

Article

# Three-Dimensional Mapping of Forest Soil Carbon Stocks Using SCORPAN Modelling and Relative Depth Gradients in the North-Eastern Lowlands of Germany

Alexander Russ <sup>1,\*</sup>, Winfried Riek <sup>1,2</sup> and Gerd Wessolek <sup>3</sup><sup>1</sup> Landesbetrieb Forst Brandenburg, 16225 Eberswalde, Germany; Winfried.Riek@hnee.de<sup>2</sup> Fachbereich für Wald und Umwelt, Hochschule für nachhaltige Entwicklung Eberswalde, 16225 Eberswalde, Germany<sup>3</sup> Institut für Ökologie, Technische Universität Berlin, 10587 Berlin, Germany; Gerd.Wessolek@tu-berlin.de

\* Correspondence: Alexander.Russ@LFB.Brandenburg.de

**Abstract:** To cope with the challenges in forest management that are contemporarily caused by climate change, data on current chemical and physical soil properties are more and more necessary. For this purpose, we present a further amalgam of depth functions and SCORPAN modelling to provide data at arbitrary depth layers. In this concept, regionalisation is split up into the modelling of plot totals and the estimation of vertical distributions. The intended benefits by splitting up are: consistency between estimates on plot level and depth layer level, avoidance of artificial depth gradients, straightforward interpretation of covariates in the sense of pedogenetic processes, and circumnavigation of the propagation of uncertainties associated with separation between horizons during field sampling. The methodology was tailored to the circumstances within the north-eastern lowlands and the utilisation of current inventory data of the National Forest Soil Inventory (NFSI) in Brandenburg (Germany). Using the regionalisation of soil organic carbon (SOC) as an example, the application is demonstrated and discussed in detail. The depth to groundwater table and terrain parameters related to the catchment area were the main factors in SOC storage. The use of kriging did not improve the model performance. The relative depth gradients of SOC were especially distinguished by tree species composition and stand age. We suppose that interesting fields of application may be found in scenario-based modelling of SOC and when SOC serves as a basis for hydrological modelling.

**Keywords:** depth functions; SCORPAN modelling; soil forming factors; forest soils; soil organic carbon; regionalisation; north-eastern lowlands; forest site mapping; Brandenburg



**Citation:** 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. <https://doi.org/10.3390/app11020714>

Received: 16 November 2020

Accepted: 5 January 2021

Published: 13 January 2021

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



**Copyright:** © 2021 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

Nowadays, regional data on forest soil properties are increasingly demanded for various questions concerning forest management practices, like tree species selection, liming, harvest intensity, or the detection of risk areas. Thus, quantitative data related to climate change and its site-specific drought effects, as well as data on current soil nutrient status, are required to support decision-making.

Soil organic carbon (SOC) profoundly influences a soil's cation exchange capacity [1–3], which is a key feature for nutrient supply to trees. Soil texture, SOC, and the usually tightly correlated bulk density [4–6] are the main factors controlling the soil hydraulic properties [7–9]. Especially in the case of the mainly sandy textured soils of the north-eastern lowlands, soil organic carbon strongly influences the spatial patterns of hydraulic properties [10]. In addition to these well-understood principles, which are widely used in modelling water and element budgets, SOC also affects the wettability of soils [11,12] and, thus, the development of preferential flow paths [13]. Implications of preferential flow paths are not limited to percolation and actual soil water storage [14], and may also include feedback effects on local climate [15].

Whereas in the glacial deposits of the north-eastern lowlands, soil texture is largely predetermined by parent material, other soil forming factors, like groundwater, vegetation/forest management, and climate, also have to be taken into consideration for soil organic carbon. In particular, due to the influence of groundwater and vegetation, the temporal variation of SOC is expected to be high (compared to soil texture). Since enrichment of soils with soil organic matter usually starts just after deposition of parent material, tighter correlations with current terrain can also be expected. On the other hand, to cope with the temporal variation, contemporary measurements of SOC are strictly needed for regionalisation.

Over the last decades, several approaches for mapping SOC have been developed, which generally make intense use of terrain parameters derived from digital elevation models. Further covariates related to soil forming factors, like climate data, geological mapping units, or data on landcover, are also frequently used. The methods usually involved range from machine learning techniques, like support vector machines [16] or artificial neural networks [17], to all kinds of statistical models, such as stepwise linear regression and decision trees [18–21]. Especially since the introduction of the SCORPAN framework for digital soil mapping by McBratney et al. [22], which extended the classic concept of soil forming factors [23] by soil and space, the additional use of geostatistics in regionalisation approaches for SOC became a widely used technique [24]. Newer approaches also use deep learning [25], data augmentation [26], and increment-averaged kriging [27] to predict multiple depths with a single model.

The majority of studies conducted are on the prediction of SOC for predefined depths of the entire soil solum or discrete soil layers. Prior to development of prediction models, it is therefore necessary that the solum depth or soil layer boundaries are well defined for the respective purposes. If the SOC of several soil layers should be mapped, fitting of the increasing number of prediction models becomes more and more time consuming. The SOC stock of every layer is then treated as unbounded random variables (see also [28]). Thus, it becomes far from straightforward to ensure consistency between SOC stored in single layers and the stocks of the entire soil solum at every single site in the resulting maps.

Analogous problems arise for the resulting depth gradients of SOC, which might adopt artificial shapes. Furthermore, fitting of individual models for respective soil layers forces the user to deal with several different estimates of variable importance. Except for soil layers with a priori well-known differences regarding the main processes for carbon enrichment, such as the organic layer and mineral soil, it will hardly be possible to logically interpret the influence of covariates in terms of soil forming factors. This may hinder the optimal selection of predictors and detection of spurious covariates to be excluded. Last but not least, individual regionalisation of depth layers may enhance the propagation of potentially large sampling errors caused by incorrect separation. Such errors especially result from mingling of the organic layer and mineral soil [29].

The usage of depth functions to describe the vertical variation of soil properties has a long tradition in soil science (see [30] for a detailed overview). The application of soil attribute depth functions in three-dimensional spatial prediction of soil properties was already proposed by Bishop et al. [31]. It has also already been shown that the concept of depth functions is beneficial in mapping SOC distribution throughout the soil profile: Equal-area quadratic smoothing splines were used to derive SOC for homogeneous depth increments from legacy profile data as a data basis for the development of artificial neural networks and regression models [21,32,33]. To obtain maps of continuous depth functions, Malone et al. [32] utilised equal-area quadratic splines for interpolation through predicted SOC contents of discrete soil layers after regionalisation. To calculate weighted means of SOC for mapping units from available legacy soil data, Odgers et al. [34] applied equal-area quadratic smoothing splines. From a more conceptual perspective, depth functions are used for two reasons within 2.5D soil mapping frameworks [35]: The first reason is the harmonisation of soil observations. Equal-area quadratic smoothing splines are then often fitted to depth interval values after mapping to derive continuous depth

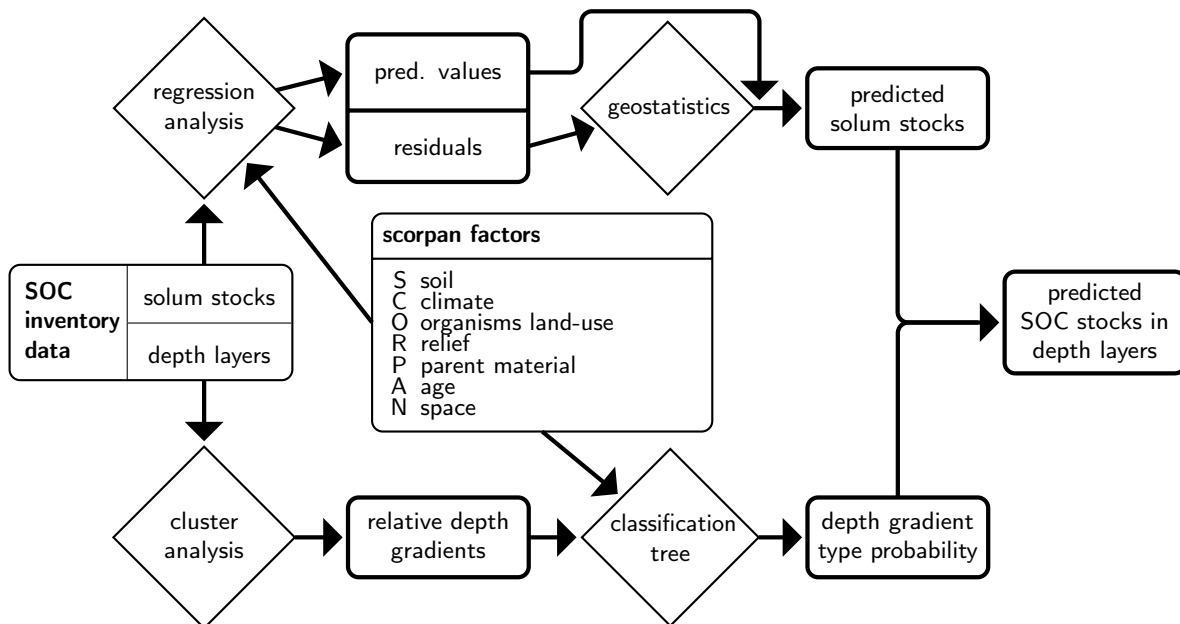
functions for each pixel mapped. The equal-area smoothing splines of Bishop et al. [31] can be considered as the state of the art for these two applications for one reason in particular: The interpolation process preserves the masses originally measured or predicted on a depth interval basis (especially for small values of  $\lambda$ ).

Direct utilisations of depth functions during the mapping process itself are given by Mishra et al. [36] and Kempen et al. [37]. Mishra et al. [36] used ordinary kriging to interpolate parameters of exponential functions fitted to single-point observations beforehand. Kempen et al. [37] predicted parameters of soil-type-specific depth functions of SOC using cokriging.

By using relative depth gradients, we try to strengthen the interpretability of SCORPAN factors in three-dimensional mapping of SOC in forest soils. Our further objectives were to ensure the consistency between solum stocks and stocks within single depth layers, to avoid the construction of artificial depth gradients, and to lessen the propagation of sample separation errors. The basic conceptual hypothesis of the method is that distinguishing between covariates that control horizontal and vertical distribution of SOC simplifies variable selection during model development. Even though increasing interpretability may decrease model performance measures, we believe that a good understanding of covariates in terms of soil forming factors is valuable for the mapping results.

## 2. Material and Methods

For the regionalisation of SOC stocks in the north-eastern lowlands, we started with the general methodology proposed by Kempen et al. [37]. The main accommodations to the requirements for mapping SOC in the north-eastern lowlands concern a simplified derivation of depth functions, which does not require expert-knowledge-based groupings like horizons and soil types. Additionally, we tried to facilitate the accessibility of the main factors controlling the horizontal and vertical distribution of SOC. To achieve this, we substituted soil-type-specific depth functions with data-driven clustering and split up the mapping process for the vertical and horizontal distribution of SOC with relative depth gradients. A flowchart of the methods and data involved is given in Figure 1.



**Figure 1.** Conceptual framework for mapping soil organic carbon (SOC) stocks in depth layers using SCORPAN modelling and relative depth gradients.

In the first step, stepwise regression analysis was used to select the most significant covariates, which determine whole solum SOC stocks at inventory plots, from a bundle of potentially relevant SCORPAN factors. Afterwards, the residuals of the fitted regression

models were examined using geostatistical methods (variogram analysis, ordinary kriging) on spatial dependencies. Hence, the regionalisation of the SOC solum stocks complied with the widely used regression kriging discussed in detail by Hengl et al. [38].

The regionalisation of depth gradients started with calculation of relative SOC stocks within the particular depth layer by expressing single stocks in proportion to solum stocks. The obtained relative depth gradients were then grouped into depth gradient types using cluster analysis. Finally, classification tree approach was used to estimate the probability that a site was in one of the distinct depth types. In addition to the purpose of regionalisation, the resulting classification tree offers insights into the main factors controlling the vertical distribution of SOC.

### 2.1. Inventory Data

To address the expected temporal variability of SOC adequately, we restricted the SOC data invoked to inventory and inventory-like data. Our primary source of SOC data was thus the second National Forest Soil Inventory (NFSI) [39]. An additional NFSI-like soil sampling campaign was conducted at a regional scale to enhance the analysis of topographic effects on SOC. Within the NFSI, field sampling was conducted during 2006 and 2009. At the regional scale, the collection of soil samples was done in the years 2008 and 2009 [40]. We did not make use of additional legacy profile data to avoid biased estimates on total SOC stocks and to prevent a mix-up of spatial and temporal effects on SOC. In contrast to the often-used legacy profile data, the NFSI comes with six well-defined uniform depth layers (organic layer, 0–5, 5–10, 10–3, 30–60, and 60–90 cm). Hence, the homogenisation or refinement of soil layers was not necessary prior to regionalisation. The volumetric SOC contents ( $\text{kg m}^{-3}$ ) were obtained by using Equation (1), where  $\rho_{fe}$ ,  $C_{org}$ , and  $CF$  denote the bulk density of the fine earth ( $\text{kg m}^{-3}$ ), the concentration of organic carbon ( $\text{kg kg}^{-1}$ ), and the volume fraction of coarse fragments (%), respectively.

$$\text{SOC} = \rho_{fe} C_{org} \left(1 - \frac{CF}{100}\right) \quad (1)$$

The volume fraction of coarse fragments was estimated in the field and measured by sieve analysis in the laboratory (depending on rock size). Mass fractions obtained from the sieve were transformed into volume fractions by assuming a rock density of  $2.65 \text{ g cm}^{-3}$ . The bulk density of the fine earth was determined on soil-core samples. Mass and volume fractions of coarse fragments ( $>2 \text{ mm}$ ) contained within the core samples were excluded from the calculation. Organic carbon was determined by dry combustion using a total analyser (LECO CNS-2000). All laboratory analyses were performed in compliance with the NFSI standards, which are documented in detail in GAFA [41].

### 2.2. Environmental Covariates

Since the preliminary considerations strongly suggested that SOC is controlled by a wide range of influences, we tried to obtain environmental covariates that address all seven SCORPAN factors appropriately. A short overview of the compiled variables is given in Tables 1 and 2.

Covariates related to parent material and soil properties were primarily derived from forest site mapping data [42,43] and by means of digital soil mapping techniques [40,44]. In addition to widely used particle size fractions, we additionally computed the geometric mean particle size  $Dg\%$ ... for numerical characterisation of the soil texture.  $Dg\%$ ... was originally proposed by Meersmans et al. [45] to improve the reflection of SOC variation due to texture. For estimates of the depth of groundwater, we additionally considered water-table maps based on monitoring wells [46].

Climatic factors were directly taken from the detailed compilation by Riek et al. [47]. The selection of indicators to be considered was guided by assumptions on dominant processes influencing the storage of SOC inside the study area. Thus,  $CWB_{win}$  (climatic water balance during winter) and  $P_{win}$  (winter precipitation) were chosen as proxies for

vertical translocation, whereas  $CWB_{sum}$  (climatic water balance in the summer half-year),  $P_{sum}$  (summer precipitation), and  $T_a$  (annual mean temperature) were used to indicate climatic effects on soil biological activity. Additionally, we calculated the annual range of temperature  $\Delta_T$  as a surrogate for continentality and the number of frost-free days. The latter was originally introduced by Ramann [48] and then revisited by Jenny [23] as Ramann's Weathering Factor, expressing the length of the 'chemical weathering season'.

