soil survey



**Citation:** Rahmani, S.R.; Ackerson, J.P.; Schulze, D.; Adhikari, K.; Libohova, Z. Digital Mapping of Soil Organic Matter and Cation Exchange Capacity in a Low Relief Landscape Using LiDAR Data. *Agronomy* **2022**, *12*, 1338. <https://doi.org/10.3390/agronomy12061338>

Academic Editor: Louis Kouadio

Received: 20 April 2022

Accepted: 30 May 2022

Published: 31 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## Core Ideas

- In a low relief environment, SOM and CEC variation were captured using lidar elevation data
- Universal kriging, random forest, and cubits predictive models performed similarly
- Predictive models provided more detailed spatial distribution of SOM and CEC compared to SSURGO

## 1. Introduction

Soil properties vary over space mainly related to factors such as climate, organisms, relief or topography, parent material and time [1]. In a relatively small area such as in our study site (~570 ha), most of these factor (also known as soil forming factors [1]) are considered constant with topography being the main driving force of soil variation. This is embodied in the catena concept [2] which states that soils follow predictable and repeatable pattern based on topography.

SOM content and CEC are important indicators for soil health and fertility. These properties influence plant growth and performance through soil water and soil nutrient

redistribution. Therefore, accurate and spatially detailed maps of SOM and CEC are needed to support research decisions regarding crops and soil management at this study site. Also, detailed and accurate spatial soil information is needed for agricultural and ecological decision-making. Conventional, polygon-based soil maps are the main data source for these applications. In the United States, the Soil Survey Geographic (SSURGO) database [3] and the State Soil Geographic (STATSGO) database [3] maintained by the Natural Resources Conservation Service (NRCS) are extensively used for many applications. These polygon-based maps were originally developed for land management and may not be suitable for spatially accurate soil property maps [4] at field or farm levels. Map unit polygons often contain more than one major soil component as well as a few minor soil components, which reduces the map unit purity. For applications such as precision crop management and high throughput phenomics research, more detailed soil maps are needed that are currently available from the SSURGO database.

Digital soil mapping (DSM) is an approach for overcoming the limitations of conventional soil polygon maps and for improving the accuracy of soil property predictions at a finer resolution [5]. Digital soil maps are generated using different techniques (including geo-statistical algorithms) and stored within a geographic information system (GIS), which allows data to be used for further analysis and interpretation [6].

In this study, our goal was to map SOM content and CEC at the Purdue University Agronomy Center for Research and Education (ACRE). SOM content and CEC are important for plant nutrient availability and soil hydraulic properties, [7,8] and influence the phenotypic response and productivity of plants [9].

Several DSM methods have been used to map SOM and CEC using point samples, remote sensing indices, and terrain attributes derived from a digital elevation model (DEM) as inputs [7–12]. For example, the CEC was predicted based on a generalized linear model using environmental variables including local relief and geomorphic units as prediction covariates [12]. Linear and multiple linear regressions (MLR) models have been widely used for spatial prediction of soil organic carbon due to their simplicity in application and ease of interpretation [7,10]. For example, [13] applied regression kriging (MLR and kriging of residuals) to predict and map field-scale variability of soil organic carbon using terrain attributes as covariates while [14] utilized legacy soil map and Landsat 5 TM variables for predicting CEC in a flat landscape based on random forest and geostatistical (i.e., Cokriging) models. Other studies used generalized linear models [15], and machine learning algorithms such as artificial neural networks [15–18], random forest [11,19,20], and Cubist [8,17,19,20] for predicting SOM and CEC.

Mapping soil properties in low relief areas can be a challenge since soil forming factors, especially topography and vegetation, may not co-vary with soil properties over space to the level at which they can be used effectively in DSM [21]. Terrain attributes derived from high-resolution elevation data, however, can capture local soil spatial variation that is caused due to interaction of water flow and topography [15]. In much of the Midwestern United States, high-resolution elevation data based on light detection and ranging (LiDAR) are widely available making lidar and lidar-based terrain attributes convenient environmental covariates for DSM.

Lidar and lidar-based terrain attributes are not the only environmental covariates used in DSM studies. Environmental covariates derived from optical remote sensing systems (e.g., satellite and unmanned aerial vehicle imagery) are often used in DSM [15,17,22]. Optical measurements are ideal covariates for DSM in locations with high correlation between land surface reflectance and soil properties. For example, the use of remotely derived vegetation indices (e.g., normalized difference vegetation index and enhance vegetation index) can improve accuracy of soil organic matter in DSM models [23]. However, there are circumstances where optical remote sensing data is less viable in soil properties mapping. One such circumstance is on agricultural research farms where experimental plots are intensively managed irrespective to the variation in soil properties such that the subsequent variation in surface reflection often poorly correlate to variation in soil properties. Using

Landsat remote sensing imagery derived covariates and terrain attributes [17] mapped soil organic carbon in a flat topographic area. The result showed a poor correlation of remote sensing covariates and strong correlation of terrain attributes with soil organic carbon.

The goals of this research were to (1) map SOM content and CEC in a low relief area using only lidar elevation-based terrain attributes, and (2) compare the prediction accuracy of SOM and CEC maps created by universal kriging, Cubist, and random forest with Soil Survey Geographic (SSURGO) database. The assumption is that at this scale (~570 ha), topography plays a major role in spatial distribution of soil properties such as SOM and CEC. In this study, we compared the performance of three predication models: universal kriging (UK), Cubist, and random forest (RF) to map SOM and CEC. We hypothesized that at this scale, terrain-driven hydrological flow patterns are the dominant process responsible for SOM and CEC differences in surface soils and that DSM models calibrated using lidar-derived terrain attributes should be capable of predicting the distribution of SOM and CEC at field and farm levels.

## 2. Materials and Methods

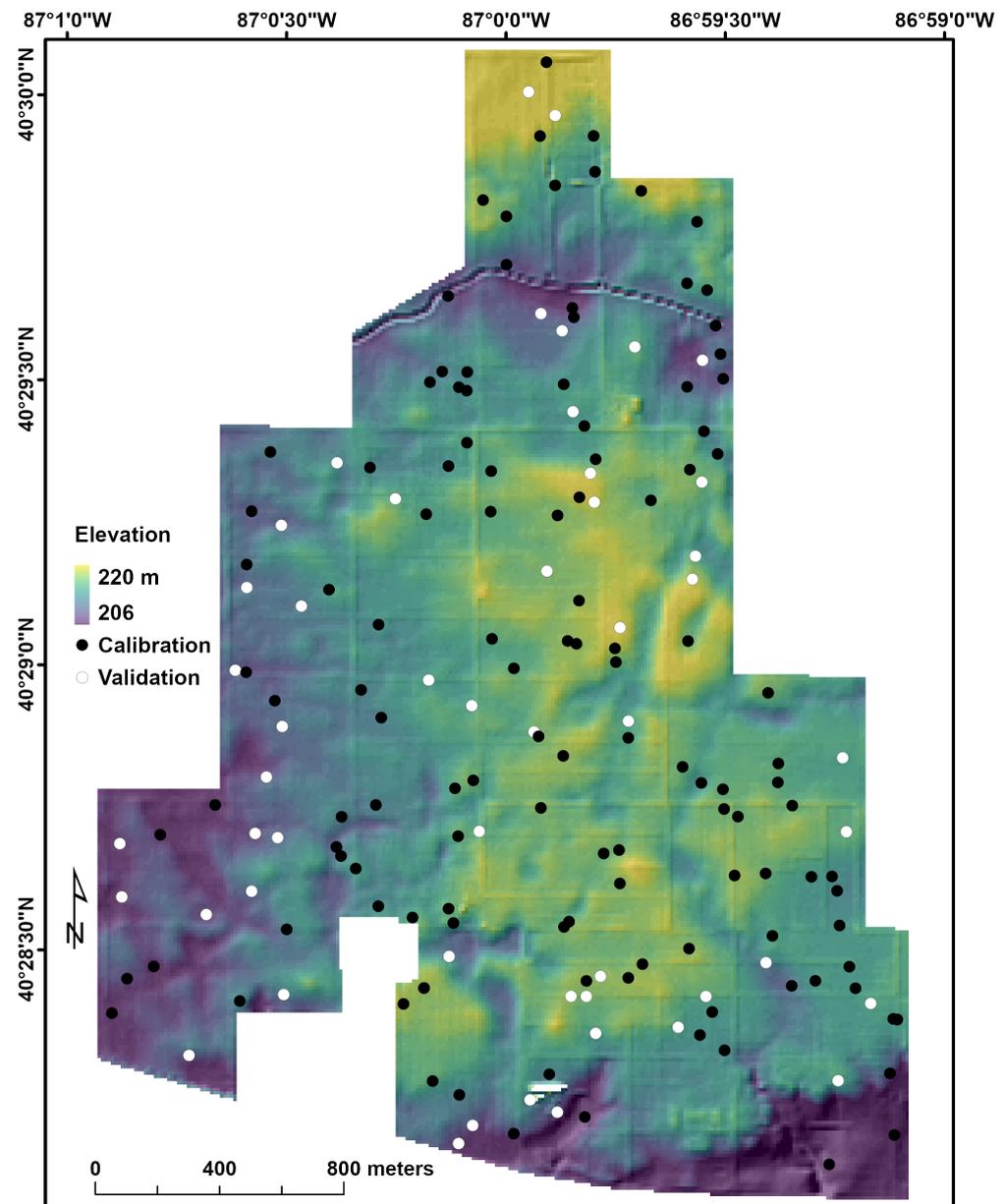
### 2.1. The Study Area

The Purdue Agronomy Center for Research and Education (ACRE) is an agronomic field research station located in Tippecanoe County Indiana, USA (40°29' N, 86°59' W) (Figure 1) and comprised 570 ha at the time this research was conducted. The mean annual temperature is 10 °C and mean total annual precipitation is 970 mm [24]. The average summer temperature (June to August) is 22.2 °C and average winter temperature (December to February) is −2.6 °C [24]. ACRE is located at the transition between the Eastern Hardwood Forests to the east and the prairies of the Great Plains to the west. Low relief, gently undulating Wisconsin age (15,000–20,000 years) till plain underlies the area [24]. The soils formed in ~50 cm of loess over loamy Wisconsin till and outwash. The site is in the mesic soil temperature regime and the udic soil moisture regime, but large areas of the study site have soils with an aquic soil moisture regime due to the presence of a seasonal high-water table [24]. Most of the soils are poorly and somewhat poorly drained. Mollisols occur over most of the study area, but Alfisols occur on the southern edge [24]. Corn and soybean are the major crops.

### 2.2. Soil Sampling and Analysis

As part of a previous, unpublished study, 174 soil samples were collected over the study area. Representative sampling locations were selected using the conditioned Latin hypercube sampling (cLHS) algorithm [22] using the *chls* package [25] in R-software 3.5.1 [26]. The cLHS method is a stratified random procedure that selects sampling locations based on the probability distribution of environmental covariates. Topographic wetness index (TWI), topographic position index (TPI), multi resolution valley bottom flatness, and ridge top flatness (MrVBF and MrRTF) that were derived from the DEM were used as environmental covariates for cLHS design.

The samples were collected from a soil depth from 0 to 10 cm by using a Push probe in July 2015, oven-dried at 40 °C, crushed, and passed through a 2 mm sieve. At each sampling location, 15 sub-samples were collected within a meter radius and combined to create a composite sample for analysis. The samples were analyzed by A&L Great Lakes Laboratories, Inc., Fort Wayne, IN, USA, following the soil test procedures for the North Central Region [27]. Briefly, SOM was determined by loss on ignition at 360 °C with a base factor of 0.97 and SOM expressed on a weight percent basis (%), CEC (cmol<sub>c</sub> kg<sup>−1</sup>) was measured by sum of cations displaced by 1 M ammonium acetate solution at pH 7 by inductively coupled plasma mass spectrometry. For model building and spatial predictions, the data were randomly split, with 70% of the samples used for model calibration and 30% used for model evaluation (Figure 1).



