

Article

Regionalization of Uncovered Agricultural Soils Based on Organic Carbon and Soil Texture Estimations

Martin Kanning ^{*,†}, Bastian Siegmann [†] and Thomas Jarmer [†]

Institute of Computer Science, Osnabrück University, Wachsbleiche 27, D-49090 Osnabrück, Germany; bsiegmann@igf.uni-osnabrueck.de (B.S.); tjarmer@igf.uni-osnabrueck.de (T.J.)

* Correspondence: makanning@uni-osnabrueck.de; Tel.: +49-541-969-3181

† These authors contributed equally to this work.

Academic Editors: Nicolas Baghdadi, Clement Atzberger and Prasad S. Thenkabail

Received: 13 July 2016; Accepted: 3 November 2016; Published: 8 November 2016

Abstract: The determination of soil texture and organic carbon across agricultural areas provides important information to derive soil condition. Precise digital soil maps can help to till agricultural fields with more accuracy, greater cost-efficiency and better environmental protection. In the present study, the laboratory analysis of sand, silt, clay and soil organic carbon (SOC) content was combined with hyperspectral image data to estimate the distribution of soil texture and SOC across an agricultural area. The aim was to identify regions with similar soil properties and derive uniform soil regions based on this information. Soil parameter data and corresponding laboratory spectra were used to calibrate cross-validated (leave-one-out) partial least squares regression (PLSR) models, resulting in robust models for sand ($R^2 = 0.77$, root-mean-square error (RMSE) = 5.37) and SOC ($R^2 = 0.89$, RMSE = 0.27), as well as moderate models for silt ($R^2 = 0.62$, RMSE = 5.46) and clay ($R^2 = 0.53$, RMSE = 2.39). The regression models were applied to Airborne Imaging Spectrometer for Applications DUAL (aisADUAL) hyperspectral image data to spatially estimate the concentration of these parameters. Afterwards, a decision tree, based on the Food and Agriculture Organization (FAO) soil texture classification scheme, was developed to determine the soil texture for each pixel of the hyperspectral airborne data. These soil texture regions were further refined with the spatial SOC estimations. The developed method is useful to identify spatial regions with similar soil properties, which can provide a vital information source for an adapted treatment of agricultural fields in terms of the necessary amount of fertilizers or water. The approach can also be adapted to wider regions with a larger sample size to create detailed digital soil maps (DSMs). Further, the presented method should be applied to future hyperspectral satellite missions like Environmental Mapping and Analysis Program (EnMap) and Hyperspectral Infrared Imager (HyspIRI) to cover larger areas in shorter time intervals. Updated DSMs on a regular basis could particularly support precision farming aspects.

Keywords: hyperspectral; stratification; sand; silt; clay; soil organic carbon (SOC); partial least squares regression (PLSR); decision-tree; digital soil maps (DSMs)

1. Introduction

Many human interests, such as food production and water quality, are fundamentally connected to soils. Numerous competitive interests, however, endanger soil conditions worldwide. To prevent soils from degradation, reasonable and sustainable soil management on both a global and local scale is mandatory [1]. Precision agriculture can essentially support these challenges by improving nutrient efficiency and productivity, thus reducing environmental damages [2]. Current agricultural machinery can provide farmers with additional information about soils and plants. This information can be

acquired from a variety of sensors and embedded information systems to enable a more precise, cost-efficient and environmentally protective tillage of agricultural fields [3]. Digital soil maps (DSMs), showing the spatial distribution of important soil attributes, can support decision-making that highly depends on up-to-date information about soil conditions at specific locations [4]. These DSMs can be produced using a variety of data sources based on different approaches. In [5], a Kriging regression was used to create soil organic carbon (SOC) maps from several single soil samples with detailed information by observing their spatial distribution. Grimm et al. [6] spatially estimated SOC by including topography, existing DSMs and geology information as predictors in a random forests regression analysis. A similar approach was developed by de Brogniez et al. [7], who combined data from the Land use/Cover Area frame statistical Survey (LUCAS) with additional environmental information to predict SOC on European scale. Another study from Villa et al. [8] applied spectral indices to Landsat TM and simulated low resolution imagery to predict SOC in large peri-urban areas. However, these maps cover large areas but provide a low spatial resolution, which results in a lack of detail, especially for small-scale applications where higher resolution of soil information is very useful. Peralta et al. [9] derived more detailed regionalized DSMs for precision agriculture by measuring apparent soil electrical conductivity for entire fields, among other parameters, to successfully predict clay, soil organic matter, cation exchange capacity and soil gravimetric water content. Remote sensing data provide this information, thereby enabling the spatial assessment of soil parameters with sufficient spatial resolution [10]. On the contrary, soil maps for precision agricultural applications are not necessarily dependent on pixel-wise numerical information. More generalized regionalized maps, derived from accurate predictions of key soil attributes, should be sufficient.

Soil texture, defined by the composition of the three particle size groups of sand, silt and clay, is one of the most important soil parameters. Since physical properties, such as consistency, rootability and storage capacity of water and air, correlate with soil texture, it is a key parameter to describe soil characteristics and plant growth conditions [11].

In recent years, several studies have been conducted based on a variety of concepts to estimate each of these compartments in relation to spectroscopy data [12]. Since the three particle size groups are closely connected, most of these studies determined each of them by using spectral data acquired under laboratory conditions [13–17]. Nanni and Demattê [13] used a multiple regression approach, which was applied to hyperspectral data, and achieved high model qualities for clay ($R^2 = 0.91$) and medium-high model qualities for sand ($R^2 = 0.79$). On the contrary, the silt model provided a poor quality ($R^2 = 0.27$). Demattê et al. [14] applied the same regression method and obtained comparable results for sand and clay, but a better model for silt ($R^2 = 0.57$). These authors do not provide any explanation for the poor performance of silt, but conclude that two well-modeled grain size fractions are sufficient, since the third can always be calculated by subtraction. Other results from Casa et al. [16], using similar hyperspectral laboratory data and partial least squares regression (PLSR), show a slightly better modeling performance for silt ($R^2 = 0.63$), but a weaker one for clay ($R^2 = 0.58$), while the model for sand leads to a similar result ($R^2 = 0.79$). Chang et al. [17] also achieved comparable results, working with principal component regression applied to hyperspectral laboratory data. In this study, the model for silt turned out to be the strongest ($R^2 = 0.84$). In summary, regressions with sand constantly performed well in all studies, while silt and clay modeling is less reliable under laboratory conditions. Nevertheless, spectral data are to a certain degree correlated with soil texture, such that spectroscopy is an important tool in soil science [10].

Soil texture assessment from image data acquired by airborne and spaceborne systems is a more difficult issue, mainly due to atmospheric distortions and lower spatial and spectral resolution of the sensors [10]. Hyperspectral airborne images acquired by HyMap were analyzed with contrary results regarding their ability to derive soil texture fractions with PLSR by [18,19]. Although Selige et al. [18] achieved an R^2 of 0.95 for sand based on data with a 6 m spatial resolution and 128 spectral bands, the same parameter only resulted in an R^2 of 0.20 using data with a 5 m spatial resolution and 124 spectral bands in a study conducted by Gomez et al. [19]. Clay again led to almost similar results

($R^2 = 0.71$ and 0.67) in both studies. Another study conducted by Hively et al. [20], where PLSR was also used, provided decent model accuracies for all three compartments. The hyperspectral image data were acquired with a 2.5 m ground sampling distance (GSD) in 178 spectral bands. Dematté et al. [14] applied their laboratory data approach to multispectral Landsat 5 images with a 30 m spatial resolution and six spectral bands, resulting in a drastic performance drop for sand ($R^2 = 0.64$) and clay ($R^2 = 0.61$) compared to the models built under laboratory conditions. Due to the aforementioned constraints, soil texture modeling from image spectra performs worse compared to laboratory spectra, but still shows high potential for further investigations.

Another important parameter in order to characterize soils is the SOC content. It has a major influence on soil quality [21]. Soils with high amounts of SOC are able to store nutrients [22] and water more efficiently [23] compared to those with less SOC. For sustainable agriculture and environmental protection, detailed information about SOC concentration is very important and has been extensively investigated using remote sensing techniques in recent years [23].

