remote sensing

MDPI

Article

Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects

Fabio Castaldi [1,*], Sabine Chabrillat [2], Axel Don [3] and Bas van Wesemael [4]

1 ILVO—Flanders Research Institute for Agriculture, Fisheries and Food, Technology and Food Science-Agricultural Engineering. Burg. Van Gansberghelaan 115 bus 1-9820 Merelbeke, Belgium
2 Helmholtz-Zentrum Potsdam—Deutsches GeoForschungsZentrum GFZ, 14473 Potsdam, Germany; chabri@gfz-potsdam.de
3 Thünen Institute of Climate-Smart Agriculture. Bundesallee 65, 38116 Braunschweig, Germany; axel.don@thuenen.de
4 Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Universite Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; bas.vanwesemael@uclouvain.be
* Correspondence: fabio.castaldi@ilvo.vlaanderen.be

Received: 1 August 2019; Accepted: 10 September 2019; Published: 12 September 2019

Abstract: Soil organic carbon (SOC) loss is one of the main causes of soil degradation in croplands. Thus, spatial and temporal monitoring of SOC is extremely important, both from the environmental and economic perspective. In this regard, the high temporal, spatial, and spectral resolution of the Sentinel-2 data can be exploited for monitoring SOC contents in the topsoil of croplands. In this study, we aim to test the effect of the threshold for a spectral index linked to soil moisture and crop residues on the performance of SOC prediction models using the Multi-Spectral Instrument (MSI) Sentinel-2 and the European Land Use/cover Area frame Statistical survey (LUCAS) topsoil database. The LUCAS spectral data resampled according to MSI/Sentinel-2 bands, which were used to build SOC prediction models combining pairs of the bands. The SOC models were applied to a Sentinel-2 image acquired in North-Eastern Germany after removing the pixels characterized by clouds and green vegetation. Then, we tested different thresholds of the Normalized Burn Ratio 2 (NBR2) index in order to mask moist soil pixels and those with dry vegetation and crop residues. The model accuracy was tested on an independent validation database and the best ratio of performance to deviation (RPD) was obtained using the average between bands B6 and B5 (Red-Edge Carbon Index: RE-CI) (RPD: 4.4) and between B4 and B5 (Red-Red-Edge Carbon Index: RRE-CI) (RPD: 2.9) for a very low NBR2 threshold (0.05). Employing a higher NBR2 tolerance (higher NBR2 values), the mapped area increases to the detriment of the validation accuracy. The proposed approach allowed us to accurately map SOC over a large area exploiting the LUCAS spectral library and, thus, avoid a new ad hoc field campaign. Moreover, the threshold for selecting the bare soil pixels can be tuned, according to the goal of the survey. The quality of the SOC map for each tolerance level can be judged based on the figures of merit of the model.

Keywords: Sentinel-2; SOC; multispectral; Normalized Burn Ratio 2; LUCAS spectral library

1. Introduction

Since the onset of agriculture, soils have lost their organic carbon to such an extent that the soil functions of many croplands are threatened. There is strong demand for mapping and monitoring critical soil properties and, in particular, soil organic carbon (SOC). After all, SOC stock is one of the three sub-indicators for the Sustainable Development Goal (SDG) 15.3.1 defining the proportion of land

that is degraded over total land area [1]. The Voluntary Guidelines for Sustainable Soil Management published by Intergovernmental Technical Panel on Soils [2] also clearly indicated the loss of SOC as one of the main causes of soil degradation and lay down a set of good practices to enhance the soil organic matter content. These practices will be used as guidelines for cross-compliance rules for common agricultural Policy (CAP) in the European Union related to the good agricultural and environmental condition (GAEC) of the land for the period post-2020 [3]. In this context, the continuous spatial and temporal monitoring of the organic carbon content in agricultural soils becomes extremely important, not only from the environmental perspective, but also in economic terms to ensure that the beneficiaries of the CAP respect their cross-compliance obligations.

Pilot studies have demonstrated the potential for remote sensing in the optical domain for SOC mapping. This technique is generally used in croplands where the following conditions prevail: (i) a uniform SOC content throughout the plough layer (usually the top 20 to 30 cm), (ii) the soil is bare and an eventual soil crust has been ploughed in just after seeding of summer crops (such as maize, sugar beet, or potato) or winter crops (such as winter cereals), and (iii) surface water content is low since the surface quickly dries out during the cloud-free conditions required for acquiring an image. Remote sensing products can contribute toward establishing national or regional reference systems for soil properties as an alternative to the costly techniques relying on traditional sampling and analysis. High resolution and up-to-date SOC maps produced by remote sensing techniques will be valuable tools to address issues related to the SDG's or the common agricultural policy at the regional scale, while, at the same time, providing within field SOC patterns. The latter can be used to optimize cropland management and reduce environmental impacts, such as reducing fertilizer input by accounting for the mineralization of soil organic matter, enhancing infiltration, and reducing local flooding.

The EU (European Union) Copernicus satellite program can give a boost to high-resolution SOC mapping in croplands covering large areas and at low costs. The Sentinel-2 mission of the EU Copernicus program is a constellation of two satellites (Sentinel-2A and Sentinel-2B) equipped with the Multi-Spectral Instrument (MSI) that collects optical imagery over land and coastal waters. The main strengths of Sentinel-2 data are: (i) their availability: everybody can freely explore and download the images, (ii) the short revisit time: five days under the same viewing angle, (iii) the fine spatial resolution: 10 m in the visible and 20 m in the near infrared (NIR) and short wave infrared (SWIR) region, (iv) the presence of 13 bands including two bands acquiring data in the SWIR region, and (v) high quality physically-calibrated data. The spatial and spectral characteristics of the MSI/Sentinel-2 sensor are, thus, promising for soil applications. In particular, the two SWIR bands, even if they are quite broad, can be particularly useful to exploit both the correlation between SOC content and soil brightness and the spectral signature that is assigned to soil organic carbon (SOC) and soil texture in this region [4]. The availability and the short revisit time of the MSI/Sentinel-2 sensor entails a huge amount of satellite data that increases the possibility to acquire reliable images throughout the year (e.g., cloud-free images). This is especially important for imaging spectroscopy of the topsoil because the time window in which we can find bare soil in croplands of the temperate regions is narrow (i.e., after seedbed preparation). At the same time, the collection of a series of satellite images for the same area allows multi-temporal analysis or tessellated processes.

Some authors already used Sentinel-2 data for soil variable prediction and mapping [5–8], and obtained encouraging results, especially for the SOC content in the plough layer. However, some issues still need to be addressed to improve a soil product based on Sentinel-2 data. One of these is the choice of the sampling strategy for a calibration dataset that is representative of the SOC variability of the investigated area. Castaldi et al. [9] demonstrated that Sentinel-2 spectral data can support algorithms to select the sample locations that ensure a wide range of SOC content. However, a new field campaign and the related soil chemical analysis are time-consuming and expensive in comparison to using an existing soil spectral library. Castaldi et al. [10,11] conceived the bottom-up approach that allows exploiting the Land Use and Coverage Area frame Survey (LUCAS) European topsoil dataset

and airborne data for SOC mapping without spectral transfer function between laboratory and remote spectral data. While this approach proved to be successful in a relatively small pilot area restricted to bare cropland soils, the main challenge for the SOC estimation from an entire Sentinel-2 tile (i.e., 100 by 100 km) is the minimization of disturbing factors, such as vegetation, residues, roughness, and soil moisture. Clearly, the search for ideal soil conditions (i.e., croplands in seedbed condition) reduces the time window to get a useful image in temperate regions. Rogge et al. [12] used temporal series of Landsat and Sentinel-2 scenes. Their approach masks photosynthetically active vegetation using normalized difference vegetation index (NDVI) as a vegetation index. This map produces, among other things, an average reflectance of bare soil pixels in croplands. Given the position of most of the MSI/Sentinel-2 bands, such an approach is effective in removing the disturbing effect of growing crops, but is less suitable to mask wet soils or crop residues. Although, the two wavelengths used in hyperspectral imagery to calculate the Normalized Soil moisture index [13] are very close to the two SWIR bands of the Sentinel-2, the capability of the multi-spectral data for dry vegetation detection and soil moisture quantification still has to be proven. In this regard, Demattê et al. [14] applied different thresholds based on Landsat spectral data in order to select bare soil during the dry season in Brazil and remove soil spectra affected by straw or vegetation residues.

