



## Article

# Assessing Soil Prediction Distributions for Forest Management Using Digital Soil Mapping

Gonzalo Gavilán-Acuna <sup>1,\*</sup>, Nicholas C. Coops <sup>1</sup>, Guillermo F. Olmedo <sup>2</sup>, Piotr Tompalski <sup>3</sup>, Dominik Roeser <sup>1</sup> and Andrés Varhola <sup>1</sup>

<sup>1</sup> Department of Forest Resources Management, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada; nicholas.coops@ubc.ca (N.C.C.)

<sup>2</sup> Investigaciones Forestales Bioforest S.A., Camino a Coronel, Km. 15, Concepción 403 0000, Chile; guillermo.olmedo@arauco.com

<sup>3</sup> Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada

\* Correspondence: ggavilan@mail.ubc.ca; Tel.: +1-23-6869-9764

**Abstract:** Texture, soil organic matter (SOM), and soil depth (SoD) are crucial properties in forest management because they can supply spatial information on forest site productivity and guide fertilizer applications. However, soil properties possess an inherent uncertainty that must be mapped to enhance decision making in management applications. Most digital soil mapping predictions primarily concentrate on the mean of the distribution, often neglecting the estimation of local uncertainty in soil properties. Additionally, there is a noticeable scarcity of practical soil examples to demonstrate the prediction uncertainty for the benefit of forest managers. In this study, following a digital soil mapping (DSM) approach, a Quantile Regression Forest (QRF) model was developed to generate high-resolution maps and their uncertainty regarding the texture, SoD, and SOM, which were expressed as standard deviation (Sd) values. The results showed that the SOM ( $R^2 = 0.61$ , RMSE = 2.03% and with an average Sd = 50%), SoD ( $R^2 = 0.74$  and RMSE = 19.4 cm), clay ( $R^2 = 0.63$ , RMSE = 10.5% and average Sd = 29%), silt ( $R^2 = 0.59$ , RMSE = 6.26% and average Sd = 33%), and sand content ( $R^2 = 0.55$ , RMSE = 9.49% and average Sd = 35%) were accurately estimated for forest plantations in central south Chile. A practical demonstration of precision fertilizer application, utilizing the predictive distribution of SOM, effectively showcased how uncertainty in soil attributes can be leveraged to benefit forest managers. This approach holds potential for optimizing resource allocation and maximizing economic benefits.

**Keywords:** soil prediction uncertainty; precision forestry; LiDAR-derived DEM; SCORPAN



**Citation:** Gavilán-Acuna, G.; Coops, N.C.; Olmedo, G.F.; Tompalski, P.; Roeser, D.; Varhola, A. Assessing Soil Prediction Distributions for Forest Management Using Digital Soil Mapping. *Soil Syst.* **2024**, *8*, 55. <https://doi.org/10.3390/soilsystems8020055>

Academic Editor:  
Jarosław Zawadzki

Received: 1 February 2024  
Revised: 29 April 2024  
Accepted: 1 May 2024  
Published: 16 May 2024



**Copyright:** © 2024 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/>).

## 1. Introduction

Environmental characteristics and soil site conditions have a significant impact on forest growth and stand development [1], and as a result, understanding them is critical for forecasting planting limits on specific forest species and their productivity [2]. Soils, in particular, play an important role in regulating, supporting, and provisioning ecosystem services, and data on soil properties represent critical information for forest planning and management schemes [3]. Key soil properties such as the texture and organic matter content are essential to estimate site productivity, which is defined as a quantitative estimate of a site's potential to produce a volume of trees in a given time [4].

Soil texture has shown to be strongly related to the movement, retention, capacity, and availability of nutrients and water content, as well as the ease with which plant roots may penetrate the ground and absorb water [5]. Previous research on forest growth has observed that soil texture is one of the most important properties for evaluating soil compaction recovery to guide future planting procedures [6]. For example, Ref. [7] found a strong correlation ( $R^2 = 0.84$ ) between the soil texture content and soil density. Soil organic matter (SOM) is an important

chemical property for evaluating forest site fertility because it has a large influence on the soil cation exchange capacity [8], which is a vital factor for nutrient supply to plants [9]. Previous research found that increasing the SOM by 1% raised the site index (the height of dominant trees per stand in a target year) by approximately 20% in *Pinus banksiana* plantations and by 7% in *Picea glauca* plantations [10]. Additionally, soil depth (SoD), described as the depth of the soil profile from the top surface to the bedrock or root barriers [11], is also correlated to nutrient capacity and the plant's available water content, and it controls biological activity [12]. This soil property has been linked to site index prediction in *Pinus* plantations [13] as well as overall productivity in forest plantations [14]. Changes in the SoD have also been linked to variations in the basal area within hardwood plantations. Specifically, the loss of soil thickness due to erosion is projected to decrease the basal area from an initial estimate of 18 to 26 m<sup>2</sup>/ha to a range of 8 to 21 m<sup>2</sup>/ha in future generations of forest plantations [15].

Understanding the interplay between physical and chemical soil properties that contribute to forest biomass production and carbon storage is vital, particularly in managed forest plantations where the goal is to maximize productivity over short rotation cycles. Enhancing the spatial characterization of soil properties not only facilitates an increased understanding of site productivity [4] but also optimizes fertilizer applications. Forests planted on sites with high nutrient contents tend to be less responsive to fertilizer applications, potentially eliminating the need for such interventions [16]. Ref. [17], for instance, observed a 400% increase in volume in *Pinus ponderosa* plantations due to fertilizer applications at sites with low concentrations of organic matter and nitrogen content. Consequently, identifying site fertility is critical for reducing fertilization management costs without significantly compromising potential growth. Furthermore, soil property data helps in optimizing heavy machinery allocation during timber harvesting and site preparation in plantations to minimize soil compaction.

Conventionally, the representation of soil properties and their relationships with environmental characteristics are traditionally obtained through field surveys, soil pit samples, and laboratory analyses [18]. Several approaches for interpolating the spatial information of soil properties, commonly employing geostatistical models, have emerged in recent decades based on field and laboratory measurements and spatially explicit environmental data [19]. The introduction of the SCORPAN theoretical model has proven to provide the best approximation for the development of digital soil mapping (DSM), improving the spatial representation of numerous soil properties. This theoretical model posits that soil properties are the result of the interactions between seven factors or variables that describe soil formation: S (previously measured soil information), C (climate), O (land cover), R (topography), P (parent material), A (soil age), and N (spatial position) [18]. This approach has been frequently used to predict and map soil properties such as depth gradients of soil organic carbon [9] and soil texture [20]. These soil maps are designed to provide decision-makers with accurate and precise information, which is necessary in forest plantation management to improve and enable site-specific management operations [21], mainly in relation to precise fertilization treatments based on granular assessments of soil nutrient deficiencies [22]. Nonetheless, it is also well recognized that the maps produced using DSM techniques are not error-free [23].

Uncertainty in digital soil mapping can originate from modeling errors and measurement inaccuracies associated with the input data [23]. Quantifying uncertainty is a crucial step because it assesses the reliability of the prediction data, typically expressed within a range of confidence intervals [24]. In the context of precision silviculture, possessing a comprehensive understanding of the confidence intervals for predicted variables, such as soil fertility, holds significant importance when making site-specific fertilization management decisions [25]. This knowledge provides valuable advantages by allowing resources to be directed towards areas based on the inherent variability in soil properties. However, the majority of digital soil mapping predictions tend to focus solely on the mean of the prediction distribution and often avoid estimating local uncertainty in soil predictions. This is due to inherent challenges, particularly with machine learning models, which are

the preferred and most common approach for DSM [26]. For instance, the Random Forest (RF) method has gained widespread use in digital soil mapping, attributable to its ensemble of regression trees that yield more robust estimations with less biased internal error estimates [27]. Nevertheless, this approach only retains the mean of the prediction observations while disregarding other valuable information [28]. Other statistical models, such as the Bayesian Maximum Entropy (BME) model for spatial prediction, address the integration of data with uncertainty into the modeling process, aiming to enhance their predictive capabilities compared with traditional estimation methods [29]. However, this is normally challenged by computational complexity, sample size limitations [30], uncertainty in interpretability, and the complexity of parameter estimation [31].

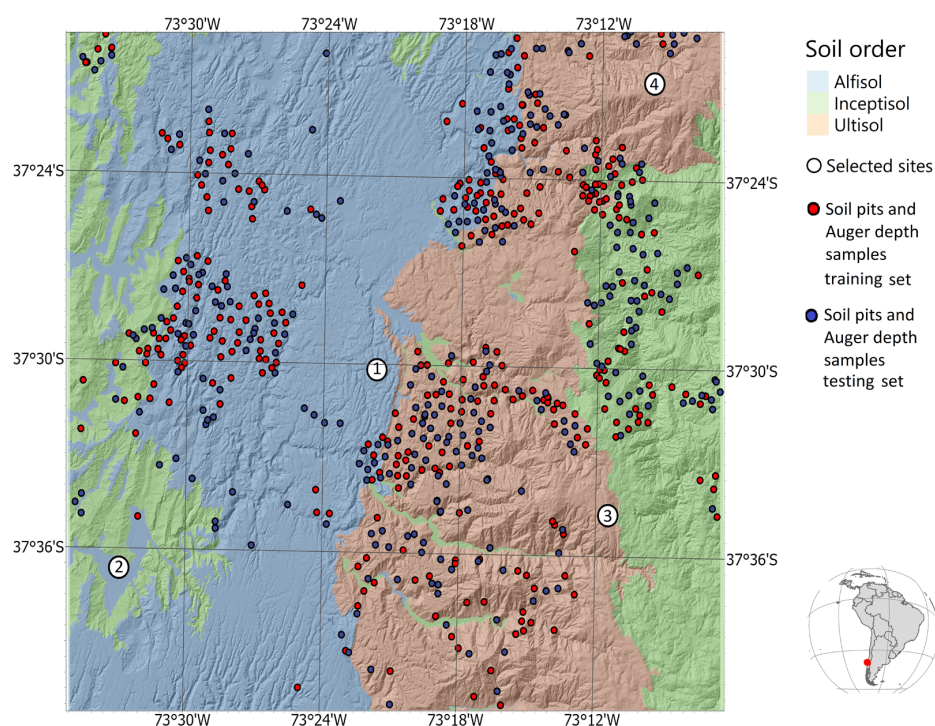
The Quantile Regression Forest (QRF) [32] approach is a tree-based ensemble method which has the advantage of allowing the measurement of prediction uncertainties for each soil property as well as the depiction of probability distributions of dependent environmental variables [28]. A QRF, akin to an RF, offers valuable information regarding both the median and the distribution of the target variable. The principal difference lies in the treatment of observations within each node and tree: while an RF retains only the mean of the observations, thereby discarding additional data, a QRF preserves the values of all observations [32]. In recent years, QRFs have been gaining popularity in soil investigations, providing accurate estimates of the SOM [23,33,34], clay content, and other soil parameters [28,35], including in-depth 3D representations of the SOM, pH, and clay content, among others [36]. While most studies have centered on comparing the QRF method with other machine learning algorithms concerning their predictive prowess—for example, for soil organic carbon [37] and soil organic matter [28,34]—with favorable results, there remains a gap in terms of its use for management decisions in the soil domain. The application of uncertainty for management is undoubtedly beneficial for decision making. Still, recent research lacks practical examples on harnessing this uncertainty for the benefit of forest managers.

Recognizing the significant gap in current research, this study introduces an innovative approach to soil science by developing probabilistic maps to aid forest management decisions. The ability to visualize a range of possible soil conditions across a given area presents a substantial advancement over conventional methods, which typically rely on less descriptive, average-based data. This probabilistic approach becomes essential for making informed, data-driven decisions crucial for sustainable forest management. By focusing on these maps, this study aims to enhance operational strategies and decision-making processes in forestry. The purpose of this study therefore is to evaluate the spatial uncertainty distribution for soil depth, texture, and organic matter, with the intention of applying these findings to forest management operations at a 10 m spatial resolution. For this, we utilized the Quantile Regression Forest (QRF) model approach. To do so, we investigated (1) the uncertainty of soil property mapping products, (2) elaborating a 3D map for the texture and SOM based on the SoD, and (3) practical management application examples that leverage fine resolution and its associated prediction distribution.