Under laboratory conditions, Patzold et al. [24] were able to predict SOC with PLSR and achieved an R^2 value of 0.93. They concluded that SOC estimations using laboratory spectroscopy lead to promising results. McCarty et al. [25] supported this assumption by presenting an R^2 value of 0.98 only based on NIR spectra. Viscarra Rossel et al. [26] were unable to achieve such a good result ($R^2 = 0.72$) with the same regression method when using fewer samples and a lower observed data range. Similar results were also presented by Jarmer et al. [27], who performed PLSR analyses with ASD FieldSpec-II spectra and obtained an R^2 value of 0.62.

Stevens et al. [28] used PLSR to predict SOC from hyperspectral airborne images acquired for a large area with the AHS sensor. Due to an extensive sampling strategy, the independently validated modeling ($R^2 = 0.75$) resulted in a reliable spatial estimation of SOC. Comparable results were achieved by Selige et al. [16] and Schwanghart et al. [29] using HyMap hyperspectral images. Moreover, Gomez et al. [30] derived SOC from Hyperion hyperspectral data. They found that data with a medium spatial resolution (30 m) only allow for spatial SOC predictions with moderate accuracy ($R^2 = 0.51$). Landsat multispectral imagery was proved to be suitable for large-scale SOC estimations ($R^2 = 0.91$) on non-agricultural fields in arid and semi-arid ecosystems [31]. Conversely, other findings from Nanni and Dematté [13] declared that Landsat data were a less reliable data source for SOC estimations ($R^2 = 0.51$).

In summary, many studies have been conducted that predict soil texture or SOC under laboratory conditions from hyperspectral data with promising results (cp. Table 1). In contrast, only a few studies exist that determine soil texture or SOC directly from hyperspectral airborne or satellite imagery. In most cases, this led to low prediction accuracies on agricultural soils. Nevertheless, this approach still holds great potential for digital soil mapping with hyperspectral data.

In this study, soil texture for each fraction and organic carbon content were derived from spatial and spectral high resolution airborne hyperspectral imagery. Furthermore, given the incoming results, an approach to generate reasonable and useful regionalization maps was developed by combining soil texture and SOC estimations, which provide vital information for precision agriculture applications.

Table 1. Overview of estimation results for sand, silt, clay and soil organic carbon (SOC) from the literature using different types of sensors and methods.

Sensor	Method	R ² Sand [%]	R ² Silt [%]	R ² Clay [%]	R ² SOC [%]	Reference
Laboratory						
IRIS	MR	0.79	0.27	0.91	0.80	[13]
IRIS	MR	0.82	0.57	0.86	0.30	[14]
ASD	PLSR	0.79	0.63	0.58	-	[16]
ASD	PLSR	-	-	-	0.93	[24]
ASD	PLSR	-	-	-	0.62	[27]
NIRSystems	PCR	0.82	0.82	0.67	-	[17]
NIRSystems	PLSR	-	-	-	0.98	[25]
Varian Cary 500	PLSR	0.75	0.52	0.67	0.72	[26]
Airborne/Spaceborne Imagery						
HyMap	PLSR	0.95	-	0.71	0.90	[18]
HyMap	PLSR	0.20	0.17	0.67	0.02	[19]
HyMap	PLSR	-	-	-	0.77	[29]
HyperSpecTIR	PLSR	0.79	0.79	0.66	0.65	[20]
AHS	PSR	-	-	-	0.75	[28]
Landsat TM	MR	0.53	-	0.68	0.51	[13]
Landsat TM	MR	0.64	0.54	0.61	0.41	[14]
Landsat TM	PLSR	-	-	-	0.91	[31]
Hyperion	PLSR	-	-	-	0.51	[30]

IRIS: interface region imaging spectrograph, ASD: analytical spectral device, AHS: airborne hyperspectral scanner, MR: multiple regression, PLSR: partial least squares regression, PCR: principal component regression, PSR: penalized-spline signal regression.

2. Materials and Methods

2.1. Study Area

The study area (11°54'E, 51°47'N) is located between the cities of Köthen and Bernburg (Saale) in the federal state of Saxony-Anhalt (Germany) and mainly characterized by agriculture. It is influenced by the rain shadow of the Harz Mountains, which is the main reason for the low precipitation rate of 500 mm a year [32]. The terrain, with an altitude of 70 m above sea level, is relatively flat. Within the study area, soil samples were taken from two fields with sizes of 20 ha (B) and 180 ha (A) (Figure 1). According to the European Soil Database [33], following the WRB soil classification rules [34], field A is characterized by Luvic Phaeozems (lvPH) in the east. The west of field A and field B are characterized by Calcaric Cambisols (caCM). Due to the characterizing humic topsoil layer, which is up to 40 cm depth, the high water storage ability and the high amount of organic material, these soils are particularly fertile. Field A is dominated by medium fine soil texture and very low SOC content, while Field B has medium soil texture and low SOC content [31].

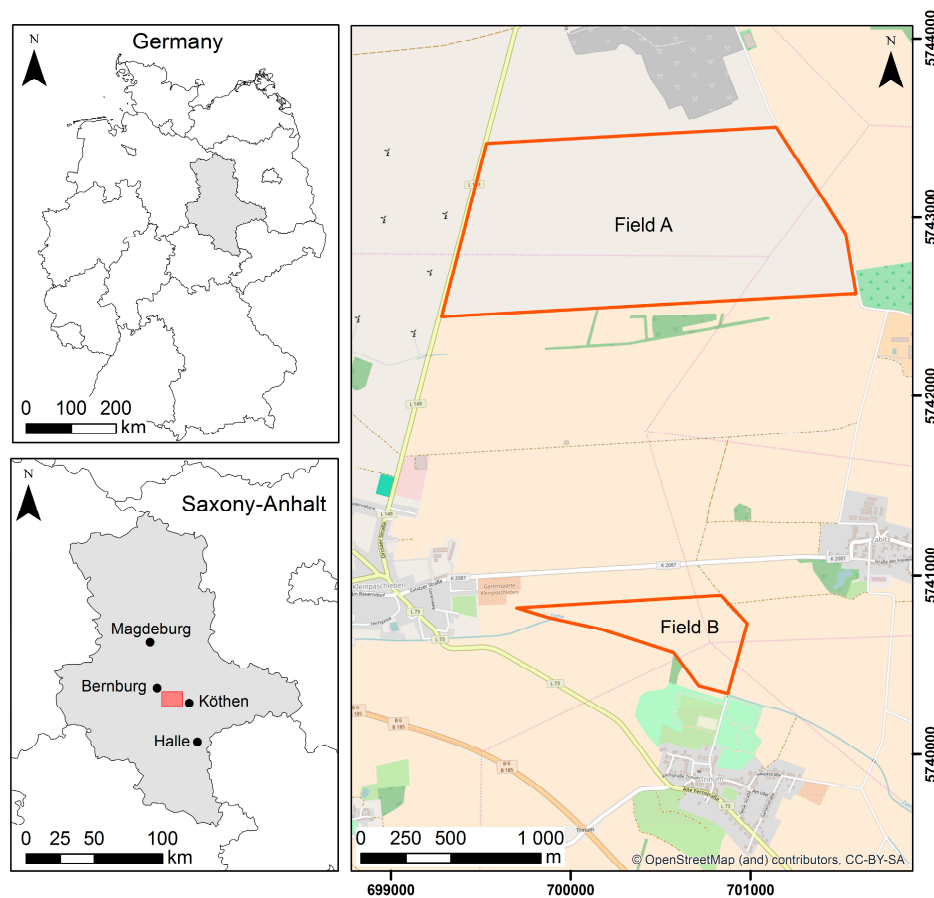


Figure 1. Location of the study site in Germany and Saxony-Anhalt (left); and the two investigated fields (red polygons) (right).