**Table 1.** Environmental covariates for the regionalisation of SOC arranged by the associated SCORPAN factors.

Soil Properties (s), Parent Material (p)		Organisms, Vegetation, Land Use (o)	
$CF_{...}$	coarse fragments [%]	$P_{deciduous}, P_{pine}, P_{beech}, P_{oak}$	proportions of tree species
$S\%_{...}, Si\%_{...}, C\%_{...}$	proportions of particle-size classes (sand, silt, clay)	<b>Topography (r)</b>	
$Dg\%_{...}$	geometric mean particle size	terrain attributes derived from the digital elevation model (DEM) (see Table 2)	
$z_{CaCO_3}$	depth of calcareous horizon	<b>Age (a)</b>	
$z_{GW}$	depth of groundwater	$series_{...}$	geochemical series (age of deposit)
<b>Climate (c)</b>		$t_{stand}$	forest stand age
$CWB_{sum}, CWB_{win}$	climatic water balance (seasonal)	<b>Spatial Position (n)</b>	
$T_a$	annual mean temperature	$WA_{200}, WA_{500}, WA_{1000}$	proportion of surrounding forest area (within search radii of 200, 500, 1000 m)
$n_{d>0^{\circ}C}$	frost-free days	$l_{edge}$	distance to nearest forest edge
$\Delta_T$	annual range of temperature (as surrogate for continentality)		
$P_a, P_{win}, P_{sum}$	precipitation (annual, seasonal)		

The proportions of tree species were derived from forest inventory data ('Datenspeicher Wald') to obtain forest-specific indicators on land-use effects. The usage of forest inventory data allowed us to calculate the respective percentages from ratios of a single-species basal area to the total basal area of the forest stands. In the rare case of missing data, CORINE landcover classes [49] were used instead (applies to mapping only).

The role of topography on SOC was examined with several terrain attributes derived from a digital elevation model (DEM). As the cell size of the underlying DEM was 25 m, the terrain attributes derived from the environmental covariates with the highest spatial resolution in our study. Seeing that the covariates derived from DEM might be able to improve the prediction of SOC at the smallest spatial scales and the general availability of DEM data, we made exhaustive use of them. Nevertheless, the selection of indices from the vast number of existing terrain attributes was once again strictly constrained with respect to the theory of topography as a soil forming factor. The computed terrain attributes are tabulated in Table 2.

The time factor on SOC was considered from a geologic and a forest management perspective. For this, we computed the mean stand age as a basal-area-weighted average on the basis of forest inventory data. To cope with the problem of soil formation time, we used the geochemical series gathered by forest site mapping as a rough estimate.

**Table 2.** Terrain attributes derived from the digital terrain analysis.

<b>Topographic Indices Derived by Evaluating the <math>3 \times 3</math> Altitude Submatrix or Elevations Within Search Radii</b>	
$\beta$	slope [50]
north, east	northness and eastness derived from cosine and sine transformation [51] of aspect [50]
$\beta_d$	downslope index [52]
$C, C_{\text{plan}}, C_{\text{prof}}$	isotropic, planform, and profile curvature [50]
$\Psi_{L...}$	negative openness (radii 100, 250, 500 m in [53])
$\Psi_S$	sky-view factor [54]
$\Psi_T$	terrain configuration factor [55]
$\Psi_{\text{wte}}$	wind exposition [54]
$\hat{S}$	direct solar beam irradiance [56]
$TPI...$	topographic position index (radii 25, 125, 500 m in [57])
$TRI...$	terrain ruggedness index (radii 50, 300 m in [58])
<b>Topographic Indices Calculated at the Catchment Scale</b>	
$A_C$	catchment area [59]
$\beta_C$	average slope [50] within catchment
$L, S, LS$	slope length, slope steepness, and topographic factors from the Revised Universal Soil Loss Equation (RUSLE in [60])
$TWI$	topographic wetness index [61]
$SPI$	stream power index [62]
<b>Terrain Attributes Derived from Drainage (DN) and Ridge Networks (RL) [63,64] at Different Spatial Scales (Large Scale (l), Small Scale (s))</b>	
$\beta_{RL}(...), \beta_{DN}(...), \beta_{\text{tot}}(...)$	uphill slope angle (to ridge network), downhill slope angle (to drainage network), and total slope angle (slope from ridgeline to drainage network) [65]
$s_{RL}(...), s_{DN}(...), s_{\text{tot}}(...)$	horizontal distances to ridgeline and drainage networks and total slope length (distance from ridgeline to drainage network) [65]
$\Delta z_{DN}(...)$	elevation above the thalweg [66]
$RSP(...)$	relative slope position [65]
$MBI(...)$	mass balance index [67]

In addition to the consideration of space by the use of geostatistics, we also computed simple distance-based indices. The created indices should serve as proxies for edge effects, like micro-climate and atmospheric deposition. Since the probability of land-use changes in the past increases with closeness to the forest edge, the indices may also act as proxies for former stand history (beyond the time horizon of forest inventory data).

During the development of the model, we tried to circumnavigate the positional errors contained in the different sources of spatial data. Thus, whenever the NFSI contained the same information, as in the case of soil texture or tree species composition, we favoured the inventory data as a source for derivation of environmental covariates. Field estimates from the NFSI of terrain features, slope, exposition, and curvature were also used to verify the spatial positioning of sample locations within the DEM. Following Grimm and Behrens [68], we also iteratively inspected alternative residual values at DEM cells surrounding the measured coordinates. Thereby, we attempted to obtain a further mitigation of uncertainties resulting from locational errors of the DEM and sample points. The assumed range of spatial uncertainty and, thus, the respective domain for virtually shifting the sample points within the DEM were constrained to the  $3 \times 3$  altitude submatrix in both cases.

### 2.3. Regression Analysis of Total Solum SOC

The amount of SOC within the soil solum was assumed to be equal to the sum of SOC stored in the six NFSI depth layers (organic layer plus mineral soil 0–90 cm). Altogether, 424 profiles could be used within the analysis. The total amount of SOC stored within the soil solum  $\text{SOC}_{\text{solum}}$  ( $\text{kg m}^{-2}$ ) was modelled using multiple linear regression equations in the usual form of Equation (2). Predicted values, the constant, regression coefficients, covariates, and the number of explanatory variables are denoted by  $\hat{Y}$ ,  $b_0$ ,  $b_j$ ,  $X_j$ , and  $J$ , respectively:

$$\hat{Y} = b_0 + \sum_{j=1}^J b_j X_j. \quad (2)$$

Distributions of solum stocks were inspected by histogram analysis. The functional relationships to predictor variables were examined using scatter plots. In cases showing evidence of non-linear relationships, appropriate intrinsically linear functions were tested. Residual values were examined for heteroscedasticity and autocorrelation in scatter plots and were additionally tested with the Durbin–Watson statistic for the latter. As a measure for the multicollinearity, the tolerance value  $T_j$  was computed. Low tolerances indicate increasing magnitudes of multicollinearity. A tolerance value of 1 denotes perfectly uncorrelated predictors. Thresholds for entry and removal of covariates into/from the regression equations within the stepwise selection process were set to probabilities of F of 0.1 and 0.15, respectively [69].

We chose stepwise linear regression analysis over the more elaborate statistical techniques that are currently used in digital soil mapping. The intended advantages by sticking to this well-established approach were the integrated variable selection during model development and the direct interpretability of covariate effects: Using multiple linear regression, the influence of each predictor on the SOC can be seen from the regression coefficients  $b_j$ . Positive coefficients indicate that SOC increases with increasing covariate values. Absolute values of standardised regression coefficients  $\hat{b}_j$  were used as an indicator for relative variable importance within the regression models. Throughout the iterative model development, we permanently examined the plausibility of the predictors that were entered. Covariates that failed to comply with the expected effects associated with the respective soil forming factors were continuously eliminated. This was especially the case if the sign of the estimated regression coefficients was not in accordance with the a priori assumed causal relationships. We also examined the stability of the direction of effects within the lower and upper confidence interval bounds,  $b_j - t_{95} s_{b_j}$  and  $b_j + t_{95} s_{b_j}$ .

Model performance was assessed by means of scatter plots and statistical indices. For the latter, the common measures—coefficient of determination  $r^2$ , adjusted coefficient of determination  $r_{\text{adj}}^2$ , norm of residuals  $s_e$ , and F-statistic  $F_{\text{emp}}$ —were calculated. In addition to the classical model performance measures computed on the calibration sample, we used cross-validation to estimate the effects of random fluctuations due to the limited calibration sample size [70]. For this purpose, we employed repeated tenfold cross-validation [71]. For each model one hundred runs of tenfold cross-validation were performed, resulting in one thousand model fits. Based on the one hundred validation data sets that were obtained, the coefficient of determination  $r_{\text{cv}}^2$  and the root mean square error  $\text{RMSE}_{\text{cv}}$  were recalculated. We also computed the least squares analysis of slope  $b_{1\text{cv}}$  and the concordance correlation coefficient of Lin [72] ( $\text{CCC}_{\text{cv}}$ ). Both statistics are common indicators in assay and instrument validations. All computations were performed using IBM SPSS Statistics 19 (IBM®).

Inspired by the stratified modelling approach of Zirlewagen and v. Wilpert [73], regionalisation of the total solum SOC was divided into two submodels. The first submodel was calibrated at terrestrial sites, while the second addressed soil development under hydromorphic conditions. The idea behind partitioning into submodels is the assumption that markedly different factors control storage of SOC in both hydrologic domains. A desired side effect was the introduction of a (nonlinear) threshold point for groundwater

effects on SOC at the terrestrial sites. To prevent discontinuities and to maintain sufficient cases, sites within a transient area from the hydromorphic regime to terrestrial soils were used in the development of both models. The overlap used ranged from 2.0 to 3.0 m depth to the groundwater table. The upper threshold for distinguishing between both hydrologic domains conforms to the classification given by the German soil mapping guidelines, ‘GWS6’ [74], whereas the lower bound of 3.0 m is taken from local forest site mapping instructions [43]. It is the cut-off above which forest sites are considered to be influenced by groundwater.

#### 2.4. Geostatistical Analysis

Following the recommendations of McBratney et al. [22], we applied variogram analysis and kriging on the residuals to examine the spatial trends within the remaining variance. Considering the density of the sampling points, we mainly intended to incorporate unknown large-scale effects, such as paleoclimate or atmospheric depositions, rather than to improve predictions in the close surroundings of single NFSI plots.

All geostatistical analyses were conducted under the assumption of isotropy. The autocorrelation coefficients  $\rho(h)$  and semivariances (Matheron’s estimator)  $\gamma(h)$  were calculated according to Webster and Oliver [75]. Since variograms and autocorrelograms are sensitive to the underlying lag increment, the first step of the analysis was the estimation of a suitable interval. In order to find an appropriate increment, we calculated the average separation between sample points and their nearest neighbours, as recommended by Webster and Oliver [75]. Finally, the potential model improvement due to the estimation of residuals by ordinary kriging was evaluated. For this purpose, we performed a leave-one-out cross-validation, as recommended by Webster and Oliver [75].

#### 2.5. Depth Gradients

To derive the relative depth gradients, the volumetric SOC contents ( $\text{kg m}^{-3}$ ) of the six NFSI layers  $i$  were expressed as depth-related percentages  $\text{SOC}\%$  ( $\% \text{ m}^{-1}$ ) of total solum SOC storage  $\text{SOC}_{\text{Solum}}$  ( $\text{kg m}^{-2}$ ) (Equation (3)). At sites where a forest floor layer was lacking, the partial SOC storage for this layer was set to zero.

$$\text{SOC}\%_i = 100 \frac{\text{SOC}_i}{\text{SOC}_{\text{Solum}}} \quad (3)$$

All depth data were grouped into depth gradient types using clustering. We used the common agglomerative hierarchical procedure of Ward [76]. As an objective function for minimising at each fusion step, the squared Euclidean distance was used. As stated by Backhaus et al. [77], Ward’s method can be considered as conservative, which means that there is little tendency to form equal or unequal group sizes. Therefore, Ward’s method is usually one of the best grouping methods for finding the actual underlying structure of the data. In particular, it is an excellent choice if the number of clusters is unknown [78]. For that reason, hierarchical grouping methods were preferred over iterative partitioning methods that require a priori assumptions on the best number of clusters [79], even if this solution may not be superior in terms of within-group homogeneity. For the calculation of Euclidean distances, the partial SOC stocks were equally weighted for each depth layer.

The estimation of the optimal cluster number was guided by examination of the scree plot (elbow method) and the stopping rule of Mojena [78]. The estimation of optimal cluster solutions by the latter is based on the threshold ranges of the standardised error sum of squares (ESS). For this, ESS values are standardised with respect to the mean value and standard deviation of the ESS across all clustering stages. Within standardised ESS values ranging from 2.75 to 3.50, Mojena [78] observed the best overall results in terms of the predicted number of clusters for the subsequent clustering stage. To obtain suitable graphical representations of the selected cluster solution, we used the equal-area quadratic smoothing splines of Bishop et al. [31]. Where available, the SOC data of the NFSI depth layer of 90–140 cm were used to improve the fitting at the bottom of the solum. According



to the findings of Bishop et al. [31], we set the the roughness parameter  $\lambda$  to 0.1. Only in cases where the fit using  $\lambda = 0.1$  obviously failed, we increased  $\lambda$  to 1.0 and refitted the the respective spline. The between-cluster distances of the obtained clusters are delineated numerically by the normalised differences between within-cluster means and total means (t-statistics). In addition to the graphical depiction, the within-cluster distances were assessed by the F-statistic, which expresses the ratio of within-group variability to total variability. Finally, depth gradient types could be obtained as arithmetic means of relative depth gradients merged to a single cluster.

### 2.6. Mapping of Relative Depth Gradient Types

To predict the probabilities of individual depth gradient types, we employed the classification and regression trees approach (CART) of Breiman et al. [80]. The usage of relative depth gradients enhances non-linear dependencies between environmental covariates and class probabilities. Therefore, we preferred tree-based approaches over parametric approaches, like stepwise linear discriminant programs. Thus, using a tree-based approach, we avoided elaborate transformations of the predictor variables. A further consideration leading to the employment of CART was the metric scale in the majority of the environmental covariates. Compared to other tree classification approaches like Chi-Squared Automatic Interaction Detection (CHAID) [81], CART is able to process interval-level variables directly without prior grouping. The Gini diversity index [82,83] was used as measure of node impurity in the growing and pruning steps of tree construction. Thus, the obtained trees were optimised to predict the probabilities of the depth gradient types (class probability tree). The most characteristic feature of CART is the determination of right-sized trees by optimal pruning instead of by using stopping rules. We employed tenfold cross-validation and selected the best-performing tree within the standard error (1 SE rule). The random samples used in tenfold cross-validation were constructed with regard to class membership. The random selection was constrained so that the proportions of clusters in each test sample equalled the respective proportions in the total sample. The only 'stopping rule' invoked was that each terminal node should contain at least a minimum of five cases, thus reducing computational effort, but still providing sufficiently large trees for subsequent pruning in almost any situation [80].