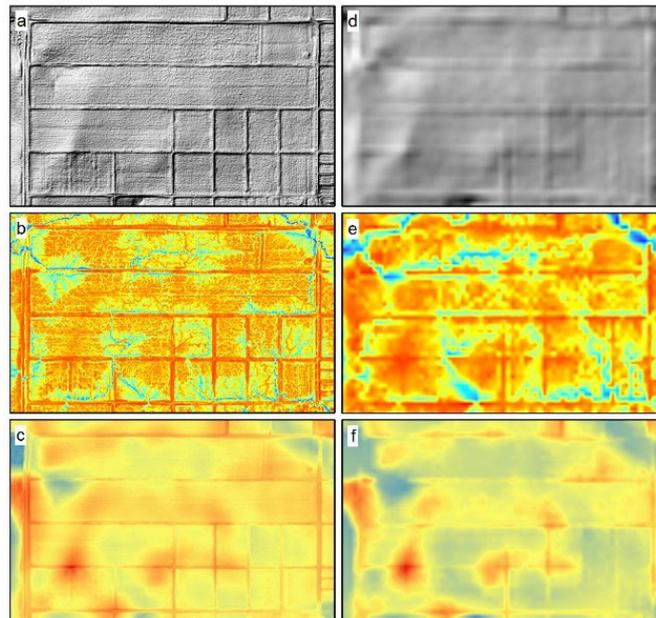
**Figure 1.** Study area and sampling locations over a lidar-derived elevation map over hillshade base map. Seventy percent of the samples were used for model calibration and 30 percent for validation.

### 2.3. Digital Elevation Model and Terrain Attributes

#### 2.3.1. Digital Elevation Model

Digital elevation model for Tippecanoe County, IN acquired as a statewide governmental funded campaign in 2013 at  $1.5 \text{ m} \times 1.5 \text{ m}$  pixel resolution or dimensions using lidar was used, and it is freely available for download from the Indiana Spatial Data Portal (<http://gis.iu.edu/>, accessed on 10 December 2020). The lidar data were collected with Leica ALS70, Leica ALS80, Optech Galaxy PRIME, and Riegl LMS-Q1560 airborne sensors (Woolpert, 2020 [28]). The DEM was re-projected from the Indiana State Plane West Coordinate System (NAD\_1983\_StatePlane\_Indiana\_West\_FIPS\_1302\_Feet) which uses dimensions in feet, to the Indiana Geospatial Coordinate System (InGCS) for the Tippecanoe and White Counties (NAD\_1983\_2011\_InGCS\_Tippecanoe-White\_m) in meters. The InGCS has lower grid vs. ground distortion when compared to the State Plane Coordinate System ( $\pm 80$  ppm) [29] and thus is more appropriate for a small area such as ACRE.

Digital elevation models with pixel resolutions on the order of 1–2 m are often too detailed and noisy for modeling field-scale soil spatial variability [8]. For example, Ref. [30] found that pixel resolutions from 5–10 m are sufficient to capture the topography for digital soil mapping of a post-glaciated landscapes in northern Indiana. Our initial evaluation showed that within ACRE, anthropogenic micro-topographic features such as roads and field boundaries, that are on average 20 cm higher than the cultivated fields, unduly affected the DEM derived indices (Figure 2). We resampled the original 1.5 m DEM to 10 m using simple mean aggregation in ArcGIS 10.6 [31] to smooth out most of the anthropogenic features (Figure 2).

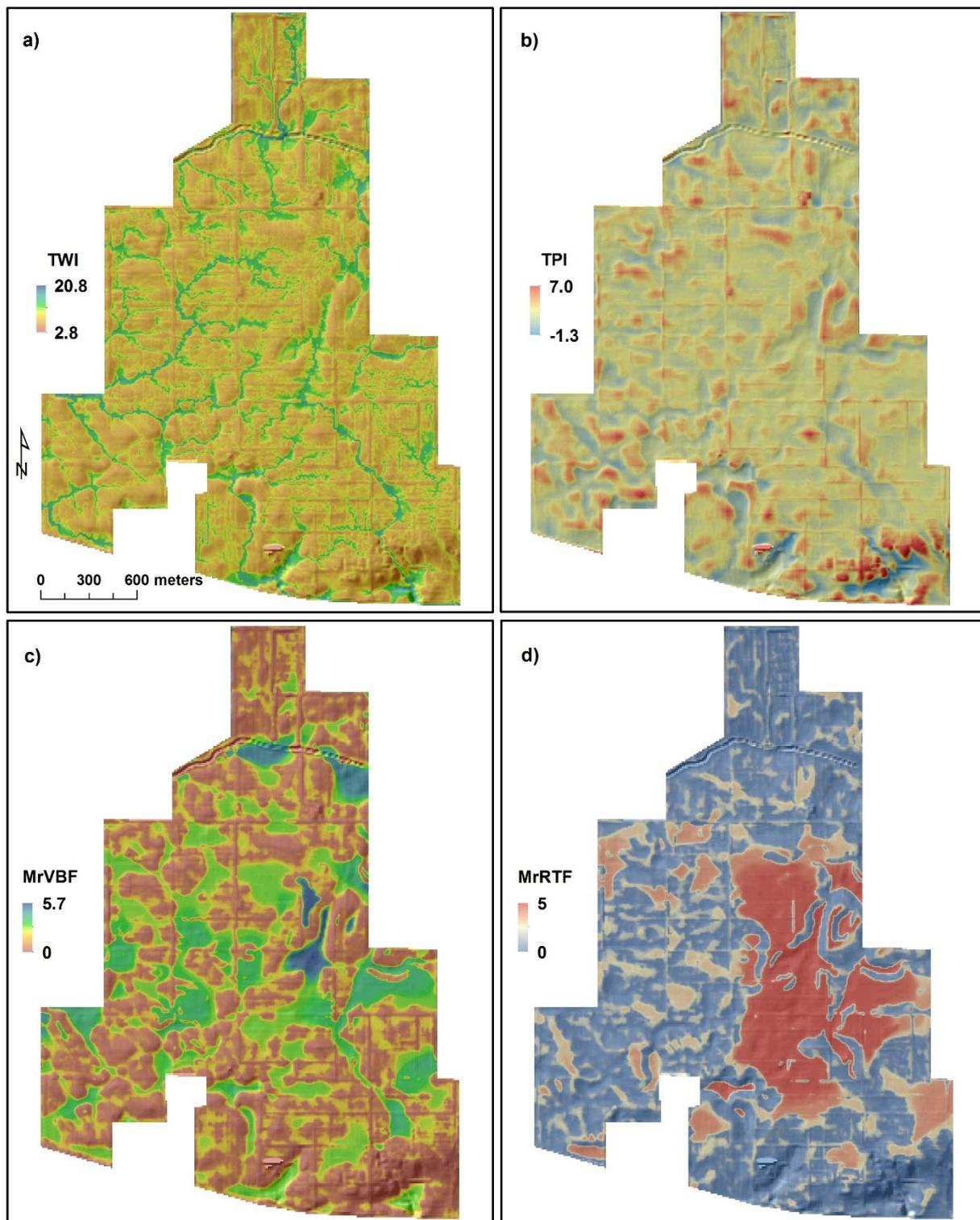


**Figure 2.** The hillshade (a) topographic wetness index (b), and topographic position index (c) from the 1.5 m lidar DEM. The hillshade (d) topographic wetness index I, and (e) topographic position index (f) from the 10 m resample lidar DEM.

The watershed contributing water to ACRE obtained from the United States Geological Survey–National Hydrography Dataset (USGS-NHD) downloaded from the United States Department of Agriculture (USDA) Geospatial Data Gateway (GDG) (<https://datagateway.nrcs.usda.gov/>, accessed on 10 December 2020) for Tippecanoe County, IN was used to clip the elevation data. The watershed boundary rather than the ACRE farm boundary was used to reduce artifacts for the derived terrain attributes. The ACRE farm boundary was then used to clip the derived terrain attributes for further analysis.

### 2.3.2. Terrain Attributes

It is possible to generate many terrain attributes from a DEM, but it is important to limit their number to avoid redundancy, model overfitting and interpretations [32]. Suleymanov et al. (2021) [33] found that of 17 generated terrain attributes, only three of them (elevation, slope, and MrRTF) were most important for predicting the spatial variability of soil properties including SOM. We focused on those terrain attributes that have a close relationship to water redistribution across a post-glaciated landscape such as at ACRE and that are commonly used in DSM. We calculated the following terrain attributes using SAGA-GIS 2.1.4 [34]: topographic wetness index (TWI), topographic position index (TPI), multi-resolution valley bottom flatness index (MrVBF), multi-resolution ridge top flatness index (MrRTF) (Figure 3).



**Figure 3.** Terrain attributes calculated from the digital elevation model. (a) Topographic wetness index (TWI), (b) topographic position index (TPI), (c) multi resolution valley bottom flatness index (MrVBF), and (d) multi resolution ridge top flatness index (MrRTF).

#### Topographic Wetness Index

The topographic wetness index quantifies potential moisture retention and redistribution properties of a landscape and shows relationship between topography and hydrological processes, mainly surface runoff in a watershed. It was calculated as  $TWI = \ln(A_s / \tan\beta)$ ,

where  $A_s$  is the specific catchment area ( $m^2/m$ ) and  $\beta$  is the slope angle, considering a multi-flow-direction algorithm. Higher values of TWI represent areas that accumulate water, such as depressions and drainage ways, while lower values represent areas that shed water, such as crests and ridges.

#### Topographic Position Index

The topographic position index [35] compares the elevation of a cell ( $Z_0$ ) to the average elevation of its surrounding cells ( $Z_\alpha$ ) in a specific area as defined by circles of arbitrary radius ( $TPI = Z_0 - Z_\alpha$ ). Positive values of TPI represent ridges, and negative values represent valleys, while values close to zero represent linear sloping areas between ridges and valleys. This index is scale dependent, and by using different radii it can differentiate small hummocks from larger ridges, as well as small depressions from wider valleys. We evaluated different radii to calculate TPI. The larger radii (150, 200, 300, and 500 m) resulted in smoothing of the landscape features, while smaller radii (30 and 60 m) generated linear artifacts and fragmented the actual landforms into small pieces. Based on visual interpretation and familiarity with the study location, a radius of a 100 m was found to best represent the landscape units of the study site.

#### Multiresolution Valley Bottom Flatness and Multiresolution Ridge Top Flatness

The MrVBF algorithm [36], identifies valley bottoms by utilizing the lowness and flatness characteristics. The lowness parameter is measured by ranking elevation with respect to a circular neighborhood area, and the flatness parameter is measured using the inverse of slope.

The slope threshold is a critical parameter in MrVBF calculations, and it depends on the DEM resolution. The suggested slope threshold for a DEM of 250 m resolution is 4%, while for a DEM of 25 m it is 16%, and for a DEM of 8 m it is 32% [36]. We compared slope thresholds of 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 16% (the default slope threshold of the algorithm) and based on our familiarity with the area and resulting terrain attributes in the field, a slope threshold of 2% was found to best represent the topography of the landscape. The MrRTF is a separate index but complementary to the MrVBF. It is derived in a similar way as MrVBF, except it identifies the upper parts of the landscape. As with MrVBF, the same slope threshold value (2%) was selected for MrRTF.

#### 2.4. Data from the Soil Survey Geographic Database

For Tippecanoe County, the SSURGO database provides soil mapping information at a scale 1:15,840 [24]. In this study, low, representative, and high values of SOM and CEC for each mapping unit from SSURGO (Table 1) were compared to the predictions from the DSM models. The SSURGO SOM and CEC values were directly acquired from the Web Soil Survey website [37], except the mean value of CEC, which was calculated as the average of the low and high CEC values. The SSURGO values of soil properties have been derived from a combination of laboratory measured data and soil scientist expert knowledge [38]. Based on a conversion factor of 1.72, SOM in the SSURGO dataset was converted from soil organic carbon of the Walkley-Black method, while CEC was determined by summation of cations, which were displaced by ammonium acetate solution [39,40]. It is important to recognize that the SOM data comparison between DSM models and SSURGO might not be consistent due to the differences in laboratory methods and the conversion factors used to convert soil organic carbon from Walkley-Black to SOM.