## 2. Study Area

The study area was located at a specific site in central south Chile, in the Región del Maule (35°14' S, 73°20' W). The diverse terrain of this region comprises the coastal Andes ranges, the central valley, and the Andes foothills. Alfisol soils, found along the coast, are described as highly fertile with high concentrations of aluminum (Al) and iron (Fe), and rich in clay content [38]. Ultisols are present on the Andes' coastal range and inceptisols occur in the valley [39] (Figure 1). Ultisols have a kandic or argillic horizon and few basic cations that have formed under forest vegetation in humid climates [40]. Inceptisols are soils with minimal horizon development, with some evidence of clay minerals, metal oxides, or humus accumulating in layers, but not enough to classify the soil into an order defined by characteristic surfaces [41]. The parent material in the study area varies from marine sediments at the coast, to metamorphic and granitic sediments in the valley. The

climate is rainy and temperate, with the annual precipitation varying widely from 1219 mm along the coast to 2835 mm along the Andes coastal range, showing dry periods in the summer and wetter periods in the rest of the year. The elevation on the study area ranges from sea level (~1 m) to 1280 m.a.s.l. A land-use map for the area is illustrated in the Appendix A section (Figure A1).



**Figure 1.** Map showing the study area distribution, with soil pit and auger information in red and blue (for training and testing sets, respectively), the four selected sites described in Figure 5 in white circles with their respective numbers, and the soil order within the ALS survey location. Additionally, the lower right image shows the delineation of the study area in the Americas.

### 3. Materials

#### 3.1. LiDAR

Airborne LiDAR technology, also known as airborne laser scanning (ALS), was used to collect data over 160,000 ha (Figure 1), spanning over a continuous region of forest plantations. Data acquisition took place between February and October of 2020–2021 following the specifications shown in Table 1.

**Table 1.** Airborne laser scanning (ALS) data acquisition parameters.

Acquisition Parameter	Value
Sensor	Optech Galaxy Prime (Teledyne Optech, Vaughan, Ontario, Canada).
Utilized plane	Tecnam P2006
Flying height	3000 m AGL
Average flying speed (knots)	115
Pulse repetition frequency (kHz)	700
Scan angle	26 deg
Returns recorded	Up to 5
Overlap	60%
Average overall point density	29.2 points/m <sup>2</sup>
Average pulse density	20.8 pulses/m <sup>2</sup>



### 3.2. Soil Pit and Auger Data

Within the ALS survey area, a soil profile legacy database collected by Forestal Arauco was used in this study, comprising 654 soil pits and 1442 soil depth samples obtained utilizing an auger in bore holes up to 4.1 m deep (auger length), with their spatial location recorded using a Garmin Handheld GPSMAP 64 2.6" GPS (Garmin Ltd., Olathe, Kansas, USA). Data for all soil profiles were collected just after harvesting between 1994 and 2018 and were distributed across *Pinus radiata* plantations.

Soil pits were 150 cm deep × 150 cm wide and their profile properties were evaluated using the recommendations provided in [17]: (i) horizon nomenclature, including master horizons and other modifiers, as well as horizon thickness (in cm); (ii) soil matrix color in the moist state using the Munsell notation; (iii) texture class determined using the field hand test; (iv) carbonates determined using the effervescence field test; and (v) stoniness in vol. %.

Soil pit samples were analyzed in the laboratory to obtain sand, clay, and silt content measurements, as well as SOM, determined using wet oxidation throughout the entire soil pit profile and horizon thickness (cm). Soil depth information was based on 2096 observations, which included both soil pits and auger samples. Table 2 contains a summary of all of the information for SoD, texture, and SOM.

**Table 2.** Soil information. Values shown represent the number of observations (N) and the mean, standard deviation, and minimum and maximum values.

Soil Property	N	Mean	Standard Deviation	Min	Max
SoD (cm)	2096	158.7	65.6	30	420
Silt%	654	26.1	9.25	6	71
Clay%	654	41.8	17.3	11	87
Sand%	654	32.1	12.9	1	64
SOM%	654	6.7	3.2	0.3	18.2

### 3.3. Climate

Long-term monthly temperature (maximum and minimum) and rainfall data from 1990 to 2020 were obtained at 500 m resolution from CR2 (Center for Climate and Resilience Research) [42], available online. This dataset encompasses the continental region of Chile on a consistent 0.05-degree latitude–longitude grid. It was developed using statistical models that calibrate various climatic variables against rigorously controlled observational data, including atmospheric variables from weather stations, topographical details, and land surface temperature data derived from the MODIS satellite sensor.

### 3.4. Existing Soil Information

The Chilean Natural Resources Information Center (CIREN) provided us with vector information on the soil morphological properties and parent material, available online [39], including soil class data following the USDA Soil Classification System at order, suborder, great-group, subgroup, family, and series levels.

### 3.5. Forest Cover

Landsat 8 OLI remote sensing imagery that covered the study area was acquired for the period from 2013 to 2022, ensuring a maximum cloud cover of no more than 20%. These images provided observations at 16-day intervals with a moderate spatial resolution of 30 m. They served as proxies for land-cover representation, and further details are discussed in Section 4.

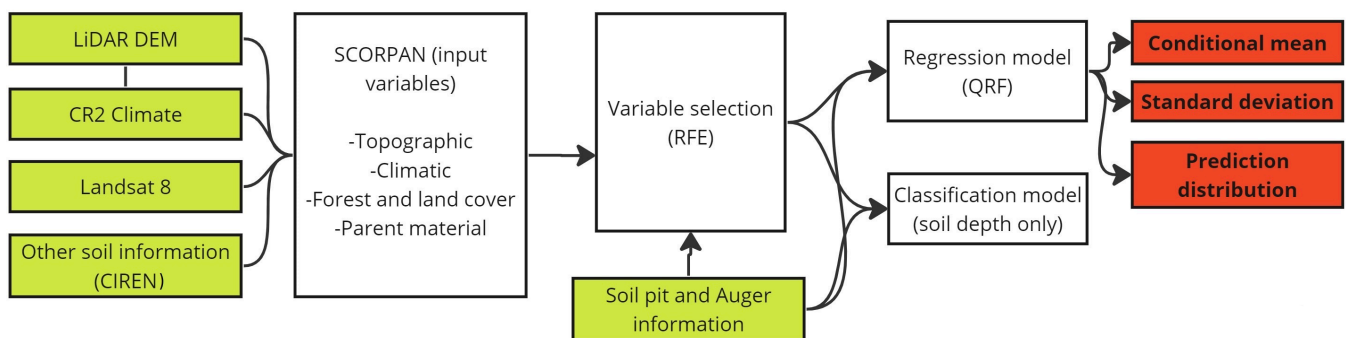
## 4. Methodology

### 4.1. Digital Elevation Model

The raw ALS data were processed using conventional routines to build a digital elevation model (DEM), which included tiling, filtering, and ground classification at a 10 m spatial resolution. For processing, the LAStools software package (version 211206) was employed [43]. The lasground algorithm was used to classify the ground data (with default parameters), and then blast2dem was used to create the DEM.

### 4.2. Modeling Soil Properties

The SCORPAN approach was used to model the SoD, SOM, and soil texture. The input variables for the SCORPAN model encompassed all of the aforementioned soil properties. For our research, we chose to produce a 10 m spatial resolution map for DSM. This decision was primarily driven by the increasing demand for site-specific precision in silviculture management and precision fertilization treatments within forest stands based on a granular assessment of the soil nutrient deficiencies. According to [44], fast-growing trees in plantation environments have a root lateral extension of 1.5 to 2.5 times the tree height, ranging from 10 m and up, depending on their age, for which a 10 m resolution map could account for individual tree interactions with chemical and physical soil properties affecting growth. An overview of the approach is showed in Figure 2.



**Figure 2.** Methodology workflow overview. The input data is highlighted in green. The models developed are shown in white boxes, and the final outcome is marked in red.

First, a set of input variables (environmental covariates) related to soil-forming factors were produced using the SCORPAN approach. Second, a variable selection method, specifically Recursive Feature Elimination, was used to determine the best variables for each soil attribute. Third, Quantile Regression Forest (QRF) [32] models were employed for digital soil mapping. Data were split into training (80%) and testing (20%) datasets, and their spatial distribution is illustrated in Figure 1. Models were assessed using coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute percent error (MAPE) values. In the fourth step, we derived a percentile distribution, focusing on the median—which offers more reliable predictions than the mean for soil attributes with outliers or skewed distributions—and assessed the uncertainty of each predicted soil attribute. In addition, by utilizing the entire range of the distribution of SOM, we informed the creation of a probabilistic map tailored for precision management applications. Further details are provided in Section 4.5.

For the soil depth data, an extra step was required due to the unbalanced distribution of the data (60% of observations were above 180 cm). First, to prevent the regression model from forecasting values predominantly over the class with the most data, the soil depth information was pre-classified using an RF classification model into either ‘above 180 cm’ or ‘below 180 cm’. The model was then validated with a confusion matrix using the testing dataset (above 180 cm only). The value of 180 was assigned for the SoD in pixels so that the model could predict the class ‘above 180 cm’. Second, a QRF model was developed for the subset of the data corresponding to soil depths below 180 cm. The final

soil depth prediction was an ensemble of both classification and regression models, with the regression output being applied to areas classified as less than 180 cm deep.

#### 4.3. Input Environmental Covariates for DSM

##### 4.3.1. Topographic Variables

Topographic attributes are important variables for soil formation [18] as they determine the pathways of surface water movement across a watershed and therefore affect watershed hydrologic responses to rainfall, as well as soil properties such as the organic matter decomposition rate and soil texture [45,46]. Primary topographic variables can be computed directly from the DEM [47]. We applied the Automated Geoscientific Analysis (SAGA) [48] approach to obtain nineteen primary topographic variables affecting soil formation properties (Table 3). All of these variables were derived directly from the DEM at a 10 m resolution.

**Table 3.** List of topographic input variables.

Code	Description
DEM	Digital elevation model
Aspect	Aspect degree (%)
CNBL	Channel network base level (m.a.s.l)
CND	Channel network distance
CI	Convergence index
DiffI	Diffuse insolation
DirI	Direct insolation
LS_factor	Slope length and steepness factor
MRRTF	Multiresolution Index of Ridge-Top Flatness
MRBVF	Multiresolution Index of Valley-Bottom Flatness
PC	Plan curvature
ProfC	Profile curvature
Slope	Slope degree
TanC	Tangential curvature
TSC	Terrain surface convexity
TWI	Topographic wetness index
TC	Total catchment area
TIns	Total insolation
ValDepth	Valley depth

##### 4.3.2. Forest Cover

Previous studies have successfully utilized satellite information and satellite indices as proxies to represent the land cover. Examples include the use of GlobeLand30 data, acquired from Landsat TM and ETM+ sensors [33], and NDVI data from Landsat 8 [49] to predict soil organic carbon. According to [50], maximum-value vegetation indices can capture the dynamics of green vegetation while reducing typical issues such as cloud contamination, surface directional reflectance, atmospheric attenuation, and view and lighting geometry. To reflect historical forest cover for the SCORPAN vegetation variables, two vegetation indices—the Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation index (NDVI)—were computed using Landsat 8 OLI at a 30 m resolution using Google Earth Engine. For this study, the EVI and NDVI were calculated from 2013 to 2022; to minimize the effect of clouds or the influence of atmospheric constituents on these vegetation indices, the maximum value per year was chosen to reflect the yearly NDVI and EVI. The final NDVI and EVI values were calculated as the 10-year average maximum values and then resampled to 10 m using the bilinear interpolation method.

##### 4.3.3. Climate Variables

Climate variables are useful in forecasting soil properties and have been applied to digital soil mapping [51]. Previous studies have used kriging with external drift (KED) in

order to interpolate and downscale temperature data using elevation as an external drift, improving the final forecast over kriging based solely on spatial information [52,53]. KED is a type of interpolation technique that combines spatial information from the data as related to an external drift defined by an auxiliary variable [54]. To verify that the climatic variables were integrated with the other SCORPAN variables (topography, parental material, and satellite vegetation indices), they were resampled to 10 m. For this, we applied the KED approach using the association between each monthly climatic variable (rainfall, minimum temperature, and maximum temperature described in the data) and the DEM as the external drift. This process was implemented using 10-fold cross-validation with the covariance function automatically chosen. The covariance function was selected automatically from either a Spherical, Exponential, Gaussian, or Matérn function via minimizing the prediction RMSE and coefficient of correlation.

These variables were then used to create 19 bioclimatic variables, calculated as a function of the monthly minimum and maximum temperatures and precipitation (mm) and elaborated with the KED approach (variables listed in Table A1), using the climate-based models provided by the United States Geological Survey (USGS).

#### 4.3.4. Parent Material and Other Soil Information

Parent material and lithology information are considered to be fundamental variables for digital soil mapping [18]. Parent material has been shown to have a considerable influence on soil properties such as texture, color, pH, and mineral composition [55]. The soil class information and parent material, available in vector format, were rasterized to a 10 m resolution.