2.2. Data and Preprocessing

2.2.1. Laboratory Data

During a field campaign on 3 May 2011, 40 topsoil samples (A = 30, B = 10) were taken under dry weather conditions from the two investigated fields. Satellite data from previous years were used prior to the field campaign to develop an adjusted strategy for allowing a representative sampling of both fields. An integrative sample was taken from the upper 2 cm of the soil profile at each sampling position, representing an area of about 1 m². The coordinates of each sample were logged with a GPS device. After the field campaign, all samples were air-dried in laboratory, freed from any plant residuals and gently crushed to pass through a 2 mm sieve. Afterwards, the samples were analyzed concerning their particle size distribution after the pipette method (ISO 11277:2009) [35]. This method allows the measurement of each fraction separately. The samples were also grounded and analyzed on their SOC content by measuring the loss on ignition (DIN 18128:2002-12) [36].

2.2.2. Airborne Data

An airborne image acquisition campaign with the imaging Airborne Imaging Spectrometer for Applications DUAL (aisaDUAL) hyperspectral scanner (Specim Ltd., Oulu, Finland) was conducted on 10 May 2011. The system consists of two separate sensors, aisaEAGLE (VIS/NIR, 400–1000 nm) and aisaHAWK (SWIR, 1000–2500 nm). The sensors acquired three images with a ground sampling distance of 3 m and 367 spectral bands.

The images were affected by along track deficient lines caused by a sensor miscalibration. These image distortions were reduced by applying the reduction of miscalibration effects (ROME) destriping

algorithm [37]. Atmospheric correction was performed with the fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH). After the atmospheric correction of the images, some spectral artifacts remained within the data. These artifacts were removed by conducting an additional empirical line correction using dark and bright ground target measurements. Afterwards, the three flight stripes were geometrically corrected, orthorectified and combined into a mosaic. Finally, to further reduce the aforementioned striping effects, a minimum noise fraction (MNF) transformation was applied to the data to identify noisy spectral bands. Subsequently, an inverse MNF transformation was performed using only the first 114 MNF bands that contained all the useful information. Although this significantly improved the image quality, it could not remove the effects completely.

2.3. Methodology

As a first step, a subset of all fallow areas within the study site was manually created by drawing polygons. As a consequence, all image pixels containing spectral information of vegetation were discarded, thereby allowing the further analysis to be performed only on pure soil surfaces. Additionally, one line of pixels on the outside borders of each field was excluded from the selection to avoid mixed spectral information. Some small vegetated areas, however, remained within the fallow areas, which were detected by calculating the normalized difference vegetation index (NDVI) [38] from the subset. Pixels with a value above 0.3, indicating the presence of photosynthetic activity, were masked [39].

Secondly, aisaDUAL pixels were extracted from the image data by using the GPS coordinates of each soil sample. The limited GPS accuracy made it necessary to average the image spectra in a 3×3 window around the pixel corresponding to the acquired coordinate. Since soil conditions do not change distinctly within such a short distance, spectral differences of neighboring pixels should be small.

The averaged image reflectance spectra for each soil sample and corresponding laboratory analysis for sand, silt, clay and SOC content were then used for the calibration and validation of PLSR models within the R statistical environment [40]. PLSR has proven to be a reliable method for the prediction of soil texture and SOC in several studies (e.g., [12,23]). For each parameter, a model was calculated, allowing a maximum of 10 latent variables. Model validation was carried out by a leave-one-out cross-validation (C_V) using R^2 and root-mean-square error (RMSE) as criteria for model quality. Since all three grain size fractions should sum up to 100%, one fraction can always be calculated by the other two fractions. Therefore, the model for the fraction, which provides the lowest R^2 and the highest RMSE, respectively, was excluded from further investigations to stabilize the prediction results.

The two more robust grain size fraction models as well as the SOC model were then applied to the image data enabling the spatial prediction of the parameters in percent for each pixel of the subset. The third grain size fraction, which provided the lowest model accuracy, was then calculated by subtracting the other two modeled fractions from 100%.

Afterwards, a decision tree was developed for soil texture classification using the Food and Agriculture Organization (FAO) scheme from 2014 [34]. According to this scheme, determining the soil texture of each pixel depends on the composition of the different fractions. Applying the decision tree to the image data allowed the classification of each pixel to a specific soil texture. The resulting classes were divided into subclasses taking into account the SOC content. The result is a map showing different soil regions, each of which is based on the combination of the occurring soil texture and SOC content class.

Furthermore, relief information of the study site was derived from a digital elevation model with a 10 m equidistance, which was converted to a raster to obtain information regarding the altitude of each pixel. Subsequently, the altitude of each sample was extracted using the corresponding GPS coordinates of the soil samples. Since the soils of lower areas are expected to consist of smaller grain size particles and higher SOC content, compared to areas of higher altitude, the quality of the regionalization was visually evaluated based on this assumption.

3. Results

3.1. Data Analysis

The investigated soils are characterized by a relatively low clay, medium sand and relatively high silt content. On average the sand fraction content is represented with 34.3%, while silt contents occur with an average of 47.4%. In comparison to the other fractions, clay provides the lowest amount with an average of 18.3%. The laboratory data show that the observed area is characterized by loam and silt loam textures. The average SOC is 1.97%. The elevation of the study area is around 74 m above sea level (m.a.s.l.). Table 2 summarizes the descriptive statistics of the observed parameters.

Table 2. Descriptive statistics for measured data of sand, silt, clay, SOC and interpolated information.

Parameter	<i>n</i>	Min.	Max.	Mean	SD
Sand [%] (0.063–2 mm)	40	9.9	48.3	34.3	11.4
Silt [%] (2–63 μm)	40	36.4	68.7	47.4	8.9
Clay [%] (<2 μm)	40	11.8	27.7	18.3	3.4
SOC [%]	40	1.06	3.91	1.97	0.82
Height [m]	40	69.85	78.24	73.89	2.69

SD: Standard Deviation.

The correlation analysis (Table 3) shows high positive (>0.80) and negative dependencies (<0.80) between nearly all observed data. Most mentionable are the high correlations of sand and organic carbon in relation to all other parameters. Height information depicts correlations with the other parameters as well.

Table 3. Correlations for measured data of sand, silt, clay, SOC and interpolated height information with a statistical significance of $\alpha < 0.01$.

	Sand	Height	Clay	Silt	SOC
Sand	1	0.82	−0.81	−0.97	−0.94
Height	−	1	−0.67	−0.81	−0.85
Clay	−	−	1	0.66	0.80
Silt	−	−	−	1	0.90
SOC	−	−	−	−	1

3.2. PLSR

The validated PLSR model for sand provided an R^2_{cv} of 0.77 and an $RMSE_{cv}$ of 5.37. The regression line was close to the 1:1 line, such that no tendency towards a systematic over- or underestimation of the sand content was observed. Maximum over- or underestimation was around 10%. The regression function also indicated a good prediction accuracy, since the slope was close to one and the intercept was lower than 2%. However, due the small number of samples with low sand content, these samples are expected to have a higher influence on model quality when over- or underestimated. The silt model showed moderate quality ($R^2_{cv} = 0.62$, $RMSE_{cv} = 5.46$). While higher values tended to be underestimated, lower values were slightly overestimated. This was also indicated by the regression coefficient of 0.88, while the offset of 5.5 is negligible for the predicted data range. Although the maximum prediction error was around 10%–15%, this error only occurred for two samples. Compared to the other models the PLSR for clay provided the lowest model accuracy with an R^2_{cv} of 0.53 and an $RMSE_{cv}$ of 2.39. The narrow data range between 12% and 28% led to unstable results. Defective estimations for data outside the observed range should be quite distinct. Nevertheless, it has to be pointed out that the absolute RMSE is relatively low, especially when considering that, apart from two samples, all prediction errors are lower than 5%. In summary, it had to be mentioned that the RMSE values for sand, silt and clay models still indicated relatively small prediction errors, since none of

them was larger than 5.5%. The PLSR for SOC resulted in a strong model ($R^2_{cv} = 0.89$, $RMSE_{cv} = 0.27$). The regression line had a very low intercept and a slope, which was close to one. Hence, the regression line provided a close match to the 1:1 line. Moreover, the RMSE of the model was very low, which was additional evidence for a robust regression model. Scatterplots of the leave-one-out cross-validated regression models for all parameters are illustrated in Figure 2.