**Table 1.** The soil survey geographic (SSURGO) soil organic matter content (SOM) and cation exchange capacity (CEC) low, representative (Rep.), mean, and high values for 0–10 cm based on the spline function. Data is for the ACRE study site.

SSURGO Soil Map Unit		SOM				CEC		
		Area %	Low	Rep. %	High	Low	Mean cmol <sub>c</sub> kg <sup>-1</sup>	High
Cm	Chalmers silty clay loam	33.9	3.5	4.9	6.5	18.2	28.5	38.8
CwB2	Crosby-Miami silt loams	0.4	1.0	2.4	3.1	2.5	7.0	12.2
Du	Drummer soils	17.6	3.3	4.9	6.5	22.4	29.2	36.2
Md	Mahalasville-Treaty complex	0.2	3.3	4.8	6.4	18.2	24.2	30.3
MsC2	Miami silt loam	0.2	1.1	2.1	3.3	4.7	9.2	16.0
Mu	Milford silty clay loam	4.2	3.5	5.7	6.7	25.4	31.1	36.7
Pg	Pella silty clay loam	1.6	4.8	4.7	7.1	22.9	28.9	34.9
Pk	Peotone silty clay loam	0.6	4.3	5.9	7.4	21.2	29.6	38.1
RcA	Raub-Brenton complex	22.1	1.9	2.9	4.2	12.5	17.4	22.3
RoB	Rockfield silt loam	4.6	1.1	1.6	2.1	6.7	12.2	17.6
SwA	Starks-Fincastle complex	3.6	0.9	2.1	2.9	7.1	12.8	18.6
TfB	Throckmorton silt loam	1.4	1.1	2.1	3.2	5.6	11.6	17.6
TmA	Toronto-Millbrook complex	8.7	1.9	2.9	3.8	10.9	16.7	22.5
Ua *	Udorthents, loamy	0.9	-	-	-	-	-	-

\* The Ua is a disturbed and removed soil from its original place and was not considered.

The analysis were performed through various packages in R-software v. 3.5.1. To generate representative SOM and CEC data from a soil depth from 0 to 10 cm, the SSURGO data were harmonized by depth using mass-preserving splines function (*ea\_spline*) of the *itthir* package 1.0 [41]. The spline function models the continuous depth distribution of soil properties based on discrete horizon observations [42]. The equal area spline function for modeling soil organic carbon content with depth has been proven to be useful in several studies [8,16,43]. For detailed information and mathematical expression of the spline function, see [16,44].

Since a map unit may have two or more components, we derived the final values of a map unit based on the weighted mean of each component. For instance, CwB2 (Crosby-Miami silt loams, 2 to 4 percent slopes, eroded) contains 64% Crosby and 33% Miami and 3% other minor components. Based on the spline function for the 0–10 cm depth, the SOM representative value for Crosby was 2.67% and 2.18% for Miami. The final SOM representative value of CwB2 for the 0–10 cm depth was derived as:

$$\text{CwB2 mapping unit mean value of SOM (\%)} = (0.64 \times 2.67) + (0.33 \times 2.18) = 2.43 \quad (1)$$

### 2.5. Spatial Prediction Models

Three different models (universal kriging, Cubist, and random forest) were used to predict SOM content and CEC. All model training and evaluation was performed in R-software v. 3.5.1 [26].

Universal kriging, also known as regression kriging [45], is a hybrid approach to modeling, meaning that the prediction of a desired variable is made based on a combination of deterministic and stochastic components. The deterministic part of the regression relies on the covariate information, while the stochastic part relies on the spatial auto-correlation of the regression residual based on a variogram [42]. We ran backwards stepwise linear models to select appropriate terrain attributes for the deterministic part of UK. The *gstat* package 2.0.2 [46] was used for UK prediction of SOM and CEC.

Cubist is a machine learning tool that uses a rule-based regression algorithm for prediction [47]. It operates based on *if, then, else* statements. If a condition is matched, the next step is a prediction of the desired soil property by using ordinary least squares regression from the covariates within that subset [48]. However, if a condition is not met, then the next node of the tree is defined by the rule and the *if, then, else* sequence is repeated.

The interpretation of a Cubist model is easy as it provides an explicit model stating the relative importance of the predictors. The *Cubist* 0.2.2 package [48] was used to predict SOM and CEC of the study area using 5 rules, 5 extrapolation, and 10 committees as suggested by (<https://www.rulequest.com/cubist-unix.html>, accessed on 10 December 2021).

The Random Forest (RF) [49], is an ensemble machine learning algorithm that predicts the property of interest based on covariates by creating multiple decision trees. The outcomes of the decision trees are then aggregated to provide the final prediction. A random and independent bootstrap sample of the training data is used to train each tree in the forest. From the bootstrap sample, a random subset (~2/3) is selected for training and the remaining points (~1/3), are used for validating the tree. Additionally, a random subset of the variables is selected to split the nodes of each tree [50]. In this study, we used 1000 trees for RF as suggested by [11,42]. The *randomForest* package 4.6.14 [51] was used to predict both SOM and CEC.

The spatial dependency of the residuals from universal kriging, cubist and RF predictions were modelled with variogram and their distributions were mapped using kriging interpolation. Kriging of residuals may capture spatial variability that was not represented by deterministic or linear models of UK. Variogram has three main parameters: nugget, sill, and range. The nugget is the variance unexplained by the variogram and referred to noise in the data or random error. Sill represents the maximum variability among the point pairs. The range represents the maximum distance of spatial autocorrelation. For UK, a spherical variogram, and for Cubist and RF, exponential variograms were fitted to kriging the residual of SOM. For kriging the CEC residuals, a spherical variogram was fitted for all three predictive models. Residual kriging was carried out with *gstat* package 2.0.2 [46]. The final estimates of SOM and CEC were derived by adding kriged residuals and the predicted values from the corresponding models.

## 2.6. Validation of Model Performance

All of the predictive models were first evaluated with the calibration dataset from which they were generated (internal evaluation). A second random-hold-back independent evaluation was conducted using 30% of the data for testing the prediction performance of each model. Root mean square error (RMSE), mean error (ME) or bias, coefficient of determination ( $R^2$ ), and Lin's concordance correlation coefficient ( $\rho_c$ ) (Equations (2)–(5)) indices were used for validation

$$\text{RMSE} = \sqrt{\left( \frac{\sum_{i=1}^n (\text{obs}_i - \text{pred}_i)^2}{n} \right)} \quad (2)$$

$$\text{ME (i.e., bias)} = \frac{\sum_{i=1}^n \text{obs}_i - \text{pred}_i}{n} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (\text{obs}_i - \overline{\text{obs}}) (\text{pred}_i - \overline{\text{pred}})}{\sqrt{\sum_{i=1}^n (\text{obs}_i - \overline{\text{obs}})^2} \sqrt{\sum_{i=1}^n (\text{pred}_i - \overline{\text{pred}})^2}} \quad (4)$$

$$\rho_c = \frac{2\rho\sigma_{\text{obs}}\sigma_{\text{pred}}}{\sigma_{\text{pred}}^2 + \sigma_{\text{obs}}^2 + (\mu_{\text{pred}} + \mu_{\text{obs}})^2} \quad (5)$$

where  $\text{obs}_i$  are the observed values and  $\text{pred}_i$  are the predicted SOM and CEC values of at location  $i$ ,  $\mu_{\text{obs}}$  is the mean of the observed values,  $\mu_{\text{pred}}$  is the mean of the predicted values,  $\sigma_{\text{obs}}^2$  is the variance of the observed values,  $\sigma_{\text{pred}}^2$  is the variance of predicted values,  $n$  is the number of the sampling locations, and  $\rho$  is the correlation coefficient among the observations and predictions [42].

The RMSE shows the accuracy of the prediction, and smaller values translate to higher accuracy. Bias shows the mean error of the prediction and is equal to zero for an unbiased prediction. The  $R^2$ , assesses the ability of the model to explain the variability in the predictions, while Pearson's correlation coefficient ( $\rho$ ), measures the precision of the

relationship between the predicted and observed values. The  $\rho_c$  [52], is a single statistic that measures both the precision and the accuracy of the relationship. The  $\rho_c$  is also known as the goodness of fit along a 45° line (1:1 line). The value of  $\rho_c$  falls between  $-1$  and  $+1$ . A value of  $-1$  indicates perfect negative agreement, while a value of  $+1$  indicates perfect positive agreement between the predicted and observed values. A  $\rho_c$  value of zero shows that there is no agreement at all [42,52]. The strength of the agreement was evaluated using  $\rho_c$  based on the proposed scale from [53]. The *goof* function of the *ithir* package 1.0 [41] was used to compute these evaluation indices.

### 3. Results and Discussion

#### 3.1. Descriptive Statistics

The mean SOM content was 4.2% and both calibration (4.2%) and validation (4.0%) mean SOM values were comparable. Similarly, the range values of SOM were comparable varying from 1.2 to 7.2% for all datasets (Table 2). The mean CEC values for the whole dataset ( $19.9 \text{ cmol}_c \text{ kg}^{-1}$ ) and the calibration ( $20.1 \text{ cmol}_c \text{ kg}^{-1}$ ) and validation ( $19.5 \text{ cmol}_c \text{ kg}^{-1}$ ) datasets were comparable as were the range values varying from 9.9 to  $30.1 \text{ cmol}_c \text{ kg}^{-1}$ .

**Table 2.** Summary statistics of soil organic matter content (SOM) and cation exchange capacity (CEC) data for the study area.

Statistical Index	SOM			CEC		
	Whole Dataset	Calibration	Validation	Whole Dataset	Calibration	Validation
		%			$\text{cmol}_c \text{ kg}^{-1}$	
Minimum	1.2	1.9	1.2	9.9	11.1	9.9
1st Quartile	3.5	3.5	3.2	16.2	16.5	14.9
Median	4.0	4.0	4.2	20.1	20.1	20.1
Mean	4.2	4.2	4.0	19.9	20.1	19.5
3rd Quartile	4.8	4.9	4.6	23.2	23.3	23.0
Maximum	7.2	7.0	7.2	30.1	30.1	29.3
Standard Deviation	1.1	1.2	1.1	4.6	4.5	4.9

As with SOM content and CEC, the differences between calibration and validation data sets for the terrain attributes were comparable (Table 3). The mean values for the calibration dataset were slightly higher compared to validation dataset for all terrain attributes. However, there were no trends with regard to other statistical parameters.

**Table 3.** Summary statistics of terrain attributes for the study area.

Terrain Attributes	Dataset	Statistical Index						
		Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	Standard Deviation
TWI	Whole dataset	5.29	7.37	8.69	9.39	10.71	18.73	2.64
	Calibration	5.29	7.37	8.72	9.58	10.81	18.73	2.84
	Validation	6.04	7.51	8.41	8.93	10.25	15.31	2.04
TPI	Whole dataset	-0.46	-0.11	-0.03	-0.01	0.09	0.71	0.18
	Calibration	-0.45	-0.11	-0.04	0.00	0.09	0.62	0.18
	Validation	-0.46	-0.10	-0.03	-0.01	0.07	0.71	0.19
MrVBF	Whole dataset	0.00	0.37	1.60	1.62	2.75	5.24	1.33
	Calibration	0.00	0.46	1.74	1.72	2.80	4.96	1.28
	Validation	0.01	0.17	0.65	1.39	2.63	5.24	1.44
MrRTF	Whole dataset	0.00	0.14	0.74	1.64	2.98	4.99	1.76
	Calibration	0.00	0.15	1.16	1.84	3.65	4.99	1.84
	Validation	0.00	0.12	0.47	1.18	1.50	4.99	1.47

### 3.2. Spatial Trend Modeling

Each model utilized different terrain attributes for SOM and CEC predictions. A backwards stepwise linear model selection was used for the UK model. Based on the following equations (Equations (6) and (7)), the backwards stepwise model selected TWI, TPI, and MrRTF for SOM, and TPI and MrVBF for CEC predictions.

$$\text{SOM}(\%) = 3.37 + 0.11 \times \text{TWI} - 2.20 \times \text{TPI} - 0.09 \times \text{MrRTF} \quad (6)$$

$$\text{CEC} (\text{cmol}_c \text{ kg}^{-1}) = 18.41 - 9.36 \times \text{TPI} + 1.02 \times \text{MrVBF} \quad (7)$$

From a pedological standpoint, Equations (6) and (7) reveal meaningful relationships between terrain and SOM or CEC. Equation (6) shows that SOM is positively correlated to TWI or wet/low-lying areas of the landscape, while it is negatively correlated to TPI and MrRTF or higher/steeper areas of the landscape. Similarly, CEC (Equation (7)) is negatively correlated with TPI, but positively correlated with MrVBF or lower landscape positions.