#### 4.4. Variable Selection

According to [56], a large number of independent variables in machine learning models can lead to poor prediction performance and overfitting. Recursive Feature Elimination (RFE) is a feature selection method that tries to find the best feature subset based on the learned model and classification accuracy by removing features that have the least effect on training errors [57]. RFE is essentially a backward predictor selection method that selects features by recursively considering smaller and smaller subsets of features, then builds a model with the remaining attributes and calculates the model's accuracy using internal cross-validation [58]. This step is critical for avoiding overfitting issues caused by a high-dimensional dataset with an excessive number of features [59]. The RFE pre-processing method was used to extract a subset of relevant variables from the 44 covariates available. The top variables were selected based on a measure of the percent increase in mean squared error. This score indicates how much each predictor variable contributes to the accuracy of the model. We used 5 folds and 5 repeats while running the RFE.

#### 4.5. Quantile Regression Forest

As previously mentioned, a QRF not only estimates the conditional mean but also retains information on other quantiles of the response variable, which can be used to generate the prediction distribution. The QRF framework employs the power of Random Forests by building multiple decision trees, which then aggregate the predicted quantiles from each tree, offering a comprehensive distributional forecast rather than a single point estimate [28]. This method proves especially valuable when the interest lies in understanding the relationship between a set of predictor variables and specific percentiles of the response variable [32]. Hereafter, these will be referred to as percentiles for clarity. For instance, a prediction at the 50th percentile corresponds to the median of the dependent variable. Similarly, predictions at the 25th and 75th percentiles provide estimates of the lower and upper quartiles, respectively, thus providing a more detailed perspective on the range of potential outcomes. In the context of soil analysis, this allows us to observe the variability in predictions of soil properties, thereby aiding in more targeted soil management strategies.



The QRF technique was used to develop a regression model for predicting the soil texture, SOM, and SoD under 180 cm.

#### 4.5.1. Uncertainty, Soil Property Predictions, and Forest Management Practical Example

For each soil property, the median (50th percentile) and standard deviation (SD) (representing uncertainty) were determined. Sd is determined based on the entire prediction distribution per pixel. Additionally, for the four selected sites across the study area (see Figure 1), various percentiles were calculated to understand the conditional distribution and variability in the response outcome. The percentiles selected for this purpose were the 10th, 25th, 50th, 75th, and 90th percentiles. These sites were selected for their variability in terms of DEM at the study site, located at 227, 8, 942, and 817 m above sea level for sites 1, 2, 3, and 4, respectively, and covered different types of soil orders.

Subsequently,  $R^2$  and RMSE values were calculated using the testing dataset, based on the median.

To incorporate predictive uncertainty and the entire distribution of soil properties for forest management applications, we constructed a practical example. In this example, we used predictions of SOM to identify sites with varying fertility levels. We established a threshold of 5% SOM, directing the allocation of fertilizers to sites where fertility levels were at or below this threshold. This 5% SOM threshold was selected arbitrarily to suit the needs of our example; however, this value can be adjusted based on fertility research that examines the relationship between SOM and forest yield, thereby establishing a more appropriate threshold for fertilizer application. Using the QRF, we modeled the SOM content across a range of quantiles, from the 1st to the 99th percentile (0.01 to 0.99). This method allowed us to estimate the conditional distribution of SOM content given the selected environmental variables (RFE). Predictions were generated for each pixel in the raster dataset, with the model outputting the predicted SOM content at specified quantiles. We specifically targeted the percentile that most closely approached the 5% SOM threshold. The results were stored spatially, enabling us to visualize and assess areas likely to meet or exceed this SOM content threshold. To represent the probability of having values equal to or above the target threshold, we used the following equation:

$$P = 1 - \text{argpmin} | Qp(Y | X) - yt | \quad (1)$$

where  $Y$  denotes the SOM content,  $X$  represents the set of the selected variables (post RFE for SOM),  $Qp(Y|X)$  is the  $p$ th percentile of the conditional distribution of  $Y$  given  $X$ ,  $yt$  is the target SOM content (5%), and  $P$  is the probability of having pixel values above or equal to the established SOM threshold.

We then set an 80% probability threshold to guide fertilizer application decisions. This level offers a pragmatic balance between certainty and practicality in decision making. Setting the threshold at 80% establishes a high level of confidence that the soil genuinely meets the required SOM content before applying fertilizers. Additionally, an 80% threshold minimizes the risk of under-fertilizing, which can occur if the threshold is set too high (e.g., 95%), potentially leading to suboptimal forest yields and economic losses.

#### 4.5.2. Three-Dimensional Soil Property Predictions

To obtain the prediction values at different depths (3D representation) for the SOM, clay, silt, and sand content, we used depth information from soil pit chemical analyses at different horizon thicknesses (in cm) as the dependent variables in the QRF model, and the depth along with the value of the selected environmental covariates as independent variables, as used in [36]. For this case, the average depth value assigned at the center of the observed depth interval per soil-pit thickness was used. This model can be generalized as follows:

$$Sa_{(x,y)} = f(\text{depth}_{(x,y)}, \text{topography}_{(x,y)}, \text{forest land cover}_{(x,y)}, \text{climate}_{(x,y)}, \text{parent material}_{(x,y)}, \text{soil}_{(x,y)}) \quad (2)$$

where  $S_a$  is the soil property of interest, depth is the depth at which this property was measured, and topography, forest land cover, climate, parent material, and soil are the selected environmental variables (code available on [60]).

These models were then used to predict the soil property over the spatial distribution of the environmental variables and values of depth between 0 and 180, every 10 cm. The prediction maps for the clay, silt, and sand content were used to classify soil texture according to the USDA textural soil classification [61]. Here, the soil texture class is derived based on the different proportions of sand, silt, and clay. For instance, an equal proportion of sand, silt, and clay results in a loam soil class, while a higher proportion of clay (30–45%) compared to sand (20–45%) and silt (20–45%) will result in a clay loam classification.

#### 4.6. Software Implementation

R (version 4.3.3) was used for all simulations and validations for producing climatic variables, regression models, and classification models. The 10-fold cross-validation for downscaling the temperature (max and min) and precipitation data using KED was performed using *autoKriging* (Version: 1.1-9) [62] from the ‘*automap*’ package (Version 1.1-9) [63]. The *biovars* function in the ‘*dismo*’ (Version 1.3-14) [64] package was used to obtain 19 bioclimatic variables based on the monthly minimum and maximum temperature and precipitation (mm). The QRF was implemented via the *quantregForest* R-package [65]. The RF classification model was implemented using the *caret* package (Version 6.0-94) [66], as well as tuning the hyperparameters, including the number of randomly selected predictors, the minimum node size, and the splitting rule. Soil texture classification was implemented using the *soiltexture* package (Version: 1.5.3) and the USDA.TT classification [67].

## 5. Results

### 5.1. Climate Variables

The cross-validation results demonstrated that KED downscaling provided accurate estimates of the climatic variables. For all months, the RMSE difference between the original and downscaled resolution was low, ranging from 0.04 to 0.62 mm for precipitation, from 0.04 to 0.09 °C for the minimum temperature, and from 0.08 to 0.12 °C for the maximum temperature. Due to an increase in rainfall throughout the winter (May to August), there was a slight increase in error for precipitation during this time period, coming to an average of 0.44 mm. For all months and variables, an  $r$  value of 0.99 resulted in the cross-validation process.

### 5.2. Variable Selection

The RFE reduced the 44 available input variables to the 10 most important for every soil property. Predominantly, these variables were generally of a topographic nature (60%), followed by climatic variables (37%), and forest land-cover variables (3%). For the prediction of all properties (SoD, SOM%, clay%, silt%, sand%), the DEM and total annual precipitation (bioclimatic variable) were selected as the top variables. The channel network base level (CNBL) was the second most important topographic variable, influencing every soil property. Furthermore, the mean annual temperature was one of the top ten input variables for the SoD and SOM, while temperature seasonality affected the SOM, clay, silt, and sand content. Table 4 summarizes the 10 input variables selected for each soil property prediction.

**Table 4.** Selected input variables for the SoD, SOM, clay content, silt content, and sand content predictions. The first section of this table contains climatic variables, followed by topographic variables, and, finally, remote sensing proxies of the forest and land cover. A more detailed description of each variable is given in Table A1.

Input Variable	SoD	SOM%	Clay%	Silt%	Sand%
Total precipitation (mm)	✓	✓	✓	✓	✓
Precipitation seasonality (%)	✓				

**Table 4.** *Cont.*

Input Variable	SoD	SOM%	Clay%	Silt%	Sand%
Mean temperature (°C)	✓	✓			
Mean diurnal range (°C)		✓			
Temperature seasonality (%)		✓	✓	✓	✓
Precipitation of the wettest month (mm)			✓	✓	✓
Temperature annual range (°C)			✓	✓	✓
Channel network distance	✓		✓	✓	
DEM	✓	✓	✓	✓	✓
Channel network base level (m.a.s.l.)	✓	✓	✓		✓
Valley depth	✓		✓	✓	✓
Aspect	✓				
Slope length and steepness factor	✓			✓	
Terrain surface convexity	✓				
Diffuse insolation		✓	✓		
Slope		✓			
Multiresolution Index of Valley-Bottom		✓		✓	✓
Flatness					
Multiresolution Index of Ridge-Top				✓	
Flatness					
Total insolation			✓		
EVI					✓
NDVI					✓

### 5.3. Validation

#### 5.3.1. Soil Depth Classification

The SoD classification model achieved an overall accuracy of 0.84, as evaluated using a confusion matrix, with a 95% confidence interval ranging from 78.99% to 87.46%. The matrix revealed 164 true positive cases of soil depths correctly predicted to be greater than 180 cm, and 100 true negative cases where soil depths were accurately identified as less than or equal to 180 cm. There were 36 commission errors and 16 omission errors, indicating instances of overestimation and underestimation, respectively (Table 5). Additionally, the Kappa statistic, which measures the agreement between predictions and references corrected for chance agreement, was 0.66.

**Table 5.** Confusion matrix for SoD classification.

Reference Prediction	Depths over 180 cm	Depths under 180 cm
Depth over 180 cm	164	36
Depth under 180 cm	16	100

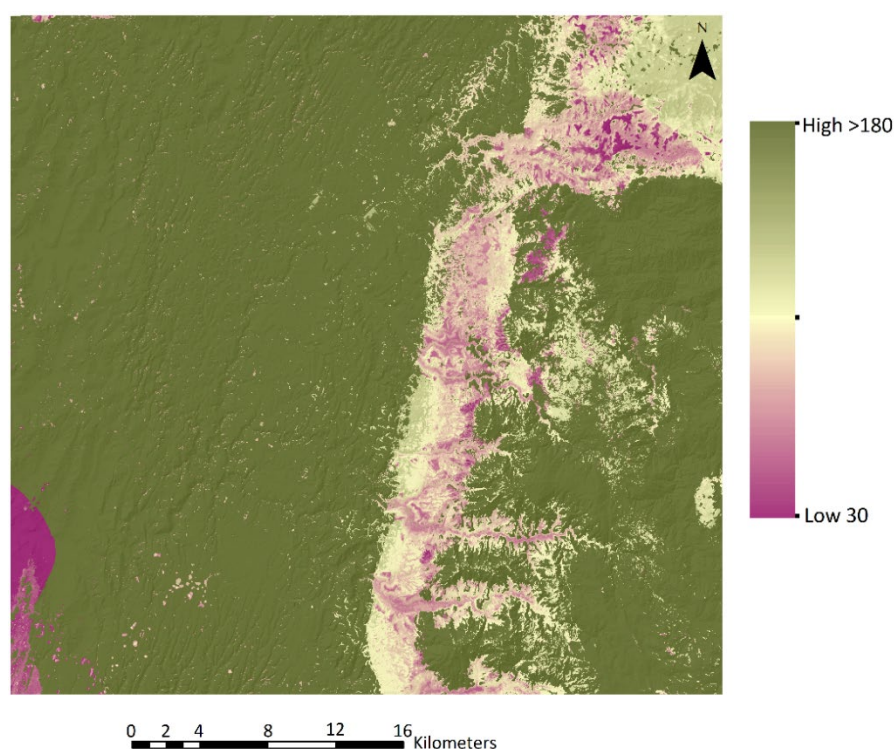
#### 5.3.2. Soil Attributes

The results of the soil attribute modeling using the QRF are presented in Table 6, which includes both the training and testing sets for validating the accuracy of the predictions. The most accurate predictions using the testing dataset were for SoD (below 180 cm), with an  $R^2$  of 0.74 and an RMSE of 19.4. In contrast, the sand predictions were the least accurate, exhibiting the lowest  $R^2$  at 0.55 and a MAPE of 42. Other variables such as SOM, clay, and silt were modeled with  $R^2$  values ranging from 0.59 to 0.61. However, the MAPE for SOM was 41.18, while for clay and silt, it fluctuated between 20.73 and 20.91. Predicted versus observed plots for SOM, clay, sand, silt, and SoD (below 180 cm) are illustrated in Appendix A.