In this case, we aim testing the effect of the threshold for a spectral index linked to soil moisture and crop residues on the performance of SOC prediction models using the Sentinel-2 and the European LUCAS topsoil database. On the one hand, such an approach will allow deriving a qualified compromise between prediction quality and spatial coverage of the map. On the other hand, exploiting the LUCAS database reduces the need for calibration data and a new ad hoc field campaign. We tested this approach within a large area (100 by 100 km) in North-eastern Germany characterized by different soil types.

2. Materials and Methods

2.1. Study Area and Sentinel-2 Data

The study area measures 10,000 km^2, which corresponds to the coverage of the Sentinel-2B image acquired over North-eastern Germany on 30 August, 2017 at 10:22:49 a.m. (UTC+1) (Figure 1), and it includes the test site of DEMMIN® (Durable Environmental Multidisciplinary Monitoring Information Network, upper left corner: 54°2′ N, 12°52′ E, lower right corner: 53°45′ N, 13°27′ E). The DEMMIN® test site was installed in 1999 by the German Aerospace Center (DLR) as a calibration and validation test site for national and international remote sensing missions. Since the installation of the Terrestrial Environmental Observatory Northeast (TERENO-NE) in 2011 [15], which is managed by the German Research Center for Geosciences Potsdam (GFZ), both institutions have cooperated in the region of Demmin. The primary objective of TERENO-NE is long-term monitoring (>15 years) and analysis of environmental change. The DEMMIN test site is located in the intensively-used agricultural state Mecklenburg-Western Pomerania, consisting of 80.3% cropland, 19.5% grassland/pastureland, and 0.2% other agrarian land use [16]. The site is within the young morainic soil-landscape of northern Germany [17] characterized by geomorphological features, including (sandy-) loamy morainic parent material of the late Pleistocene, kettle holes, and slightly undulating relief. Stagnosols, Luvisols, and their transitional types evolved from unconsolidated glacial till and colluvial soils from eroded loamy material or deposits covering former bogs [17]. The hummocky region shows a complex variability at a short distance of eroded summits to colluvial foot slopes [18]. The truncated soil profiles are mainly classified as Luvisols and Calcaric Regosols, while the colluvial soils are Gleyic-Colluvic Regosols and Terric Histosols. Tillage erosion is the dominant process of soil redistribution due to industrial farming on the very large fields. The large Borentin field (107 ha) illustrates the complex topography and its relation with the spatial patterns in soil properties (Figure 1c,d). The soils of the undulating ground moraine are truncated, while, in the center of the field, a flat depression occurs with a few kettle holes and low moor areas that are partly covered by colluvium [16].

The Sentinel-2B image acquired on 30 August, 2017 was downloaded from the Copernicus open access hub. This image was selected because most of it is cloud-free and it was acquired in a time window, end of the summer, in which soils are likely to be in a seedbed condition in the temperate regions in North-eastern Germany. The atmospheric correction was carried out using the Sen2cor plugin (v. 2.5.5) [19] integrated into Sentinel Application Platform (SNAP) software obtaining the level 2-A. We selected nine bands: B2 (central wavelength = 492.1 nm, bandwidth = 98 nm), B3 (559 nm; 46 nm), B4 (664.9 nm, 39 nm), B5 (703.8 nm, 20 nm), B6 (739.1 nm, 18 nm), B7 (779.7 nm, 28 nm), B8 (832.9 nm, 133 nm), B11 (1610.4 nm, 141 nm), and B12 (2185.7 nm, 238 nm).

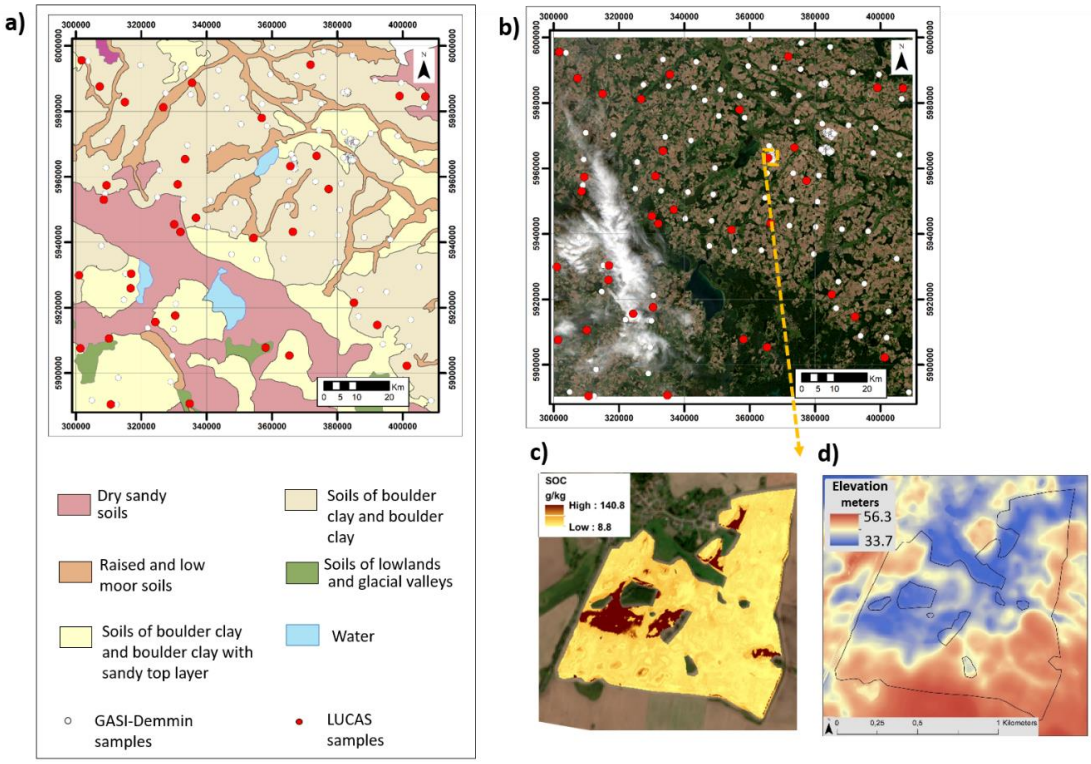

Figure 1. Soil association map (**a**) and RGB (Red: 665 nm, Green: 560 nm, Blue: 490 nm) satellite image (**b**) acquired by Multi-spectral Instrument (MSI) Sentinel-2 sensor. Zoom of the Borentin field illustrating the spatial patterns in SOC (**c**) (modified from Castaldi et al. [6]) and the elevation map obtained from a digital terrain model (**d**). All panels are in the Demmin area (Terrestrial Environmental Observatory Northeast test site) in North-eastern Germany. Coordinates are in WGS 84/UTM zone 33N.

2.2. SOC Calibration Models, Validation, and Mapping

The LUCAS topsoil dataset is a large European soil spectral library (about 20,000 soil samples from 25 countries), which was collected in the framework of the European land use/cover area frame statistical survey in 2009 [20]. The topsoil sampling locations were selected using a latin hypercube-base stratified sampling design from the LUCAS master sample grid of 2 km by 2 km. The LUCAS dataset consists of laboratory spectra and 12 chemical and physical variables for each soil sample. The spectra were acquired with an XDS Rapid Content Analyzer (FOSS NIR Systems Inc., Laurel, MD, USA) spectrometer between 400 and 2500 nm. For the calibration dataset, we selected from the entire LUCAS dataset all samples (35) collected in croplands within the study area (Figure 1, Table 1). The laboratory spectra of the selected LUCAS soil samples were re-sampled according to the spectral bands of the MSI/Sentinel-2B sensor (Figure 2).