## 3. Results

### 3.1. Mapping Total Solum SOC

The best covariates for predicting solum SOC, which were determined by stepwise regression, are given in Table 3. The signs of  $RSP(s)$  (relative slope position) and  $s_{DN}(s)$  (horizontal distance to drainage network) show that SOC increases at low relative slope positions and at toe-slope positions in close proximity to the drainage line. In terms of soil forming processes, this indicates lateral inputs of organic matter due to flooding and depositions by overland flow from higher slope positions. Similar explanations can be given for the increase of SOC with the steepness of catchment slopes  $\beta_C$  and the decrease with the stream power index  $SPI$ . Erosion within the catchments causes a corresponding deposition at the sites, while at high erosive potential ( $SPI$ ), the SOC is lowered by topsoil losses or loss of surface inputs (litter). In addition to these covariates related to erosion by overland flow, the terrain parameter  $east \times \beta$  (product of local eastness and slope) can be considered as an indicator for the redistribution of organic matter by the prevailing westward winds.

An indicator for the vertical translocation of SOC within the soil profile can be seen in the winter precipitation  $P_{win}$ . Especially during winter season, when the soil is wet, rainfall strengthens podzolization (e.g., [84,85]), and thus enhances SOC sequestration in deeper soil layers.

Covariates focusing on in situ soil development are  $z_{CaCO_3}$  (depth of calcareous horizon),  $CF_{0...90}$  (proportions of coarse fragments) and  $\hat{S}$  (direct solar beam irradiance). At first glance, the positive correlation coefficient of  $z_{CaCO_3}$  appears contradictory to the in situ

effects of calcium. Actually, calcium bridges organic matter and clay minerals and stabilises aggregates [86]. However, in acidic soils without near-surface or even any calcareous layers within the solum (e.g.  $z_{CaCO_3} > 3.0$  m), soil respiration is often hindered due to the low activity of soil microbial communities under acid conditions. Of course, acid conditions induced by high  $z_{CaCO_3}$  may also enhance podzolization processes. On the contrary, soil respiration usually increases with increasing soil temperatures [87,88] under intense direct insolation  $\hat{S}$ . The reduction of SOC stocks at high rock contents  $CF_{0...90}$  indicates that the corresponding loss of soil volume is not balanced by higher concentrations of SOC in the remaining fine-earth fraction (see [89]).

**Table 3.** Selected covariates  $X_j$  for estimation of SOC stocks in the soil solum and the corresponding statistical parameters: regression coefficient  $b_j$  ( $10^{-4}$ ), standard error of the regression coefficients  $s_{b_j}$  ( $10^{-4}$ ), standardised regression coefficients  $\hat{b}_j$ , t-statistic  $t_j$ , p-value  $p_j$ , confidence interval bounds  $b_j - t_{95}s_{b_j}$  and  $b_j + t_{95}s_{b_j}$  ( $10^{-4}$ ), and tolerance value  $T_j$ .

Submodel for Terrestrial Soils ( $z_{GW} > 200$ cm)								
$X_j$	$b_j$	$s_{b_j}$	$\hat{b}_j$	$t_j$	$p_j$	$b_j - t_{95}s_{b_j}$	$b_j + t_{95}s_{b_j}$	$T_j$
$b_0$	120,171.553	23,687.729		5.073	0.000	73,589.574	166,753.532	
$RSP(s)$	21,833.983	3433.882	0.316	6.358	0.000	15,081.245	28,586.720	0.489
$z_{CaCO_3}$	56.971	6.384	0.354	8.923	0.000	44.416	69.526	0.768
$\hat{S}$	-83.837	11.125	-0.267	-7.536	0.000	-105.714	-61.960	0.962
$\beta_C$	3523.879	351.225	0.452	10.033	0.000	2833.195	4214.564	0.596
$SPI$	-724.272	102.116	-0.281	-7.093	0.000	-925.084	-523.460	0.771
$P_{win}$	350.321	61.723	0.207	5.676	0.000	228.942	471.700	0.907
$s_{DN}(s)$	-157.999	32.076	-0.247	-4.926	0.000	-221.076	-94.922	0.482
$CF_{0...90}$	-986.206	229.538	-0.157	-4.296	0.000	-1437.593	-534.819	0.906
$east \times \beta$	1385.928	330.355	0.165	4.195	0.000	736.285	2035.571	0.781
$P_{deciduous}$	-65.345	21.793	-0.126	-2.999	0.003	-108.200	-22.490	0.685
$series_I$	4550.659	2083.234	0.092	2.184	0.030	453.973	8647.344	0.674
Submodel for Soil with a Near-Surface Groundwater Table ( $z_{GW} \leq 300$ cm)								
$X_j$	$b_j$	$s_{b_j}$	$\hat{b}_j$	$t_j$	$p_j$	$b_j - t_{95}s_{b_j}$	$b_j + t_{95}s_{b_j}$	$T_j$
$b_0$	467,508.956	46,026.281		10.157	0.000	376,026.739	558,991.173	
$\ln(z_{GW})$	-74,129.674	9063.024	-0.613	-8.179	0.000	-92,143.415	-56,115.933	0.957
$A_C$	0.718	0.158	0.342	4.541	0.000	0.404	1.033	0.950
$z_{CaCO_3}$	68.062	33.571	0.154	2.027	0.046	1.336	134.789	0.937
$MBI(I)$	-356,426.489	196,452.978	-0.135	-1.814	0.073	-746,898.055	34,045.078	0.975

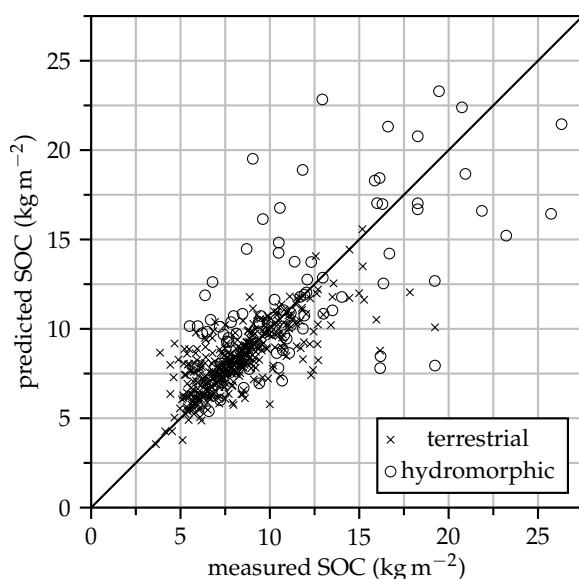
The land-use effects on SOC are captured by the proportions of deciduous trees  $P_{deciduous}$ . The negative regression coefficient implies the well-known risk of litter loss by wind blowing for leaf litter. On the other hand, this might also signal that SOC stored in the generally thicker forest floor layers at coniferous sites is not transferred to mineral soil at deciduous sites to the same extent. The raised SOC storage at juvenile glacial deposits ( $series_I$ ) appears as a result of the higher ecosystem productivity in these nutrient-rich soils.

The increase of SOC stocks in the presence of high groundwater tables ( $z_{GW} \leq 300$  cm) is in line with expectations. Carbon mineralisation is more and more inhibited due to the development of anaerobic conditions with rising groundwater levels at these sites. To describe the nonlinear response to groundwater levels  $z_{GW}$  properly, the best fit was found for the logarithmic transformation. Larger catchments also result in higher SOC stocks. In addition to the mean depth of groundwater, a large catchment area  $A_C$  indicates stable, less-fluctuating groundwater levels, thus causing higher ecosystem productivity at low water tables and reduced mineralisation (especially at high water tables). The effect of  $z_{CaCO_3}$  is similar to that of the terrestrial sites. Acidic conditions inhibit soil microbial activity. The decrease of SOC with increasing mass balance index  $MBI(I)$  is in good agreement with net loss due to the lateral transport processes associated with positive  $MBI(I)$  values.

Especially for the various terrain attributes, the occurrence of multicollinearity, masking the individual variable importance, cannot be completely neglected. In particular, within the submodel for the terrestrial soils, some smaller tolerance values were observed. Despite these restrictions, the inspection of standardised regression coefficients  $\hat{b}_j$  clearly indicates a generally high importance of terrain parameters in the submodel for terrestrial soils. The highest absolute value of  $\hat{b}_j$  was found for the slope within the catchment  $\beta_C$ . The soil property  $z_{CaCO_3}$  is also of particular importance in this submodel. The distinctly most important predictor in the submodel for soils with near-surface groundwater was, of course, the depth of groundwater  $z_{GW}$  itself. The standardised coefficient obtained for the terrain parameter  $A_C$  must be seen in terms of the groundwater regime rather than terrain effects. Although the number and importance of included variables are quite different, predictors related to nearly every SCORPAN factor are incorporated into the regression equations. The only exception to this is the space factor, indicating that edge effects and stand history are of minor importance.

The  $p$ -values  $p_j$  derived from the t-statistic are small for the majority of the regression coefficients. Increased  $p$ -values occur only for coefficients in the submodel for sites with near-surface groundwater. However, even in these cases, the sign and, thus, the direction of effects do not shift within the lower and upper confidence interval bounds,  $b_j - t_{95} s_{b_j}$  and  $b_j + t_{95} s_{b_j}$ . All in all, the results of the t-statistic suggest statistically reliable regression equations.

The model performance measures are given in Figure 2. As expected, the  $p$ -values derived from the F-distribution signal high significance. The number of explanatory variables  $J$  entered into both equations is distinctly below the threshold of  $10 N/J$  that was recommended by Hengl et al. [38] to prevent overfitting. Particularly in terms of absolute errors, the norm of residuals  $s_e$  confirms that the performance of the groundwater submodel is slightly poorer. The increased differences in this domain can also be seen clearly from the scatter plot of the measured and predicted values (Figure 2). These differences between both submodels mainly result from the high variability of SOC in the group of soils with near-surface groundwater. Thus, the coefficients of determination,  $r^2$  and  $r^2_{adj}$ , are quite similar in both models. The scatter plot shows no extreme outliers or systematic deviations. Merely by viewing the data of both models at once, some heteroscedasticity appears. Within the single models, though, the observed overall heteroscedasticity has no effect on the estimation of confidence intervals [77].

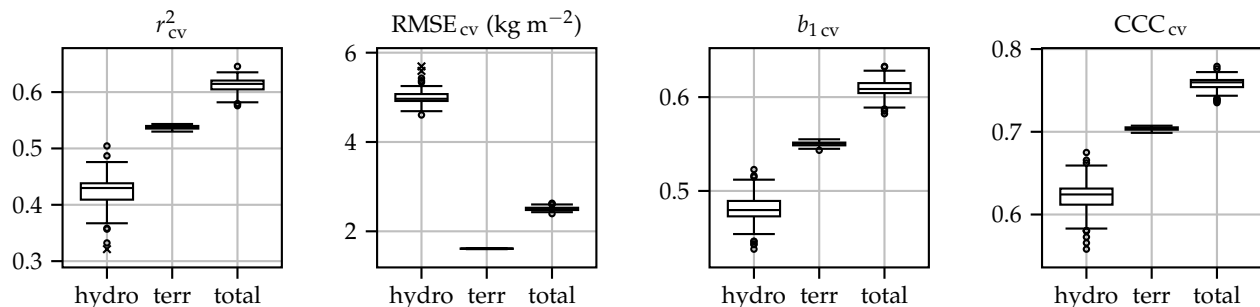


Submodel	$N$	$J$	$r^2$	$r^2_{adj}$	$s_e$	$s_{soc}$	$F_{emp}$	$p$
hydromorphic	92	4	0.532	0.510	4.60	6.58	24.7	0.000
terrestrial	376	11	0.560	0.547	1.60	2.37	42.1	0.000
total	424	15	0.632	0.618	2.02	3.30	46.7	0.000

**Figure 2.** Measured SOC stocks versus values predicted by regressions models for terrestrial soils and soils with near-surface groundwater. Model performance measures for the regressions models: number of observations  $N$ , number of explanatory variables  $J$ , coefficient of determination  $r^2$ , adjusted coefficient of determination  $r^2_{adj}$ , norm of residuals  $s_e$ , standard deviations of observed SOC stocks  $s_{soc}$ , F-statistic  $F_{emp}$ , and  $p$ -value  $p$ .

The assessment of random fluctuations due to the limited size of the calibration sample is shown in Figure 3. The cross-validation estimates of  $r_{cv}^2$  derived for the terrestrial soil models are in good agreement with the values obtained from the calibration sample. In the case of models for hydromorphic conditions, the agreement between the calibration sample and cross-validation estimates of  $r_{cv}^2$  is slightly poorer. In comparison to the cross-validation results for the group of terrestrial sites, the fluctuations of  $r_{cv}^2$  across the repetitions are distinctly higher for the hydromorphic sites. Quite similar relations can be discovered by inspecting  $s_e$  and  $RMSE_{cv}$ . The cross-validation estimates  $RMSE_{cv}$  are almost as small as the respective calibration sample estimates  $s_e$ . The fluctuation of  $RMSE_{cv}$  across validation runs is nearly negligible. Again, the situation is different for the cross-validation estimates derived for hydromorphic sites, showing slightly increased values that vary substantially over the performed repetitions. Altogether, these observations are in line with the findings of Browne [70] regarding the use of cross-validation in regression applications: If the calibration sample is sufficiently large, the model performance measures yielded by calibration and cross-validation are essentially the same. On the other hand, the cross-validation clearly unveils the slightly optimistic impression of model performance that was originally derived for the small sample of hydromorphic sites. The indicators  $b_{1cv}$  and  $CCC_{cv}$  adopted from assay and instrument validation also reflect this dependence on sample size. By definition, the slope  $b_1$  between the observed and predicted values equals the coefficient of determination when this performance measure is used in conjunction with regression analysis (applies to the calibration sample).

Thus, the cross-validation estimates  $r_{cv}^2$  and  $b_{1cv}$  also reach a comparable level, as long as the sample size is large.  $CCC_{cv}$  extends the Pearson correlation coefficient by an additional penalty for deviations of  $b_{1cv}$  from the 45° line. As a consequence of this, squared values of  $CCC_{cv}$  are distinctly lower than  $r_{cv}^2$  each time.



**Figure 3.** Box-and-whisker plot showing the results of model evaluation by one hundred runs of tenfold cross-validation: coefficient of determination  $r_{cv}^2$ , root mean square error  $RMSE_{cv}$ , least squares analysis of slope  $b_{1cv}$ , and the concordance correlation coefficient of Lin [72]  $CCC_{cv}$ .