The spatial representation of the SoD data is illustrated in Figure 3. This map combines the regression output applied to areas classified as less than 180 cm deep in the classification prediction map. Deep soils are located on the foothills of the Andes coastal range, whereas shallow soils (less than 180 cm deep) are found in areas with a slope steepness greater than 35° and channel network areas.

**Table 6.** Accuracy of the modeled soil properties.

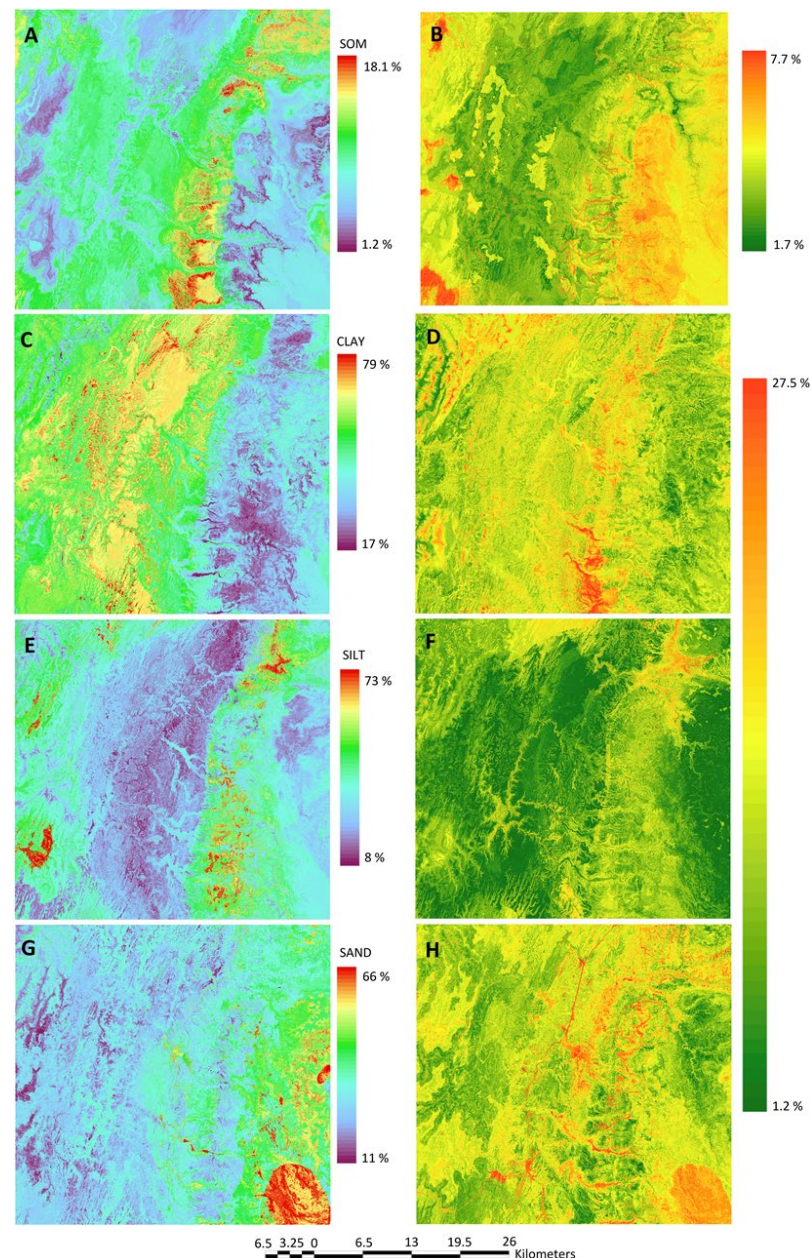
Variable	R <sup>2</sup>	RMSE	MAPE
Using the testing dataset:			
SoD	0.74	19.4	10.53
SOM%	0.61	2.03	41.18
Clay%	0.63	10.5	20.91
Silt%	0.59	6.26	20.73
Sand%	0.55	9.49	42
Using the training dataset:			
SoD	0.86	10.4	7.6
SOM%	0.84	1.29	21.8
Clay%	0.78	8.75	16.9
Silt%	0.68	6.08	18.4
Sand%	0.73	8.21	40

**Figure 3.** SoD predictions in cm, obtained from an ensemble of soil depth maps using a classification model (above and below 180 cm) and a regression model.

#### 5.4. Mapping the Uncertainty and Soil Property Predictions

Figure 4 depicts the spatial representations of the SOM, clay, silt, and sand predictions. In the upper soil layer, there is a noticeably high content of SOM across the study area, with values peaking at 18.1%. Some pockets of land in the coastal range exhibit increased uncertainty (Figure 4B). A significant portion of the study area has a high clay content, reaching up to 79%, especially near the Andes foothills (in areas with ultisols) and within some parts of the valley (alfisols). However, there is a higher uncertainty associated with this soil property in the coast and along channel network pathways (Figure 4D). Silt is primarily concentrated in the northern region of the Andes foothills, while sand is more prevalent at higher altitudes. Across all soil properties, with the exception of silt, there is greater uncertainty along the channel network located within the valley.



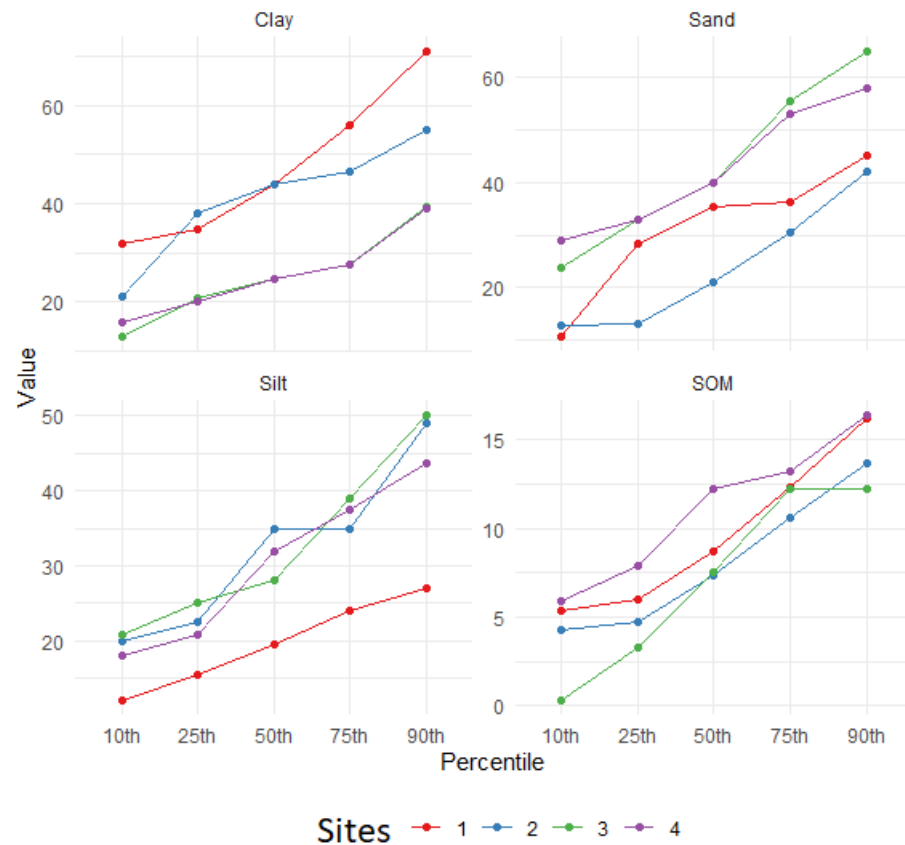


**Figure 4.** The 50th percentile predictions for the (A) soil organic matter content, (C) clay%, (E) silt%, and (G) sand% at the surface level (0 cm). Additionally, this figure displays the standard deviation of these predictions for the (B) soil organic matter content, (D) clay%, (F) silt%, and (H) sand%.

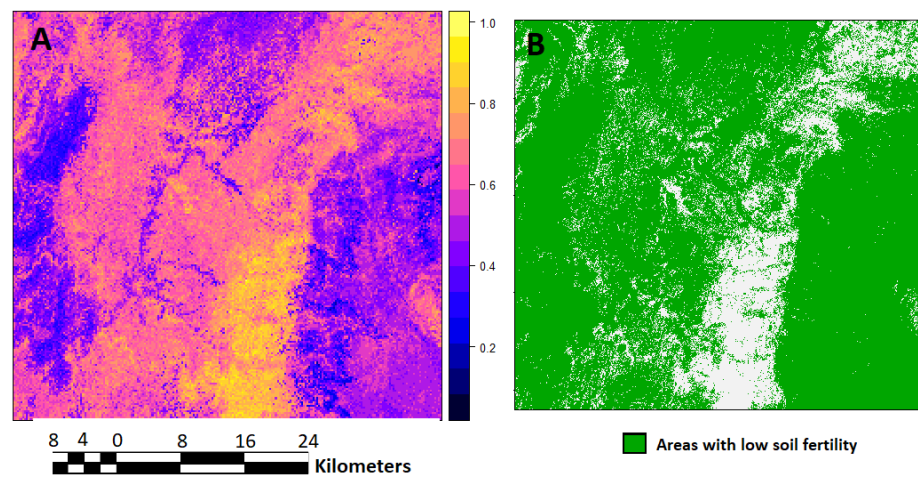
Figure 1 displays four sites selected based on their variability within the study area. The distribution of these four sites is detailed in Figure 5, presenting the 10th, 25th, 50th, 75th, and 90th percentiles. From this, we can observe the variability in soil properties. Generally, sites 2 and 3 exhibit lower SOM levels compared to other areas. Meanwhile, sites 3 and 4 demonstrate more balanced concentrations of sand, silt, and clay.

In our practical example, we created a probabilistic map indicating areas with 5% SOM content, as illustrated in Figure 6A. This analysis facilitated the development of a probabilistic map based on the entire range of distribution for the SOM, which was used to identify areas of high or low fertility for targeted fertilizer application. Figure 6B highlights the designated zones for fertilization: regions with at least an 80% probability of surpassing 5% SOM content are colored green, while those with a 5% or lower probability are colored purple and marked for fertilizer application.