Cubist utilized all four terrain attributes for SOM, and only TPI and MrVBF for CEC predictions. Out of ten models generated by Cubist for SOM and CEC predictions, we selected the models with the lowest prediction error. For example, the SOM model (Equation (8)) was applicable in 123 locations where the average SOM was 4.09% with a prediction error of 0.70%. The CEC model (Equation (10)) applied to the 123 training locations had a mean value of 20.12  $\text{cmol}_c \text{ kg}^{-1}$  and prediction error of 2.84  $\text{cmol}_c \text{ kg}^{-1}$ . The Cubist model provided slightly different models for SOM prediction based on the combination of four terrain attributes but produced identical models for CEC. Following are examples of Cubist models for SOM and CEC predictions.

$$\begin{aligned} & \text{If } \text{TWI} \leq 13.76 \\ & \text{then } \text{SOM} (\%) = 3.79 + 0.21x\text{MrVBF} - 1.17x \text{TPI} - 0.11x\text{MrRTF} \end{aligned} \quad (8)$$

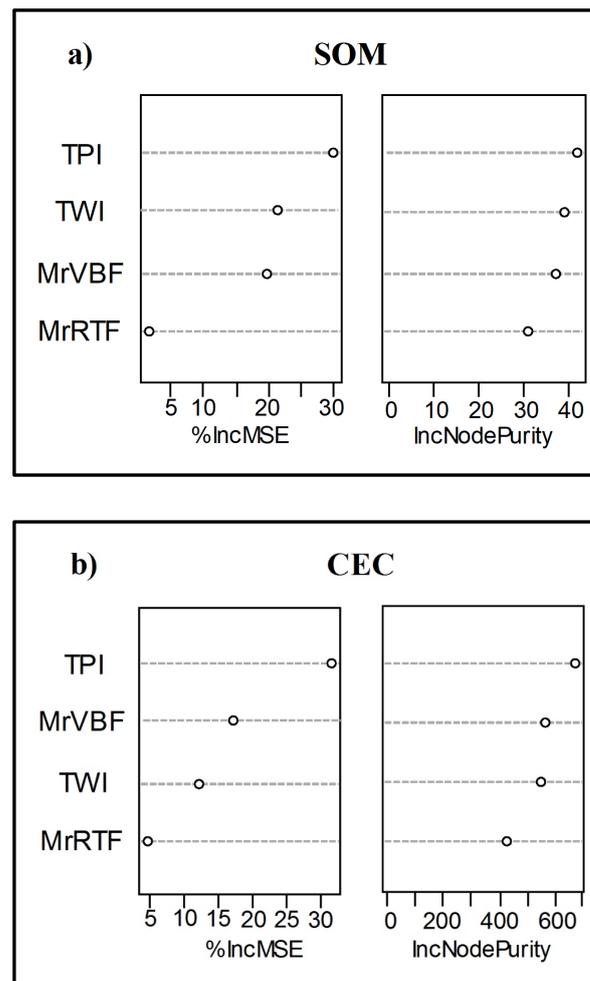
$$\begin{aligned} & \text{If } \text{TWI} > 13.76 \\ & \text{then } \text{SOM} (\%) = 7.74 + 0.02x\text{MrVBF} - 0.32x\text{MrRTF} \end{aligned} \quad (9)$$

$$\text{CEC} (\text{cmol}_c \text{ kg}^{-1}) = 18.84 - 9.30x\text{TPI} + 1.03x\text{MrVBF} \quad (10)$$

The Cubist model also provided the relative importance (RI) of the terrain attributes in developing conditions rules (*if then else* rules) and in developing multivariate linear function. In SOM prediction, the RI of the four terrain attributes was 54% (TPI), 37% (MrRTF), 35% (TWI), and 34% (MrVBF) in the MLR function and 54% for TWI in rule conditions. Therefore, TWI is the most important predictor for the Cubist model for SOM prediction.

For CEC, the Cubist model did not estimate conditional rules and returned only a single regression equation (Equation (10)). Not surprisingly, due to the lack of conditional rules, the equation from Cubist is almost identical with the one from stepwise linear regression (Equation (7)). Subsequently, the UK and Cubist predictive models produced similar results for CEC (Table 3). This suggests that for our study area, the additional complexity of Cubist models is unwarranted as the relationship between soil properties and landscape can be described by a single regression equation without the need for complex conditional rules.

Random forest used all four terrain attributes for predicting both SOM and CEC. The *varImpPlot* function in the *randomForest* package 4.6.14 [51] shows the importance of terrain attributes for SOM and CEC predictions. For the RF predictions the most important terrain attributes for SOM predictions were TPI and TWI, while for CEC predictions TPI and MrVBF (Figure 4). Overall, TPI was the most important variable and MrRTF was the least important variable in all selected models.



**Figure 4.** Random forest generated importance plots of terrain attributes, (a) for soil organic matter content (SOM) and (b) for cation exchange capacity (CEC) prediction. The %IncMSE shows the mean decrease in accuracy. The IncNodePurity shows the decrease in node purity at the end of the tree. The higher %IncMSE and IncNodePurity values show that a particular variable is highly important and if removed the prediction accuracy and node purity will be affected.

Both the UK and Cubist models show that SOM and CEC increase as TWI and MrVBF increase and decrease as TPI and MrRTF values increase. Even though, RF utilized all four terrain attributes for SOM and CEC predictions, TPI and MrVBF were among the most important variables for both SOM and CEC, similar to UK and Cubist (Equations (6)–(10)). Ref. [17] reported that among the five terrain attributes (elevation, TWI, plan curvature, total catchment area, and channel network base level), TWI and total catchment area closely related to the soil organic carbon content in flat slope areas. They also stated that besides these two terrain attributes, soil nutrient indicators were other environmental covariates that showed close correlation with soil organic carbon content. Ref. [43] found that wetness index and MrVBF were among the most important terrain attributes that influenced spatial distribution of SOM. They also reported that beside terrain attributes, land use, soil type and precipitation were other environmental variables that influence SOM distribution. Ref. [54] also documented the influence of land use, soil type, and geology on soil organic carbon distribution. They also noted that after elevation, TWI was the second most important terrain attribute in soil organic carbon prediction. The influence of soil type, geology, elevation, and slope on soil organic carbon distribution in wet cultivated fields was also reported by [11] who found that most of the soil organic carbon variation

at the surface (0–10 cm) was explained by topographic attributes, while in the subsurface (10–50 cm) it was best estimated by soil texture classes.

TWI had correlation of 0.49 with SOM, 0.43 with CEC,  $-0.56$  with TPI, 0.01 with MrRTE, and 0.67 with MrVBF. TPI had correlation of  $-0.51$  with SOM,  $-0.53$  with CEC,  $-0.52$  with MrVBF, and 0.19 with MrRTE. MrVBF had correlation of 0.46 with SOM, 0.52 with CEC, and 0.04 with MrRTE. MrRTE showed a correlation of  $-0.21$  with SOM,  $-0.06$  with CEC. Suleymanov et al. (2021) [33] reported 0.5 correlation between SOM and MrRTE and MrRTE was a key attribute for mapping soil organic carbon and thickness of humus. Ref. [15] also observed a correlation of 0.5 between soil properties including soil organic carbon and auxiliary information. Ref. [20] reported lower correlation ( $<-0.38$ ) between DEM and soil attributes including CEC, and moderate to strong correlation ( $<0.60$ ) between satellite bands of synthetic soil image and soil attributes. The legacy soil map showed better correlation (0.5) with CEC compared to Landsat 5 TM derived attributes ( $<0.44$ ) in the flat landscape of a semiarid region of Brazil [14]. Similar to [17], we found that based on the whole dataset, SOM has a strong correlation with CEC (0.73).

### 3.3. Variogram of Model Residuals

Variogram parameters of the UK, Cubist and RF model residuals of SOM and CEC are presented in the Table 4. The nugget is the variance unexplained by the variogram and referred to noise in the data or random error. Compared to SOM, all DSM models, particularly UK and Cubist demonstrated highest nugget values for CEC. Sill represents the maximum variability among the point pairs. The range represents the maximum distance of spatial autocorrelation. The range values were increased in order of UK, Cubist, and RF for both SOM and CEC. UK showed lower range or spatial autocorrelation for both SOM and CEC while, RF shows higher spatial autocorrelation for the stated soil properties. The nugget to sill ratio determines the strength of spatial autocorrelation [55–58]. According to [55], a nugget to sill ratio of less than 0.25 shows strong, 0.25–0.75 shows moderate, and greater than 0.75 shows weak spatial dependence or autocorrelation. According to the nugget:sill ratio, SOM showed moderate while, CEC showed weak spatial dependency. According to [58], higher nugget:sill ratio and longer ranges, indicates greater variability in soil properties and a need for more observations for better prediction.

**Table 4.** Variogram parameters of universal kriging (UK), Cubist, and random forest (RF) model residuals to predict soil organic matter content (SOM) and cation exchange capacity (CEC).

Soil Property	DSM Model	Semi-Variogram Model	Nugget	Sill	Nugget:Sill	Range (m)
SOM	UK	Spherical	0.22	0.43	0.51	126.12
	Cubist	Exponential	0.18	0.39	0.46	175.00
	RF	Exponential	0.08	0.11	0.72	176.29
CEC	UK	Spherical	5.46	5.70	0.96	229.00
	Cubist	Spherical	5.23	5.86	0.89	307.60
	RF	Spherical	1.16	1.41	0.82	405.50

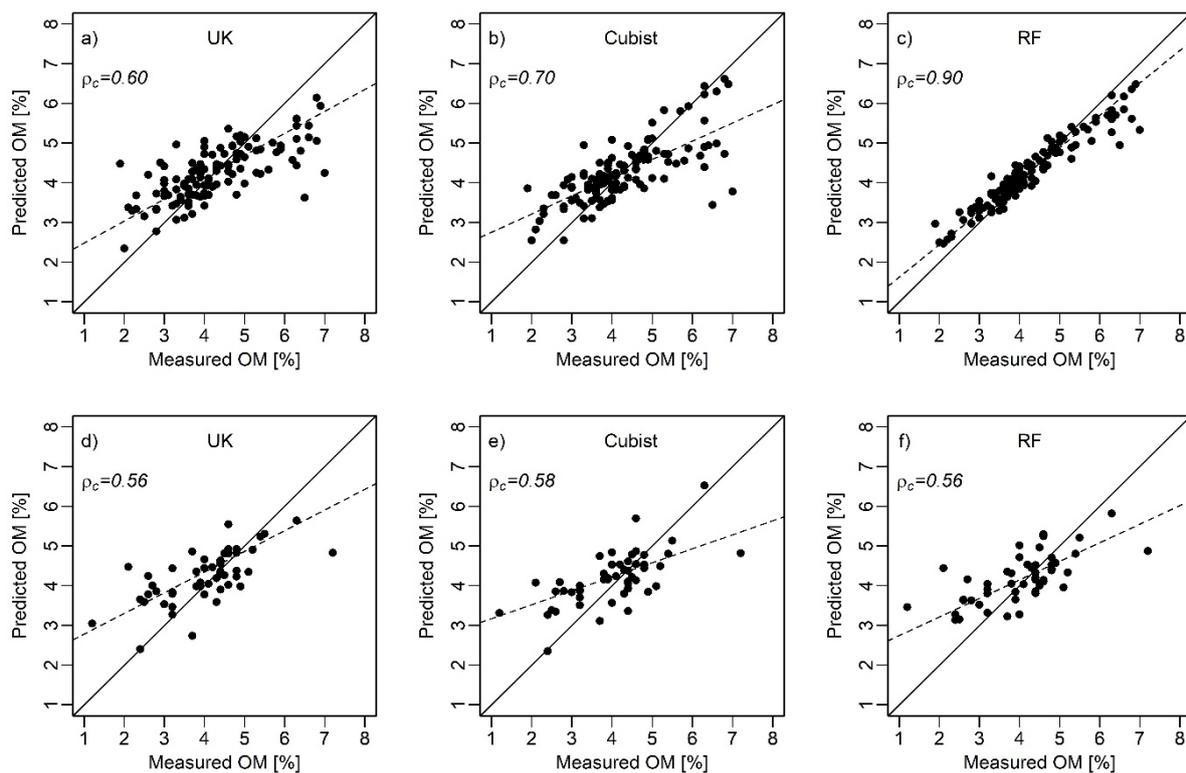
### 3.4. Predictive Model Performance

Based on the 4 evaluation metrics (Table 5) and scatter plots of measured versus predicted SOM (Figure 5) and CEC (Figure 6), we found no major differences in the prediction performance of all three models.