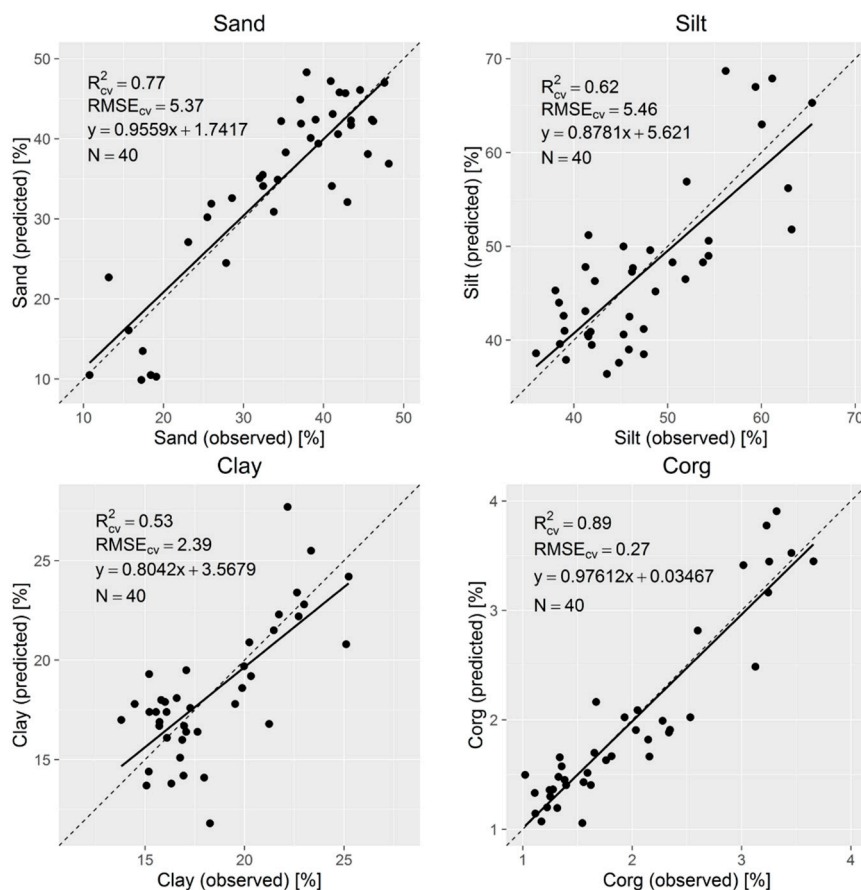


Figure 2. Scatterplots of leave-one-out cross-validated partial least squares regression (PLSR) models for sand, silt, clay and soil organic carbon (SOC) with the regression line (continuous) and the 1:1 line (dashed).

The factor loadings of the PLSR models, illustrated in Figure 3, show the importance for all wavelengths within each model. Sand has its most relevant negative factor loadings at around 680, 1350 and 2000 nm, and most positive ones at around 450 and 750 to 900 nm. The important negative loadings for silt are located at around 450 and 750 to 900 nm. Positively loaded wavelengths for this model are at 680, 1350 and 2000 nm. Clay is significant at around 490, 750 to 950, 2000 and 2200 nm with negative loadings and at around 500, 1000 to 1150 and 1350 nm with positive loadings. SOC has distinct negative loadings at 500, 680, 1500 and 2000 nm. Important positive loadings are at 450, 550, 750 and 1300 nm. Overall, most of the relevant spectral features are located between 400 and 1000 nm. Overall, despite the algebraic sign of the loadings the important wavelengths are quite similar for all the observed parameters. However, the clay model also has very significant bands beyond 1000 nm. All the other models show very noisy loadings with almost no significant peaks wavelengths beyond 2000 nm. The loadings for silt and sand are almost identical but mirrored.

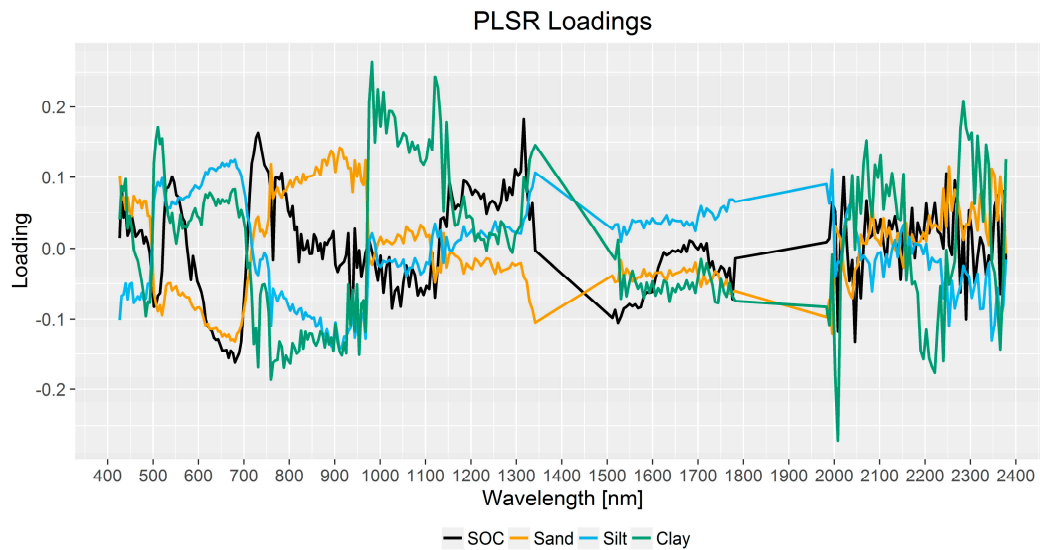


Figure 3. Partial least squares regression (PLSR) factor loadings for sand, silt, clay and soil organic carbon (SOC) models.

The spatial predictions (Figure 4) for the large observed field in the north (A) showed a dominance of sand. While a small part in the south and southwest possessed medium sand contents (approximately 30%), the rest of the field showed distinctly higher concentrations of sand by up to 65%, which increased from southwest to northeast. Silt showed the same pattern, but in opposite direction. The southwest had a higher silt content (approximately 50%) than the east (approximately 30%) and the northeast (approximately 35%). The entire field was characterized by a rather homogeneous clay content, which was represented by around 15%, with a slight increase to the south and the southwest up to roughly 20%. SOC showed increasing content from the northeast (approximately 1%) to the southwest (approximately 2%). Variations in SOC are strongly related to the elevation of field A, which increased from the southwest (72.5 m.a.s.l.) to the north and the northeast (77.5 m.a.s.l.).

The smaller observed field (B) had a relatively constant elevation at 70 m.a.s.l. with a slight depression in the center of the area. Sand was substantially less prevailing in contrast to field A with values ranging from approximately 5% in the west and 30% in the east. In contrast, silt dominated this field much more. The western part showed very high (approximately 70%) silt concentrations, while the eastern part showed medium (approximately 50%) silt concentrations. Clay ranged from roughly 30% in the west to approximately 20% in the east, meaning that it was more present than in the observed northern field (A). Related to the lower elevation, the SOC content was distinctly higher with a range between 2% and 4% from the east to the west.

Besides the two sampled fields, the model was also applied to non-observed neighboring fields with uncovered soils. The predicted patterns in the southwest of field A were continuing on the neighboring field. Sand content increased with higher elevation from the north to the south, while silt, clay and SOC decreased simultaneously. The two long narrow fields close to the local hill in the center of the study site were dominated by high sand content (approximately 60%), but had less silt and clay. Predictions for silt in these fields were slightly heterogeneous, since they varied between 30% and 50% within a small area. Additionally, they were characterized by lower SOC content.

The three small fields north of field B showed similar properties for all parameters, as predicted for the neighboring observed field. However, they had fewer similarities with the field north of them, which had a more sandy soil texture. The field east of field B mostly had the same elevation, while the predicted pattern from field A was continuing smoothly. Elevation increased to the east, which also resulted in a coarser texture. This increase was oriented to the shape of the contour line.