Table 1. Descriptive statistics of the calibration and GASI+Demmin (German Agricultural Soil Inventory) datasets.

		SOC g kg^{-1}			
Dataset	**N°**	**Mean**	**Min**	**Max**	**Std**
Calibration	35	30.0	5.2	384.0	70.0
GASI+Demmin	253	17.7	6.0	211.0	25.6

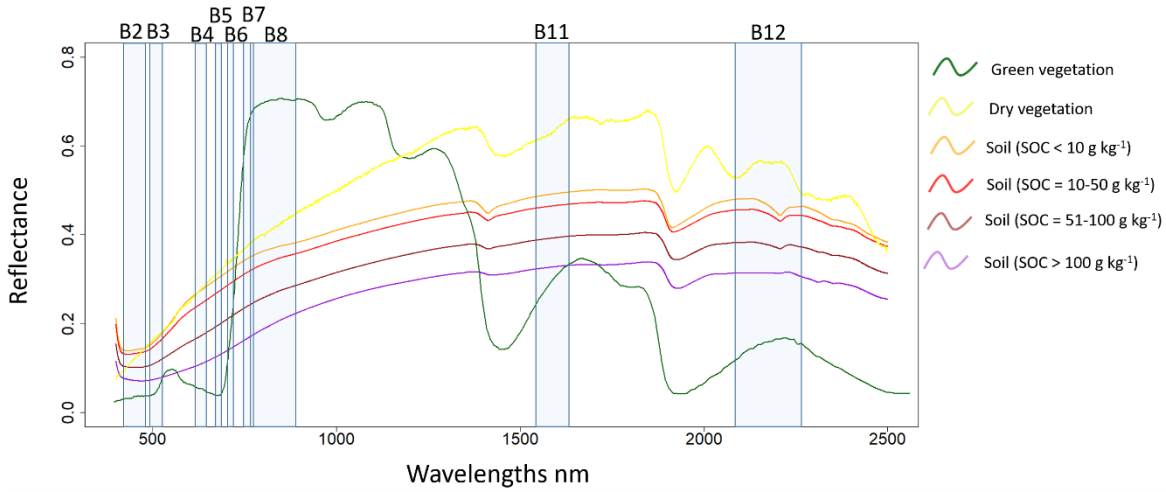

Figure 2. Laboratory soil spectra (400–2500 nm) from LUCAS topsoil database with different organic carbon content compared to dry and green vegetation spectra. The light blue rectangles highlight the spectral position and range of the nine MSI/Sentinel-2 bands used in this work.

The soil albedo decreases with increasing SOC content (Figure 2). This effect is particularly developed in the MSI/Sentinel-2 NIR and SWIR bands. Consequently, in order to have a proxy of the soil brightness, we re-sampled the spectra to MSI/Sentinel-2 bands and calculated the correlation of all combinations of two bands within the SOC content (Equation (1)). These band (B) combinations were used as SOC indices (CI) to fit an exponential regression model to estimate SOC.

$$CI = \frac{B_i + B_m}{2}, \quad (1)$$

These models were later applied to real MSI/Sentinel-2 spectra extracted from pixels in the satellite image. The prediction accuracy of the SOC maps was evaluated on a completely independent validation dataset. The validation dataset was retrieved from the GASI+Demmin (German Agricultural Soil Inventory) dataset, that consists of 253 soil samples (Figure 1, Table 1). The ratio of performance to deviation (RPD), i.e., the ratio between the standard deviation and the root mean square error (RMSE), was used as a measure for the estimation accuracy. Additionally, 78 samples of the GASI+Demmin dataset were collected in the framework of the German Agricultural Soil Inventory [21]. These samples were collected from a soil profile in cropland sites at 0–10 cm depth using three 200 cm^3 soil cores. The samples were dried at 40 °C and sieved at 2 mm. The SOC content was determined for a sample aliquot that was finely ground and analyzed via a dry combustion (TRUMAC, LECO, Saint Joseph, USA). If inorganic carbon was present in the samples, a thermos gradient combustion was used and the organic carbon was determined until a combustion temperature of 550 °C.

The remaining 175 GASI+Demmin samples were taken in the north-eastern part of the study area. The soil samples were collected from the topsoil (0 cm to 10 cm) using a gouge auger in 2017. Each sample consisted of five sub-samples taken within an area of a 5-m radius. The samples were air-dried and sieved at 2 mm and the total carbon was measured by dry combustion using a Carbon Nitrogen (CN) analyzer (VarioMax, ElementarGmbh, Hanau, Germany). If present, inorganic carbon

was obtained by measuring the pressure of CO_2 emitted as a result of hydrochloric acid addition. The SOC content was determined by subtracting the inorganic carbon to total carbon.

The SOC distribution of the GASI+Demmin and calibration dataset is right skewed and most samples have an SOC content below 30 g kg^{-1}, while only a few samples show a very high SOC content (Figure 3).

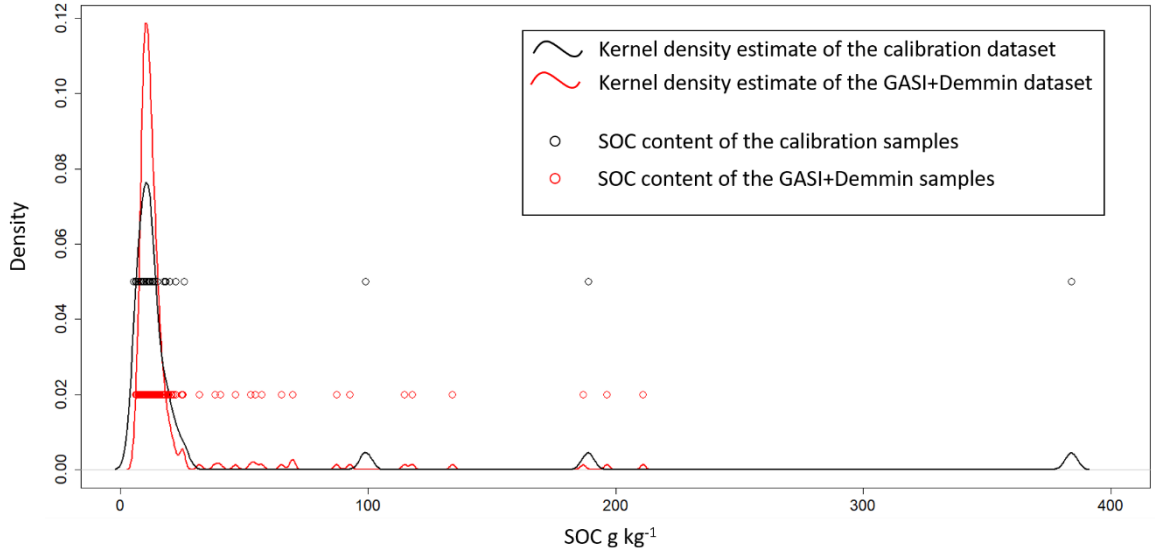

Figure 3. Kernel density estimate plot for the measured Soil Organic Carbon (SOC) values of the calibration (**black line**) and GASI+Demmin (**red line**) datasets. The black and red circles represent, respectively, the SOC content of the calibration and GASI+Demmin samples. Their position along the Y-axis was chosen only for visualization purposes.

The SOC prediction models based on Sentinel-2 indices were applied to the exposed bare soil pixels of the Sentinel-2 image, i.e., the pixels that meet all the following criteria (Figure 4).