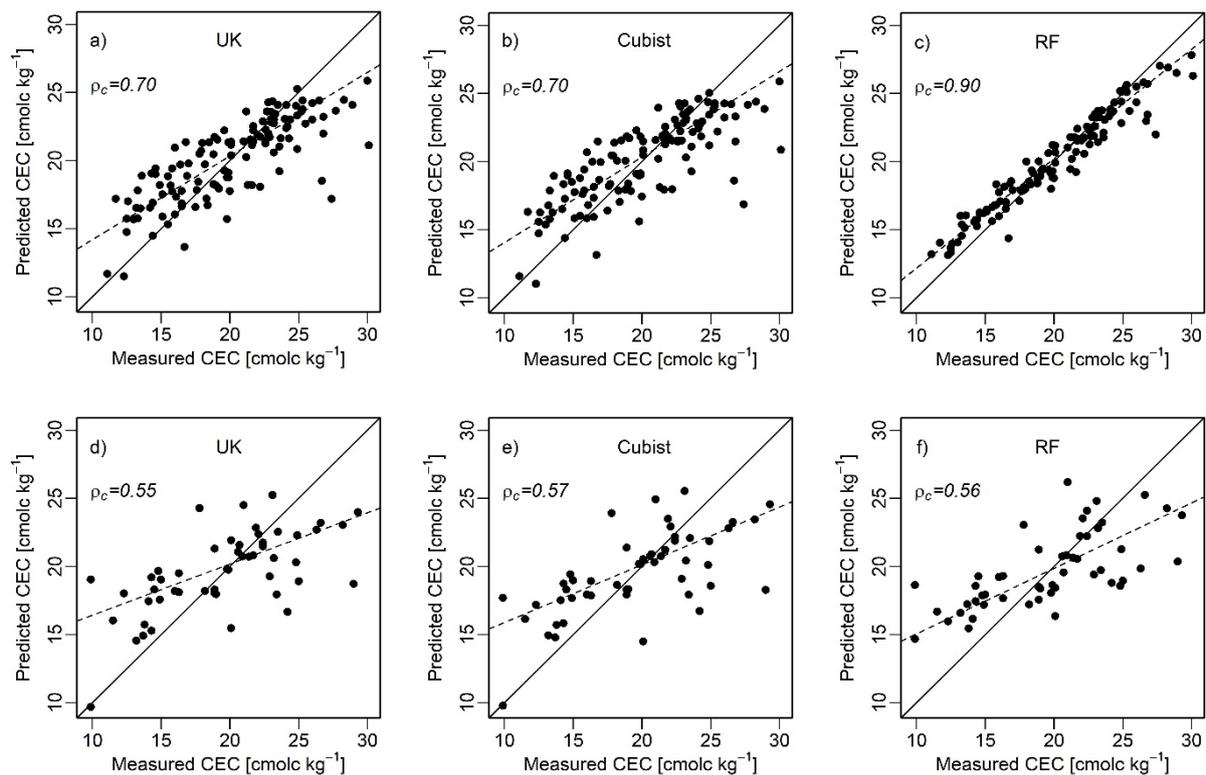
**Table 5.** Universal kriging (UK), Cubist, and random forest (RF) accuracy assessment for organic matter content (SOM) and cation exchange capacity (CEC) predictions with calibration and validation datasets.

Prediction Model		SOM				CEC			
		R <sup>2</sup>	Bias	RMSE	$\rho_c$	R <sup>2</sup>	Bias	RMSE	$\rho_c$
		%				cmol <sub>c</sub> kg <sup>-1</sup>			
UK	Calibration	0.50	0.00	0.80	0.60	0.60	0.00	2.80	0.70
	Validation	0.44	0.22	0.83	0.56	0.39	0.05	3.74	0.55
Cubist	Calibration	0.50	0.00	0.80	0.70	0.60	0.00	2.80	0.70
	Validation	0.45	0.17	0.80	0.58	0.41	0.00	3.68	0.57
RF	Calibration	0.90	0.00	0.40	0.90	0.90	0.00	1.40	0.90
	Validation	0.45	0.17	0.80	0.56	0.44	0.17	3.62	0.56
SSURGO	Calibration	0.31	-0.14	1.11	0.56	0.51	3.88	6.06	0.53
	Validation	0.40	-0.03	0.99	0.62	0.56	3.85	5.78	0.58

Also [19] found no major difference and reported similar R<sup>2</sup> and RMSE values for soil organic carbon estimation based on Cubist and RF. Similar conclusion was reported by [17], who noted that RF and Cubist showed similar predictive performance in soil organic carbon estimation.



**Figure 5.** Scatter plots of measured vs. predicted soil organic matter (SOM) content based on calibration (a–c) and validation data (d–f) for Universal kriging (UK), Cubist, and random forest (RF) models. The solid line indicates a line of concordance or a 1:1 relationship. The dashed line indicates the line of best fit.



**Figure 6.** Scatter plots of measured vs. predicted CEC based on calibration (a–c) and validation data (d–f) for Universal kriging (UK), Cubist, and random forest (RF) models. The solid line indicates a line of concordance or a 1:1 relationship. The dashed line indicates the line of best fit.

For the calibration data, RF had lower RMSE, lower bias, higher  $R^2$ , and higher concordance than UK and Cubist for both SOM and CEC predictions. With RF, this was expected due to the ensemble approach, which can result in low bias and variance [59,60]. Ref. [17] reported that predicted and observed values of soil organic carbon on the RF scatter plot were closer to the central or 1:1 line compared to Cubist, artificial neural network, support vector machine, and multiple linear regression models. For the RF models, however, there was a significant change in model performance between calibration and evaluation data. Thus, the performance decreased by half during the evaluation and was comparable with the performance of UK and Cubist. For example, for SOM predictions, the  $R^2$  of the RF prediction were 0.90 for the calibration dataset and decreased to 0.45 for the evaluation dataset which was similar to UK (0.44) and Cubist (0.45). This significant change between calibration and evaluation performance is strong evidence that the RF models may be over optimistic. This highlights one of the issues with modeling for DSM: if models are not evaluated rigorously (i.e., using independent evaluation rather than leave-one-out), model performance estimates may be overly optimistic. Ref. [43] also documented higher  $R^2$  and lower RMSE based on calibration dataset compared to validation dataset. Therefore, we agree that an independent validation dataset is necessary for DSM prediction performance evaluation [54,61].

According to the scatter plots for SOM (Figure 5a–c) and CEC (Figure 6a–c), all three models tended to over predict at low values and under predict at high values. This behavior is less pronounced for RF models on the calibration data, but it is apparent for all models on the evaluation dataset. This lack of performance for all models may be due to several factors, which are discussed below.

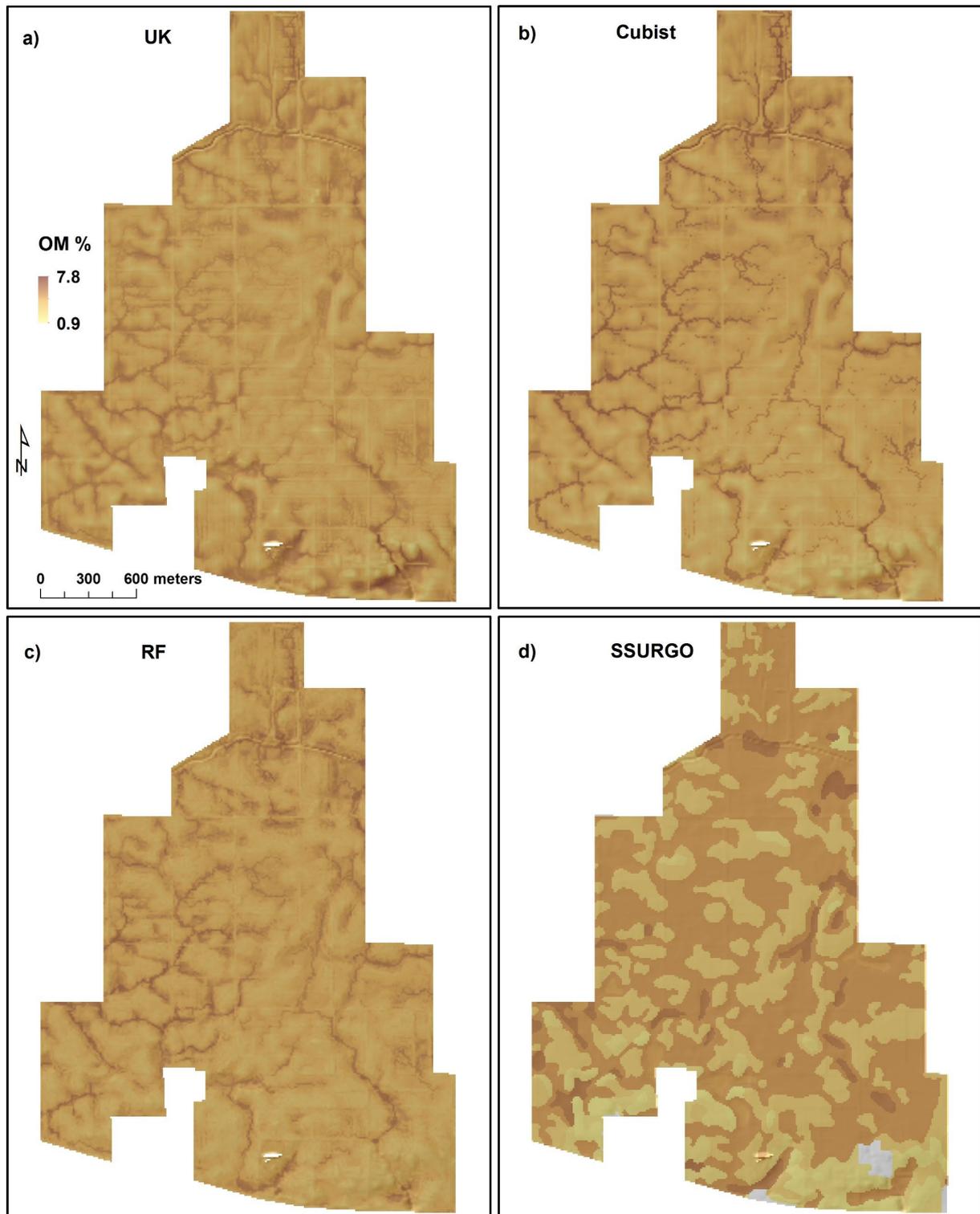
One reason for this poor correlation may be due to the history of the study location. ACRE serves as a research and education facility and consists of many smaller individual fields that are managed under highly variable practices (e.g., multiple tillage systems,

nutrient application rates, and crop rotations). For example, ACRE is crisscrossed by a grid of roads and grassed field boundaries that are on average 20 cm higher than the adjoining fields, and by a dense network of underground drainage tiles. The impact of the roads on terrain attributes is evident as linear features in the terrain attribute maps (Figure 2). The high variation in management may lead to higher soil variability from field to field than expected. Training models using samples from highly variable fields can limit their predictive performance outside the sample areas [62]. To account for the effects of variable management history, we would need to incorporate environmental covariates that describe previous management of each field into our modeling framework. This sort of detailed field-by-field management history has not been recorded for the entire farm. Recent research from [63] conducted on a long-term soil monitoring network in Switzerland has highlighted the role of land use change, crop rotations and site conditions on soil organic carbon dynamics. Further research is needed to identify suitable covariates that describe agriculture management history [64].