**Figure 5.** Percentile distribution of SOM, clay, silt, and sand contents for the four sites depicted in Figure 1.

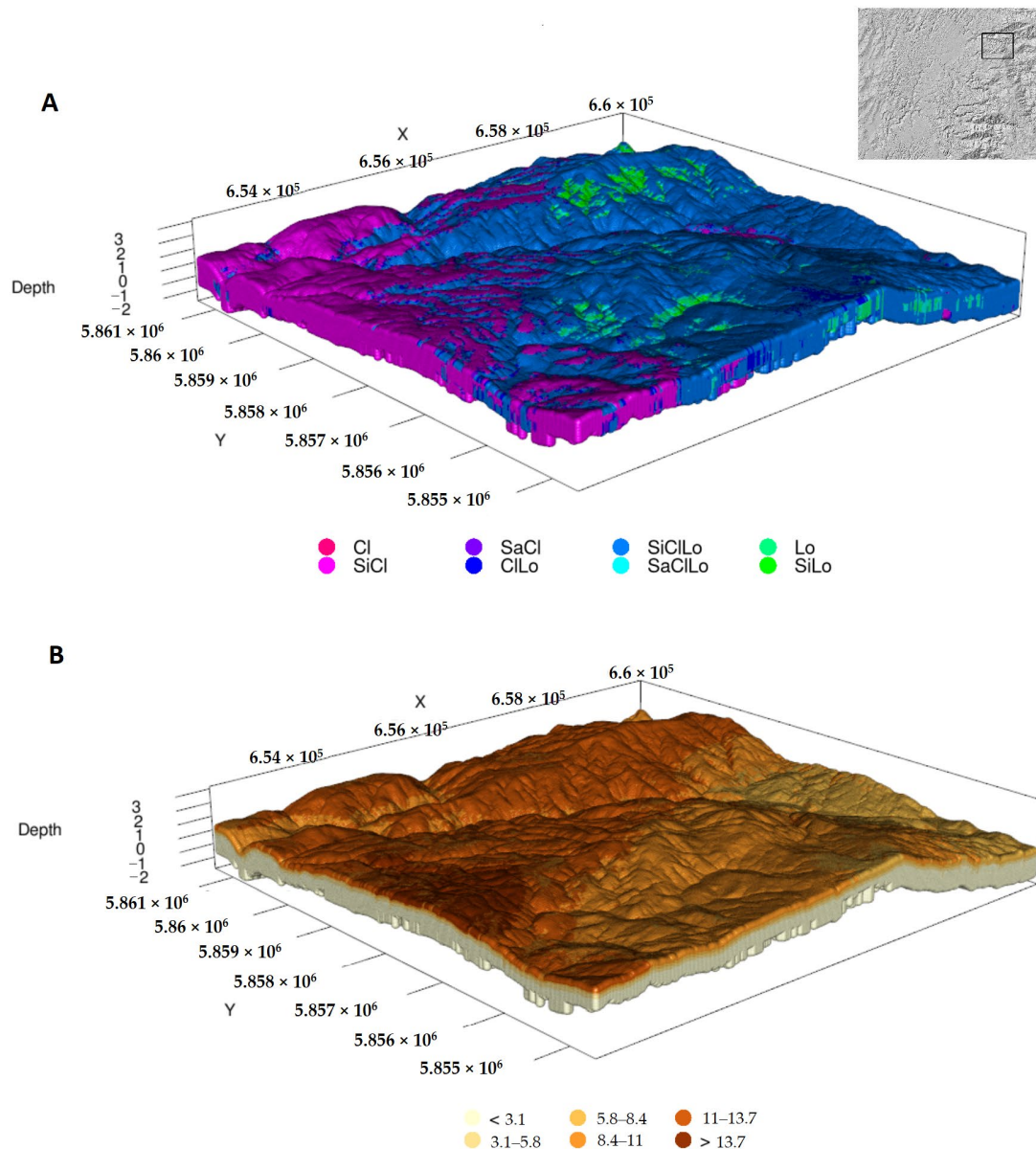


**Figure 6.** Probability maps indicating 5% SOM content. In (A), the probability distribution ranges from 1 to 100% (represented in the figure as 0.01 to 1). In (B), the highlighted areas are those designated for fertilization based on the 5% threshold (low soil fertility).

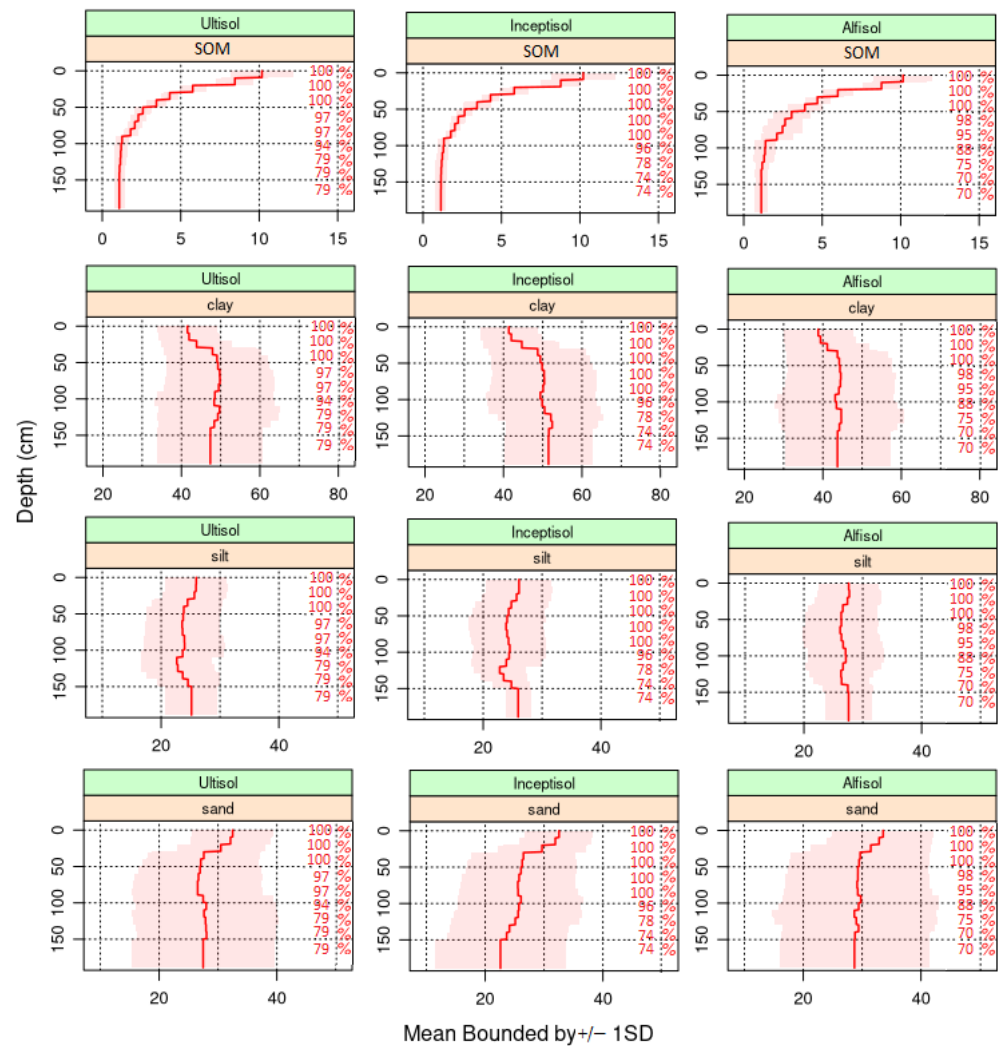
### 5.5. Three-Dimensional Soil Map

The QRF provided a 3D representation of the SOM and texture class (as a combination of clay, silt, and sand) predictions across different soil depth layers. Figure 7 displays this 3D visualization of the soil properties at a sample site selected for its topographic relief variations. Figure 8 illustrates the entire depth variation of the SOM, clay, silt, and sand contents in the study area by soil order, which aids in better illustrating depth changes. In Figure 7, the predicted SoD is used as a base layer to represent the soil thickness. The spatial information for the SOM demonstrates that as depth increases, the amount of

organic content decreases, with the highest concentrations occurring only in the first 40 cm of soil depth. Figure 7B shows the soil texture classification based on different proportions of clay, silt, and sand. Within the study area, the soil texture class is predominantly clay (61.4%) and clay loam (28%). The observed soil classifications show only small variations at different depths. In Figure 8, the median values of clay, silt, and sand show minimal variation, consistent with the observations for the soil texture classes. Most notably, there is an increase in sand content beyond 70 cm in the inceptisol layer.



**Figure 7.** Three-dimensional digital soil maps at various depths for (A) the soil texture (CI: clay; SaCl: sandy clay; SiClLo: silty clay loam; Lo: loam; SiCl: silty clay; ClLo: clay loam; SaClLo: sandy clay loam, and SiLo: silty loam) and (B) the SOM%. The upper-right image shows the delineation of this zoomed-in area within the study area.



**Figure 8.** Soil organic matter (SOM), clay, silt, and sand contents at different depths per soil layer across the entire study area, with continuous vertical lines representing the mean value every 10 cm, the shaded red area representing the standard variation, and the percentage on the right representing the number of pixels falling into the mentioned depth class.

## 6. Discussion

This study presents a method for obtaining spatial predictions and their associated uncertainties for digital soil mapping. Additionally, it demonstrates how to integrate the distributions of these predictions of soil properties into forest management operations.

Our results show that the soil properties were well represented spatially when using the QRF. The chosen 10 m spatial resolution, selected to observe soil property changes within the forest stand, provided accurate results when compared to other studies on SoD prediction using a regression model ( $R^2 = 0.74$ , MAPE = 10.53%, and RMSE = 19.4 cm). Although SoD is a difficult soil property to predict, we obtained a significant improvement by considering all of the SCORPAN input variables when compared to other studies using only topographic variables [68], indicating an overall good performance in line with the results presented in [69].

In the validation analysis, the SOM map showed reliable results, with  $R^2 = 0.61$ , MAPE = 41.18%, and RMSE = 2.03%. When compared to previous studies, we can observe that the general level of validation accuracy for SOM predictions ranges between  $R^2 = 0.45$  at 10 m using topographic attributes only [70],  $R^2 = 0.51$  at 30 m [71], and  $R^2 = 0.53$  using a QRF at 30 m resolution [34]. In our study, the observed standard deviation of the prediction distribution ranged from 1.7% to 7.7%. In most of the valley and the Andes foothills,

the variability hovered at around 17% of the mean. However, in the Andes, it rose to approximately 50%. This trend can potentially be attributed to the fact that as topographic heterogeneity increases, soil properties vary significantly [72].

Soil texture also showed accurate results given the successful validation of the clay, silt, and sand contents. Clay had the higher accuracy, followed by silt and lastly by sand (Table 5), with mean standard deviations of 29%, 33%, and 35%, respectively. Similar results were reported by other studies [20,73,74], which account for their good prediction accuracy based on the strength of the used input variables of the SCORPAN approach, emphasizing the importance of integrating key soil formation variables for reliable predictions of soil texture.

Observing the spatial distribution of standard deviation values and the spread of prediction distributions for the four selected sites depicted in Figure 1 and explained in Figure 5, it becomes apparent that there is significant variability in the prediction distributions across all predicted soil properties. This variability is particularly pronounced, as anticipated, on the foothills of the Andes coastal range, where the topography is highly heterogeneous and maximum variation occurs [75]. For example, the SOM ranges from 3.2% (25th percentile) to 12.2% (75th percentile). Surprisingly, even in the valley, where the topographic relief is relatively consistent, a notably variable prediction distribution of soil properties persists. Following the SOM example, it varies from 7.9% to 13.2% in the 25th and 75th percentiles, respectively. Capturing the spatial predictive distribution of soil properties is crucial to reducing uncertainty and enhancing the quality of digital soil mapping (DSM) products. As Ref. [25] noted, one direct application of soil prediction uncertainty is in optimizing soil sampling designs, with a focus on areas of high uncertainty. Furthermore, the use of prediction variability and the entire distribution of soil properties is also beneficial to managers in reducing uncertainty in forest growth responses to fertilizer application, as illustrated in the practical management application example. This phenomenon, where responses to fertilizer application are heightened in areas of low soil fertility, has been substantiated by previous research [16,17]. Relying solely on the mean for decision-supporting purposes, as exemplified in Figure 6, and using the same 5% SOM as a threshold to identify areas for fertilization, would imply that in 50% of cases, values will exceed the decided threshold. This translates to 50% of areas exhibiting a low response or no response to fertilizer application. As demonstrated in the example, utilizing the entire distribution range can ensure positive responses of forest growth to management applications, depending on the level of probability chosen for decisions (80% in our example), thereby minimizing resource allocation and maximizing the economic benefits.

As discussed in previous research [12], detecting the presence of shallow soils is crucial for soil protection management in forested areas. It is vital to understand the forest roots' capacity to reach available nutrients and to enhance our knowledge of the plants' available water content and the variation in nutrient availability at different soil depths for tree root uptake. As shown in Figure 7, the 3D map demonstrates a significant variation in SOM content at different depths. The decrease in SOM with increasing soil depths has been mentioned several times in the literature [76,77]. This occurs because SOM is produced near the surface due to superficial leaf decomposition in forested areas and fine-root decomposition in grasslands. This shift in distribution is also linked to varying microbial community structures at different soil depths, which influence the rate of organic matter decomposition [78]—a reduction rate that can be visualized with the methodology presented in this study. Although there is some variation in soil texture at different depths, it is not significant. Similar results found in [79] were explained by the soil's age since most of the horizons were still forming, as may be the case in our study area. Furthermore, particularly around the coast, the soil is classified as clay through several depth levels, likely because they are on alfisols—classified by the USDA soil taxonomy as clay-enriched with a relatively high natural fertility [61].

Regarding the SCORPAN variables selected for this research, DEM has emerged as one of the top variables related to soil texture and depth. This correlation can primarily be attributed to its association with soil erosion and redistribution, as well as its impact on the



SOM accumulation cycle [75]. Previous research has found that elevation, the topographic wetness index (TWI), plan curvature, the total catchment area, and the channel network base level are the key topography characteristics most closely connected to soil organic carbon concentration in flat-slope locations [80]. Furthermore, for hydrologic and geomorphic purposes, the Multiresolution Index of Ridge-Top Flatness (MRRTF), Multiresolution Index of Valley-Bottom Flatness (MRVBF), slope, and valley depth have shown strong links with sediment deposits and influence deposit depth [81]. As observed in [82], the distance to water channels provides essential information on sediment accumulation, altering the SOM content, SoD, and soil texture, which are also reflected in this study. Besides elevation, this variable has also been recognized as one of the most critical topographic parameters for the digital soil mapping of SOM [80]. According to the literature, one of the most important environmental variables influencing the rate of weathering and organic decomposition, which cause new layers of soil to accumulate at depth, is the mean annual temperature [83,84]. Furthermore, precipitation notably impacts hydrological processes such as surface runoff and groundwater flow, which are vital for organic matter decomposition rates and infiltration from the litter layer to mineral soil, as extensively reported [45,46]. This notion is corroborated in this study, wherein these two climatic variables are selected as some of the most significant predictors of among all of the predicted soil properties.