- Zero clouds probability. For this purpose, the clouds probability layer provided by European Space Agency (ESA) was used to exclude pixels with a cloud probability higher than zero.
- NDVI (normalized difference between B8 and B4) lower than 0.35 to exclude green vegetation, growing crops, and mixed pixels.
- Differences between B3 and B2 (Green Vegetation Index 1 = GVI1) and between B4 and B3 (Green Vegetation Index 2 = GVI2) both higher than 0. The application of these filters improves the soil mask [14].
- Lastly, different Normalized Burn Ratio 2 (NBR2) index thresholds (from 0.025 to 0.35 in steps of 0.025) were tested in order to exclude spectra affected by high soil moisture content or crop residues. This index was computed as the normalized difference between B11 and B12 (Equation (2)). These two bands are strongly correlated with soil moisture [22] and their differences allow discriminating between dry soil spectra, moist soil spectra, and spectra influenced by straw, other residues, or dry vegetation [14]. The difference between the reflectance around 1600 nm and that around 2100 nm is close to 0 for the soil, while the presence of straw increases this difference (Figure 2).

The choice of the NBR2 index threshold affects the number of pixels for which SOC can be mapped and, consequently, the number of GASI+Demmin samples that can be used for the validation of the SOC models.

$$NBR2 = \frac{B_{11} - B_{12}}{B_{11} + B_{12}}, \quad (2)$$

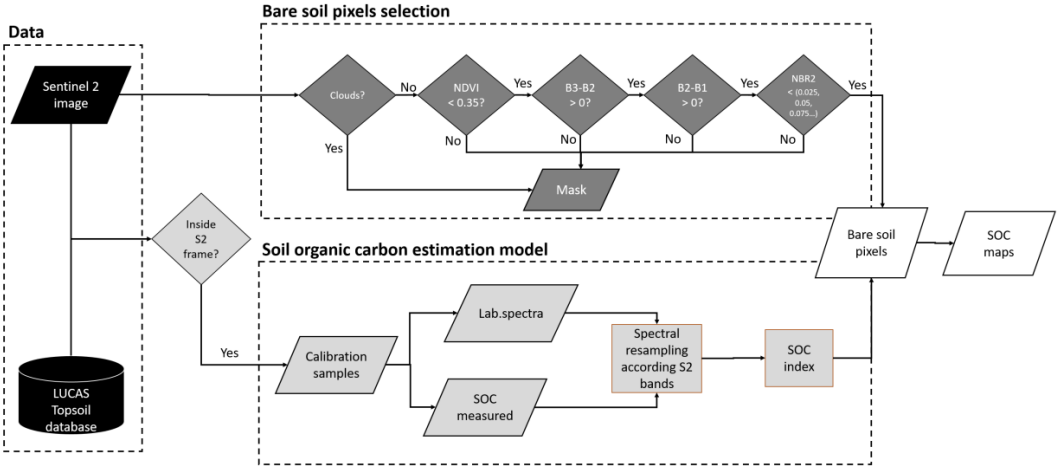

Figure 4. Flowchart concerning soil organic carbon mapping approach using the LUCAS topsoil database and the bare soil pixel selection.

3. Results

3.1. SOC Indices

We explored the linear correlation (Pearson correlation coefficient) between the reflectance indices using the laboratory spectra of the LUCAS dataset and the SOC content. The results showed a strong negative correlation between all indices and SOC. The best values were observed for the indices involving B11 and B12 (Figure 5a), while the combination of bands using B2 and B3 provided the weakest correlations (−0.71). Although the correlation results of Figure 5a refer to a linear correlation, they are consistent with the cross-validation RPD values of the SOC exponential regression models (Figures 5b and 6). The highest RPD values were obtained using B12 in combination with other bands (RPD > 5, Figure 5b).

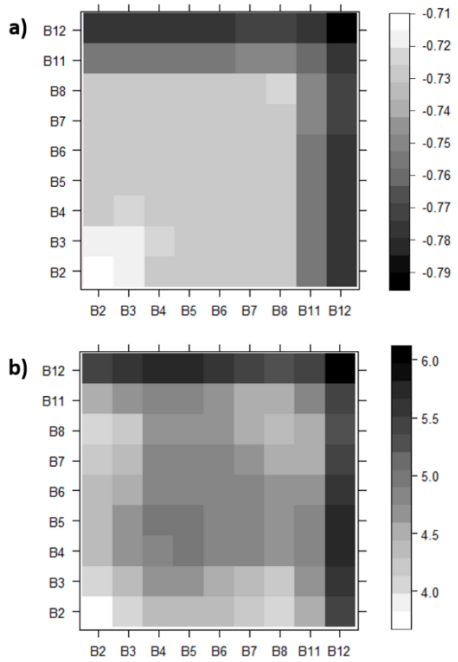

Figure 5. Pearson correlation coefficient between SOC values of the calibration dataset and SOC indices (**a**) and the cross-validation ratio of performance to deviation (RPD) for the calibration models (**b**) obtained by the resampled LUCAS spectra, according to the MSI/Sentinel-2 bands.

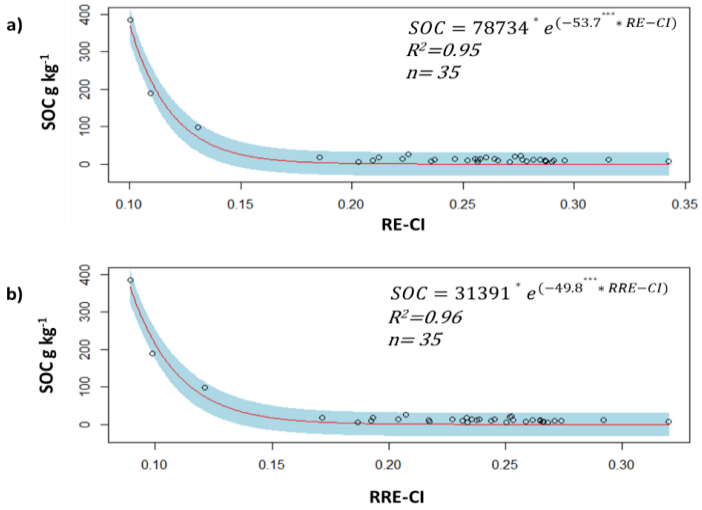

Figure 6. Exponential fitted models (red lines) for soil organic carbon (SOC) prediction using the Red-Edge Carbon Index (RE-CI) (**a**), and the Red-Red-Edge Carbon Index (RRE-CI) (**b**) index. The blue area represents the prediction limits. The black circles represent the measured SOC values of the calibration dataset. Asterisks refer to the significance level: *** p-value < 0.01, * p-value < 0.05.

3.2. SOC Maps and Validation

Furthermore, 12.5% of the pixels within the Sentinel-2 image were characterized by cloud cover, and only 26.9% of the non-cloud pixels were classified as 'without green vegetation' using the NDVI, GVI1, and GVI2 thresholds (Figure 7). Thus, 34 out of the 253 GASI+Demmin samples were localized within crop pixels or cloudy areas (Tables 1 and 2). An NBR2 threshold from 0.35 to 0.25 did not affect the number of selected pixels (26.8%) and the validation samples remaining in a suitable condition to apply the SOC models (219). The selected pixels decreased markedly when an NBR2 threshold of 0.1 was reached (58 remaining validation samples). For the most severe NBR2 threshold (0.05), only 10% of the area was selected (Figure 7) and only 17 validation samples remained (Table 2). Clearly, the NBR2 threshold affects the number of the pure pixels selected and, consequently, the surface that can be mapped. Without applying NBR2, we can map 26.9% of the investigated area, while using NBR2, we reduce the mapped area to 10% of the entire Sentinel-2 image. The mapped area in one restricted framework of the Sentinel-2 image decreases when the number of pixels to which an SOC estimation model is applied is restricted using an NBR2 threshold of 0.05 compared to a threshold of 0.075 (Figure 8). On the one hand, the decrease of the NBR2 threshold reduces the area for which the SOC is mapped (Figure 7) and, on the other hand, a decrease in NBR2 improves the SOC validation accuracy (Table 2). In Table 2 and Figure 9 (black solid line), the best RPD value among all the band combinations for different NBR2 thresholds is displayed. Moreover, we showed the results for two indices: the average value between B5 and B6 (Red-Edge Carbon Index: RE-CI) and the average between B4 and B5 (Red-Red-Edge Carbon Index: RRE-CI). The SOC validation accuracy, in terms of RPD, decreases with the NBR2 threshold (Figure 9). Thus, the highest RPD (4.4) was detected for an NBR2 value of 0.05 using the RE-CI (Table 2). The RRE-CI calibration model showed RPD values larger than 1.5 for NBR2 values below 0.2. Overall, RE-CI performed less well and already reached an RPD of 1.5 for the NBR2 value larger than 0.075. Increasing the NBR2 threshold from 0.05 to 0.1 yielded an underestimation of the additional samples involved in the validation (Figure 10).