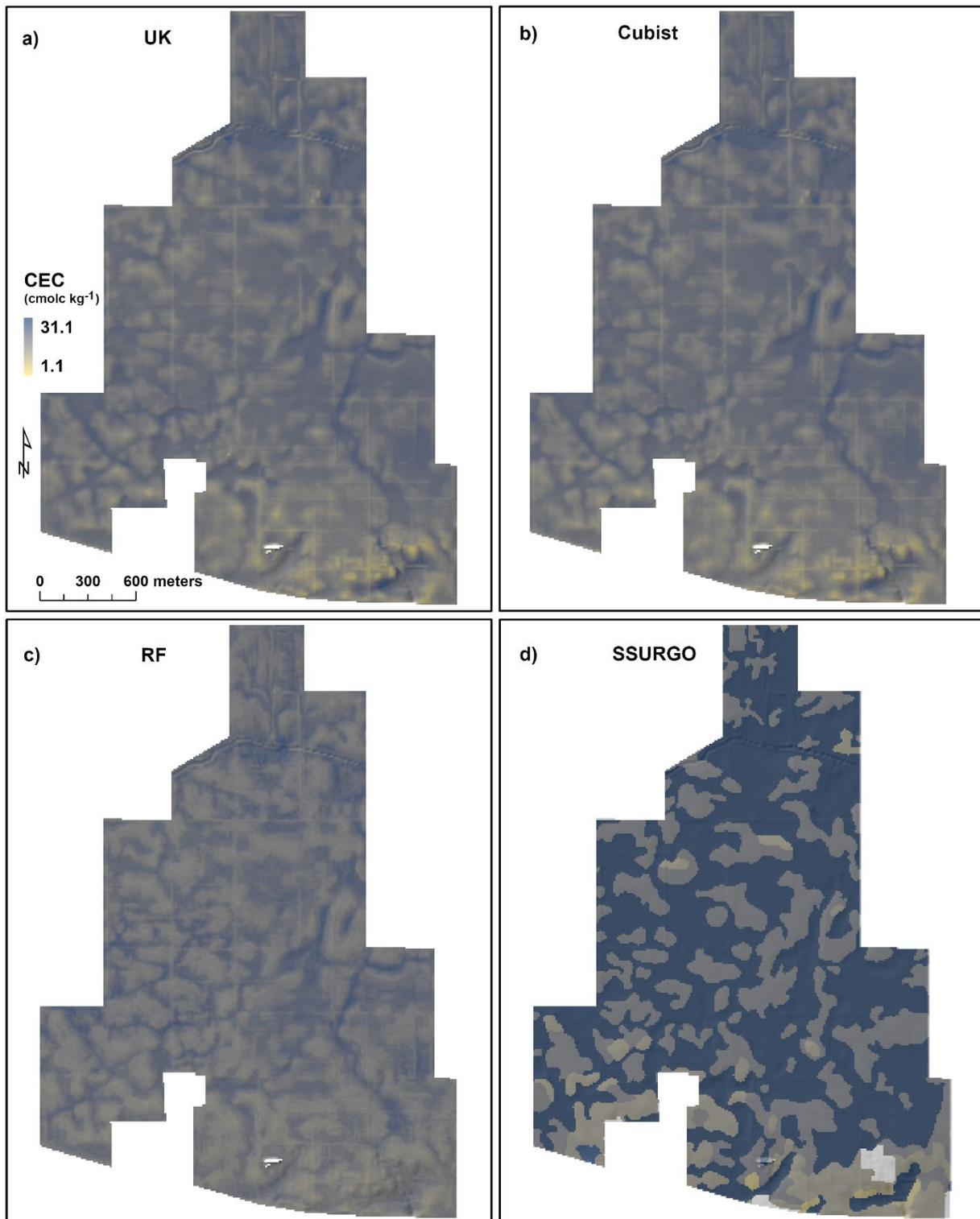
Compounding the land use and management history, the relatively flat topography with subtle topographic variation (on average 1% slope based on a  $3 \times 3$  pixel window) may also have contributed to relatively poor performance of all models. In many environments, chemical properties of surface soil and SOM are highly variable spatially, and distinct variations are often found within short distances of meters and/or decimeters [65,66]. Thus, the intensive land use and management history combined with relatively flat terrain may have diluted the influence of terrain in the distribution of SOM and CEC leading to poorer than expected model performance. Although the  $R^2$  for evaluation were relatively low, ranging from 0.39 to 0.45 in our study, they were comparable with other studies that considered terrain/climatic data only [16,50,67]. Ref. [16], argues that for quantitative soil spatial models,  $R^2$  values of 0.5 and less are not uncommon. The fact that models were based only on terrain attributes supports the idea that topography at this scale is still one of the major factors for predicting soil properties despite management. The RMSE values of all three models were lower compared to [11], which used RF for soil organic carbon prediction. They found a higher RMSE (1.72%) for the surface soil (0–10 cm) and lower (0.43) at the subsurface (10–50 cm). This suggests that the spatial distribution patterns of soil organic carbon particularly at the topsoil are highly variable. Ref. [14] reported that based on validation dataset, geostatistical model (i.e., Cokriging) provided better results ( $R^2 = 0.57$  and  $RMSE = 7.22$ ) compared to the RF ( $R^2 = 0.47$  and  $RMSE = 7.89$ ) in CEC prediction. However a higher performance of kriging over RF and cubist models on the prediction of soil organic carbon at field-scale was also reported by [68]. Unlike in our study, a high-sampling density would have favored kriging in their case.

### 3.5. Organic Matter Content and Cation Exchange Capacity Distribution in the Landscape

DSM model predictions of SOM (Figure 7) and CEC (Figure 8) were consistent with the distributions of SOM and CEC within the landscape based on theoretical and pedological principles. The maps of predicted SOM and CEC for all three models indicate higher values for SOM and CEC in lower landscape positions (i.e., foot and toeslopes), and lower values at higher and steeper landscape positions (i.e., shoulders and summits). Ref. [10] also reported higher accumulation of soil organic carbon in lower or concave areas. Lower areas receive more overland flow of nutrients and crop residue from the steeper areas leading to an increase in SOM and CEC over time. The steeper regions are subject to erosion and loss of nutrients and soil organic matter to the lower parts of the landscape. On the other hand, waterlogging in lower areas reduces the rate of soil organic matter decomposition and results in higher SOM and nutrient accumulation [69].



**Figure 7.** Soil organic matter content (SOM) prediction. (a) Universal kriging (UK), (b) Cubist, (c) random forest (RF), and (d) soil survey geographic (SSURGO).



**Figure 8.** Cation exchange capacity (CEC) prediction. (a) Universal kriging (UK), (b) Cubist, (c) random forest (RF), and (d) soil survey geographic (SSURGO).

### 3.6. Predictive Models versus SSURGO

When comparing maps of SSURGO SOM and CEC to DSM maps, all maps generally show a similar trend: high SOM and CEC occurred on lower landscape positions (Figures 7 and 8). Where these maps differ from SSURGO is in the extent of regions of high SOM and CEC and the level of detail within SSURGO map units. In SSURGO, the regions

or map units of high SOM and CEC are larger in extent compared to the DSM maps. Generally, SSURGO overrepresented the areas with high SOM and CEC (Figures 7 and 8). For example, SSURGO representative values had a median SOM of 4.9% compared to 4.1 and 4.2% for DSM maps (Table 6). Similarly, the SSURGO mean values for CEC had a median of 28.5% while DSM maps had a median between 19.4 and 20.0% (Table 6). Additionally, the standard deviation of SSURGO SOM (1.2%) and CEC (6.6%) maps are higher compared to DSM Maps, which is less than 0.8% for SOM and 2.8–3.3% for CEC (Table 6). Some of the major reasons for the overrepresentation of the areas with high SOM and CEC are the scale 1:15,840 [24] of SSURGO mapping and the design of map units. The density of the point data combined with high resolution topographic data provides a more detailed map of the SOM and CEC distribution compared to SSURGO. Also, SSURGO SOM and CEC capture the variability of the soil property within a larger and much more generalized map units thus tend to overrepresent the ranges and their extent within individual soil polygons.

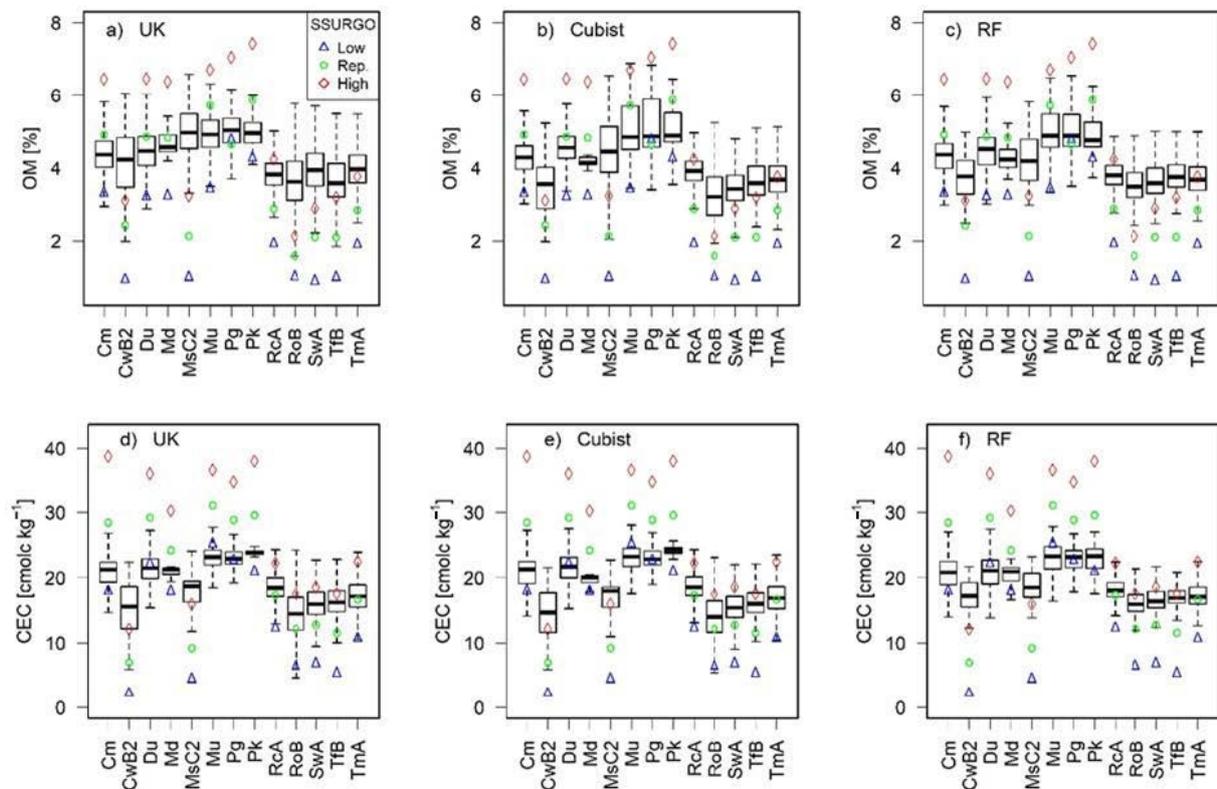
**Table 6.** Summary statistics of universal kriging (UK), Cubist, random forest (RF), and soil survey geographic (SSURGO) organic matter content (SOM) and cation exchange capacity (CEC) maps.