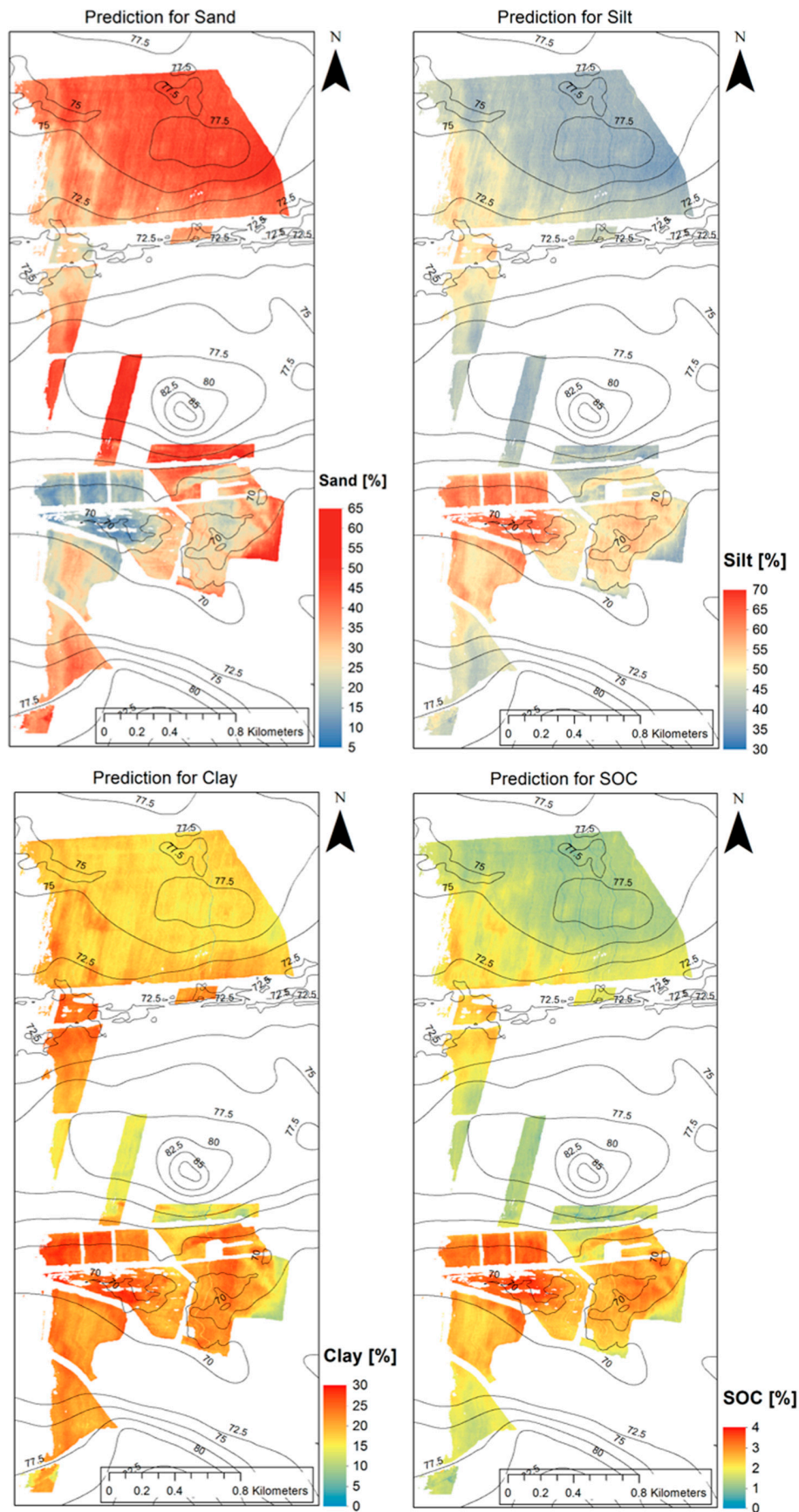


Figure 4. Spatial predictions of sand, silt, clay and SOC for the investigated and neighboring uncovered fields within the study site based on PLSR.

Overall, the prediction images of all parameters contained slightly higher and lower extreme values in comparison to the covered data range, which was used in PLSR model building. Nevertheless, areas with a similar elevation nearly showed the same parameter combinations as in the observed fields.

The three grain size fractions for sand, silt and clay have to sum up to 100%. This is why spatial predictions of sand, silt and clay were stacked to a single image to identify regions deviating from 100%. The result is shown in Figure 5. The histogram of the observed fields A and B showed a minimum of 95.65% and a maximum of 103.89%, respectively. The slight underestimations are mainly located in the image errors in field A. The median of 99.73% stated that the predictions of the parameters were very close to the expected result. Overall, 98.87% of all pixels within fields A and B had a grain size composition between 98% and 102%. By observing the grain size composition deviation for all fields, the minimum decreased to 95.07% and the maximum increased to 109.33%. However, the relatively extreme overestimations of >105% only occurred for 0.02% (60 pixels) of the data and tended to be located in non-observed areas, which were further away from the observed fields, A and B. Overall, the mean of 100.32% indicated a very slight overestimation of the non-observed fields, which occurred as small isolated patterns. However, 91.46% of the pixels had a grain size between 98% and 102%.

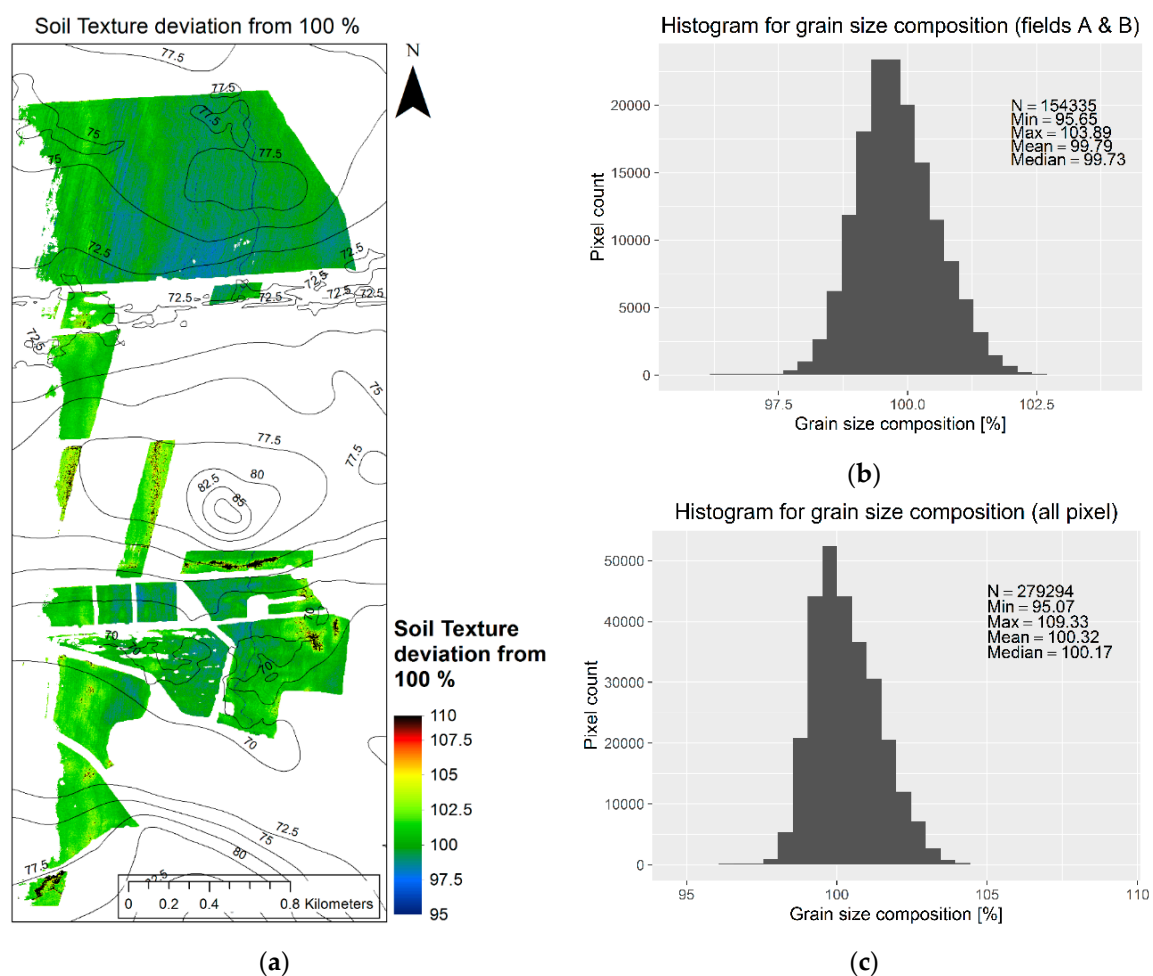


Figure 5. Cumulated spatial prediction sand, silt and clay predictions and deviation from 100%: (a) For all pixels in the image; (b) The histogram of grain size composition for pixels only from the observed fields A and B; and (c) The histogram of grain size composition for all pixels.