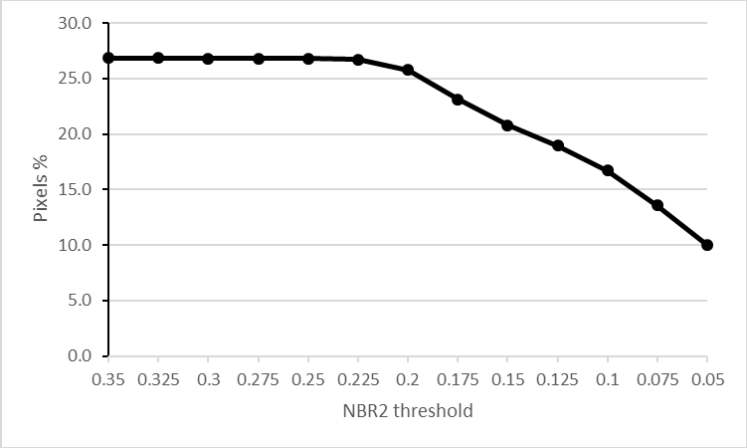

Figure 7. Percentage of the selected pixels of the Sentinel-2 image for different normalized burn ratio 2 (NBR2) thresholds.

Table 2. Descriptive statistics of the validation datasets retrieved from the GASI+Demmin dataset after the use of different Normalized Burn Ratio2 index (NBR2) thresholds and the corresponding root mean square error (RMSE) and ratio of performance to deviation (RPD) of validation. RE-CI: Red-Edge Carbon Index; RRE-CI: Red-Red-Edge Carbon Index.

		SOC g kg^{-1}				Best Validation Results		RE-CI		RRE-CI	
NBR2	N	Mean	Min	Max	Std	RMSE g kg^{-1}	RPD	RMSE g kg^{-1}	RPD	RMSE g kg^{-1}	RPD
0.05	17	18.0	7.7	134.1	30.0	6.8	4.4	6.8	4.4	10.3	2.9
0.075	36	14.7	6.3	134.1	21.1	8.1	2.6	8.1	2.6	9.2	2.3
0.1	58	17.7	6.0	134.1	23.6	14.8	1.6	18.2	1.3	15.7	1.5
0.125	125	20.8	6.0	196.4	30.8	13.4	2.3	30.8	1.0	18.1	1.7
0.15	164	18.9	6.0	196.4	27.2	12.4	2.2	27.2	1.0	17.0	1.6
0.175	186	18.3	6.0	196.4	25.7	12.2	2.1	25.7	1.0	17.1	1.5
0.2	207	17.7	6.0	196.4	24.5	24.5	1.0	49.0	0.5	49.0	0.5
0.225	217	17.4	6.0	196.4	24.0	26.7	0.9	48.0	0.5	48.0	0.5
0.25	219	17.4	6.0	196.4	23.9	26.6	0.9	47.8	0.5	47.8	0.5
0.275	219	17.4	6.0	196.4	23.9	26.6	0.9	47.8	0.5	47.8	0.5
0.3	219	17.4	6.0	196.4	23.9	26.6	0.9	47.8	0.5	47.8	0.5
0.325	219	17.4	6.0	196.4	23.9	26.6	0.9	47.8	0.5	47.8	0.5
0.35	219	17.4	6.0	196.4	23.9	26.6	0.9	47.8	0.5	47.8	0.5

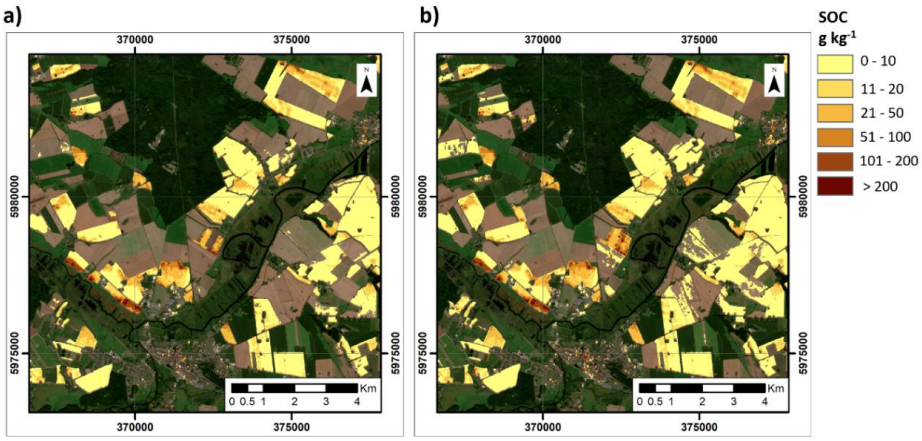

Figure 8. Zoom on a restricted area of the soil organic carbon maps obtained applying the Red-Edge Carbon Index RE-CI model to the pure pixels selected using a Normalized Burn Ratio 2 (NBR2) threshold of 0.05 (**a**) and 0.075 (**b**). The RGB (red: 665 nm, green: 560 nm, blue: 490 nm). Sentinel-2 image acquired on 30 August, 2017 was used as background. Coordinates are in the WGS 84/UTM zone 33N.

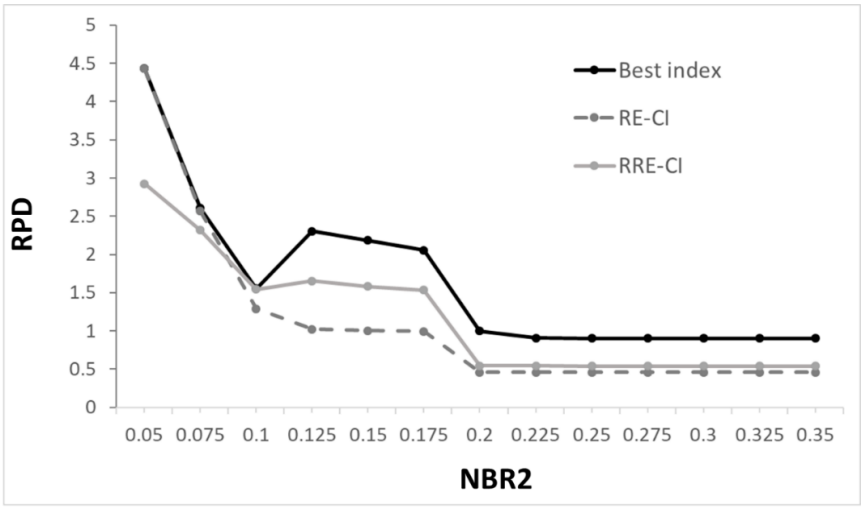

Figure 9. Ratio of performance to deviation (RPD) of validation using different Sentinel-2 indices and Normalized Burn Ratio 2 (NBR2) index thresholds.