The most significant advantage of the proposed methodology is that it enables the prediction of accurate high-spatial-resolution maps of soil properties and their predictive distributions, based on soil formation factors (SCORPAN). The application of QRFs in DSM, as used in this study, demonstrates the utility of this methodology for the spatial interpolation of key soil properties and its uncertainty within the forest stand, which has been previously applied in SoilGrids250m [23]. As mentioned in previous research [28,35], the QRF model is a promising alternative for observing the probabilistic distribution of soil predictions, especially for precision silviculture tasks. Nonetheless, the primary advantage of using a QRF for high-resolution soil properties lies in its ability to capture variable distributions. This aspect is frequently overlooked in conventional soil prediction models and management applications.

Similar to previous soil property mapping efforts [85,86], spatial soil information has proven to be a valuable tool for enhancing our understanding of site productivity variations within forested areas, especially when utilizing the spatial resolution chosen for this study (10 m). This approach to digital soil property analysis, along with its associated uncertainties, enhances fertilizer prescription capabilities by identifying areas with fertility issues, thereby facilitating precision in fertilizer application during forest operations, as demonstrated in the example provided. Moreover, granulometric information at a fine spatial resolution using the QRF method, alongside SoD data, can be employed to improve assessments of water-holding capacity and soil fertility ratings. These enhancements are crucial for process-based models used in predicting forest growth, as well as for precision planning in timber harvesting and site preparation to mitigate soil compaction risks. Integrating predictive soil property distribution into decision management strategies can further contribute to reductions in fertilization costs and increases in productivity. This methodology holds promise for transferability, as it is applicable to various soil properties, soil-related management practices, and diverse locations or countries. Future research should use high-resolution soil property information and its predictive distribution combined with ALS individual tree metrics to develop tree growth models in order to better understand the effects of site productivity on growth.

## 7. Conclusions

Quantile Regression Forests have proven to be both accurate and reliable in predicting soil properties and their distribution, especially in digital soil mapping applications. The methodology showcases spatial precision, making it invaluable for site-specific precision in silviculture management and precision fertilization treatments within forest stands. Incon-



porating prediction distribution is vital, as it plays a crucial role in minimizing uncertainty in decision-supporting applications within forest management. This is particularly true for operations such as fertilizer application, road construction planning, and forest growth modeling, especially in the context of site-specific management.

**Author Contributions:** Conceptualization, G.G.-A. and N.C.C.; data curation, G.G.-A.; formal analysis, G.G.-A. and G.F.O.; methodology, G.G.-A. and G.F.O.; validation, N.C.C. and P.T.; visualization, N.C.C., P.T., D.R. and A.V.; supervision, N.C.C.; writing—original draft preparation, G.G.-A.; writing—review and editing, N.C.C., P.T., D.R. and A.V. All authors have read and agreed to the published version of the manuscript.

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

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets generated and (/or) analyzed during the current study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors would like to thank Forestal Arauco for their role in specially providing the ALS and study-site data to Rodrigo Ahumada and Fernando Bustamante. Special thanks to Michael Burnett for reviewing the content of this document.

**Conflicts of Interest:** The authors declare that there are no competing interests or personal relationships that could have appeared to influence the work reported in this paper. No commercial interest apply on this manuscript.

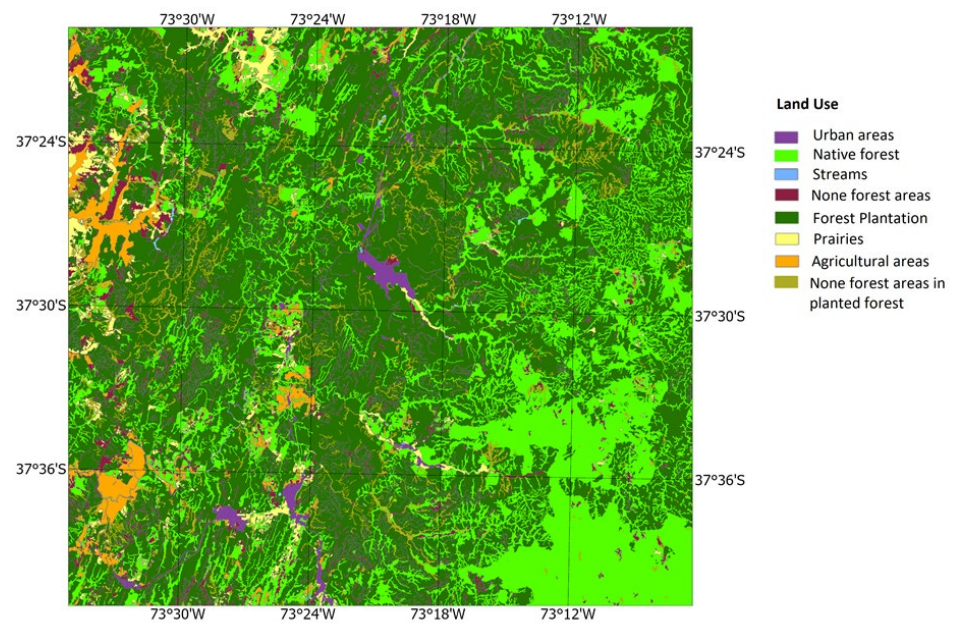
## Appendix A

**Table A1.** List of environmental input raster variables.

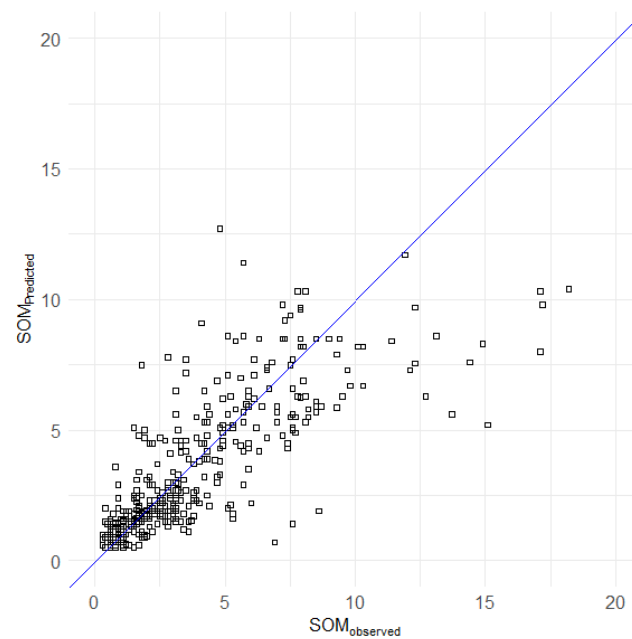
Code	Category	Description
biovar.1	Climate	bio1 = Mean annual temperature (°C)
biovar.2	Climate	bio2 = Mean diurnal range (mean of max temp - min temp) (°C)
biovar.3	Climate	bio3 = Isothermality (bio2/bio7) (× 100) (%)
biovar.4	Climate	bio4 = Temperature seasonality (standard deviation × 100) (%)
biovar.5	Climate	bio5 = Max temperature of warmest month (°C)
biovar.6	Climate	bio6 = Min temperature of coldest month (°C)
biovar.7	Climate	bio7 = Temperature annual range (bio5-bio6) (°C)
biovar.8	Climate	bio8 = Mean temperature of the wettest quarter (°C)
biovar.9	Climate	bio9 = Mean temperature of driest quarter (°C)
biovar.10	Climate	bio10 = Mean temperature of warmest quarter (°C)
biovar.11	Climate	bio11 = Mean temperature of coldest quarter (°C)
biovar.12	Climate	bio12 = Total (annual) precipitation (mm)
biovar.13	Climate	bio13 = Precipitation of wettest month (mm)
biovar.14	Climate	bio14 = Precipitation of driest month (mm)
biovar.15	Climate	bio15 = Precipitation seasonality (coefficient of variation) (%)
biovar.16	Climate	bio16 = Precipitation of wettest quarter (mm)
biovar.17	Climate	bio17 = Precipitation of driest quarter (mm)
biovar.18	Climate	bio18 = Precipitation of warmest quarter (mm)
biovar.19	Climate	bio19 = Precipitation of coldest quarter (mm)
DEM	Topography	Digital elevation model
Aspect	Topography	Aspect Degree (%)
CNBL	Topography	Channel Network Base Level (m.a.s.l)
CND	Topography	Channel Network Distance
CI	Topography	Convergence Index
DiffI	Topography	Diffuse Insolation
DirI	Topography	Direct Insolation
LS_factor	Topography	Slope Length and Steepness Factor
MRRTF	Topography	Multiresolution Ridge Top Flatness
MRBVF	Topography	Multiresolution Index of Valley Bottom Flatness
PC	Topography	Plan curvature
ProfC	Topography	Profile curvature
Slope	Topography	Slope Degree (%)
TanC	Topography	Tangential curvature

**Table A1.** *Cont.*

Code	Category	Description
TSC	Topography	Terrain Surface Convexity
TWI	Topography	Topographic Wetness Index
TC	Topography	Total catchment area
TIns	Topography	Total Insolation
ValDepth	Topography	Valley Depth
NDVI	Vegetation	Normalized Difference Vegetation index
EVI	Vegetation	Enhanced Vegetation Index
PRM	Soil Morphology	Parent rock material
SC	Soil Morphology	Soil Class