3.3. Classification/Regionalization

The spatial predictions of sand and silt were combined with a calculated clay prediction image (clay = 100% – %sand – %silt), since the clay PLSR model provided the lowest model performance.

These data sets were used as input data for the decision tree and classified according to the FAO soil texture scheme. Each image pixel was then assigned to a single soil texture class. As apparent from Figure 6a, the higher elevated northern field A was dominated by loam (L) and the lower lying southern field B by silt loam (SiL). The surrounding fields showed a similar relief-oriented pattern. Only a few pixels were classified as silt clay loam (SiCL) and sandy loam (SL). This initial regionalization image was further refined with the spatial information about SOC (Figure 6b). This additional information enabled the division of each soil texture region into subregions for any occurring SOC concentration. The large loam-dominated field (A) in the north, which had a higher elevation, was mainly characterized by a single subregion with 1% to 2% SOC content. Only the slightly lower elevated parts of this field showed a higher SOC content, which was between 2% and 3%. The smaller silt loam-dominated field B in the south was divided into several subregions, showing a more pronounced SOC heterogeneity. The slight depression in the center of the field showed a high SOC content of 3% to 4%, which decreased to the east and the south where the elevation increases. The field in the east of field B showed a similar behavior. The center of the field also had a local sink with a high SOC content (3% to 4%), while SOC content decreased in all higher elevated surrounded pixels, while the shape of SOC class changes was very similar to the trend of the contour lines.

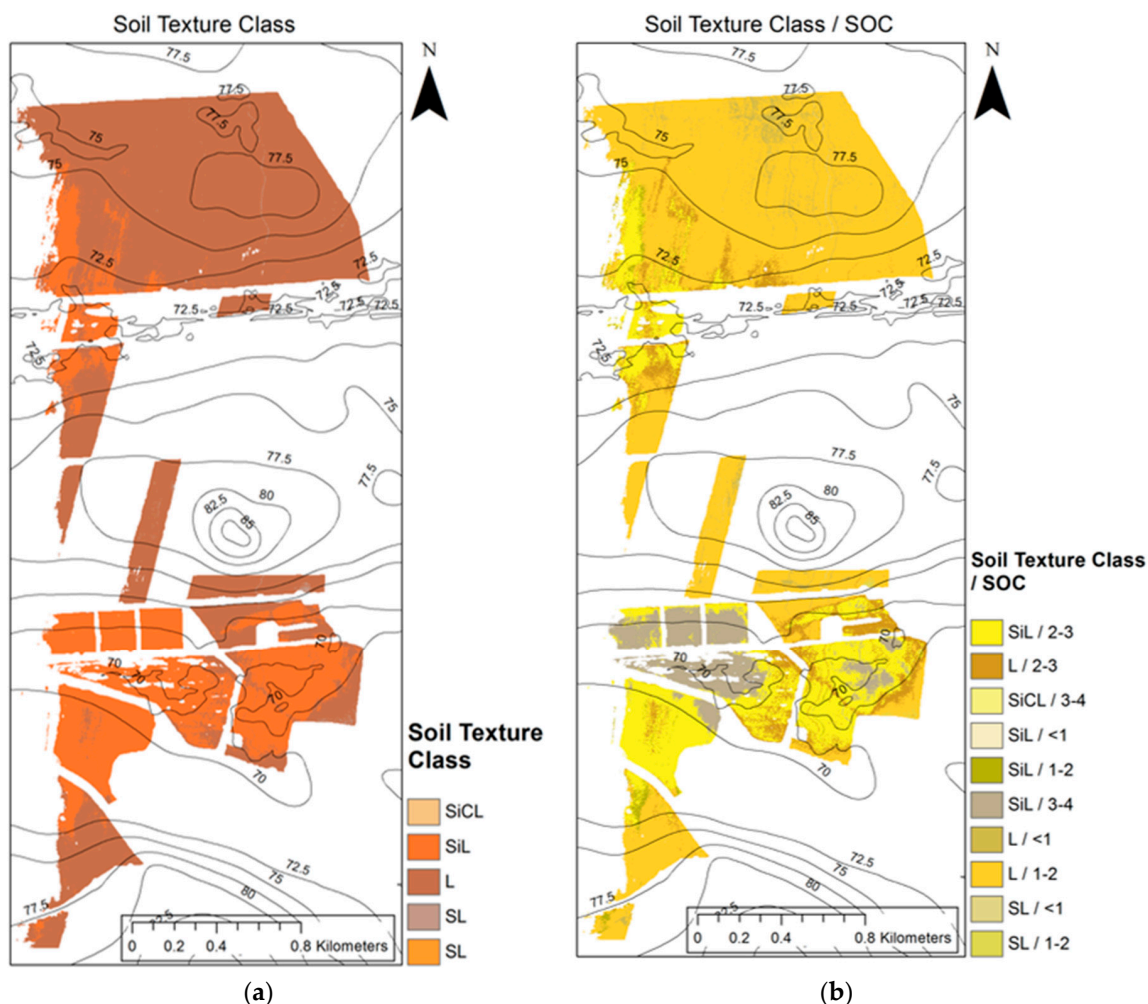


Figure 6. Regionalization results based on decision tree classification of (a) Soil texture and (b) Soil texture subdivided by SOC (SiL: silt loam; L: loam; SL: sand loam; SiCL: silt clay loam).

4. Discussion

The presented results indicated a strong connection between all observed parameters. It is noteworthy that the observed grain size fractions and SOC concentrations strongly depend on the relief. This fact was illustrated by the performed correlation analysis (cf. Table 3) and is also significantly in line with findings presented by Sinowski and Auerswald [41]. Samples collected from local sinks or bottoms of slopes contained more fine soil textures and higher SOC concentration than samples taken from higher regions. This phenomenon could be explained in terms of the result of transport processes caused by water, wind and man-made actions. The effect that fine soil particles and nutrients are more dominant in lower elevated areas was also observed by Moore et al. [42] and Lyles and Tatarko [43].

Furthermore, the observed parameters were tested regarding their predictivity using image spectra with PLSR. The model for sand turned out to be the most robust one of the three grain size fraction models. Mulder et al. [8] presented comparable results for predicting sand content under laboratory conditions. Compared to other studies focusing on image data, the modeling results correspond with a study from Hively et al. [20], who used similar input data. However, the results could not compete with findings from Selige et al. [18]. Nevertheless, the achieved model underlined the high correlation between sand particles and aisaDUAL soil spectra. The predictive model for silt can be graded as relatively robust in comparison to laboratory studies providing lower model accuracies [13,14]. However, Hively et al. [20] presented results based on image data, which led to similar results. The model quality of clay was the lowest in this study. It was also lower in comparison to other work [18,19]. The low prediction accuracy can be attributed to the relatively small data range for clay. Moreover, clay has the lowest share of the three texture compartments in our study area, such that it may have been less distinct within spectral reflectance. The SOC model provided results comparable to those achieved by Patzold et al. [24] and Stevens et al. [28]; indeed, it even slightly exceeded their quality.