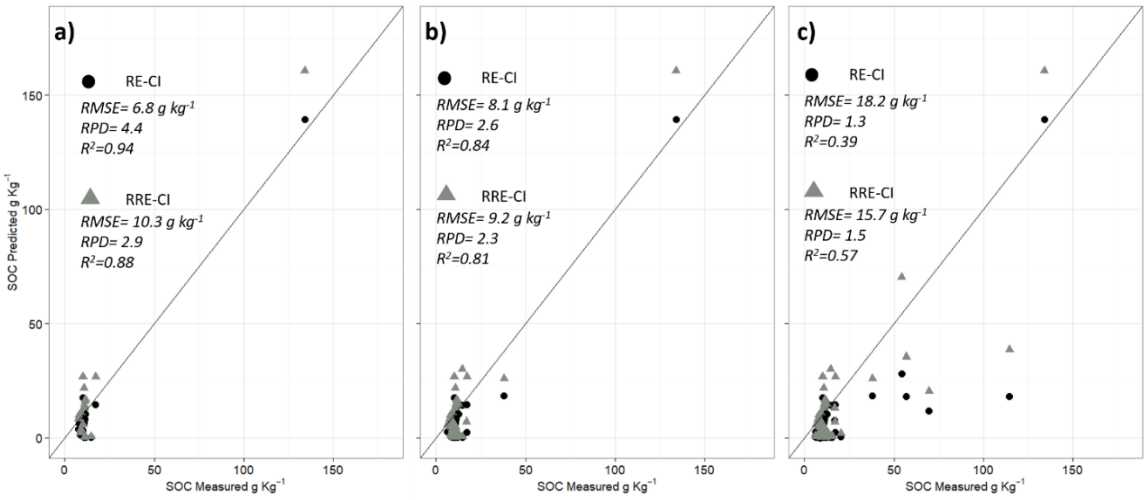

Figure 10. Plots of measured vs. predicted soil organic carbon values for the Red-Edge Carbon Index (RE-CI) and Red-Red-Edge Carbon Index (RRE-CI) index using an Normalized Burn Ratio 2 (NBR2) threshold of 0.05 (**a**), 0.075, (**b**) and 0.1 (**c**).

4. Discussion

Several differences exist between a laboratory spectrum of a dried and sieved soil sample and the spectrum of the satellite sensors representing the pixel where the sample was taken. One of the most evident differences is related to the distance between the sensor and the target. The atmospheric disturbance and lower signal-to-noise ratio of the remote sensor affect the spectral output, and, consequently, the estimation accuracy of the soil property. The B12 (2185 nm) and B11 (1610 nm) of the MSI/Sentinel-2 sensor, in combination with other bands, showed the highest correlation with SOC using the LUCAS spectra acquired in the laboratory (Figure 5). We observed a weakening of the correlations removing the two calibration samples with the highest SOC values (Pearson's coefficient between 0.5 and 0.62). This decrease in correlation especially concerns the combination of bands including B11 and B12. These two bands proved to be very important for calibration sets with a large SOC variability [6]. In other words, the differences in terms of soil brightness become more evident.

However, when we applied these indices involving the SWIR bands (B11 and B12) to real Sentinel-2 data, the accuracy was always unsatisfactory and it decreased with an increasing NBR2 value. The loss of prediction capability of the SWIR bands from the laboratory to the Sentinel-2 data is mainly due to

three factors: i) the large bandwidth of the two bands (141 nm for B11 and 238 nm for B12) and their lower spatial resolution (20 m) as compared to the other Sentinel-2 bands, ii) the lower signal-to-noise ratio (SNR) as compared to that of the visible and infrared bands [6], iii) the mixed influence of several disturbing factors on the spectral signal of the SWIR bands due to variable soil texture and mineralogy (clay or carbonate content), soil moisture [23], or dry vegetation cover [24].

Generally, the best validation results were obtained using a combination of red edge bands (RE-CI) or red/red edge (RRE-CI, Table 2, Figure 11). The exponential calibration models of Figure 6a,b showed how a small variation of the index values between 0.10 and 0.12 corresponds to abrupt changes in SOC content, while the two indices are less sensitive for SOC values lower than 100 g kg^{-1}. Many authors reported a wide absorption feature related to SOC around 664 nm [4,25], which includes the spectral range of B4 and B5. The MSI/Sentinel-2 band centered at 740 nm (B6) is spectrally close to the absorption feature associated with the overtone of the N-H bond [25,26]. This is a chemical bond that characterizes some compounds of soil organic matter, e.g., amino acids [27]. Castaldi et al. [6] analyzed the relative variable importance for estimating SOC in the Demmin area using the MSI/Sentinel-2 bands as variables in a random forest regression model. Their results showed the most important bands to be the B4 and B5 as well as the two SWIR bands (B11 and B12).

The structural complexity of the organic matter entails a large heterogeneity of the components, which, in turn, involves a large variability of spectral responses and, consequently, the need to calibrate 'local' SOC estimation models based on geographical or spectral proximity. After all, these empirical models can hardly be applied in areas characterized by different soil types. The SOC indices mainly exploit the link between SOC content and soil brightness. This link can be masked or weakened as a result of variation in mineralogy that affects the soil color as well as the type of clay minerals and iron oxide contents. Consequently, the transferability of these spectral indices is likely to be more straightforward for young soil types, where variability in clay minerals and iron oxide content is not yet pronounced.

The availability of a sufficient number of LUCAS soil samples within the study area and their wide spatial distribution (Figure 1) together with the large variability in terms of SOC content (Table 1) allowed calibrating reliable SOC models [9]. The only drawback in the LUCAS database is that it is based on a stratified random sampling strategy in geographical space (rather than feature space) [6]. Since the area covered by peaty topsoil in cropland is limited to the center of the kettle holes [28], the number of high SOC content samples in the LUCAS database is very limited (Table 1). It is likely that the LUCAS samples in other European regions are not numerous enough [29] or not representative of the soil variability of the investigated area. Until now, remote sensing or airborne spectroscopy was used for small areas that are mostly rather homogeneous in parent material and soil forming factors [30]. Castaldi et al. [9] demonstrated that sampling strategies based on the feature space, where the spectral bands were used as ancillary data, were most efficient when the Demmin area was stratified, according to 'soil scapes,' i.e., distinguishing sandy and clay topsoil (Figure 1). Other factors that are likely to improve the SOC prediction models are the absence of a soil crust (since it does not represent the spectral signal of the topsoil [30] and the large variability in SOC content). Gomez et al. [31] reported that, in contrast to clay and calcium carbonate content, SOC could not be predicted in Mediterranean environments due to the very low and rather constant SOC content. It should be noted that only 10% to 15% of the landscape is not affected by erosion [28]. The fact that the landscape is derived from relatively young glacial sediments explains that differences in pedogenic oxides, leaching of calcium carbonate, and differentiation in clay minerals have not yet had enough time to develop and are, therefore, unlikely to greatly influence the soil spectra. Consequently, the SOC indices can effectively exploit the link between SOC content and soil brightness in young soil types where this link is not masked or weakened from variation in mineralogy due to the pedogenesis age.

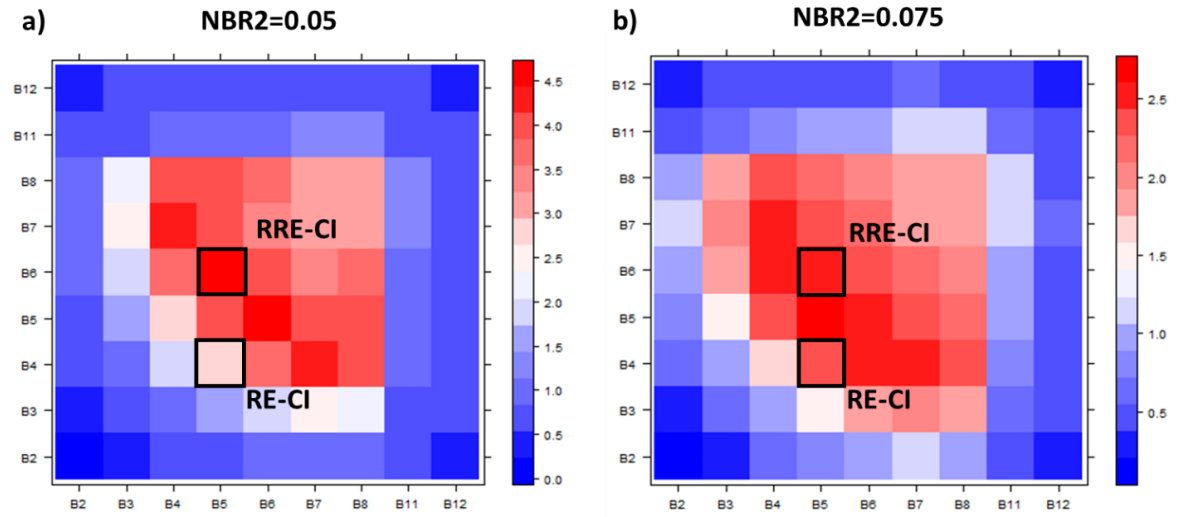

Figure 11. Ratio of performance to deviation (RPD) of validation using a Normalized Burn Ratio 2 (NBR2) index thresholds of 0.05 (**a**) and 0.075 (**b**). The black squares highlight the Red-Red edge carbon index (RRE-CI) and the Red edge carbon index (RE-CI).