### 3.2. Geostatistical Analysis of Residuals

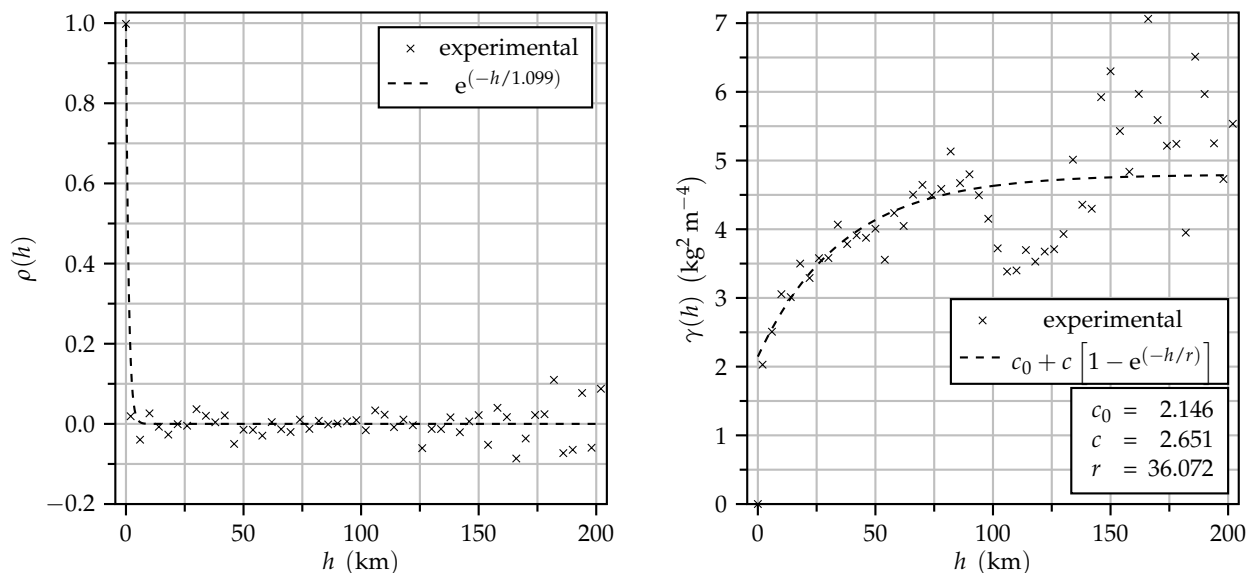
After selecting the shortest separations from all 89,676 distances between the 424 sample points, the obtained mean of 3.684 km served as a starting point to choose the lag increment. For practical reasons, e.g., plotting the variogram, the lag increment was finally set to 4.0 km. In the range 0 to 200 km, the mean number of observations contained in each of the 51 lags was 1752. Only at lags > 150 km did the numbers of cases fall below 1000.

The experimental autocorrelation coefficients  $\rho(h)$  are shown in Figure 4. The autocorrelogram shows an immediate decline of autocorrelation coefficients within the first lag interval. At all subsequent lag distances  $h \geq 2$  km, the autocorrelation coefficients show a narrow range of scatter around zero. As a consequence of the considerable number

of negative coefficients, it is not possible to compute the autocorrelation length with the exponential autocorrelation function (Equation (4)) commonly used for this purpose [90].

$$\sum_{n=1}^{i=1} [\rho_i(0) - e^{(-h/\lambda)}] = 0 \tag{4}$$

Even as the autocorrelation length reaches close to zero, the sum of coefficients cannot be balanced by the exponential autocorrelation function. Figure 4 shows an alternative fit that assumes all negative coefficients as zero, yielding an autocorrelation length of 1.1 km. These findings imply that autocorrelation is of minor importance at the scale analysed. Likewise, this can be taken as an indication that the main influences affecting SOC in the study area are already captured by the regression models. This is very similar to the findings of Zirlewagen and von Wilpert [91] on the regionalisation of SOC in the forest soil of Baden-Wuerttemberg (Germany). They also observed only very weak autocorrelation coefficients for residuals remaining after regression analysis.



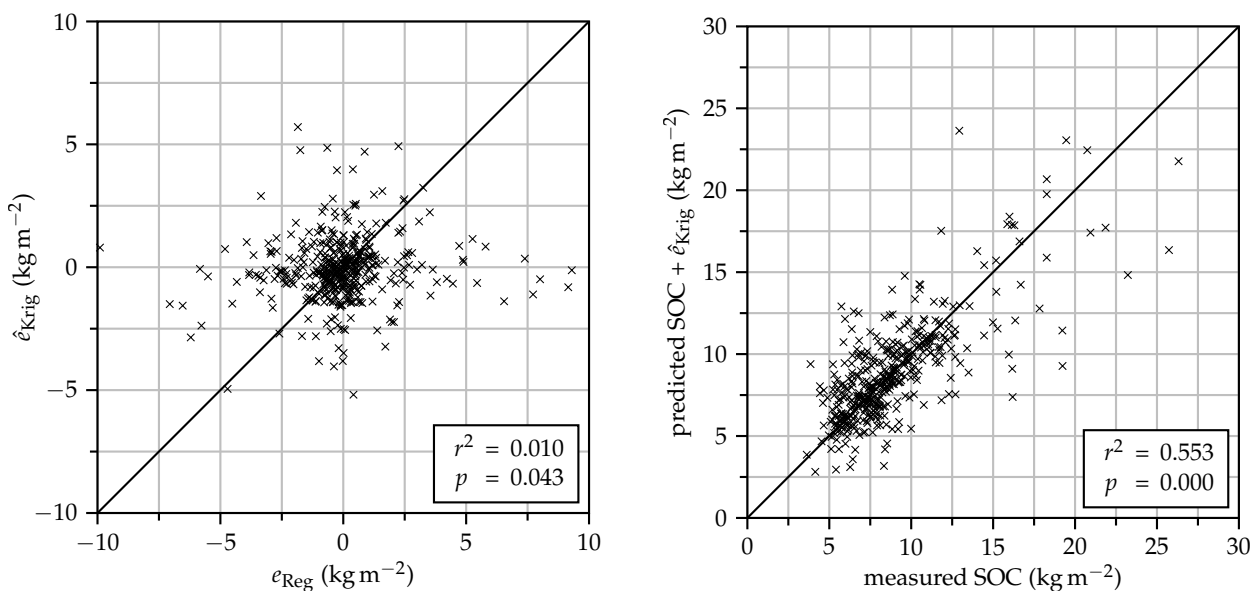
**Figure 4.** Autocorrelogram, sample variogram, and fitted variogram model (nugget + exponential) for residuals of the regression equations.

The values of the sample variogram (Figure 4) reach the sill within a range of 100 km. After reaching this maximum, the semivariance shows a hole effect, which typically arises from regular repetitions [75] or spatial variation occurring at several scales [90]. The observed regularly oscillating semivariances are possibly caused by the periodically arranged glacial deposits in the north-eastern lowlands resulting from successive ice advances. Especially within the study area, the distance between moraines and related landscape elements left by the successive glacial stages is quite similar to the observed periodicity of the semivariogram [92,93]. Although variables like  $z_{CaCO_3}$  and  $CF_{0...90}$  were entered into the regression equations, these effects may be enhanced by a slight under-representation of the corresponding predictors (e.g., particle size classes) within the regression models.

For modelling the variogram, we selected an exponential function with an additional nugget component (Figure 4). The range for fitting the function to the experimental values was restricted to 100 km, thus neglecting the hole effect at larger lag distances. Following Webster and Oliver [75], we did not fit the variogram model to the hole effect, as there was no clear theory for the observed periodicity. In addition, the use of increased weight factors at high distances in the kriging procedure would not support the desired incorporation of unknown large-scale effects. The combined variogram model achieved a high goodness of fit ( $R^2 = 0.874$ ). The parameters estimated using the Levenberg–Marquardt

method [94] indicate that pairs of residuals are spatially correlated up to a distance of 108 km (ca  $3 \times r$ ) until the model approaches the maximum semivariance (sill  $c$ ). However, nearly half of the spatially dependent variation occurs at distances below the lag increment or is contributed by measurement errors, as can be seen from the fairly high nugget component  $c_0$  [90].

Although modelling the variogram yielded a properly fitted equation with high goodness of fit, solving the kriging equations based on it had only a low predictive power. Plotting the scatter of residuals retained from regression analysis  $e_{\text{Reg}}$  against their estimates obtained from leave-one-out cross-validation showed only a poor correlation (Figure 5). When the kriging estimates are used to improve the SOC stocks predicted by regression equations, the conducted correction actually brings a slight deterioration in model performance. Analogously to autocorrelogram analysis, these results show that a high degree of large-scale variation is already explained by regression. In this study, the use of geostatistical estimates can thus be omitted.

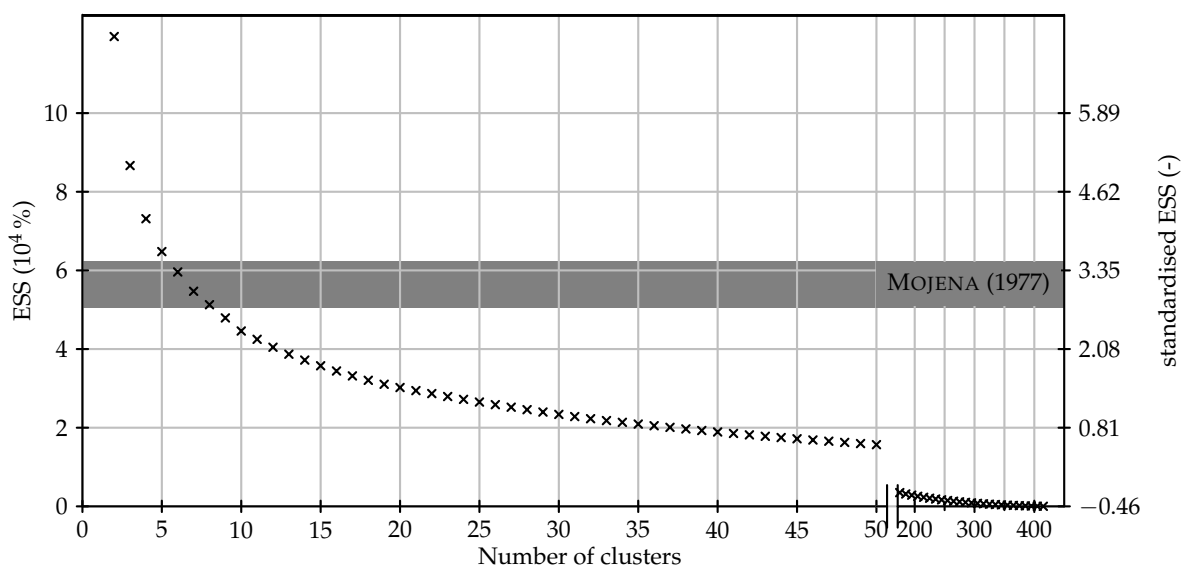


**Figure 5.** Scatter diagrams of residuals retained from regression analysis  $e_{\text{Reg}}$  against residuals estimated by ordinary kriging  $\hat{e}_{\text{Krig}}$  and measured SOC stocks versus predictions improved by kriging estimates of residuals obtained during leave-one-out cross-validation.

### 3.3. Depth Gradients

In the first step, relative depth gradients of the individual inventory plots are grouped into depth gradient types by clustering. Throughout the stagewise proceeding of the clustering process, the number of clusters decreases, starting with the number of observations (424), until all sites and clusters are unified (Figure 6).

While the the number of clusters decreases, the ESS based on squared Euclidean distances increases. After discarding the initial partition of 424 clusters and the final one-cluster solution, 422 possible cluster solutions remain. A first approximation of the optimal number of clusters with the elbow method suggests a solution with five clusters, since the most distinct increase in the ESS appears after the solutions with four clusters. Regarding the standardised ESS values obtained by clustering the relative depth gradients of SOC, the stopping rule of Mojena [78] indicates an optimal solution in the range of 5 to 8 clusters. Finally, taking into account practical considerations on handy cluster sizes that facilitate interpretation and prediction of the resulting depth gradient types, we selected the solution with five clusters for the subsequent analyses.



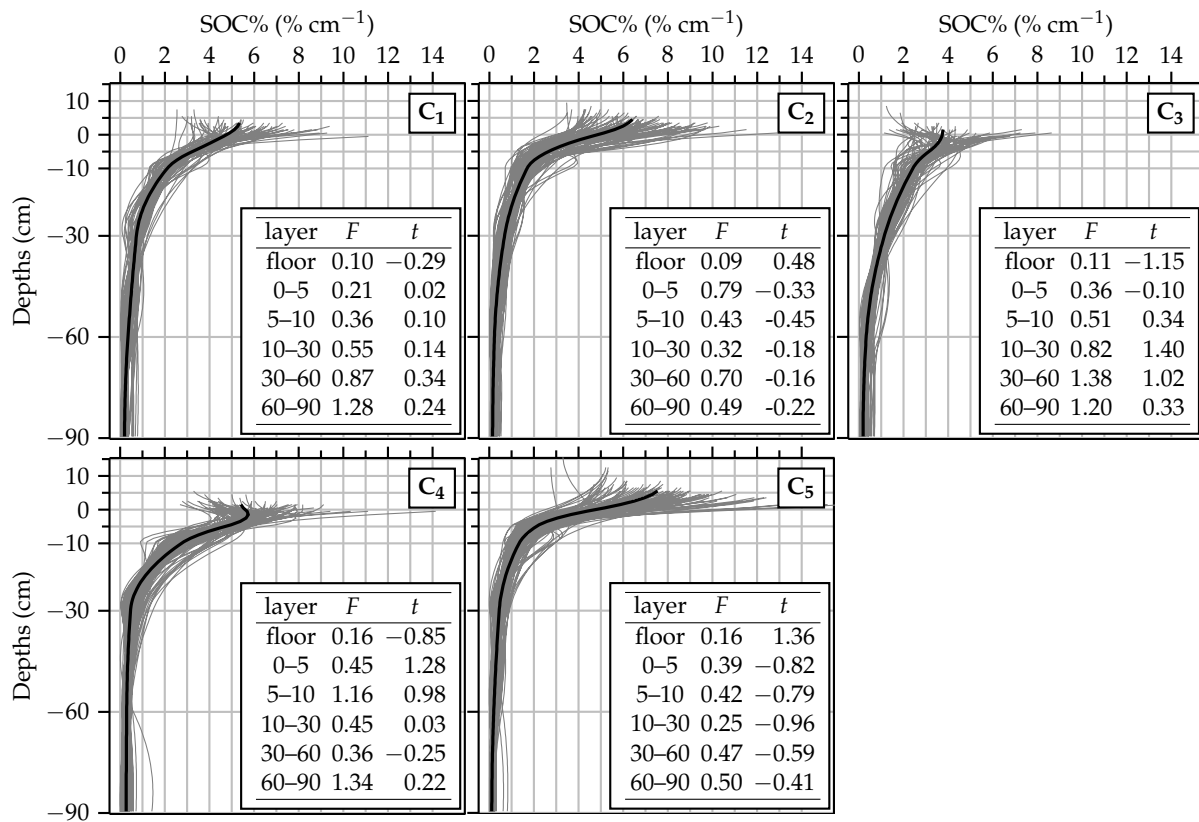
**Figure 6.** Increase in the error sum of squares (ESS) with a decreasing number of clusters during hierarchical grouping of depth gradients. The grey bar indicates the range of standardised ESS according to the stopping rule of Mojena [78].