It was shown that the correlations of the laboratory measurements of sand, silt clay and SOC show strong connections between all the observed parameters (Table 3). It would be expected that the factor loadings (Figure 3) for each model should be correlated as well. However, this could not be fully achieved in this case. This can be explained with the fact that the image spectra completely depict the complex structure and condition of the soil. In general, the overall reflectance decreases with increasing SOC. For example, on the one hand, coarser soil particles are reducing the overall reflectance due to shadowing effects. In addition, these soils tend to store less SOC which is increasing the overall reflectance. On the other hand, finer particles show a brighter overall reflectance, since shadows are smaller. Moreover, SOC is more likely to be higher on these soils which on the contrary reduces reflectance. It can be concluded, that the spectral relationships between the observed parameters are not headed into one direction but are influencing each other in a very complex manner. Keeping in mind, that several other parameters (inorganic carbon, iron oxides, water content) are influencing the soil reflectance spectrum as well, a direct transfer of relationships between isolated parameter analyses to spectral behavior is therefore not possible.

The spatial prediction of all four parameters showed a plausible distribution, since there were smooth transition areas of increasing and decreasing contents within the entire study site. Furthermore, the spatial predictions supported the aforementioned correlation of the parameters to the terrain. The northern field is characterized by a slope to the southwest. This was also clearly visible by spatial changes of texture and SOC in the prediction images. While the sand fraction decreased, silt, clay and SOC contents increased. The same effect was observed in the sampled field and neighboring fields in the south. This can also be regarded as evidence for reasonable predictions. Only the areas affected by the aforementioned striping effects in the aisaDUAL image data showed clearly visible prediction errors. In particular, the clay map illustrated a deficient detector line of the sensor, leading to incorrect predictions that were close to 0%, while contents of 15% were estimated for surrounding pixels. The overall effect of mispredicted values of sand, silt and clay was investigated by the summation of the three spatial prediction maps. Since the clay model provided the lowest accuracy, it was substituted

by calculating the clay fraction by subtracting the spatial predictions of sand and silt from 100%. Thus, the classification error was reduced.

The transfer of the calibrated models to non-observed neighboring fields led to plausible results. On the one hand, the transitions between observed and non-observed fields are mostly very smooth, especially when the fields are separated by vegetation (e.g., from field B to the east), given that the transport of materials is not interrupted in this case. On the other hand, the three rectangle-shaped fields north of field B and the narrow-shaped small field north of them showed very different characteristics, as already described in Section 3. It turned out that these fields are separated by a larger paved road, which may have had an influence on material transportation dynamics. These topographical circumstances clearly indicated the quality of the modeling results in the spatial domain.

The developed models, however, can only be applied to unknown data in the same region where similar soil conditions prevail. The over- and underestimations within the neighboring non-observed fields are most likely due to soil conditions, which were not represented by the taken samples. Especially for the silt model, whose regression function had a lower slope, this led to mispredictions in some rare cases. For example, the two fields close to the center of the study site, which were west and east of the local hill, showed heterogeneous predictions for silt. One probable reason might be that these fields are more highly elevated than any of the observed areas and therefore contain silt concentrations, which were not covered during the field campaign. This led to some of the most intense overestimations in our study. For these two fields, the usage of the clay model, instead of the silt model, may have resulted in a more accurate prediction in this case. However, for the complete study site, the exclusion of the clay model was found to be more reasonable. Nevertheless, the presented histograms (Figure 5) only depicted a maximum of small prediction errors of $\pm 2\%$. For most pixels, these deviations would not lead to a misclassification in soil texture, since each class covers at least 10% of each particle group.

The classification of soil texture, based on the developed decision tree, enabled the generation of a regionalized map of the study site, which was dominated by loam (L) and silt loam (SiL). Other occurring texture classes were mainly originated from the striping effects in the image. The two main regions were well-separated and less affected by image distortion effects. Patterns occurring in one field can be seen to continue in neighboring fields. This fact was an indicator for a reliable soil texture classification. The further refinement of the soil texture regionalization by including SOC predictions led to a classification result with 10 different classes. At this point, the strong connection between soil texture, terrain and SOC became very clear. Silt loam regions, as the finer texture class, contained more SOC than the slightly coarser loam regions. The subdivided regions were also homogenous, while the result map (Figure 6b) showed smooth transition zones between the occurring classes. To use these maps for precision farming in actuality, the calculated zones may have to be smoothed to reduce some salt and pepper effects, such that they are easier to handle for automatic agricultural machinery and provide a better readability for the user as well.

5. Conclusions

The approach presented in this study allowed for a successful determination of soil texture and SOC from hyperspectral image data. Using image spectra and soil information as input data to build PLSR models resulted in a quality that was sufficient to generate maps with regions of similar soil conditions. The regionalized digital soil maps can serve as an information source, which supports farmers by providing precise spatial information about current soil conditions. The classification based on a decision tree was applied to identify regions of different soil texture, meaning that they can easily be adapted to other areas. Since sand, silt and clay contents sum up to 100%, only the two best predicting models for sand and silt were needed for classifying soil texture classes. Our results indicate that the regionalization procedure might be transferable to non-observed areas within a region with similar soil conditions. In comparison to geostatistical approaches, where predictions are made by the spatial distribution of the sample points, a hyperspectral pixel-wise prediction is probably much more

accurate and reliable. On the contrary, high quality imagery need to be available and requires complex pre-processing. However, in the future, the availability and costs for hyperspectral image data will probably decrease with the launch of hyperspectral satellite mission. Concerning the principle of optically reflective remote sensing the methodology is only suited for assumptions from the upper 2 cm of the soil. In future studies, the presented approach should be tested in a larger and more heterogeneous study area using an extended number of analyzed soil samples, which allows for an independent validation of the regression models to be performed. For a wider field of applications in the future, the approach also needs to be tested on hyperspectral satellite data acquired by upcoming hyperspectral satellite missions, such as the Environmental Mapping and Analysis Program (EnMAP) (planned launch for 2018) and the Hyperspectral Infrared Imager (HyspIRI) (planned launch for 2022). Further, soil water conditions during image acquisition have influence on the spectral behavior of soils. Data from airborne flight campaigns or satellite missions should therefore only be used when there was no rainfall in the days before acquisition. However, despite this constraint, our approach represents a cost effective possibility to spatially predict soil parameters. The adaptation to hyperspectral satellite data will allow for the regionalization of soils for larger areas, thereby forming the basis for practical applications in precision agriculture.

Acknowledgments: This work was funded by the German Aerospace Center (DLR), using the financial resources of the Federal Ministry of Economics and Technology on the basis of a decision by the German Parliament, under grant number 50EE1014. We would like to thank the Helmholtz Centre for Environmental Research (UFZ) in Leipzig for aisaDUAL image data acquisition. Special thanks go to Wagner and the Wimex GmbH, owner of the investigated fields, for their cooperation and support. Additionally, we want to thank Daniel Doktor, Holger Lilienthal, Nicole Richter, Thomas Selige and Anne Bodemann for their assistance in the field during data collection and data preprocessing. We acknowledge support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of Osnabrück University.

Author Contributions: All authors contributed equally to research, writing and editing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Food and Agriculture Organization of the United Nations (FAO). World Soil Charter. 2015. Available online: <http://www.fao.org/3/a-mn442e.pdf> (accessed on 17 December 2015).
2. Food and Agriculture Organization of the United Nations (FAO). Status of the World's Soil Resources. 2015. Available online: <http://www.fao.org/3/a-i5199e.pdf> (accessed on 17 December 2015).
3. Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, *327*, 828–831. [[CrossRef](#)] [[PubMed](#)]
4. Adamchuk, V.I.; Hummel, J.W.; Morgan, M.T.; Upadhyaya, S.K. On-the-go soil sensors for precision agriculture. *Comput. Electron. Agric.* **2004**, *44*, 71–91. [[CrossRef](#)]
5. Adhikari, K.; Hartemink, A.E.; Minasny, B.; Bou Kheir, R.; Greve, M.B.; Greve, M.H. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE* **2014**, *9*, e105519. [[CrossRef](#)] [[PubMed](#)]
6. Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, A. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using random forests analysis. *Geoderma* **2008**, *146*, 102–113. [[CrossRef](#)]
7. De Brogniez, D.; Ballabio, C.; Stevens, A.; Jones, R.J.A.; Montanarella, L.; van Wesemael, B. A Map of the topsoil organic carbon content of Europe generated by a generalized additive model. *Eur. J. Soil Sci.* **2015**, *66*, 121–134. [[CrossRef](#)]
8. Villa, P.; Malucelli, F.; Scalenghe, R. Carbon Stocks in Peri-Urban Areas: A case study of remote sensing capabilities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4119–4128. [[CrossRef](#)]
9. Peralta, N.R.; Costa, J.L.; Balzarini, M.; Angelini, H. Delineation of management zones with measurements of soil apparent electrical conductivity in the southeastern Pampas. *Can. J. Soil Sci.* **2013**, *93*, 205–218. [[CrossRef](#)]
10. Mulder, V.L.; de Bruin, S.; Schaepman, M.E.; Mayr, T.R. The Use of Remote sensing in soil and terrain mapping—A review. *Geoderma* **2011**, *162*, 1–19. [[CrossRef](#)]