Straw, other crop residues, or dry vegetation are disturbing factors for soil imaging spectroscopy. Currently, several methods exist to map dry vegetation fractional cover based on hyperspectral imagery. Such methods can be used to separate pure soils from soils partly covered by vegetation [24]. We selected green vegetation pixels using NDVI and two other indices based on the visible region and applied the NBR2 index to detect dry vegetation or crop residues. Demattê et al. [14] already showed that the NBR2 values are correlated with the amount of dry vegetation at the soil surface. In their study, the index provided negative, or close to zero, values for pixels with a 100% soil signal, while the NBR2 values increased with straw cover. Therefore, it is reasonable to adopt a very low NBR2 threshold in order to select pure soil pixels for remote sensing applications (Figure 9). We tested different NBR2 thresholds, and sought to understand how they affect the SOC estimation accuracy. The best validation accuracy was obtained using an NBR2 threshold of 0.05. While using a less restrictive threshold, the estimation accuracy sharply decreased (Figure 9). The SOC prediction of the RE-CI and RRE-CI indices decreases with an increasing NBR2 threshold (Figure 10). This is mainly due to an underestimation of the low SOC content. Although the statistical accuracy in terms of R^2, RPD, and RMSE is still satisfactory, the SOC underestimation already appears for an NBR2 of 0.075 and becomes striking for a NBR2 of 0.1 (Figure 10b,c). Demattê et al. [14] remarked that the NBR2 index is not particularly effective for pixels with a residue cover between 80% and 20% and that a stricter NBR2 threshold (<0.05) does not improve detection of pure soil pixels. On the contrary, a less severe threshold (>0.05) will result in the selection of pixels that are almost completely covered by dry vegetation or residues.

The spectrum of a bare soil pixel can also be affected by soil moisture, which entails a nonlinear reduction of the reflectance over the entire spectrum [32]. This challenge in predicting soil moisture effects made many researchers refrain from the use of laboratory spectral libraries for the calibration of remote sensing data or incited them to find alternative solutions to the problem by spectral transfer methods [33] or by a bottom-up approach [11]. We faced the issue from another perspective, detecting and selecting only the bare soil pixels mimicking the conditions of a dry soil sample. Several methods exist that could be applied to mask moist soil pixels based on hyperspectral imagery [34,35]. The adaptation of these methods to multispectral satellites has yet to be investigated. Some authors investigated the suitability of normalized soil moisture indices to be applied for remote sensing [13,34]. Both the normalized indices calibrated by Haubrock et al. [13] and Castaldi et al. [34] exploited the combination of two SWIR wavelengths: 1800 nm and 2119 nm in Haubrock et al. [13] (NSMI = normalized soil moisture index) and between 1700 nm and 2100 nm in Castaldi et al. [34] (SMIR_A = soil moisture index for remote sensing). The SWIR bands of the MSI/Sentinel-2 sensor used for the

NBR2 are very close to those employed for the NSMI and SWIR_A index. B11 is centered at 1610 nm and it has the upper boundary at 1682 nm, while B12 is centered at 2185 nm and has the lower boundary at 2078 nm. Thus, the NBR2 can be used as proxy for soil moisture characterization.

Soil roughness is another disturbing factor for SOC estimation by satellite data. Generally, its influence is limited when the satellite images are acquired for smooth soils in a seedbed condition. However, this particular condition occurs only twice yearly and lasts several days to a week in temperate regions. Therefore, the time window for acquiring Sentinel-2 images is rather restricted. Usually, the periods of the year when it is more probable to find soil in a seedbed condition is spring when summer crops such as sugar beet, potatoes, or maize are sown or the end of the summer when winter cereals are sown.

Taking into account, the above-mentioned restrictions, on the one the hand, entails an improvement of the SOC estimation accuracy and, on the other hand, decreases the extent of the mapped surface (Figures 7 and 8). Nevertheless, the pixel selection process can be tuned according to the goal of the survey, which gives a warning on the reliability of the SOC map for each NBR2 threshold.

5. Conclusions

We tested the effect of the threshold for a spectral index linked to soil moisture and crop residues on the performance of SOC prediction models in croplands using the European LUCAS database and Sentinel-2 data. The resampling procedure of the LUCAS spectra according the MSI/Sentintel-2 bands allowed pointing out the strong correlation between the reflectance in the SWIR region and SOC content. However, the SOC estimation models using SWIR bands showed a consistent decrease of their performances when they were applied to real Sentinel-2 spectra due to the poorer spatial and spectral resolution and lower SNR as compared to the other bands, combined with the confounding effect of crop residues and soil moisture on the SWIR bands. The red edge MSI/Sentinel-2 bands and a combination of red and red edge bands provided the best estimation accuracies for the validation dataset. These results showed that: (i) the application of an SOC prediction model using visible and red edge spectra acquired in the laboratory to real Sentinel-2 data, and (ii) the high signal quality of the three Sentinel-2 bands involved in the RE-CI and RRE-CI indices (B4, B5, and B6). The pure pixel selection based on an NRB2 threshold has proven to be necessary to obtain reliable SOC maps. In particular, an NBR2 threshold of 0.05 allowed selecting (nearly) bare soil pixels from the Sentinel-2 image. It was observed that higher NBR2 threshold values allowed mapping SOC over larger areas at the expense of a decreasing estimation accuracy, especially for NBR2 values above 0.175.

Author Contributions: Conceptualization, F.C. and B.v.W.; methodology, F.C. and B.v.W.; software, F.C.; validation, S.C., F.C., A.D., and B.v.W.; formal analysis, F.C. and B.v.W.; investigation, S.C., F.C., A.D., and B.v.W.; resources, B.v.W., S.C. and A.D.; data curation, F.C. and B.v.W.; writing—original draft preparation, S.C., F.C., A.D., and B.v.W.; writing—review and editing, F.C. and B.v.W.; visualization, F.C.; supervision, B.v.W.; project administration, B.v.W.; funding acquisition, B.v.W.

Funding: The Belgian Federal Science Policy Office (BELSPO) as part of the PROSOIL project "The evaluation of forthcoming satellites for mapping topsoil organic carbon in croplands" (Contract SR/10/327) funded the research.

Acknowledgments: We thank the Terrestrial Environmental Observatory Northeast (TERENO NE) and D. Spengler with the TERENO remote sensing research group at Deutsches GeoForschungsZentrum GFZ for the introduction of the Demmin area and for the chance to perform our research activities in this project area.

Conflicts of Interest: The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. United Nations Conventions to Combat Desertification (UNCCD). Sustainable Indicator Goal 15.3.1. Available online: https://knowledge.unccd.int/topics/sustainable-development-goals-sdgs/sdg-indicator-1531 (accessed on 2 September 2019).
2. FAO 2017. Voluntary Guidelines for Sustainable Soil Management Food and Agriculture Organization of the United Nations Rome, Italy. Available online: http://www.fao.org/3/a-bl813e.pdf (accessed on 30 July 2019).