The five relative depth gradient types obtained and the sites assigned to the respective clusters are shown in Figure 7. The most distinct differences between the five clusters can be seen for the forest floor and top soil layers. This is also indicated by the t-statistics. In these depth increments, the t-statistics attain the highest absolute values. In contrast, close to the bottom of the soil solum, the derived depth gradient types seem very similar, which is also expressed by small t-values in these depth increments. The cluster  $C_5$  stands out as having the highest proportions of SOC stored in the organic layer of any cluster. Consequently,  $C_5$  is also characterised by the smallest SOC portions inside the mineral soil with t-statistics that are invariably negative. An opposite depth gradient is typified by cluster  $C_3$ , which contains the lowest proportions of SOC inside the organic layer. It also shows the highest t-values for any mineral soil layer below 10 cm depth. Cluster  $C_4$  is in close accordance with  $C_3$  concerning the forest floor, whereas the mineral soil SOC is closely bunched in the upper parts of the profile. The clusters  $C_1$  and  $C_2$  can be seen as intermediate types between  $C_3$  and  $C_5$  in which increasing proportions of SOC are stored in the forest floor layer.

It does not seem far fetched to bring the derived depth gradients in line with soil types and humus forms. Actually, the clusters  $C_3$  and  $C_4$  often coincide with the humus forms of mull and mull-like moder, respectively. The humus forms of mor humus and mor-humus-like moder appear most frequently within the clusters  $C_2$  and  $C_5$ . Soil types were merely in rough accordance with the derived depth gradient types. Lessives tend to occur preferentially in cluster  $C_4$ . Additionally, the frequencies of arenosols and podzols are increased in the clusters associated with mor humus forms  $C_2$  and  $C_5$ . The loose relationship between clusters and soil types may be a consequence of the fact that vertical distribution of SOC is not a diagnostic criterion for many soil types. On the other hand, the observed differences between soil types and humus forms indicate that the vertical distribution of SOC throughout the soil profile is often altered by current land-use and stand history.

The plotted depth gradients (Figure 7) show a reasonable homogeneity. The assessment of within-group variability by the computed F-statistics supports this impression: Overall, with F-statistics below 1 in the majority of cases, the obtained clusters can be considered as homogeneous ([77], p. 447). The 'mor humus cluster'  $C_5$  is by far the most homogeneous cluster. The occurrence of increased F-statistics is limited to deeper soil layers. Due to the small proportions of SOC stored in these depth increments, the higher variability can be widely neglected. In fact, the increased F-statistics within the deeper layers can be seen as a result of the equally weighted distances used for hierarchical grouping.

Compared to the variability observed within the upper parts of the soil, also these layers are very homogeneous, as shown by the line graphs in Figure 7.



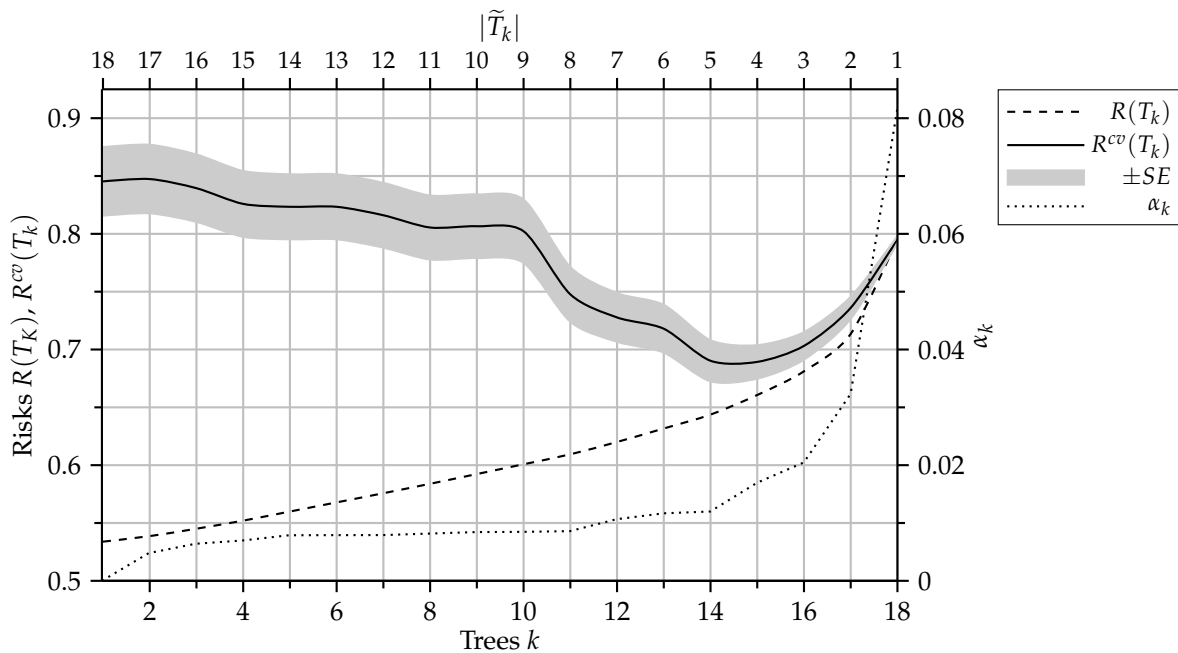
**Figure 7.** Equal-area quadratic smoothing splines fitted to the relative depth gradients of SOC percentage. The black solid line depicts the mean value of all sites assigned to the respective clusters. Additional f- and t-statistics assess the homogeneity within each cluster and the distinctness between the clusters, respectively.

### 3.4. Mapping of Relative Depth Gradient Types

The selection of an optimum-sized classification tree for the prediction of relative depth gradients is shown in Figure 8. The starting point of the selection process is the initially grown much-too-large tree with 18 terminal nodes  $|\tilde{T}_k|$  and the 10 respective auxiliary trees grown on subsamples for cross-validation. Subsequently, the weakest subbranches are incrementally pruned off by minimal cost-complexity pruning with respect to the complexity parameter  $\alpha_k$ . This results in a sequence of 18 subtrees (and 180 auxiliary subtrees) with decreasing complexity, finally ending at the root node ( $|\tilde{T}_k| = 1$ ). Since  $\alpha$  is used as a penalty for complexity in the pruning procedure, the complexity parameter  $\alpha_k$  consequently increases at every stage. As one might expect, the resubstitution estimates for the risks  $R(T_k)$ , which directly express the diversity within the terminal nodes, monotonically increase with each pruning stage. In contrast, the misclassification rate of the auxiliary trees is highest when sequential pruning is started. Hence, after a phase of decreasing cross-validation estimates  $R^{cv}(T_k)$  and reaching of the minimum 0.689 at  $k = 15$ ,  $R^{cv}(T_k)$  increases again and equals  $R(T_k)$  at the end, when all subbranches of the auxiliary tree are pruned off and just the root nodes remain. In addition to using  $R^{cv}(T_k)$  as an honest estimate of the misclassification rate, Breiman et al. [80] introduced the 1 SE rule, additionally invoking the standard error SE, to guide the user towards a more stable tree selection. Using the 1 SE rule, the subtree of the smallest complexity within the range of  $\pm SE$  around the minimum value of  $R^{cv}(T_k)$  should be selected. Keeping close to this rule, the tree  $T_{16}$ , which contains just three terminal nodes  $|\tilde{T}_k|$ , should be selected. It is obvious that a classification tree pruned back to three terminal nodes is not able to predict five clusters. Therefore, we used a less rigid interpretation of this rule and selected tree  $T_{14}$ .



This is exactly the smallest tree within the range of  $\pm SE$  around the minimum of  $R^{cv}(T_k)$  that makes it feasible to predict any distinguished depth gradient type.

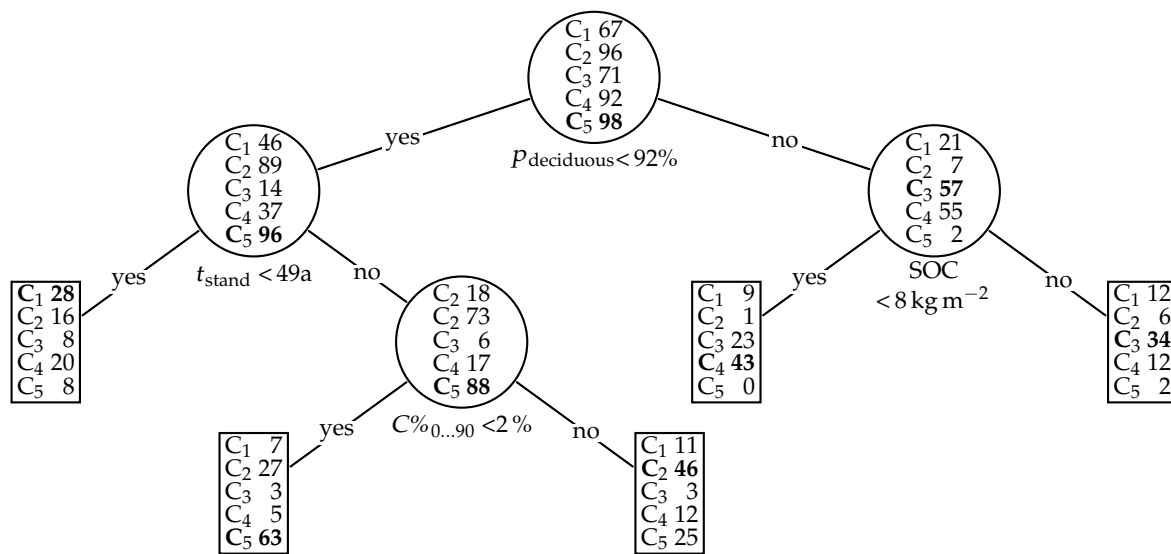


**Figure 8.** Selection of the best pruned subtree by tenfold cross-validation: resubstitution  $R(T_k)$  and cross-validation  $R^{cv}(T_k)$  estimates of risks for the sequence  $k$  of subtrees and auxiliary subtrees derived by minimal cost–complexity pruning according to the complexity parameter  $\alpha_k$ .  $|\tilde{T}_k|$  denotes the number of terminal nodes of the respective subtrees.  $SE$  is the estimate for the standard error of  $R^{cv}(T_k)$ .

The obtained classification tree  $T_{14}$  is shown in Figure 9. In addition to the splitting rules determined, the frequencies of the clusters contained in each node are also given. The frequencies shown for the root node thus equal the number of clusters inside the total sample. A quite distinct split between clusters associated with thick organic layers,  $C_2$  and  $C_5$  and the clusters with high fractions of SOC contained in mineral soil,  $C_3$  and  $C_4$ , can thus be achieved by the proportions of deciduous trees within the forest stand. Actually, in sum just nine sites of the types  $C_2$  and  $C_5$  remained within the right subbranch ( $p_{\text{deciduous}} > 92\%$ ). The clear recognition of depth gradient types associated with mor-humus-like humus forms by the proportions of deciduous trees is in line with the well-known preference of coniferous stands for developing thicker organic layers.

The next and final split on the right is the SOC stored in the soil solum, and it separates cluster  $C_3$  from cluster  $C_4$ . This separation reflects the usually less distinct gradients of SOC across the mineral soil layers in soils rich in organic matter, like gleysols, or soils containing colluvic material.

The left branch, which contains the sites with higher proportions of deciduous trees, is subsequently split by stand age  $t_{\text{stand}}$ . In particular, the ‘mor humus cluster’  $C_5$  is rarely observed in juvenile coniferous stands. Presumably, this must be seen in terms of stand history and the more recent forest plantings. These were often combined with tillage procedures, removing the organic layer, or transferring it to the mineral soil. In the past, clear-cut logging was an established technique widely used throughout the north-eastern lowlands. Especially since the 1960s, the combination of clear-cut logging with intense tillage procedures became a popular method [95]. To facilitate the ploughing, tillage was often preceded by stump removal with bulldozers. Even in cases where no-tillage planting was performed, the necessary clearings stimulate the mineralisation of the organic layer through warming of the forest floor and increased light exposure.



**Figure 9.** Optimally pruned class probability tree for prediction of relative depth gradient types using environmental covariates. The most frequent clusters within each node are shown in bold print.

Finally, the subbranch of more mature coniferous forest stands is further differentiated by clay content within the mineral soil  $C\%_{0...90}$ . At sites that are extremely poor in clay, the highest proportions of SOC are stored inside the organic layer ( $C_5$ ). If at least some clay is present within the soil profile, the cluster  $C_2$  is most frequently found. As described above, the main difference between the two clusters is the less steep gradient of SOC from the organic layer to the mineral soil. The observed shift of SOC from the organic layer to mineral soil with increasing clay content can be seen as a result of higher soil biological activity. In these usually less acidic and dry soils, an increased bioturbation by earthworms can be expected. Furthermore, this may point to the reduced ability of clay-poor mineral soils to stabilise SOC in organic–mineral complexes [86,96,97]. In mineral soils containing at least some clay, clay minerals increase the proportions of the pore space, where carbon is protected against bacteria and enzymes.

In addition to assessing the accuracy of the tree predictions with risks  $R(T_{14})$  and  $R^{cv}(T_{14})$ , based on the node impurities and, hence, expressing the risk for erroneous class probabilities, we also computed the error matrix of Congalton and Green [98]. For building the error matrix, we used the most frequent cluster within the individual terminal nodes of the tree for classifications (Figure 10). As one might expect, the  $p$ -values clearly show that the classifications of each individual cluster ( $p_{K_i}$ ), as well as for the entire error matrix ( $p_{\hat{K}}$ ), are significantly better than random results. Referring to the nomenclature of Landis and Koch [99], the relative strength of agreement associated with the derived values of conditional KHAT statistic  $\hat{K}_i$  can be considered as moderate in the majority of cases (KHAT is an estimate of Kappa). The classification achieved at least fair accordances for any of the clusters. For the overall classification,  $\hat{K}$  also indicates a fair strength of agreement between the reference and classified data. The poorest results in terms of  $\hat{K}_i$  are produced for cluster  $C_1$ . This can be seen as a result of the anthropogenic origin assumed for this depth gradient, making predictions challenging. In total, the hit rate of the classification tree expressed by overall accuracy (OA) slightly exceeds 50%. The depth gradient with the highest individual hit rates (producer’s accuracy (PA) and user’s accuracy (UA)) is the most frequently observed cluster  $C_5$ , which is common in mature pine stands on sandy soils. The ratios of row- ( $n_{i+}$ ) and column-sums ( $n_{+j}$ ) are reasonably balanced, so the predictions are unbiased and no depth gradient is over-predicted.