Statistical Index	SOM				CEC			
	UK	Cubist	RF	SSURGO	UK	Cubist	RF	SSURGO
	%				cmol <sub>c</sub> kg <sup>-1</sup>			
Minimum	0.9	1.9	2.4	1.6	1.1	5.2	11.9	7.0
1st Quartile	3.8	3.8	3.7	2.9	17.7	17.6	17.5	17.4
Median	4.2	4.2	4.1	4.9	20.0	19.9	19.4	28.5
Mean	4.2	4.2	4.2	4.0	19.6	19.6	19.7	23.5
3rd Quartile	4.7	4.6	4.6	4.9	22.0	22.1	21.8	28.5
Maximum	7.8	7.2	6.5	5.9	28.6	28.1	27.8	31.1
Standard Deviation	0.7	0.8	0.7	1.2	3.2	3.3	2.8	6.6

One interesting area of agreement between SSURGO and DSM maps is for CEC predictions in the southern quarter of the study area. In this area, both SSURGO and DSM models predicted the lowest CEC values. Even though sampling points were not concentrated at this part of the study site (Figure 1), DSM models still managed to predict these regions of low CEC. Low DSM-derived CEC predictions likely resulted from the low TPI in the study areas (see the spatial trend modeling section). While SSURGO was not developed using TPI specifically, SSURGO mapping did rely heavily on relationships between soils and slope positions, which TPI captures numerically. Agreement between DSM-predicted CEC and SSURGO maps highlights the importance of soil-landscape relationships in soil spatial distributions, even at field scale.

We compared SOM and CEC predicted by DSM techniques to SOM and CEC from the SSURGO soil map. According to Table 5, the prediction performance (i.e.,  $R^2$  and concordance) of SSURGO was analogous to the DSM prediction, however, SSURGO has higher bias and RMSE particularly for CEC prediction. It is also interesting that SSURGO show slightly better results for validation data when compared to calibration data. We also compared SOM and CEC contents predicted by DSM to the SOM and CEC contents within each map unit from SSURGO (Figure 9). Both SOM and CEC show that the three predictive models follow similar prediction trends in each of the SSURGO mapping units. Based on visual interpretation of boxplots (Figure 9a–c), the results of our models for SOM are consistent with the estimates from eight SSURGO mapping units; exceptions were CwB2, McS2, RoB, SwA, Tfb, and TmA. SSURGO underrepresented the SOM for these map units while the other models predicted greater concentrations of SOM. Generally, SSURGO had a wider range in SOM and CEC values when compared to the prediction models (Figure 9). This was particularly the case for CEC estimates. The prediction of our models for CEC is consistent with only few of the SSURGO mapping units see: RcA, RoB, SwA, Tfb, and

TmA (Figure 9d–f). Based on visual inspection of the boxplots (Figure 9d–f), however, for most of the mapping units, our models either over- or under-predicted CEC.



**Figure 9.** Box-plots of model predicted values of soil organic matter content (SOM) and cation exchange capacity (CEC) for each SSURGO map unit. The boxplots of (a–c) show SOM and (d–f) show CEC prediction based on universal kriging (UK), Cubist, and random forest (RF), respectively. Triangles show low (Low), rhombuses show high (High) and circles show representative (Rep) values of SOM and CEC for each SSURGO map unit.

There are several reasons for the inconsistencies of model predictions with SSURGO. First, SSURGO has inherent limitations; the soil variability is represented using aggregated polygon map units with one to four named components plus inclusions of other soils or non-soils areas that do not explicitly capture the underlying spatial variability of soils within polygons [69]. Thus, these inclusions reduce the purity of the map units and impact interpretation and modeling [70]. Second, the procedure for SOM analysis differed between the datasets. The Walkley-Black method was used for the SSURGO data, while the loss-on-ignition (LOI) method was used for our collected data. Due to incomplete digestion of soil organic carbon, the Walkley-Black method usually underestimates SOM [71,72]. Additionally, the SSURGO values might have been impacted by errors introduced by the spline interpolation. Third, the SSURGO database was developed based on historical soil survey data and may not accurately reflect the current status of soil properties, particularly SOM and CEC, which are relatively dynamic and altered by various factors such as land management, climate change, and seasonal variability [9,73,74]. Additionally, the data were produced over different time periods and therefore inherit inconsistencies [4]. A fourth reason for the inconsistency is that the surveyors who collected data for SSURGO may not have had enough soil observations for building their mental models of soil variability as the detailed scale that lidar may suggest. A fifth reason for the inconsistency is that SSURGO values are not purely derived from laboratory analysis, instead the data may have resulted from a combination of laboratory measurements and field observations of expert soil scientists [38,75]. Due to these shortcomings, using SSURGO data in quantitative

modeling and/or for monitoring soil carbon stocks sequestration could be misleading, particularly at the farm scale for highly variable soils in post-glaciated landscapes.

#### 4. Conclusions

In this study we developed digital soil maps of organic matter and cation exchange capacity for a 570-ha research farm, the Purdue Agronomy Center for Research and Education (ACRE). Digital soil maps were developed using only terrain attributes derived from lidar based digital elevation model. We compared three model approaches for predicting spatial distribution of SOM and CEC: universal kriging (UK), Cubist, and Random Forest. All models performed similarly for both SOM and CEC. Considering the high variability in farm management practices and nutrient application, the prediction accuracies were considered reasonable. All three predictive models showed similar spatial patterns that were comparable to the SSURGO map unit extents. This highlights the strength of the soil-landscape models as captured by predictive models or the soil mental models of soil surveyors (SSURGO) for mapping SOM and CEC. Overall SSURGO had a wider range of predicted soil properties when compared to the DSM models. However, DSM showed a finer-resolution view of the distribution of soil properties at ACRE and would be more informative for field precision management than the comparatively coarse-scale SSURGO data.

The results of this study demonstrate that lidar data can be used to adequately predict SOM and CEC at the farm scale in this post-glaciated landscape. This result has important implication for DSM in areas where certain environmental covariates may not be suitable or available. Provided there is a coupling between terrain attributes and soil properties, and that sampling locations capture the relationships between landscape and soil properties, even relatively simple empirical models such as universal kriging, can adequately predict spatial distribution of SOM and CEC.

**Author Contributions:** Conceptualization, S.R.R., J.P.A., D.S. and Z.L.; methodology, S.R.R., J.P.A., D.S. and Z.L.; software, S.R.R.; validation, S.R.R.; formal analysis, S.R.R.; investigation, S.R.R.; resources, D.S.; data curation, S.R.R.; writing—original draft preparation, S.R.R.; writing—review and editing, S.R.R., J.P.A., D.S., K.A. and Z.L.; visualization, S.R.R.; supervision, D.S. and Z.L.; project administration, D.S.; funding acquisition, D.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Jenny Sutherland for extensively editing and proof reading of the manuscript. Zamir Libohova is a member of Research Consortium GLADSOILMAP, supported by LE STUDIUM Loire Valley Institute for Advanced studies. USDA is an equal opportunity provider and employer.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Abbreviations

ACRE, Agronomy Center for Research and Education; CEC, cation exchange capacity; cLHS, conditioned Latin hypercube; DEM, digital elevation model; DSM, digital soil mapping; GIS, geographic information system; MrRTF, multi resolution ridge top flatness index; MrVBF, multi resolution valley bottom flatness index; NRCS, Natural Resources Conservation Service; RF, random forest; SOM, organic matter content; SSURGO, Soil Survey Geographic database; STATSGO, state soil geographic database; TPI, topographic position index; TWI, topographic wetness index; UK, universal kriging.

#### References

1. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*, 1st ed.; McGraw-Hill Publications in the Agricultural Sciences; McGraw-Hill: New York, NY, USA, 1941.
2. Milne, G. Some suggested units of classification and mapping particularly for East African soils. *Soil Res. Victoria* **1935**, *4*, 183–198.

3. Soil Survey Staff. Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Available online: <https://sdmdataaccess.sc.egov.usda.gov> (accessed on 21 May 2021).
4. Nauman, T.W.; Thompson, J.A.; Odgers, N.P.; Libohova, Z. Fuzzy disaggregation of conventional soil maps using database knowledge extraction to produce soil property maps. In *Digital Soil Assessments and Beyond*; Minasny, B., Malone, B.P., McBratney, A.B., Eds.; Taylor & Francis Group of CRC Press: London, UK, 2012; pp. 203–208.
5. McBratney, A.B.; Santos, M.M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
6. Minasny, B.; McBratney, A.B.; Malone, B.P.; Wheeler, I. Digital mapping of soil carbon. *Adv. Agron.* **2013**, *118*, 1–47. [[CrossRef](#)]
7. Arrouays, D.; Vion, I.; Kicin, J.L. Spatial analysis and modeling of topsoil carbon storage in temperate forest humic loamy soils of France. *Soil Sci.* **1995**, *159*, 191–198. [[CrossRef](#)]
8. Lacoste, M.; Minasny, B.; McBratney, A.; Michot, D.; Viaud, V.; Walter, C. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma* **2014**, *213*, 296–311. [[CrossRef](#)]
9. Grigal, D.F.; Vance, E.D. Influence of soil organic matter on forest productivity. *N. Z. J. For. Sci.* **2000**, *30*, 169–205.
10. Florinsky, I.V.; Eilers, R.G.; Manning, G.R.; Fuller, L.G. Prediction of soil properties by digital terrain modelling. *Environ. Model. Softw.* **2002**, *17*, 295–311. [[CrossRef](#)]
11. Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, H. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma* **2008**, *146*, 102–113. [[CrossRef](#)]
12. McKenzie, N.J.; Austin, M.P. A quantitative Australian approach to medium and small scale surveys based on soil stratigraphy and environmental correlation. *Geoderma* **1993**, *57*, 329–355. [[CrossRef](#)]
13. Adhikari, K.; Hartemink, A.E. Soil organic carbon increases under intensive agriculture in the Central Sands, Wisconsin, USA. *Geoderma Reg.* **2017**, *10*, 115–125. [[CrossRef](#)]
14. Chagas, C.D.S.; Carvalho Júnior, W.D.; Pinheiro, H.S.K.; Xavier, P.A.M.; Bhering, S.B.; Pereira, N.R.; Calderano Filho, B. Mapping soil cation exchange capacity in a semiarid region through predictive models and covariates from remote sensing data. *Rev. Bras. Ciência Solo* **2018**, *42*. [[CrossRef](#)]
15. Mosleh, Z.; Salehi, M.H.; Jafari, A.; Borujeni, I.E.; Mehnatkesh, A. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environ. Monit. Assess.* **2016**, *188*, 195. [[CrossRef](#)]
16. Malone, B.P.; McBratney, A.B.; Minasny, B.; Laslett, G.M. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **2009**, *154*, 138–152. [[CrossRef](#)]
17. John, K.; Abraham, I.I.; Michael, K.N.; Okon, A.E.; Chapman, A.P.; Marcus, A.S. Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. *Land* **2020**, *9*, 487. [[CrossRef](#)]
18. Behrens, T.; Förster, H.; Scholten, T.; Steinrücken, U.; Spies, E.D.; Goldschmitt, M. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil. Sci.* **2005**, *168*, 21–33. [[CrossRef](#)]
19. Fatholouloumi, S.; Vaezi, A.R.; Alavipanah, S.K.; Ghorbani, A.; Saurette, D.; Biswas, A. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. *Sci. Total Environ.* **2020**, *721*, 137703. [[CrossRef](#)] [[PubMed](#)]
20. Mendes, W.D.S.; Demattê, J.A.M.; Salazar, D.F.U.; Amorim, M.T.A. Geostatistics or machine learning for mapping soil attributes and agricultural practices. *Rev. Ceres* **2020**, *67*, 330–336. [[CrossRef](#)]
21. Zhang, G.L.; Feng, L.I.U.; Song, X.D. Recent progress and future prospect of digital soil mapping: A review. *J. Integr. Agric.* **2017**, *16*, 2871–2885. [[CrossRef](#)]
22. Minasny, B.; McBratney, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* **2006**, *32*, 1378–1388. [[CrossRef](#)]
23. Peng, Y.; Xiong, X.; Adhikari, K.; Knadel, M.; Grunwald, S.; Greve, M.H. Modeling soil organic carbon at regional scale by combining multi-spectral images with laboratory spectra. *PLoS ONE* **2015**, *10*, e0142295. [[CrossRef](#)]
24. USDA–NRCS. *Soil Survey of Tippecanoe County, Indiana*; United States Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA, 1998.
25. Roudier, P.; Beaudette, D.E.; Hewitt, A.E. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In *Digital Soil Assessments and Beyond*; CRC Press: Boca Raton, FL, USA, 2012; pp. 227–231.
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018. Available online: <https://www.R-project.org/> (accessed on 30 May 2022).
27. NCR. *Recommended Chemical Soil Test Procedure for the North Central Region*; North Central Region Publication No. 221; Missouri Agriculture Experiment Station: Columbia, MO, USA, 1998.
28. Woolpert. Indiana Statewide Lidar 2017 B17 West—Airborne Lidar Report. Available online: <https://lidar.jinha.org/> (accessed on 10 June 2021).
29. INDOT. *Indiana Geospatial Coordinate System, Version 1.05*; Indiana Department of Transportation (INDOT), Aerial and Land Survey Office: Indianapolis, IN, USA, 2016.
30. Winzeler, H.E.; Owens, P.R.; Joern, B.C.; Camberato, J.J.; Lee, B.D.; Anderson, D.E.; Smith, D.R. Potassium fertility and terrain attributes in a Fragiudalf drainage catena. *Soil. Sci. Soc. Am. J.* **2008**, *72*, 1311–1320. [[CrossRef](#)]
31. Redlands, California, USA, C.E.S.R.I. ArcGIS Desktop: Release 10.6. 2011. Available online: <https://esri.com> (accessed on 30 May 2022).