3. Panos, P.; Jones, A.; Aguglia, L.; Legowski, M.; Van Doorslaer, B.; Masson, J.; Laureau, D. CAP Specific Objectives Explained—Brief No 5: Efficient Soil Management. Available online: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/key_policies/documents/cap-specific-objectives-brief-5-soil_en.pdf (accessed on 3 September 2019).
4. Ben-Dor, E.; Inbar, Y.; Chen, Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sens. Environ.* **1997**, *61*, 1–15. [CrossRef]
5. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [CrossRef]
6. Castaldi, F.; Hueni, A.; Chabrillat, S.; Ward, K.; Buttafuoco, G.; Bomans, B.; Vreys, K.; Brell, M.; van Wesemael, B. Evaluating the capability of the Sentinel-2 data for soil organic carbon prediction in croplands. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 267–282. [CrossRef]
7. Vaudour, E.; Gomez, C.; Fouad, Y.; Lagacherie, P. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sens. Environ.* **2019**, *223*, 21–33. [CrossRef]
8. Taghadosi, M.M.; Hasanlou, M.; Eftekhari, K. Retrieval of soil salinity from Sentinel-2 multispectral imagery. *Eur. J. Remote Sens.* **2019**, *52*, 138–154. [CrossRef]
9. Castaldi, F.; Chabrillat, S.; van Wesemael, B. Sampling Strategies for Soil Property Mapping Using Multispectral Sentinel-2 and Hyperspectral EnMAP Satellite Data. *Remote Sens.* **2019**, *11*, 309. [CrossRef]
10. Castaldi, F.; Chabrillat, S.; Chartin, C.; Genot, V.; Jones, A.R.; van Wesemael, B. Estimation of soil organic carbon in arable soil in Belgium and Luxembourg with the LUCAS topsoil database. *Eur. J. Soil Sci.* **2018**, *69*, 592–603. [CrossRef]
11. Castaldi, F.; Chabrillat, S.; Jones, A.; Vreys, K.; Bomans, B.; van Wesemael, B. Soil organic carbon estimation in croplands by hyperspectral remote APEX data using the LUCAS topsoil database. *Remote Sens.* **2018**, *10*, 153. [CrossRef]
12. Rogge, D.; Bauer, A.; Zeidler, J.; Mueller, A.; Esch, T.; Heiden, U. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sens. Environ.* **2018**, *205*, 1–17. [CrossRef]
13. Haubrock, S.-N.; Chabrillat, S.; Lemmnitz, C.; Kaufmann, H. Surface soil moisture quantification models from reflectance data under field conditions. *Int. J. Remote Sens.* **2008**, *29*, 3–29. [CrossRef]
14. Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.* **2018**, *212*, 161–175. [CrossRef]
15. Zacharias, S.; Bogena, H.; Samaniego, L.; Mauder, M.; Fuß, R.; Pütz, T.; Frenzel, M.; Schwank, M.; Baessler, C.; Butterbach-Bahl, K.; et al. A Network of Terrestrial Environmental Observatories in Germany. *Vadose Zone J.* **2011**, *10*, 955. [CrossRef]
16. Blasch, G.; Spengler, D.; Itzerott, S.; Wessolek, G. Organic Matter Modeling at the Landscape Scale Based on Multitemporal Soil Pattern Analysis Using RapidEye Data. *Remote Sens.* **2015**, *7*, 11125–11150. [CrossRef]
17. NutzungsdifferenzierteBodenübersichtskarte der Bundesrepublik Deutschland 1:1.000.000 (BÜK 1000 N2.3)—Auszugskarten Acker. Available online: https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK1000/buek1000_node.html (accessed on 6 June 2019).
18. Heinrich, I.; Balanzategui, D.; Bens, O.; Blasch, G.; Blume, T.; Böttcher, F.; Borg, E.; Brademann, B.; Brauer, A.; Conrad, C.; et al. Interdisciplinary Geo-ecological Research across Time Scales in the Northeast German Lowland Observatory (TERENO-NE). *Vadose Zone J.* **2018**, *17*. [CrossRef]
19. Mueller-Wilm, U.; Devignot, O.; Pessiot, L. Sen2Cor Configuration and User Manual. S2-PDGS-MPC-L2A-SUM-V2.5.5. 2018. Available online: http://step.esa.int/thirdparties/sen2cor/2.5.5/docs/S2-PDGS-MPC-L2A-SUM-V2.5.5_V2.pdf (accessed on 1 July 2019).
20. Tóth, G.; Jones, A.; Montanarella, L. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environ. Monit. Assess.* **2013**, *185*, 7409–7425. [CrossRef] [PubMed]
21. Vos, C.; Don, A.; Hobley, E.U.; Prietz, R.; Heidkamp, A.; Freibauer, A. Factors controlling the variation in organic carbon stocks in agricultural soils of Germany. *Eur. J. Soil Sci.* **2019**, *70*, 550–564. [CrossRef]

22. Musick, H.B.; Pelletier, R.E. Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands. *Remote Sens. Environ.* **1988**, *25*, 167–184. [CrossRef]
23. Nocita, M.; Stevens, A.; Noon, C.; van Wesemael, B. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma* **2013**, *199*, 37–42. [CrossRef]
24. Steinberg, A.; Chabrillat, S.; Stevens, A.; Segl, K.; Foerster, S. Prediction of Common Surface Soil Properties Based on Vis-NIR Airborne and Simulated EnMAP Imaging Spectroscopy Data: Prediction Accuracy and Influence of Spatial Resolution. *Remote Sens.* **2016**, *8*, 613. [CrossRef]
25. Rossel, R.A.V.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
26. Katuwal, S.; Knadel, M.; Norgaard, T.; Greve, M.H.; De Jonge, L.W. Visible-Near-Infrared Spectroscopy can predict Mass Transport of Dissolved Chemicals through Intact Soil. *Sci. Rep.* **2018**, *8*, 11188. [CrossRef] [PubMed]
27. Senwo, Z.N.; Tabatabai, M.A. Amino acid composition of soil organic matter. *Biol. Fertil. Soils* **1998**, *26*, 235–242. [CrossRef]
28. Kappler, C.; Kaiser, K.; Tanski, P.; Klos, F.; Fülling, A.; Mrotzek, A.; Sommer, M.; Bens, O. Stratigraphy and age of colluvial deposits indicating Late Holocene soil erosion in northeastern Germany. *CATENA* **2018**, *170*, 224–245. [CrossRef]
29. Stevens, A.; Nocita, M.; Tóth, G.; Montanarella, L.; van Wesemael, B. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE* **2013**, *8*, e66409. [CrossRef] [PubMed]
30. Chabrillat, S.; Ben-Dor, E.; Cierniewski, J.; Gomez, C.; Schmid, T.; van Wesemael, B. Imaging Spectroscopy for Soil Mapping and Monitoring. *Surv. Geophys.* **2019**, *40*, 361–399. [CrossRef]
31. 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]
32. Lobell, D.B.; Asner, G.P. Moisture Effects on Soil Reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66*, 722. [CrossRef]
33. Nouri, M.; Gomez, C.; Gorretta, N.; Roger, J.M. Clay content mapping from airborne hyperspectral Vis-NIR data by transferring a laboratory regression model. *Geoderma* **2017**, *298*, 54–66. [CrossRef]
34. Castaldi, F.; Palombo, A.; Pascucci, S.; Pignatti, S.; Santini, F.; Casa, R. Reducing the influence of soil moisture on the estimation of clay from hyperspectral data: A case study using simulated PRISMA data. *Remote Sens.* **2015**, *7*, 15561–15582. [CrossRef]
35. Fabre, S.; Briottet, X.; Lesaignoux, A. Estimation of Soil Moisture Content from the Spectral Reflectance of Bare Soils in the 0.4–2.5 µm Domain. *Sensors* **2015**, *15*, 3262–3281. [CrossRef]

© 2019 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/).