		Reference					$n_{i+}$	UA	$\hat{K}_i$	$p_{K_i}$
		C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>				
Classification	C <sub>1</sub>	28	16	8	20	8	80	0.350	0.228	0.000
	C <sub>2</sub>	11	46	3	12	25	97	0.474	0.320	0.000
	C <sub>3</sub>	12	6	34	12	2	66	0.515	0.418	0.000
	C <sub>4</sub>	9	1	23	43	0	76	0.566	0.445	0.000
	C <sub>5</sub>	7	27	3	5	63	105	0.600	0.480	0.000
	$n_{+j}$	67	96	71	92	98	424	OA	$\hat{K}$	$p_{\hat{K}}$
PA	0.418	0.479	0.479	0.467	0.643		0.505	0.378	0.000	

**Figure 10.** Cross-tabulation of predicted vs. observed depth gradient types. Accuracy measures of the error matrix are denoted as: row total  $n_{i+}$ , column total  $n_{+j}$ , producer’s accuracy (PA), user’s accuracy (UA), overall accuracy (OA), KHAT statistic  $\hat{K}$ , conditional KHAT statistic  $\hat{K}_i$ , and  $p$ -values  $p_{\hat{K}}$  of respective KHAT statistics.

By combining the tree estimates of the relative depth gradient probabilities with the estimates of the SOC stocks within the entire soil solum derived by regression analysis, the volumetric SOC content of single layers can be calculated. The application of the complete methodology for making predictions for NFSI depth layers then results in the estimates presented in Table 4 and Figure 11. In addition to the total model performance measures, values based on single layers—thus excluding depth as an independent variable—are also given.

**Table 4.** Model performance measures obtained from the calibration sample and repeated tenfold cross-validation: number of observations  $N$ , number of explanatory variables  $J$ , coefficient of determination  $r^2$ , adjusted coefficient of determination  $r^2_{adj}$ , norm of residuals  $s_e$ , standard deviations of observed volumetric SOC contents  $s_{soc}$ , F-statistic  $F_{emp}$ ,  $p$ -value  $p$ , root mean square error  $RMSE_{cv}$ , least squares analysis of slope  $b_{1cv}$ , and concordance correlation coefficient  $CCC_{cv}$ . The variations of cross-validation estimates (doubled standard deviations) within the repeated runs are given in parentheses.

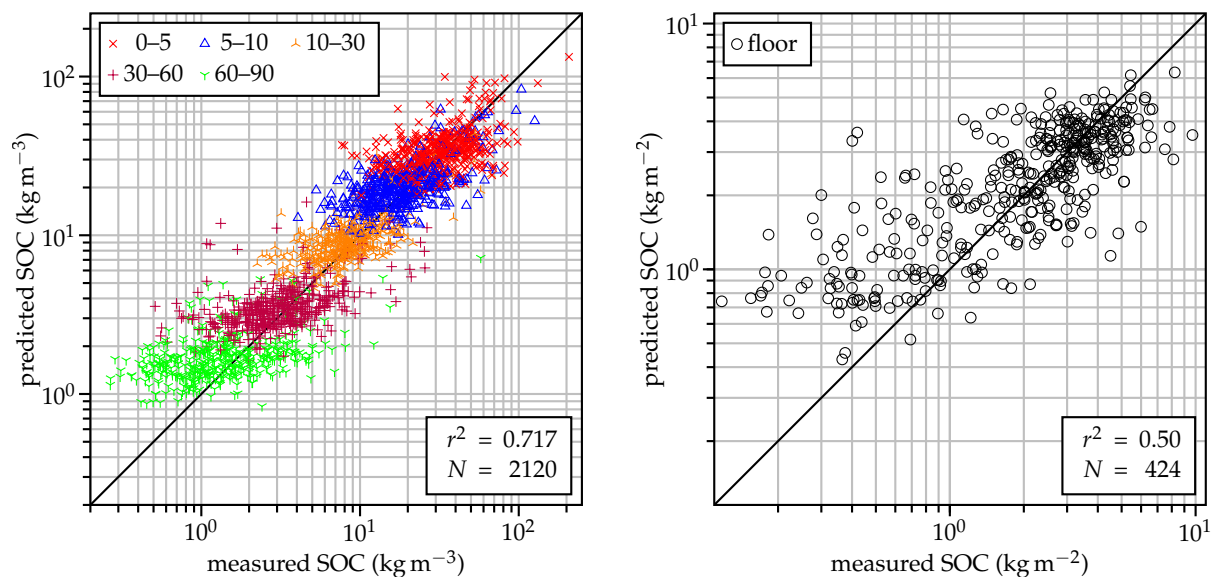
Layer	$N$	$J$	$r^2$	$r^2_{adj}$	$s_e$	$s_{soc}$	$F_{emp}$	$p$	$r^2_{cv}$	$RMSE_{cv}$	$b_{1cv}$	$CCC_{cv}$
floor *	424	18	0.500	0.478	1.2	1.7	22.5	0.00	0.50 (±0.01)	1.21 (±0.01)	0.51 (±0.01)	0.67 (±0.01)
0–5 cm	424	18	0.296	0.265	16.0	18.6	9.5	0.00	0.26 (±0.02)	16.48 (±0.35)	0.37 (±0.02)	0.48 (±0.02)
5–10 cm	424	18	0.487	0.464	9.9	13.8	21.4	0.00	0.46 (±0.02)	10.10 (±0.18)	0.42 (±0.01)	0.61 (±0.01)
10–30 cm	424	18	0.554	0.534	4.9	7.3	27.9	0.00	0.53 (±0.02)	5.05 (±0.08)	0.45 (±0.01)	0.65 (±0.01)
30–60 cm	424	18	0.286	0.254	2.9	3.4	9.0	0.00	0.28 (±0.01)	2.91 (±0.02)	0.26 (±0.01)	0.42 (±0.01)
60–90 cm	424	18	0.229	0.194	2.9	3.2	6.7	0.00	0.18 (±0.03)	2.92 (±0.03)	0.10 (±0.01)	0.19 (±0.02)
0–90 cm	424	18	0.638	0.622	2.6	4.3	39.7	0.00	0.59 (±0.03)	2.76 (±0.08)	0.54 (±0.02)	0.72 (±0.02)
mineral	2120	19	0.719	0.717	12.7	24.0	339.9	0.00	0.70 (±0.01)	9.12 (±0.17)	0.75 (±0.01)	0.83 (±0.01)

\* Values for forest floor  $s_e$ ,  $s_{soc}$ , and  $RMSE_{cv}$  are given in  $kg\ m^{-2}$ ; values off all other layers are given in  $kg\ m^{-3}$ .

The  $p$ -values derived from the F-statistics signal highly significant prediction of SOC in every layer considered. As one might expect, the scatter plot (Figure 11) shows a clear decrease of absolute errors with increasing depth, since SOC generally decreases with increasing soil depth. This decrease of absolute errors and variability of SOC is also confirmed by the norm of residuals and the standard deviations of volumetric SOC. Both measures are highest within the forest floor layer and lowest in the 60–90 cm layer of the NFSI.

In terms of the predictable variation expressed by the coefficients of determination  $r^2$  and  $r^2_{adj}$ , the model performance is highest within the 10–30 cm layer. It then decreases towards the the mineral soil surface, as well as with increasing soil depth. The relatively low coefficients of determination within the upper parts of the mineral soil probably result from steep vertical gradients of SOC. Inside these thin near-surface layers, a high degree of

disturbance can also often be expected. Furthermore, the results within these layers suffer most from sampling artefacts related to segregation of the organic layer from the mineral soil. Close to the bottom of the soil solum, the coefficients of determination are quite low. However, for the majority of practical applications, this may be balanced by the small SOC contents within these layers in total.

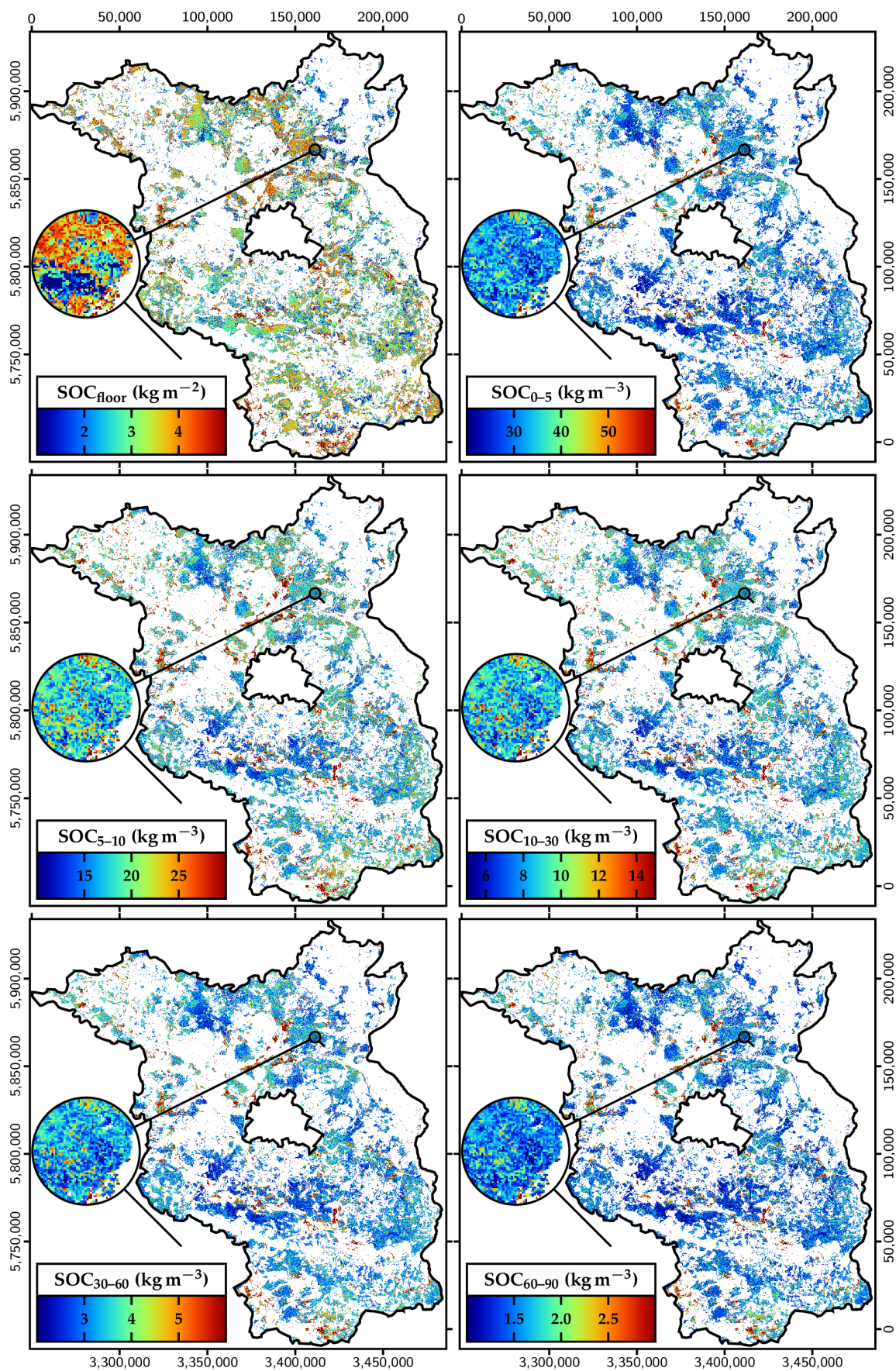


**Figure 11.** Scatter diagrams of measured versus predicted values of volumetric SOC at National Forest Soil Inventory (NFSI) depth layers of the mineral soil (0–5, 5–10, 10–30, 30–60, and 60–90 cm) and SOC stocks of the organic layer.

Within the organic layer, the coefficients of determination are again fairly high. However, especially in the range of small SOC contents, several outliers that overestimate the SOC stored within the organic layer occur (Figure 11). Facing thin forest floor layers, the distinction between the mineral soil and the organic layer during sampling becomes more and more challenging. In extreme cases, very thin organic layers are ‘overlooked’ during sampling and are finally sampled together with the first layer of the mineral soil. Thus, the slightly biased estimation observed at low amounts of SOC within the organic layer must not necessarily be taken as a sign of poor model performance. Due to the generally high predictive power of soil depth on SOC content, the coefficients of determination increase markedly if these are computed using all observations jointly. Model performance also increases with increasing layer thickness. If the SOC content within the total mineral soil (0...90 cm) is predicted, the coefficients of determination are similar to the values calculated for total solum stocks.

All in all, the model performance measures obtained from the calibration sample are well confirmed by the respective cross-validation estimates. In particular, the concordance between  $s_e$  and  $RMSE_{cv}$  is promising for an estimation of the prediction errors that is not overly optimistic. Larger discrepancies between the calibration and cross-validation estimates are limited to the depth layers at 0–5 and 60–90 cm. As indicated by the doubled standard deviations, the results of the repeated runs of cross-validation are also less stable within these depth layers. Since  $CCC_{cv}$ , which comprises an additional assessment of slope, is a much stricter model performance indicator than  $r$  is, the squared values of  $CCC_{cv}$  are again distinctly lower than  $r^2_{cv}$ .

The application of the derived models for mapping SOC stored in the forest soils of Brandenburg is shown in Figure 12. As examples, the geographic distributions of SOC within the six NFSI depth layers used for model development are shown. Nevertheless, whenever required, the usage of the relative depth gradient types enables the user to predict any depth layer of interest.



**Figure 12.** Cartograms displaying the spatial distribution of SOC within the NFSI depth layers obtained by regionalisation using relative depth gradients for a 100 m grid of the forested area of Brandenburg.

The cartograms are based on estimates derived for a grid resolution of 100 m. To prevent the estimates from drifting beyond physical bounds, the range of values of the predictor variables was constrained. Cut-offs were set with respect to the limits that are present in the NFSI data used for model development. This became necessary because, in case of the grid, which contains one million individual grid cells, the sample size was drastically enlarged compared to the NFSI sample. Associated with this is a broadened range of values for the environmental covariates, resulting in extreme combinations of predictors in single grid cells. Similarly, according to the soil types that were present in the training sample, histosols, which were not within the scope of the conducted analysis, were excluded from mapping. To obtain a proper visualisation of the spatial distribution for the majority of sites, the colour map legends of the final cartograms were made to span between the 5th and 95th percentiles.

The cartograms for every depth layer drawn show a high spatial variation. Especially at small spatial scales, a high heterogeneity can be observed. As can be seen within the example landscape extract under the magnifying glass, even at the smallest distances, changes of SOC equivalent to the total range of values appear. In particular, the SOC contents of the three upper mineral soil layers attain magnitudes that may significantly contribute to soil water storage and nutrient status. The SOC contents also vary in a range wide enough to substantially alter the soil's physical and chemical properties, which underlines the practical importance of regionalisation. Pursuant to the classification given by Arbeitsgruppe Boden [74], the equivalent soil organic matter contents within the 0–5 cm layer range from medium (h3) to very high (h5) values. Even within the 10–30 cm layer, the h2 class still occurs frequently. The SOC contents of the two subsoil layers are generally very low (h1), and are thus usually only of minor importance for ecological site characteristics. On the other hand, only small quantities of hydrophobic compounds are required to form water-repellent particle coatings [100]. In particular, under acidic conditions, the SOC within the subsoil layers may still contribute to prolongation of preferential flow paths. However, in terms of carbon sequestration, the amounts of SOC actually stored within these thick subsoil layers are essential reservoirs (see below).