32. Soil Science Division Staff. Soil survey manual. In *USDA Handbook 18*; Ditzler, C., Scheffe, K., Monger, H.C., Eds.; Government Printing Office: Washington, DC, USA, 2017.
33. Suleymanov, A.; Abakumov, E.; Suleymanov, R.; Gabbasova, I.; Komissarov, M. The Soil Nutrient Digital Mapping for Precision Agriculture Cases in the Trans-Ural Steppe Zone of Russia Using Topographic Attributes. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 243. [[CrossRef](#)]
34. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model. Dev. Discuss.* **2015**, *8*, 2271–2312. [[CrossRef](#)]
35. Weiss, A. Topographic position and landforms analysis. In Proceedings of the Annual Meeting of ESRI User Conference, San Diego, CA, USA, 9–13 July 2001. Available online: <http://www.jennessent.com/downloads/tpi-poster-tnc18%C3%9722.pdf> (accessed on 12 December 2020).
36. Gallant, J.C.; Dowling, T.I. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* **2003**, *39*. [[CrossRef](#)]
37. Soil Survey Staff. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online: <https://websoilsurvey.nrcs.usda.gov/> (accessed on 12 December 2020).
38. Libohova, Z.; Odgers, N.P.; Ashtekar, J.; Owens, P.R.; Thompson, J.A.; Hempel, J. Some challenges on quantifying soil property predictions uncertainty for the GlobalSoilMap using legacy data. In *Digital Soil Mapping Across Paradigms, Scales and Boundaries*; Zhang, G.L., Brus, D., Liu, F., Song, X.D., Lagacherie, P., Eds.; Springer: Singapore, 2016; pp. 131–140. [[CrossRef](#)]
39. Franzmeier, D.P.; Steinhardt, G.C.; Crum, J.R.; Norton, L.D. *Soil Characterization in Indiana: I. Field and Laboratory Procedures*; West Lafayette, Indiana with cooperation of the Soil Conservation Service, U.S. Department of Agriculture. Research Bulletin No. 943; Department of Agronomy, Agriculture Experiment Stations, Purdue University: West Lafayette, IN, USA, 1977.
40. Soil Survey Staff. *Kellogg Soil Survey Laboratory Methods Manual*; Soil Survey Investigations Report No. 42, Version 5.0.; Burt, R., Soil Survey Staff, Eds.; U.S. Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA, 2014.
41. R Package, Version 1.0. *Ithir: Soil Data and Some Useful Associated Functions*, R Package: Sydney, Australia, 2018.
42. Malone, B.P.; Minasny, B.; McBratney, A.B. *Using R for Digital Soil Mapping*; Springer International Publishing: Basel, Switzerland, 2017. [[CrossRef](#)]
43. Adhikari, K.; Hartemink, A.E.; Minasny, B.; Bou Kheir, R.; Greve, M.B.; Greve, M.H. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE* **2014**, *9*, e105519. [[CrossRef](#)]
44. Bishop, T.F.; McBratney, A.B.; Laslett, G.M. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **1991**, *91*, 27–45. [[CrossRef](#)]
45. Hengl, T.; Heuvelink, G.B.; Rossiter, D.G. About regression-kriging: From equations to case studies. *Comput. Geosci.* **2007**, *33*, 1301–1315. [[CrossRef](#)]
46. Gräler, B.; Pebesma, E.; Heuvelink, G. Spatio-Temporal Interpolation using gstat. *RFID J.* **2016**, *8*, 204–218. [[CrossRef](#)]
47. Quinlan, J.R. Learning with continuous classes. In Proceedings of the Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 16–18 November 1992; pp. 343–348.
48. Kuhn, M.; Quinlan, R. Cubist: Rule and Instance Based Regression Modeling. R Package Version 0.2.2. Available online: <https://CRAN.R-project.org/package=Cubist> (accessed on 12 December 2020).
49. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
50. Forkuor, G.; Hounkpata, O.K.; Welp, G.; Thiel, M. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. *PLoS ONE* **2017**, *12*, e0170478. [[CrossRef](#)]
51. Liaw, A.; Wiener, M. Classification and regression by random Forest. *R News* **2002**, *2*, 18–22.
52. Lawrence, I.; Lin, K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1989**, *45*, 255–268. [[CrossRef](#)]
53. McBride, G.B. *A Proposal for Strength-of-Agreement Criteria for Lin's*; Concordance Correlation Coefficient; National Institute of Water and Atmospheric Research Ltd.: Hamilton, New Zealand, 2005. Available online: <https://www.medcalc.org/download/pdf/McBride2005.pdf> (accessed on 12 December 2020).
54. Wiesmeier, M.; Barthold, F.; Blank, B.; Kögel-Knabner, I. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil* **2011**, *340*, 7–24. [[CrossRef](#)]
55. Cambardella, C.A.; Moorman, T.B.; Novak, J.M.; Parkin, T.B.; Karlen, D.L.; Turco, R.F.; Konopka, A.E. Field-scale variability of soil properties in central Iowa soils. *Soil Sci. Soc. Am. J.* **1994**, *58*, 1501–1511. [[CrossRef](#)]
56. Keskin, H.; Grunwald, S. Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma* **2018**, *326*, 22–41. [[CrossRef](#)]
57. Rossi, J.; Govaerts, A.; De Vos, B.; Verbist, B.; Vervoort, A.; Poesen, J.; Muys, B.; Deckers, J. Spatial structures of soil organic carbon in tropical forests—A case study of Southeastern Tanzania. *Catena* **2009**, *77*, 19–27. [[CrossRef](#)]
58. Greve, M.H.; Greve, M.B.; Kheir, R.B.; Bøcher, P.K.; Larsen, R.; McCloy, K. Comparing decision tree modeling and indicator Kriging for mapping the extent of organic soils in Denmark. In *Digital Soil Mapping*; Springer: Dordrecht, The Netherlands, 2010; pp. 267–280.
59. Nabiollahi, K.; Eskandari, S.; Taghizadeh-Mehrjardi, R.; Kerry, R.; Triantafyllis, J. Assessing soil organic carbon stocks under land-use change scenarios using random forest models. *Carbon Manag.* **2019**, *10*, 63–77. [[CrossRef](#)]

60. Nussbaum, M.; Spiess, K.; Baltensweiler, A.; Grob, U.; Keller, A.; Greiner, L.; Schaepman, M.E.; Papritz, A. Evaluation of digital soil mapping approaches with large sets of environmental covariates. *Soil* **2018**, *4*, 1–22. [[CrossRef](#)]
61. Rentschler, T.; Werban, U.; Ahner, M.; Behrens, T.; Gries, P.; Scholten, T.; Teuber, S.; Schmidt, K. 3D mapping of soil organic carbon content and soil moisture with multiple geophysical sensors and machine learning. *Vadose Zone J.* **2020**, *19*, e20062. [[CrossRef](#)]
62. Rossel, R.V.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [[CrossRef](#)]
63. Guillaume, T.; Bragazza, L.; Lévassieur, C.; Libohova, Z.; Sinaj, S. Long-term soil organic carbon dynamics in temperate cropland-grassland systems. *Agric. Ecosyst. Environ.* **2021**, *305*, 107184. [[CrossRef](#)]
64. Caley, G.K.; Hengl, T.; Gräler, B.; Meyer, H.; Magney, T.S.; Brown, D.J. Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D+ T: The Cook Agronomy Farm data set. *Spat. Stat.* **2015**, *14*, 70–90. [[CrossRef](#)]
65. Schöning, I.; Totsche, K.U.; Kögel-Knabner, I. Small scale spatial variability of organic carbon stocks in litter and solum of a forested Luvisol. *Geoderma* **2006**, *136*, 631–642. [[CrossRef](#)]
66. Trangmar, B.B.; Yost, R.S.; Uehara, G. Application of geostatistics to spatial studies of soil properties. *Adv. Agron.* **1985**, *38*, 45–94. [[CrossRef](#)]
67. Mason, E.; Sulaeman, Y. Comparison of three models for predicting the spatial distribution of soil organic carbon in Boalemo Regency, Sulawesi. *J. Ilmu. Tanah. Dan. Lingkungan.* **2016**, *18*, 42–48. [[CrossRef](#)]
68. Pouladi, N.; Møller, A.B.; Tabatabai, S.; Greve, M.H. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma* **2019**, *342*, 85–92. [[CrossRef](#)]
69. Nauman, T.W.; Thompson, J.A. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma* **2014**, *213*, 385–399. [[CrossRef](#)]
70. Geza, M.; McCray, J.E. Effects of soil data resolution on SWAT model stream flow and water quality predictions. *J. Environ. Manag.* **2008**, *88*, 393–406. [[CrossRef](#)]
71. De Vos, B.; Lettens, S.; Muys, B.; Deckers, J.A. Walkley–Black analysis of forest soil organic carbon: Recovery, limitations and uncertainty. *Soil Use Manag.* **2007**, *23*, 221–229. [[CrossRef](#)]
72. Shamrikova, E.V.; Kondratenok, B.M.; Tumanova, E.A.; Vanchikova, E.V.; Lapteva, E.M.; Zonova, T.V.; Lu-Lyan-Min, E.I.; Davydova, A.P.; Libohova, Z.; Suvannang, N. Transferability between soil organic matter measurement methods for database harmonization. *Geoderma* **2022**, *412*, 115547. [[CrossRef](#)]
73. Bot, A.; Benites, J. *The Importance of Soil Organic Matter: Key to Drought-Resistant Soil and Sustained Food Production*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2005; Volume 80.
74. Toni, L.R.M.; de Santana, H.; Zaia, C.T.B.V.; Zaia, D.A.M. Seasonal and depth effects on some parameters of a forest soil. *Semin. Exact Technol. Sci.* **2009**, *30*, 19–32. [[CrossRef](#)]
75. Libohova, Z.; Seybold, C.; Adhikari, K.; Wills, S.; Beaudette, D.; Peaslee, S.; Lindbo, D.; Owens, P.R. The anatomy of uncertainty for soil pH measurements and predictions: Implications for modelers and practitioners. *Eur. J. Soil Sci.* **2019**, *70*, 185–199. [[CrossRef](#)]