**Figure A1.** Land-use map within the study area. Source available online [87].



**Figure A2.** *Cont.*

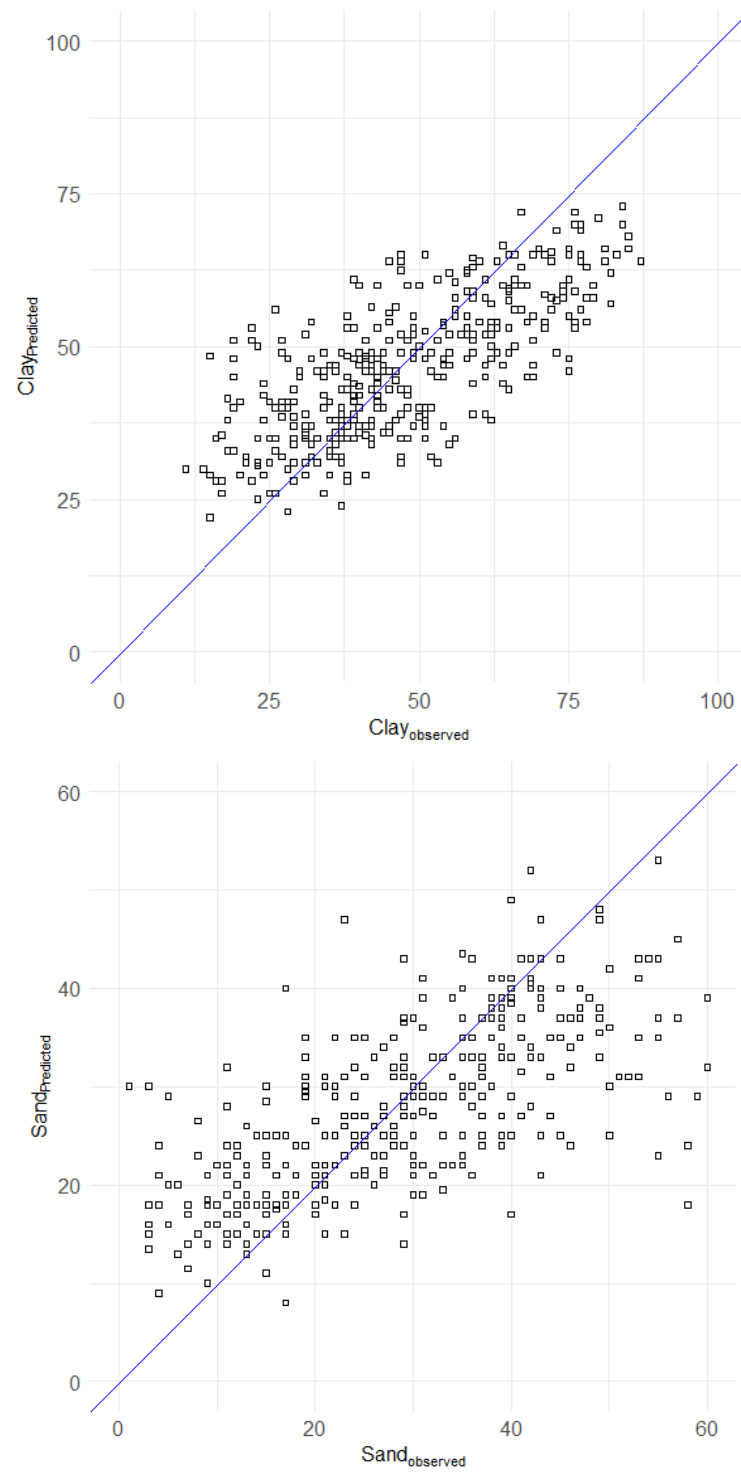
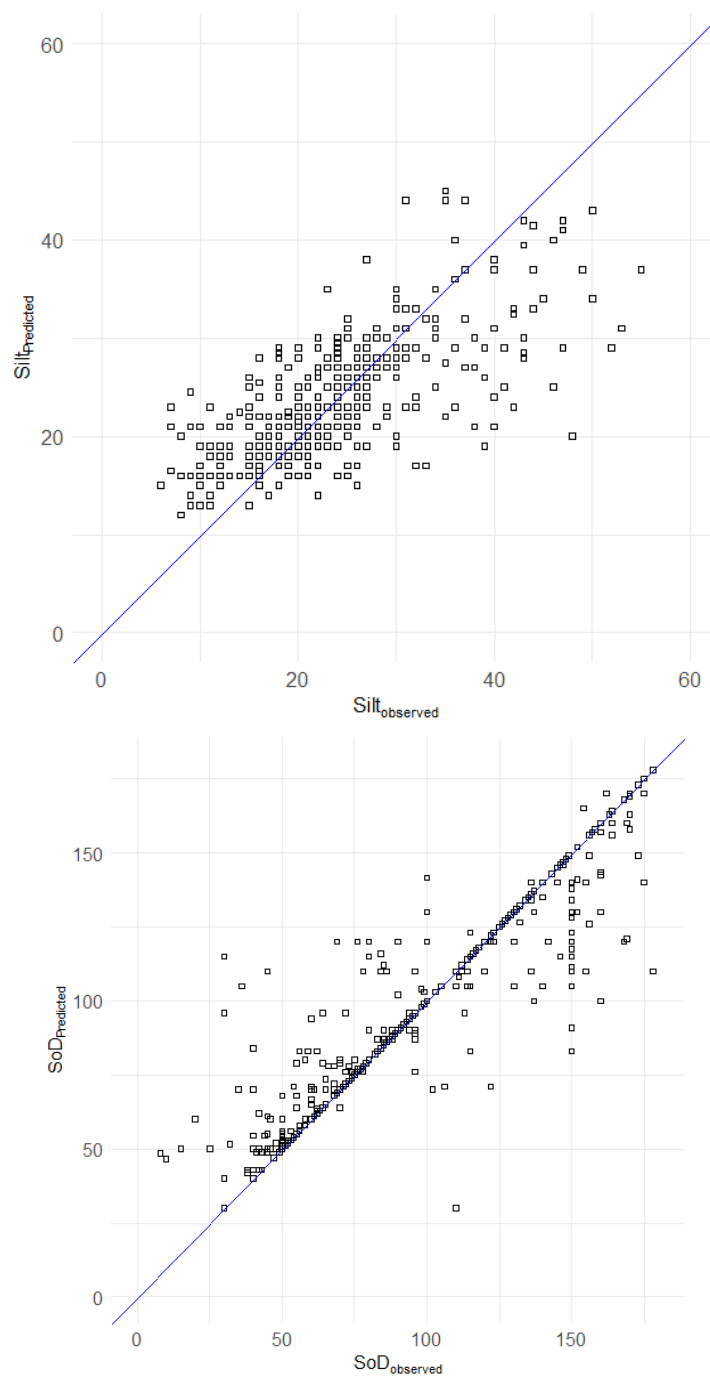


Figure A2. Cont.



**Figure A2.** Predicted versus observed values using the 50th percentile on the testing dataset.

## References

1. Eckhart, T.; Pötzelsberger, E.; Koeck, R.; Thom, D.; Lair, G.J.; van Loo, M.; Hasenauer, H. Forest Stand Productivity Derived from Site Conditions: An Assessment of Old Douglas-Fir Stands (*Pseudotsuga menziesii* (Mirb.) Franco Var. *menziesii*) in Central Europe. *Ann. For. Sci.* **2019**, *76*, 19. [[CrossRef](#)] [[PubMed](#)]
2. Worrell, R.; Malcolm, D.C. Productivity of Sitka Spruce in Northern Britain. *Forestry* **1990**, *63*, 105–118. [[CrossRef](#)]
3. Horst, T.Z.; Dalmolin, R.S.D.; Caten, A.T.; Moura-Bueno, J.M.; Cancian, L.C.; Pedron, F.D.A.; Schenato, R.B. Edaphic and Topographic Factors and Their Relationship with Dendrometric Variation of *Pinus taeda* L. in a High Altitude Subtropical Climate. *Rev. Bras. Ciênc. Solo* **2018**, *42*, e0180023. [[CrossRef](#)]
4. Skovsgaard, J.P.; Vanclay, J.K. Forest Site Productivity: A Review of the Evolution of Dendrometric Concepts for Even-Aged Stands. *Forestry* **2008**, *81*, 13–31. [[CrossRef](#)]
5. Phogat, V.; Tomar, V.; Dahiya, R. Soil Physical Properties. *Adv. Soil Dyn.* **2013**, *1*, 21–254. [[CrossRef](#)]

6. Gier, J.M.; Kindel, K.M.; Page-Dumroese, D.S.; Kuennen, L.J. *Soil Disturbance Recovery on the Kootenai National Forest, Montana*; General Technical Report; U.S. Department of Agriculture, Forest Service: Fort Collins, CO, USA, 2018; Volume 2018, pp. 1–31.
7. Wagner, L.E.; Ambe, N.M.; Ding, D. Estimating a Proctor Density Curve from Intrinsic Soil Properties. *Trans. Am. Soc. Agric. Eng.* **1994**, *37*, 1121–1125. [[CrossRef](#)]
8. Horn, A.L.; Düring, R.A.; Gäth, S. Comparison of the Prediction Efficiency of Two Pedotransfer Functions for Soil Cation-Exchange Capacity. *J. Plant Nutr. Soil Sci.* **2005**, *168*, 372–374. [[CrossRef](#)]
9. Russ, A.; Riek, W.; Wessolek, G. Three-Dimensional Mapping of Forest Soil Carbon Stocks Using Scorpan Modelling and Relative Depth Gradients in the North-Eastern Lowlands of Germany. *Appl. Sci.* **2021**, *11*, 714. [[CrossRef](#)]
10. Grigal, D.F.; Vance, E.D. Influence of Soil Organic Matter on Forest Productivity. *N. Z. J. For. Sci.* **2000**, *30*, 169–205.
11. Horst-Heinen, T.Z.; Dalmolin, R.S.D.; Caten, A.T.; Moura-Bueno, J.M.; Grunwald, S.; Pedron, F.d.A.; Rodrigues, M.F.; Rosin, N.A.; da Silva-Sangoi, D.V. Soil Depth Prediction by Digital Soil Mapping and Its Impact in Pine Forestry Productivity in South Brazil. *For. Ecol. Manag.* **2021**, *488*, 118983. [[CrossRef](#)]
12. Dharumarajan, S.; Vasundhara, R.; Suputhra, A.; Lalitha, M.; Hegde, R. Prediction of Soil Depth in Karnataka Using Digital Soil Mapping Approach. *J. Indian Soc. Remote Sens.* **2020**, *48*, 1593–1600. [[CrossRef](#)]
13. Gavilán-Acuña, G.; Olmedo, G.F.; Mena-Quijada, P.; Guevara, M.; Barría-Knopf, B.; Watt, M.S. Reducing the Uncertainty of *Pinus radiata* Pine Site Index Maps Using an Spatial Ensemble of Machine Learning Models. *Forests* **2021**, *12*, 77. [[CrossRef](#)]
14. Romanyà, J.; Vallejo, V.R. Productivity of *Pinus radiata* Plantations in Spain in Response to Climate and Soil. *For. Ecol. Manag.* **2004**, *195*, 177–189. [[CrossRef](#)]
15. Fralish, J.S. The Effect of Site Environment on Forest Productivity in the Illinois Shawnee Hills. *Ecol. Appl.* **1994**, *4*, 134–143. [[CrossRef](#)]
16. Ramírez Alzate, M.V.; Rubilar, R.A.; Montes, C.; Allen, H.L.; Fox, T.R.; Sanfuentes, E. Mid-Rotation Response to Fertilizer by *Pinus radiata* D. Don at Three Contrasting Sites. *J. For. Sci.* **2016**, *62*, 153–162. [[CrossRef](#)]
17. McFarlane, K.J.; Schoenholtz, S.H.; Powers, R.F. Plantation Management Intensity Affects Belowground Carbon and Nitrogen Storage in Northern California. *Soil Sci. Soc. Am. J.* **2009**, *73*, 1020–1032. [[CrossRef](#)]
18. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
19. Heuvelink, G.B.M.; Webster, R. Modelling Soil Variation: Past, Present, and Future. *Geoderma* **2001**, *100*, 269–301. [[CrossRef](#)]
20. Laborczy, A.; Szatmári, G.; Takács, K.; Pásztor, L. Mapping of Topsoil Texture in Hungary Using Classification Trees. *J. Maps* **2016**, *12*, 999–1009. [[CrossRef](#)]
21. Taylor, S.E.; McDonald, T.P.; Veal, M.W.; Corley, F.W.; Grift, T.E. Precision Forestry: Operational Tactics for Today and Tomorrow. In Proceedings of the 25th Annual Meeting of the Council of FOREST Engineers, Auburn, AL, USA, 16–20 June 2002; p. 6.
22. Taylor, S.E.; McDonald, T.P.; Fulton, J.P.; Shaw, J.N.; Corley, F.W.; Brodbeck, C.J. Precision Forestry in the Southeast US. In Proceedings of the 1st International Precision Forestry Symposium, Stellenbosch, South Africa, 5–10 March 2006; pp. 397–414.
23. Kasraei, B.; Heung, B.; Saurette, D.D.; Schmidt, M.G.; Bulmer, C.E.; Bethel, W. Quantile Regression as a Generic Approach for Estimating Uncertainty of Digital Soil Maps Produced from Machine-Learning. *Environ. Model. Softw.* **2021**, *144*, 105139. [[CrossRef](#)]
24. Tavazza, F.; Decost, B.; Choudhary, K. Uncertainty Prediction for Machine Learning Models of Material Properties. *ACS Omega* **2021**, *6*, 32431–32440. [[CrossRef](#)]
25. Stumpf, F.; Schmidt, K.; Goebes, P.; Behrens, T.; Schönbrodt-Stitt, S.; Wadoux, A.; Xiang, W.; Scholten, T. Uncertainty-Guided Sampling to Improve Digital Soil Maps. *Catena* **2017**, *153*, 30–38. [[CrossRef](#)]
26. Verrelst, J.; Rivera, J.P.; Moreno, J.; Camps-Valls, G. Gaussian Processes Uncertainty Estimates in Experimental Sentinel-2 LAI and Leaf Chlorophyll Content Retrieval. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 157–167. [[CrossRef](#)]
27. Adeniyi, O.D.; Brenning, A.; Bernini, A.; Brenna, S.; Maerker, M. Digital Mapping of Soil Properties Using Ensemble Machine Learning Approaches in an Agricultural Lowland Area of Lombardy, Italy. *Land* **2023**, *12*, 494. [[CrossRef](#)]
28. Vaysse, K.; Lagacherie, P. Using Quantile Regression Forest to Estimate Uncertainty of Digital Soil Mapping Products. *Geoderma* **2017**, *291*, 55–64. [[CrossRef](#)]
29. Zhang, G.-L.; Liu, F.; Song, X.-D.; Zhao, Y.-G. *Digital Soil Mapping Across Paradigms, Scales, and Boundaries: A Review*; Springer: Berlin/Heidelberg, Germany, 2016; ISBN 9789811004148.
30. Brus, D.J.; Bogaert, P.; Heuvelink, G.B.M. Bayesian Maximum Entropy Prediction of Soil Categories Using a Traditional Soil Map as Soft Information. *Eur. J. Soil Sci.* **2008**, *59*, 166–177. [[CrossRef](#)]
31. D’Or, D. *Spatial Prediction of Soil Properties, the Bayesian Maximum Entropy Approach*; University Catholique de Louvain: Ottignies-Louvain-la-Neuve, Belgium, 2003.
32. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
33. Yigini, Y.; Olmedo, G.F.; Reiter, S.; Baritz, R.; Viatkin, K.; Vargas, R. *Soil Organic Carbon Mapping: Cookbook*, 2nd ed.; FAO: Rome, Italy, 2018.
34. Nikou, M.; Tziachris, P. Prediction and Uncertainty Capabilities of Quantile Regression Forests in Estimating Spatial Distribution of Soil Organic Matter. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 130. [[CrossRef](#)]
35. Schmidinger, J.; Heuvelink, G.B.M. Validation of Uncertainty Predictions in Digital Soil Mapping. *Geoderma* **2023**, *437*, 116585. [[CrossRef](#)]