In addition to the predominant small-scale patterns caused by terrain and land cover, some large-scale spatial gradients of mineral soil SOC can also be found. The SOC contents are slightly increased in the climatically wetter regions in the north-west and south of the study area. Amongst the large- and small-scale variations, some spatial clusters also become apparent at medium scale. These areas reflect the effects of parent material, which is usually closely related to the glacial deposits within the north-eastern lowlands. Increased SOC contents occur in the area of the post-glacial floodplains, whereas outwash plains are very poor in SOC.

On the small scale as well as on the large scale, the spatial patterns of SOC stored within the organic layer are highly independent from those observed for any layer of the mineral soil. Even the vertically directly adjacent mineral soil layer at 0–5 cm shows a widely different spatial distribution. At first glance, this is not a surprise, since SOC storage in both compartments underlies distinct mechanisms. Whereas the amounts of SOC within the forest floor layer are highly controlled by short-term influences, such as tree species compositions or forest management and stand history in general, the SOC storage is usually more closely related to long-term factors, like topography or parent material. On the other hand, these findings underline that the depth gradient types approach is flexible enough to cope with these different sources of variation, despite the fact that the horizontal distribution of the total SOC stocks is done using the same regression model.

Table 5 gives a small example of potential utilisations within the context of carbon sequestration: Across all NFSI depth increments, the highest amounts of SOC are stored in the organic layer by far. In total, more than one-third of the solum SOC is stored within the forest floor. However, due to the thickness of the subsoil layers, the deeper NFSI depth increments also substantially contribute to the stock. Although sites influenced by a near-surface groundwater table merely cover 12.5% of the forested area of Brandenburg,

considerable proportions of SOC are stored in these soils. Especially when the stable carbon pools within the mineral soil are considered, the proportion of 20.9% of stored SOC is distinctly larger than the respective area percentage.

**Table 5.** Carbon storage in forest soils of Brandenburg at terrestrial sites (Terr.) and sites influenced by a near-surface groundwater table (Hydro.), as well as its distribution among depth layers.

		Carbon Storage in NFSI Depth Increments ( $10^9$ kg)					Sums ( $10^9$ kg)		
		Floor	0–5 cm	5–10 cm	10–30 cm	30–60 cm	60–90 cm	0–90 cm	Solum
Hydro.	(1302 km <sup>2</sup> )	6.073	3.576	2.087	4.084	2.321	1.075	13.143	19.216
Terr.	(9139 km <sup>2</sup> )	26.373	13.921	7.971	15.197	8.668	4.078	49.834	76.207
Total	(10,441 km <sup>2</sup> )	32.447	17.497	10.058	19.281	10.988	5.153	62.977	95.424

#### 4. Discussion

Since soils and the responsible soil forming factors are unique to the spatial domain under consideration, objective comparisons with studies conducted elsewhere are likely to fail. These difficulties are enhanced by the fact that the available environmental covariates also strongly depend on administrative units. This is especially the case for soil and geological maps. However, despite these limitations, the underlying principles of SOC storage in soils are, of course, transferable to other regions, and thus, some similarities can be observed.

A strong influence of terrain attributes associated with catchment area on SOC was already found by the pioneers in the field of soil attribute prediction through digital terrain analysis [18]. Similar findings have also often been reported in more recent studies. The *TWI* (topographic wetness index) was one of the best predictors for mapping soil organic matter in Croatia [38]. A high regression coefficient of *TWI* was also found for mapping mineral soil SOC within the pre-alpine lowlands [91]. Concerning terrain parameters, the most similar results to those of this study were reported by Grimm et al. [20], who found  $A_C$ ,  $RSP(\dots)$ ,  $LS$  ( $LS$  factor from the Revised Universal Soil Loss Equation), and *TWI* as variables with the highest importances for mapping SOC within the topsoil layer on Barro Colorado Island. Slightly different rankings of variable importance were given by Ballabio [16]: For the prediction of SOC stored in A horizons within the Italian Alps, elevation,  $\hat{S}$ , and  $\Psi_S$  (sky-view factor) had the highest relative influence, whereas *TWI* and  $LS$  were ranked third and fifth. The findings of Zushi [101], that for the prediction of topsoil SOC stocks in Japanese cedar plantations with multiple linear regression analysis only  $\beta$  and  $\Psi_S$  were selected by stepwise variable selection, while *TWI* and *north* (northness derived from local aspect) had no significant influence, seem even more contradictory to the variable importances observed here.

Beyond that, there are also landscapes where topography appears to be behind other soil forming factors in general. In addition to studies that focused on soil properties and land-use effects [45], in some cases, the influence of terrain on mineral soil SOC was low, even where it was explicitly tested [19,21,102]. Increased influences of landcover classes and tree species compositions were generally observed when SOC within topsoil or forest floor layers was predicted [21,91,102]. This is in good agreement with the superior importance of  $p_{\text{deciduous}}$  in the CART model for the prediction of relative depth gradients. In studies where data on the depth of the groundwater table were available, it was naturally one of the main factors that controlled SOC at the respective sites [45]. The fairly high influence of  $z_{\text{CaCO}_3}$  in our regression models must be interpreted in relation to the often-observed influences of mapping units [19,21,102], since this covariate seems to be a property unique to local forest site mapping.

The achieved model performance is in a range comparable to that of other studies dealing with the prediction of SOC for numerous depth layers [32,36,37]. With low values

near the upper and lower boundaries of the soil solum, the coefficients of determination also often exhibit similar depth dependences [20,33].

The reported improvement of predictions by kriging the residuals varies widely throughout the literature. As the improvement of regression estimates by subsequent geostatistical interpolation is the underlying idea of regression kriging, corresponding improvements are usually found [32,38,103]. However, there were also studies that found no substantial improvements [91,104]. Similarly to our results, these studies were conducted in highly structured landscapes with relatively low sampling densities. As already stated by McKenzie and Ryan [105], in these situations, the weak spatial dependence diminishes the advantages of kriging. With this in mind, it is not surprising that even the advantage of regression may disappear if the landscape of the study area is homogeneous and the sampling density is high [106,107]. In these situations, pure spatial interpolation methods become superior to regression (kriging). In addition to landscape properties and sampling densities, these relations are, of course, strongly affected by the availability of suitable environmental covariates at a proper spatial scale.

## 5. Conclusions

The concept of relative depth gradients offers a straightforward framework to ensure consistency of depth gradients and model interpretability, whereas other methodologies, such as SCORPAN modelling of depth function parameters, require further constraints.

We suppose that our amalgam of soil depth functions and digital soil mapping can be a handy methodology whenever coherent depth gradients are of special importance for the subsequent analyses. It may also offer a useful methodological framework if the interpretability of covariates that control the horizontal and vertical distribution of SOC is of particular interest. Since SOC stocks were estimated as a whole and an acceptable model performance for single layers was still achieved, the usage of relative depth gradients may help to lessen the propagation of sample separation errors. Therefore, the concept of the relative depth gradients can be beneficial if the regionalisation of SOC serves as a basis for hydrological or geochemical modelling.

Further interesting fields of application may be found in scenario-based modelling of the land-use and climate effects on SOC storage and distribution.

**Author Contributions:** Conceptualisation, W.R. and G.W.; Methodology, A.R., W.R., and G.W.; Validation, A.R., W.R., and G.W.; Formal analysis, A.R.; Investigation, A.R. and W.R.; Resources, W.R. and G.W.; Data curation, A.R. and W.R.; Writing—original draft preparation, A.R.; Writing—review and editing, W.R. and G.W.; Visualisation, A.R.; Supervision, W.R. and G.W.; Project administration, W.R. and G.W.; Funding acquisition, W.R. and G.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was partially supported through a doctoral scholarship from the Faculty of Forest and Environment of Eberswalde University for Sustainable Development ‘Arbeitsgruppe Forschung’.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We acknowledge Dieter Kühn and Albrecht Bauriegel in the Department of Soil Geology at the State Agency for Mining, Geology, and Resources in Brandenburg (LBGR) for providing the digital elevation data ‘DGM 25-3-LBGR’ used in this study.

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

## References

1. Renger, M. Berechnung der Austauschkapazität der organischen und anorganischen Anteile der Böden. *Z. Pflanzenernähr. Düng. Bodenkd.* **1965**, *110*, 10–26. [[CrossRef](#)]
2. Curtin, D.; Rostad, H. Cation exchange and buffer potential of Saskatchewan soils estimated from texture, organic matter and pH. *Can. J. Soil Sci.* **1997**, *77*, 621–626. [[CrossRef](#)]
3. 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](#)]



4. Kaur, R.; Kumar, S.; Gurung, H.P. A pedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. *Aust. J. Soil Res.* **2002**, *40*, 847–858. doi:10.1071/SR01023. [[CrossRef](#)]
5. De Vos, B.; Van Meirvenne, M.; Quataert, P.; Deckers, J.; Muys, B. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Sci. Soc. Am. J.* **2005**, *69*, 500–510. [[CrossRef](#)]
6. Hollis, J.M.; Hannam, J.; Bellamy, P.H. Empirically-derived pedotransfer functions for predicting bulk density in European soils. *Eur. J. Soil Sci.* **2012**, *63*, 96–109. [[CrossRef](#)]
7. Beke, G.J.; MacCormick, M.J. Predicting volumetric water retentions for subsoil materials from Colchester County, Nova Scotia. *Can. J. Soil Sci.* **1985**, *65*, 233–236. [[CrossRef](#)]
8. Vereecken, H.; Maes, J.; Feyen, J.; Darius, P. Estimating the soil moisture retention characteristic from texture, bulk density and carbon content. *Soil Sci.* **1989**, *148*, 389–403. [[CrossRef](#)]
9. Zacharias, S.; Wessolek, G. Excluding organic matter content from pedotransfer predictors of soil water retention. *Soil Sci. Soc. Am. J.* **2007**, *71*, 43–50. [[CrossRef](#)]
10. Rawls, W.J.; Pachepsky, Y.A.; Ritchie, J.C.; Sobecki, T.M.; Bloodworth, H. Effect of soil organic carbon on soil water retention. *Geoderma* **2003**, *116*, 61–76. [[CrossRef](#)]
11. Bisdom, E.; Dekker, L.; Schouite, J. Water repellency of sieve fractions from sandy soils and relationships with organic material and soil structure. *Geoderma* **1993**, *56*, 105–118. [[CrossRef](#)]
12. Buczko, U.; Bens, O.; Fischer, H.; Hüttl, R.F. Water repellency in sandy luvisols under different forest transformation stages in northeast Germany. *Geoderma* **2002**, *109*, 1–18. [[CrossRef](#)]
13. Greiffenhagen, A.; Wessolek, G.; Facklam, M.; Renger, M.; Stoffregen, H. Hydraulic functions and water repellency of forest floor horizons on sandy soils. *Geoderma* **2006**, *132*, 182–195. [[CrossRef](#)]
14. Wessolek, G.; Schwarzel, K.; Greiffenhagen, A.; Stoffregen, H. Percolation characteristics of a water-repellent sandy forest soil. *Eur. J. Soil Sci.* **2008**, *59*, 14–23. [[CrossRef](#)]
15. Schonsky, H.; Peters, A.; Wessolek, G. Effect of soil water repellency on energy partitioning between soil and atmosphere: A conceptual approach. *Pedosphere* **2014**, *24*, 498–507. [[CrossRef](#)]
16. Ballabio, C. Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma* **2009**, *151*, 338–350. [[CrossRef](#)]
17. Malone, B.; McBratney, A.; Minasny, B. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* **2011**, *160*, 614–626. [[CrossRef](#)]
18. Moore, I.D.; Gessler, P.E.; Nielsen, G.A.; Peterson, G.A. Soil attribute prediction using terrain analysis. *Soil Sci. Soc. Am. J.* **1993**, *57*, 443–452. [[CrossRef](#)]
19. Bui, E.; Henderson, B.; Viergever, K. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Glob. Biogeochem. Cycles* **2009**, *23*, 1–15. [[CrossRef](#)]
20. Grimm, R.; Behrens, T.; Maerker, 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](#)]
21. Collard, F.; Saby, N.; Richer de Forges, A.; Lehmann, S.; Paroissien, J.B.; Arrouays, D. Spatial prediction of soil organic carbon at different depths using digital soil mapping. In *GlobalSoilMap—Basis of the Global Spatial Soil Information System*; Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., McBratney, A.B., Eds.; CRC Press; Taylor & Francis Group: Boca Raton, FL, USA, 2014; pp. 181–184. [[CrossRef](#)]
22. McBratney, A.; Mendonça-Santos, M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
23. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*, 1st ed.; McGraw-Hill: New York, NY, USA, 1941; p. 281.
24. Keskin, H.; Grunwald, S. Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma* **2018**, *326*, 22–41. [[CrossRef](#)]
25. Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning for digital soil mapping. *Soil* **2019**, *5*, 79–89. [[CrossRef](#)]
26. Roudier, P.; Burge, O.R.; Richardson, S.J.; McCarthy, J.K.; Grealish, G.J. National scale 3D mapping of soil pH using a data augmentation approach. *Remote Sens.* **2020**, *12*, 2872. [[CrossRef](#)]
27. Orton, T.; Pringle, M.; Bishop, T.; Menzies, N.; Dang, Y. Increment-averaged kriging for 3-D modelling and mapping soil properties: Combining machine learning and geostatistical methods. *Geoderma* **2020**, *361*, 114094. [[CrossRef](#)]
28. Lark, R.M.; Bishop, T.F.A. Cokriging particle size fractions of the soil. *Eur. J. Soil Sci.* **2007**, *58*, 763–774. [[CrossRef](#)]
29. Vanguelova, E.I.; Bonifacio, E.; De Vos, B.; Hoosbeek, M.R.; Berger, T.W.; Vesterdal, L.; Armolaitis, K.; Celi, L.; Dinca, L.; Kjønaas, O.J.; et al. Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales—Review and recommendations. *Environ. Monit. Assess.* **2016**, *188*, 630. [[CrossRef](#)] [[PubMed](#)]
30. Minasny, B.; Stockmann, U.; Hartemink, A.E.; McBratney, A.B. *Digital Soil Morphometrics*; Progress in Soil Science; Springer: Cham, Switzerland, 2016; Chapter Measuring and Modelling Soil Depth Functions, pp. 225–240. [[CrossRef](#)]
31. Bishop, T.; McBratney, A.; Laslett, G. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **1999**, *91*, 27–45. [[CrossRef](#)]
32. 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](#)]
33. Dorji, T.; Odeh, I.O.; Field, D.J.; Baillie, I.C. Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. *For. Ecol. Manag.* **2014**, *318*, 91–102. [[CrossRef](#)]