11. Blume, H.-P.; Brümmer, G.; Horn, R.; Kandeler, E.; Kögel-Knabner, I.; Kretschmar, R.; Stahr, K.; Wilke, B.-M. *Scheffer/Schachtschabel Lehrbuch der Bodenkunde*, 16th ed.; Spektrum Akademischer Verlag: Heidelberg, Germany, 2010.
12. Minasny, B.; McBratney, A.B. Digital soil mapping—A brief history and some lessons. *Geoderma* **2016**, *264*, 301–311. [[CrossRef](#)]
13. Nanni, M.R.; Demattê, J.A.M. Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Sci. Soc. Am. J.* **2006**, *70*, 393–407. [[CrossRef](#)]
14. Demattê, J.A.M.; Fiorio, P.R.; Araújo, S.R. Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. *Remote Sens.* **2010**, *2*, 1998–2016. [[CrossRef](#)]
15. Demattê, J.A.M.; Alves, M.R.; Gallo, B.C.; Fongaro, C.T.; e Souza, A.B.; Romero, D.J.; Sato, M.V. Hyperspectral remote sensing as an alternative to estimate soil attributes. *Revista Ciência Agronômica* **2015**, *46*, 223–232. [[CrossRef](#)]
16. Casa, R.; Castaldi, F.; Pascucci, S.; Palombo, A.; Pignatti, S. A Comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma* **2013**, *197–198*, 17–26. [[CrossRef](#)]
17. Chang, C.-W.; Laird, D.; Mausbach, M.J.; Hurburgh, C.R. Near-infrared reflectance spectroscopy—Principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [[CrossRef](#)]
18. Selige, T.; Böhner, J.; Schmidhalter, U. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* **2006**, *136*, 235–244. [[CrossRef](#)]
19. Gomez, C.; Lagacherie, P.; Coulouma, G. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. *Geoderma* **2012**, *189–190*, 176–185. [[CrossRef](#)]
20. Hively, W.D.; McCarty, G.W.; Reeves, J.B., III; Lang, M.W.; Oesterling, R.A.; Delwiche, S.R. Use of airborne hyperspectral imagery to map soil properties in tilled agricultural fields. *Appl. Environ. Soil Sci.* **2011**. [[CrossRef](#)]
21. Reeves, D.W. The role of soil organic matter in maintaining soil quality in continuous cropping systems. *Soil Tillage Res.* **1997**, *43*, 131–167. [[CrossRef](#)]
22. Shatar, T.M.; McBratney, A.B. Empirical modelling of relationships between sorghum yield and soil properties. *Precis. Agric.* **1999**, *1*, 249–276. [[CrossRef](#)]
23. Ladoni, M.; Bahrami, H.A.; Alavipanah, S.K.; Norouzi, A.A. Estimating soil organic carbon from soil reflectance—A review. *Precis. Agric.* **2010**, *11*, 82–99. [[CrossRef](#)]
24. Patzold, S.; Mertens, F.M.; Bornemann, L.; Koleczek, B.; Franke, J.; Feilhauer, H.; Welp, G. Soil heterogeneity at the field scale—A challenge for precision crop protection. *Precis. Agric.* **2008**, *9*, 367–390. [[CrossRef](#)]
25. McCarty, G.W.; Reeves, J.B., III; Reeves, V.B.; Follett, R.F.; Kimble, J.M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. *Soil Sci. Soc. Am. J.* **2002**, *66*, 640–646.
26. Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemsta, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [[CrossRef](#)]
27. Jarmer, T.; Vohland, M.; Lilienthal, H.; Schnug, E. Estimation of some chemical properties of an agricultural soil by spectroradiometric measurements. *Pedosphere* **2008**, *18*, 163–170. [[CrossRef](#)]
28. Stevens, A.; Udelhoven, T.; Denis, A.; Tychon, B.; Liroy, R.; Hoffmann, L.; van Wesemael, B. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma* **2010**, *158*, 32–45. [[CrossRef](#)]
29. Schwanghart, W.; Jarmer, T. Linking spatial patterns of soil organic carbon to topography—A case study from south-eastern Spain. *Geomorphology* **2011**, *126*, 252–263. [[CrossRef](#)]
30. Gomez, C.; Viscarra Rossel, R.A.; McBratney, A.B. Soil organic carbon prediction by hyperspectral remote sensing and field Vis-NIR spectroscopy—An Australian case study. *Geoderma* **2008**, *146*, 403–411. [[CrossRef](#)]
31. Jarmer, T.; Hill, J.; Lavée, H.; Sarah, P. Mapping Topsoil organic carbon in non-agricultural semi-arid and arid ecosystems of Israel. *Photogramm. Eng. Remote Sens.* **2010**, *75*, 85–94. [[CrossRef](#)]
32. Deutscher Wetterdienst. Climate Data Center. Available online: ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/monthly/kl/historical/ (accessed on 22 April 2016).
33. European Commission and the European Soil Bureau Network. *The European Soil Database Distribution Version 2.0*; CD-ROM, EUR 19945 EN; European Commission: Brussels, Belgium; European Soil Bureau Network: Ispra, Italy, 2004.

34. Food and Agriculture Organization of the United Nations (FAO); IUSS Working Group WRB. *World Reference Base for Soil Resources 2014 (Update 2015) International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; World Soil Resources Reports, No. 106; FAO: Rome, Italy, 2015.
35. International Organization of Standardization (ISO). *11277:2009 Soil Quality—Determination of Particle Size Distribution in Mineral Soil Material—Method by Sieving and Sedimentation*; International Organization of Standardization (ISO): Geneva, Switzerland, 2009.
36. Deutsches Institut für Normung e.V. (DIN). *Soil—Investigation and Testing—Determination of Ignition Loss*; 18128:2002-12; Deutsches Institut für Normung e.V. (DIN): Coblenz, Germany, 2002.
37. Rogaß, C.; Spengler, D.; Bochow, M.; Segl, K.; Lausch, A.; Doktor, D.; Roessner, S.; Behling, R.; Wetzel, H.U.; Kaufmann, H. Reduction of Radiometric miscalibration—Application to pushbroom sensors. *Sensors* **2011**, *11*, 6370–6395. [[CrossRef](#)] [[PubMed](#)]
38. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, Washington, DC, USA, 10–14 December 1973; Volume 1, pp. 309–317.
39. Price, J.C. Estimating leaf area index from satellite data. *IEEE Trans. Geosci. Remote Sens.* **1993**, *31*, 727–734. [[CrossRef](#)]
40. Mevik, B.-H.; Wehrens, R.; Liland, K.H. pls: Partial least squares and principal component regression. R package Version 2.5-0. 2015. Available online: <https://CRAN.R-project.org/package=pls> (accessed on 21 April 2016).
41. Sinowski, W.; Auerswald, K. Using relief parameters in a discriminant analysis to stratify geological areas with different spatial variability of soil properties. *Geoderma* **1999**, *89*, 113–128. [[CrossRef](#)]
42. Moore, I.D.; Gessler, P.E.; Nielsen, G.A.; Peterson, G.A. Soil attribute prediction using terrain analysis. *Soil Sci. Soci. Am. J.* **1993**, *57*, 443–452. [[CrossRef](#)]
43. Lyles, L.; Tatarko, J. Wind erosion effects on soil texture and organic matter. *J. Soil Water Conserv.* **1986**, *41*, 191–193.



© 2016 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 (<http://creativecommons.org/licenses/by/4.0/>).