36. Ma, Y.; Minasny, B.; McBratney, A.; Poggio, L.; Fajardo, M. Predicting Soil Properties in 3D: Should Depth Be a Covariate? *Geoderma* **2021**, *383*, 114794. [[CrossRef](#)]
37. Veronesi, F.; Schillaci, C. Comparison between Geostatistical and Machine Learning Models as Predictors of Topsoil Organic Carbon with a Focus on Local Uncertainty Estimation. *Ecol. Indic.* **2019**, *101*, 1032–1044. [[CrossRef](#)]
38. Kumar, V.; Kant, S.; Shikha; Kumar, A. Morphological and Pedological Features of Alfisols. *Agric. Sci.* **2018**, *4*, 159–167.
39. Minagri Suelos Agrológicos. Available online: <https://www.ciren.cl/productos/suelos-agrologicos/> (accessed on 23 May 2022).
40. West, L.T.; Beinroth, F.H.; Sumner, M.E.; Kang, B.T. Ultisols: Characteristics and Impacts on Society. *Adv. Agron.* **1997**, *63*, 179–236. [[CrossRef](#)]
41. Staff, S.S. Soil Taxonomy. A Basic System of Soil Classification for Making and Interpreting Soil Surveys. In *Agriculture Handbook 436*; United States Department of Agriculture: Washington, DC, USA, 1999; p. 869.
42. Boisier, J.P. CR2MET: A High-Resolution Precipitation and Temperature Dataset for the Period 1960–2021 in Continental Chile [v2.5]; Zenodo: Geneva, Switzerland, 2023.
43. Isenburg, M. *LASTools: Efficient LiDAR Processing Software*; Rapidlasso GmbH: Gilching, Germany, 2023; Available online: <https://lastools.github.io/> (accessed on 28 April 2024).
44. Sudmeyer, R.A.; Speijers, J.; Nicholas, B.D. Root Distribution of *Pinus pinaster*, *P. radiata*, *Eucalyptus globulus* and *E. kochii* and Associated Soil Chemistry in Agricultural Land Adjacent to Tree Lines. *Tree Physiol.* **2004**, *24*, 1333–1346. [[CrossRef](#)] [[PubMed](#)]
45. Heisler, J.L.; Weltzin, J.F. Variability Matters: Towards a Perspective on the Influence of Precipitation on Terrestrial Ecosystems. *New Phytol.* **2006**, *172*, 189–192. [[CrossRef](#)]
46. O'Brien, S.L.; Jastrow, J.D.; Grimley, D.A.; Gonzalez-Meler, M.A. Moisture and Vegetation Controls on Decadal-Scale Accrual of Soil Organic Carbon and Total Nitrogen in Restored Grasslands. *Glob. Chang. Biol.* **2010**, *16*, 2573–2588. [[CrossRef](#)]
47. Gallant, J.C.; Wilson, J.P. Primary Topographic Attributes. In *Terrain Analysis: Principles and Applications*; Wilson, J.P., Gallant, J.C., Eds.; John Wiley & Sons: New York, NY, USA, 2000; pp. 51–85.
48. 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.* **2015**, *8*, 1991–2007. [[CrossRef](#)]
49. Zhang, Y.; Guo, L.; Chen, Y.; Shi, T.; Luo, M.; Ju, Q.L.; Zhang, H.; Wang, S. Prediction of Soil Organic Carbon Based on Landsat 8 Monthly NDVI Data for the Jiangnan Plain in Hubei Province, China. *Remote Sens.* **2019**, *11*, 1683. [[CrossRef](#)]
50. Holben, B.N. Characteristics of Maximum-Value Composite Images from Temporal AVHRR Data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434. [[CrossRef](#)]
51. Forkuor, G.; Hounkpatin, O.K.L.; 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](#)]
52. Hudson, G.; Wackernagel, H. Mapping Temperature Using Kriging with External Drift: Theory and an Example from Scotland. *Int. J. Climatol.* **1994**, *14*, 77–91. [[CrossRef](#)]
53. Laaha, G.; Sköien, J.O.; Nobilis, F.; Blöschl, G. Spatial Prediction of Stream Temperatures Using Top-Kriging with an External Drift. *Environ. Model. Assess.* **2013**, *18*, 671–683. [[CrossRef](#)]
54. Hengl, T.; Heuvelink, G.; Stein, A. Comparison of Kriging with External Drift and Regression-Kriging. Technical Note ITC 2003, 17. Available online: [https://webapps.itc.utwente.nl/librarywww/papers\\_2003/misca/hengl\\_comparison.pdf](https://webapps.itc.utwente.nl/librarywww/papers_2003/misca/hengl_comparison.pdf) (accessed on 28 April 2024).
55. Lenka, B.; Divya, R.K. *Advances in Agriculture Sciences*; AkiNik Publications: Delhi, India, 2020. [[CrossRef](#)]
56. Jeon, H.; Oh, S. Hybrid-Recursive Feature Elimination for Efficient Feature Selection. *Appl. Sci.* **2020**, *10*, 3211. [[CrossRef](#)]
57. Akkaya, B. The Effect of Recursive Feature Elimination with Cross-Validation Method on Classification Performance with Different Sizes of Datasets. In Proceedings of the IV International Conference on Data Science and Applications (ICONDATA'21), Pristina, Kosovo, 4–6 June 2021; pp. 4–6.
58. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification Using Support Vector Machine. *Adv. Data Min. Appl.* **2002**, *5139*, 66–72. [[CrossRef](#)] [[PubMed](#)]
59. Weston, J.; Elisseeff, A.; Schölkopf, B.; Tipping, M. Use of the Zero-Norm with Linear Models and Kernel Methods. *J. Mach. Learn. Res.* **2003**, *3*, 1439–1461.
60. Olmedo, G.F.; Gavilan, G. Code for QRF in Digital Soil Mapping. Available online: <https://github.com/ggavilan468/DSM3Dqrf/blob/main/DSM3Dqrf.R> (accessed on 15 November 2023).
61. United States Department of Agriculture. *USDA Textural Soil Classification; Soil Mechanics Level I Module*; United States Department of Agriculture: Washington, DC, USA, 1987; pp. 1–53.
62. Hiemstra, A.P. *Package 'Automap'*; Version: 1.1-9; CRAN: Vienna, Austria, 2022.
63. Hiemstra, P.; Skoien, J.O. *Package 'Automap'*; Version 1.1-9; CRAN: Vienna, Austria, 2023; pp. 1–16.
64. Hijmans, R.J.; Phillips, S.; Leathwick, J.; Elith, J. *Dismo: Species Distribution Modeling*; R Package Version 1.1-4; CRAN: Vienna, Austria, 2017; p. 55.
65. Meinshausen, N. *Package "quantregForest"—Quantile Regression Forests*; Version 4.3.3; CRAN: Vienna, Austria, 2017; pp. 1–8.
66. Kuhn, M. *The Caret Package*; Version 6.0-94; CRAN: Vienna, Austria, 2012.
67. Marcondes, R.; Souza, S.; Lucas, J. *Package 'Soiltexure'*; Version: 1.5.3; CRAN: Vienna, Austria, 2022.

68. Penížek, V.; Borůvka, L. Soil Depth Prediction Supported by Primary Terrain Attributes: A Comparison of Methods. *Plant Soil Environ.* **2006**, *52*, 424–430.
69. Chen, S.; Mulder, V.L.; Martin, M.P.; Walter, C.; Lacoste, M.; Richer-de-Forges, A.C.; Saby, N.P.A.; Loiseau, T.; Hu, B.; Arrouays, D. Probability Mapping of Soil Thickness by Random Survival Forest at a National Scale. *Geoderma* **2019**, *344*, 184–194. [[CrossRef](#)]
70. 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. [[CrossRef](#)]
71. Zeng, P.; Song, X.; Yang, H.; Wei, N.; Du, L. Digital Soil Mapping of Soil Organic Matter with Deep Learning Algorithms. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 299. [[CrossRef](#)]
72. Cavazzi, S.; Corstanje, R.; Mayr, T.; Hannam, J.; Fealy, R. Are Fine Resolution Digital Elevation Models Always the Best Choice in Digital Soil Mapping? *Geoderma* **2013**, *195–196*, 111–121. [[CrossRef](#)]
73. Dharumarajan, S.; Hegde, R. Digital Mapping of Soil Texture Classes Using Random Forest Classification Algorithm. *Soil Use Manag.* **2022**, *38*, 135–149. [[CrossRef](#)]
74. Adhikari, K.; Kheir, R.B.; Greve, M.B.; Böcher, P.K.; Malone, B.P.; Minasny, B.; McBratney, A.B.; Greve, M.H. High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Sci. Soc. Am. J.* **2013**, *77*, 860–876. [[CrossRef](#)]
75. Berhe, A.A.; Harden, J.W.; Torn, M.S.; Harte, J. Linking Soil Organic Matter Dynamics and Erosion-Induced Terrestrial Carbon Sequestration at Different Landform Positions. *J. Geophys. Res. Biogeosci.* **2008**, *113*, G04039. [[CrossRef](#)]
76. Wynn, J.G.; Bird, M.I.; Wong, V.N.L. Rayleigh Distillation and the Depth Profile of <sup>13</sup>C/<sup>12</sup>C Ratios of Soil Organic Carbon from Soils of Disparate Texture in Iron Range National Park, Far North Queensland, Australia. *Geochim. Cosmochim. Acta* **2005**, *69*, 1961–1973. [[CrossRef](#)]
77. Hobley, E.U.; Wilson, B. The Depth Distribution of Organic Carbon in the Soils of Eastern Australia. *Ecosphere* **2016**, *7*, e01214. [[CrossRef](#)]
78. Fierer, N.; Schimel, J.P.; Holden, P.A. Variations in Microbial Community Composition through Two Soil Depth Profiles. *Soil Biol. Biochem.* **2003**, *35*, 167–176. [[CrossRef](#)]
79. Özalp, M.; Turgut, B.; Erdoğan Yüksel, E.; Yıldırım, S. Changes on Soil Properties Associated with Soil Depth in Eroded Areas: A Case Study of Pamukcular Watershed. *Int. Cauc. For. Symp.* **2013**, *1*, 103–107.
80. John, K.; Isong, I.A.; Kebonye, N.M.; Ayito, E.O.; Agyeman, P.C.; Afu, S.M. 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](#)]
81. Gallant, J.C.; Dowling, T.I. A Multiresolution Index of Valley Bottom Flatness for Mapping Depositional Areas. *Water Resour. Res.* **2003**, *39*, 1347–1359. [[CrossRef](#)]
82. Graf-Rosenfellner, M.; Cierjacks, A.; Kleinschmit, B.; Lang, F. Soil Formation and Its Implications for Stabilization of Soil Organic Matter in the Riparian Zone. *Catena* **2016**, *139*, 9–18. [[CrossRef](#)]
83. Conant, R.T.; Ryan, M.G.; Ågren, G.I.; Birge, H.E.; Davidson, E.A.; Eliasson, P.E.; Evans, S.E.; Frey, S.D.; Giardina, C.P.; Hopkins, F.M.; et al. Temperature and Soil Organic Matter Decomposition Rates—Synthesis of Current Knowledge and a Way Forward. *Glob. Chang. Biol.* **2011**, *17*, 3392–3404. [[CrossRef](#)]
84. Tripathy, D.B.; Raha, S. Formation of Soil. *Them. J. Geogr.* **2019**, *8*, 144–150. [[CrossRef](#)]
85. Bontemps, J.D.; Bouriaud, O. Predictive Approaches to Forest Site Productivity: Recent Trends, Challenges and Future Perspectives. *Forestry* **2014**, *87*, 109–128. [[CrossRef](#)]
86. Guo, L.; Shi, T.; Linderman, M.; Chen, Y.; Zhang, H.; Fu, P. Exploring the Influence of Spatial Resolution on the Digital Mapping of Soil Organic Carbon by Airborne Hyperspectral VNIR Imaging. *Remote Sens.* **2019**, *11*, 1032. [[CrossRef](#)]
87. IDE Minagri IDE Minagri. Available online: <https://ide.minagri.gob.cl/geoweb/descargas/> (accessed on 22 April 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.