34. Odgers, N.P.; Libohova, Z.; Thompson, J.A. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma* **2012**, *189*, 153–163. [[CrossRef](#)]
35. 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](#)]
36. Mishra, U.; Lal, R.; Slater, B.; Calhoun, F.; Liu, D.; Van Meirvenne, M. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. *Soil Sci. Soc. Am. J.* **2009**, *73*, 614–621. [[CrossRef](#)]
37. Kempen, B.; Brus, D.J.; Stoorvogel, J.J. Three-dimensional mapping of soil organic matter content using soil type-specific depth functions. *Geoderma* **2011**, *162*, 107–123. [[CrossRef](#)]
38. Hengl, T.; Heuvelink, G.B.M.; Rossiter, D.G. About regression-kriging: From equations to case studies. *Comput. Geosci.* **2007**, *33*, 1301–1315. [[CrossRef](#)]
39. Wellbrock, N.; Ahrends, B.; Bögelein, R.; Bolte, A.; Eickenscheidt, N.; Grüneberg, E.; König, N.; Schmitz, A.; Fleck, S.; Ziche, D. *Status and Dynamics of Forests in Germany—Results of the National Forest Monitoring*; Springer: Cham, Switzerland, 2019; Chapter Concept and Methodology of the National Forest Soil Inventory, pp. 1–28. [[CrossRef](#)]
40. Russ, A. *Regionalisierung des Wasser—und Kohlenstoffhaushalts für die Waldfläche Brandenburgs—Entwicklung von Regionalisierungsansätzen unter besonderer Berücksichtigung der forstlichen Standortskarte*; Bodenökologie und Bodengenese, Selbstverlag Technische Universität: Berlin, Germany, 2015; Volume 44, p. 164. [[CrossRef](#)]
41. GAFA. *Handbuch Forstliche Analytik—Eine Loseblatt-Sammlung der Analysemethoden im Forstbereich*; Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft: Bonn, Germany, 2014.
42. Kopp, D.; Jäger, K.D.; Succow, M. *Naturräumliche Grundlagen der Landnutzung*; Akademie-Verlag: Berlin, Germany, 1982; p. 339.
43. Kopp, D.; Schwanecke, W. *Standörtlich-naturräumliche Grundlagen ökologischer Forstwirtschaft*; Deutscher Landwirtschaftsverlag: Berlin, Germany, 1994; p. 248.
44. Russ, A.; Riek, W.; Wessolek, G. Entwicklung statistischer Vorhersagemodelle zur Identifizierung von Legendeneinheiten forstlicher Standortskarten—Beispiel Nordostdeutsches Tiefland. In *Gefährdungen der Ökosystemdienstleistungen von Wäldern. Fachtagung vom 9. bis 10. Oktober 2014 in Göttingen*; Beiträge aus der NW-FVA; Universitätsverlag Göttingen: Göttingen, Germany, 2016; Volume 14, pp. 133–146. [[CrossRef](#)]
45. Meersmans, J.; De Ridder, F.; Canters, F.; De Baets, S.; Van Molle, M. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma* **2008**, *143*, 1–13. [[CrossRef](#)]
46. Hannappel, S.; Riek, W. Berechnung des Flurabstandes des oberflächennahen Grundwassers der Waldfläche Brandenburgs für hydrologisch typische Zeiträume. *Hydrol. Und Wasserwirtsch.* **2011**, *55*, 4–15.
47. Riek, W.; Kallweit, R.; Russ, A. Analyse der Hauptkomponenten des Wärmehaushalts brandenburgischer Wälder auf der Grundlage von regionalen Klimaszenarien. *Wald. Landschaftsforschung Und Naturschutz* **2013**, *13*, 17–32.
48. Ramann, E. *Bodenkunde*, 3rd ed.; Julius Springer: Berlin, Germany, 1911; p. 619.
49. Keil, M.; Bock, M.; Esch, T.; Metz, A.; Nieland, S.; Pfitzner, A. *CORINE Land Cover Aktualisierung 2006 für Deutschland*; Deutsches Zentrum für Luft- und Raumfahrt: Oberpfaffenhofen, Germany, 2010; p. 55.
50. Zevenbergen, L.W.; Thorne, C.R. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* **1987**, *12*, 47–56. [[CrossRef](#)]
51. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* **1999**, *143*, 107–122. [[CrossRef](#)]
52. Seibert, J.; Stendahl, J.; Sørensen, R. Topographical influences on soil properties in boreal forests. *Geoderma* **2007**, *141*, 139–148. [[CrossRef](#)]
53. Yokoyama, R.; Michio, S.; Pike, R.J. Visualizing topography by openness: A new application of image processing to digital elevation models. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 257–265.
54. Böhner, J.; AntoniĆ, O. *Land-Surface Parameters Specific to Topo—Climatology*; Geomorphometry—Concepts, Software, Applications; Developments in Soil Science; Hengl, T., Reuter, H.I., Eds.; Elsevier: Amsterdam, The Netherlands, 2009; Volume 33, pp. 195–226. [[CrossRef](#)]
55. Dozier, F.; Frew, J. Rapid calculation of terrain parameters for radiation modeling from digital elevation data. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 963–969. [[CrossRef](#)]
56. Oke, T.R. *Boundary Layer Climates*, 2nd ed.; Methuen: London, UK; New York, NY, USA, 1987; p. 435.
57. Weiss, A.D. Topographic position and landforms analysis. In Proceedings of the ESRI International User Conference, San Diego, CA, USA, 26–30 June 2000.
58. Riley, J.S.; DeGloria, S.D.; Elliot, R. A Terrain Ruggedness Index that quantifies topographic heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
59. Seibert, J.; McGlynn, B.L. A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resour. Res.* **2007**, *43*, W04501. [[CrossRef](#)]
60. Renard, K.; Foster, G.; Wessies, G.; McCool, D.; Yoder, D. Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE). In *Agriculture Handbook*; United States Department of Agriculture: Washington, DC, USA, 1996; Volume 703.
61. Beven, K.J.; Kirkby, M.J. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]

62. Moore, I.D.; Grayson, R.B.; Ladson, A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* **1991**, *5*, 3–30. [[CrossRef](#)]
63. O'Callaghan, J.F.; Mark, D.M. The extraction of drainage networks from digital elevation data. *Comput. Vision Graph. Image Process.* **1984**, *28*, 323–344. [[CrossRef](#)]
64. Muckell, J.; Andrade, M.; Franklin, W.R.; Cutler, B.; Inanc, M.; Xie, Z.; Tracy, D.M. Drainage network and watershed reconstruction on simplified terrain. In *17th Fall Workshop on Computational Geometry*; IBM T.J. Watson Research Center: Hawthorne, NY, USA, 2007.
65. Zirlewagen, D. *Regionalisierung Bodenchemischer Eigenschaften in Topographisch Stark Gegliederten Waldlandschaften*; Freiburger Forstliche Forschung Volume 19; Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg: Freiburg im Breisgau, Germany 2003; p. 154.
66. Böhner, J.; Köthe, R. Bodenregionalisierung und Prozessmodellierung: Instrumente für den Bodenschutz. *Petermanns Geogr. Mitteilungen* **2003**, *147*, 72–82.
67. Möller, M.; Volk, M.; Friedrich, K.; Lymburner, L. Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. *J. Plant Nutr. Soil Sci.* **2008**, *171*, 419–430. [[CrossRef](#)]
68. Grimm, R.; Behrens, T. Uncertainty analysis of sample locations within digital soil mapping approaches. *Geoderma* **2010**, *155*, 154–163. [[CrossRef](#)]
69. Rubinfeld, D.L. *Reference Manual on Scientific Evidence*; The National Academies Press: Washington, DC, USA, 2011; Chapter Reference Guide on Multiple Regression, pp. 303–357. [[CrossRef](#)]
70. Browne, M.W. Cross-Validation Methods. *J. Math. Psychol.* **2000**, *44*, 108–132. [[CrossRef](#)]
71. Burman, P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika* **1989**, *76*, 503–514. [[CrossRef](#)]
72. Lin, L.I.K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1989**, *45*, 255–268. [[CrossRef](#)] [[PubMed](#)]
73. Zirlewagen, D.; von Wilpert, K. Regionalization of soil physical input variables for soil hydraulic pedotransfer functions. *For. Ecol. Landsc. Res. Nat. Conserv.* **2011**, *12*, 73–83.
74. Arbeitsgruppe Boden. *Bodenkundliche Kartieranleitung*, 5th ed.; Schweizerbart'sche Verlagsbuchhandlung: Stuttgart, Germany, 2005; p. 437.
75. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; Statistics in Practice; John Wiley & Sons, Ltd.: Chichester, UK, 2007; p. 315.
76. Ward, J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [[CrossRef](#)]
77. Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 13th ed.; Springer: Berlin, Germany, 2011; p. 583.
78. Mojena, R. Hierarchical grouping methods and stopping rules: An evaluation. *Comput. J.* **1977**, *20*, 359–363. [[CrossRef](#)]
79. Blashfield, R.K.; Aldenderfer, M.S. Computer programs for performing iterative partitioning cluster analysis. *Appl. Psychol. Meas.* **1978**, *2*, 533–541. [[CrossRef](#)]
80. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; The Wadsworth Statistics/Probability Series; Wadsworth: Belmont, CA, USA, 1984; p. 358.
81. Kass, G.V. An exploratory technique for investigating large quantities of categorical data. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1980**, *29*, 119–127. [[CrossRef](#)]
82. Gini, C. Sulla Misura Della Concentrazione e Della Variabilità dei Caratteri. In *Variabilità e Concentrazione*; Memorie di Metodologia Statistica; Gini, C., Ed.; Dott. A. Giuffrè: Milano, Italy, 1939; Volume I, pp. 359–408. Reprint of: Atti del Istituto Veneto di Scienze, Lettere ed Arti, 1913–1914. Tomo LXXIII, Parte, II, page 1202–1248.
83. Light, R.J.; Margolin, B.H. An analysis of variance for categorical data. *J. Am. Stat. Assoc.* **1971**, *66*, 534–544. [[CrossRef](#)]
84. Schaetzl, R.J.; Mikesell, L.R.; Velbel, M.A. Soil characteristics related to weathering and pedogenesis across a geomorphic surface of uniform age in Michigan. *Phys. Geogr.* **2006**, *27*, 170–188. [[CrossRef](#)]
85. Schaetzl, R.J.; Luehmann, M.D.; Rothstein, D. Pulses of podzolization: The relative importance of spring snowmelt, summer storms, and fall rains on Spodosol development. *Soil Sci. Soc. Am. J.* **2015**, *79*, 117–131. [[CrossRef](#)]
86. Oades, J.M. The retention of organic matter in soils. *Biogeochemistry* **1988**, *5*, 35–70. [[CrossRef](#)]
87. Lloyd, J.; Taylor, J.A. On the temperature dependence of soil respiration. *Funct. Ecol.* **1994**, *8*, 315–323. [[CrossRef](#)]
88. Curtin, D.; Beare, M.H.; Hernandez-Ramirez, G. Temperature and moisture effects on microbial biomass and soil organic matter mineralization. *Soil Sci. Soc. Am. J.* **2012**, *76*, 2055–2067. [[CrossRef](#)]
89. Johnson, D.; Walker, R.; Glass, D.; Miller, W.; Murphy, J.; Stein, C. The effect of rock content on nutrients in a Sierra Nevada forest soil. *Geoderma* **2012**, *173–174*, 84–93. [[CrossRef](#)]
90. Nielsen, D.R.; Wendroth, O. *Spatial and Temporal Statistics: Sampling Field Soils and Their Vegetation*; GeoEcology Textbook; Catena Verlag: Reiskirchen, Germany, 2003; p. 398.
91. Zirlewagen, D.; von Wilpert, K. Upscaling of environmental information: Support of land-use management decisions by spatio-temporal regionalization approaches. *Environ. Manag.* **2010**, *46*, 878–893. [[CrossRef](#)]
92. Lippstreu, L.; Brose, F.; Marcinek, J. Brandenburg. In *Das Quartär Deutschlands*; Benda, L., Ed.; Gebrüder Bornträger: Berlin, Germany, 1995; pp. 116–147.
93. LGRB. *Geologische Übersichtskarte des Landes Brandenburg 1:300,000*; Landesamt für Geowissenschaften und Rohstoff Brandenburg in Zusammenarbeit mit Landesvermessungsamt Brandenburg. Landesvermessungsamt Brandenburg: Potsdam, Germany, 1997.

94. Marquardt, D.W. An algorithm for least-squares estimation of on nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [[CrossRef](#)]
95. Milnik, A.; Heyde, M.; Schult, R.; Bormeister, H.J.; Helbig, K.; Köhler, J.; Römpler, W.; Schwartz, E.; Wagenknecht, E.; Weber, T. *Verantwortung für den Wald—Die Geschichte der Forstwirtschaft in der Sowjetischen Besatzungszone und der DDR*; Brandenburgisches Ministerium für Ernährung, Landwirtschaft und Forsten: Potsdam, Germany, 1998.
96. Manjajiah, K.M.; Kumar, S.; Sachdev, M.S.; Sachdev, P.; Datta, S.C. Study of clay–organic complexes. *Curr. Sci.* **2010**, *98*, 915–921.
97. Schöning, I.; Grüneberg, E.; Sierra, C.A.; Hessenmöller, D.; Schrumpf, M.; Weisser, W.W.; Schulze, E.D. Causes of variation in mineral soil C content and turnover in differently managed beech dominated forests. *Plant Soil* **2013**, *370*, 625–639. [[CrossRef](#)]
98. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data—Principles and Practices*, 2nd ed.; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2009; p. 183.
99. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]
100. Wang, X.Y.; Zhao, Y.; Horn, R. Soil wettability as affected by soil characteristics and land use. *Pedosphere* **2010**, *20*, 43–54. [[CrossRef](#)]
101. Zushi, K. Spatial distribution of soil carbon and nitrogen storage and forest productivity in a watershed planted to Japanese cedar (*Cryptomeria japonica* D. Don). *J. For. Res.* **2006**, *11*, 351–358. [[CrossRef](#)]
102. Baritz, R.; Zirlwagen, D.; Van Ranst, E. *Multi-Source Inventory Methods for Quantifying Carbon Stocks and Stock Changes in European Forests—CarboInvent—Final Report to the EC: Methodological Standards to Detect Forest Soil Carbon Stocks and Stock Changes Related to Land Use Change and Forestry: Part II—Landscape Level*; Joanneum Research: Graz, Austria, 2006.
103. Bonfatti, B.R.; Hartemink, A.E.; Giasson, E.; Tornquist, C.G.; Adhikari, K. Digital mapping of soil carbon in a viticultural region of Southern Brazil. *Geoderma* **2016**, *261*, 204–221. [[CrossRef](#)]
104. Nussbaum, M.; Papritz, A.; Baltensweiler, A.; Walthert, L. Estimating soil organic carbon stocks of Swiss forest soils by robust external-drift kriging. *Geosci. Model Dev.* **2014**, *7*, 1197–1210. [[CrossRef](#)]
105. McKenzie, N.J.; Ryan, P.J. Spatial prediction of soil properties using environmental correlation. *Geoderma* **1999**, *89*, 67–94. [[CrossRef](#)]
106. Hoffmann, U.; Hoffmann, T.; Jurasinski, G.; Glatzel, S.; Kuhn, N. Assessing the spatial variability of soil organic carbon stocks in an alpine setting (Grindelwald, Swiss Alps). *Geoderma* **2014**, *232–234*, 270–283. [[CrossRef](#)]
107. 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](#)]